

Casper Tybjerg, Jonatan Bruun Borring, and Luan Nhu Vu

The Digitization of Silent Films and the Teaching of Film Historiography: Entanglements and Opportunities

Introduction

The present chapter presents research results concerning the cutting rates of Danish silent films, results obtained through the use of automated scene detection. Perhaps more importantly, the article also presents reflections on two methodological issues: first, the importance of integrating quantitative style analysis more closely with historical and archival research and, second, the potential of integrating historical and stylistic research with teaching. The article has grown out of a strategic initiative at the University of Copenhagen to advance the integration of research into teaching. More specifically, it emerges from a methods seminar in the first year of the MA program in Film and Media Studies.¹

The most common form of qualitative style analysis has been the examination of cutting rates.² In our estimation, quantitative style analysis is most valuable if it can be carried out at scale. Establishing how fast a particular film is cut will not be particularly informative unless it can be compared to some sort of broader pattern. For instance, a more qualified discussion of the stylistic *norms* shaping Danish silent cinema would require a fairly large number of movies to be stylistically quantified. Comparisons between, say, different national traditions would require even more data.³

¹ To underscore the collaborative character of this article, we largely refer to ourselves as “we” in the rest of the article. In practice, the bulk of the research work, particularly the development of the PySceneDetect parameters, was carried out by Borring and Vu (students in the methods seminar), whereas the framework discussion of research-teaching integration and entangled film history has largely been written by Tybjerg (teacher of the methods seminar).

² The most important resource for this kind of work remains the Cinemetrics website (www.cinemetrics.uchicago.edu), founded by Yuri Tsivian. See Yuri Tsivian, “Cinemetrics: Part of the Humanities’ Cyberinfrastructure,” in *Digital Tools in Media Studies*, ed. Michael Ross, Manfred Grauer, and Bernd Freisleben (Bielefeld: Transcript, 2009). The research group headed by András Bálint Kovács has also made significant contributions; see András Bálint Kovács, “Shot Scale Distribution: An Authorial Fingerprint or a Cognitive Pattern?” *Projections* 8, no. 2 (2014): 50–70; Mattia Savardi et al., “CineScale: A Dataset of Cinematic Shot Scale in Movies,” *Data in Brief* 36 (2021): 107002, <https://doi.org/10.1016/j.dib.2021.107002>.

³ Radomír D. Kokeš, “Norms, Forms and Roles: Notes on the Concept of Norm (Not Just) in Neoformalist Poetics of Cinema,” *Panoptikum* 22 (2019): 52–78.

The proposal for the first iteration of the integrated course, taught in the fall of 2021, had foreseen that students would be able to gather the necessary data easily using digital tools. If each two-student team could quantify four or five films, a sizable dataset could be assembled, especially if the exercise was repeated across several annual iterations of the course. Numerous practical issues arose, however, involving both access to the films and the performance of the digital tools; together, they made the proposed approach seem altogether impracticable. Because of this, that kind of research was given less attention in the second iteration of the course. Even so, Borring and Wu were able to solve at least some of the technical issues involved, opening up the possibility of doing further work along these lines in subsequent iterations of the course.

Since this chapter has emerged from work on integrating research with teaching, we decided to start out by briefly reviewing the established models for such research integration, allowing us to reflect on the degree to which our collaboration fits these models. An important result of our work is a heightened awareness of the importance of the way digital archival materials are curated and presented, what kinds of metadata are made available, and so on. To approach these issues, we have found the *Entangled Film History* approach useful, and we discuss it in the second section. While entangled film history is generally focused on transnational topics, we have drawn on its self-reflexive component and attention to the context of the research. The actual research conducted was a quantitative style analysis of a limited number of Danish silent films. The third section will provide a brief introduction to this kind of research, while the fourth examines some of the practical pitfalls encountered while conducting the research, to which the freeware program PySceneDetect has provided a good practical solution. The results of the research are presented in the fifth section, and our conclusion sums up the methodological lessons we believe we can take from our work.

Integrating Research in Teaching

In this section, we will briefly look at the conceptualization of research integration in teaching underlying the plan for the course and discuss where our research collaboration can fit within that model.

In Denmark, “research-based teaching” is written into the University Law, the legislation governing the operations of the Danish universities, as a foundational principle.

§2. It is the task of the university to conduct research and offer research-based education up to the highest international level within its disciplinary subjects.⁴

There is broad agreement among stakeholders that “research-based teaching” is foundational to the role of the universities as educational institutions; accordingly, there has been a good deal of discussion about what it actually entails. In practice, it has generally been assumed that the requirement was satisfied if a substantial proportion of courses were taught by active researchers. However, because of its definitional importance, the University of Copenhagen has made it a strategic goal to “Further develop models for student involvement in research activities and make it a credit-bearing element of their programme.”⁵ To advance this goal, a certain amount of funding was made available to faculty who wanted to experiment with a higher degree of research integration in the classes they taught.

The fact that the university’s leadership has made research integration a strategic goal underscores that it is not a neutral concept. As was pointed out some years ago in a working paper surveying the literature on research integration, it is assumed to be a good thing from the outset – which also leads to the assumption that the more of it, the better: “In the main the literature is characterized by a normative perspective which argues that there is strong value in enhancing the teaching-research nexus in terms of improving student learning and in other areas.”⁶ These assumptions are also present to a certain degree in the models developed by those who research and write on the practice of university teaching. As part of the University of Copenhagen strategic initiative, a group of researchers conducted an ethnological investigation of how research was integrated into teaching in practice, and what students and researchers thought about it.⁷ Based on this research, they developed a model of research integration, which was proposed as an alternative to models de-

⁴ Universitetsloven, LBK no. 778, August 7, 2019, accessed October 23, 2023, <https://www.retsinformation.dk/eli/lt/2019/778> (translation by the authors).

⁵ “Talent and Cooperation.” University of Copenhagen 2023 Strategy document, accessed April 11, 2024, <https://web.archive.org/web/20231128025913/https://about.ku.dk/strategy2023/education/>.

⁶ Paul Trowler and Terry Wareham, “Tribes, Territories, Research and Teaching: Enhancing the Teaching-Research Nexus” (January 1, 2008), Working Paper, 13, accessed October 23, 2023, https://www.researchgate.net/publication/252423791_Tribes_territories_research_and_teaching_Enhancing_the_teaching-research_nexus.

⁷ Tine Damsholt and Marie Sandberg, *Af lyst eller nød: En etnologisk undersøgelse af integration mellem forskning og undervisning i praksis* (Copenhagen: University of Copenhagen, 2018), accessed February 20, 2023, https://curis.ku.dk/ws/files/213594738/AF_LYST_ELLER_N_D_ELEKTRO_NISK_VERSION_2.pdf.

veloped elsewhere, particularly in Great Britain.⁸ The description of the Danish model (Figure 1) explicitly stresses that it is intended to convey the message that different types of research integration are *equal*: “The types of research-based teaching and research integration are regarded equally as the selected types must fit the individual degree programme and the students’ level.”⁹ The model replaced an earlier one that looked like a staircase, with the various types of research integration drawn as steps, signaling a hierarchical progression from lower to higher. However, the new model does not entirely avoid this trap either. The colors in the image of the model deepen as one moves upwards and to the right, which still at least implicitly suggests a hierarchy: the deeper the color, the better.

In its first iteration, the course was intended to include different levels of research integration. For the purposes of Tybjerg’s plan to investigate the aesthetic norms of Danish silent cinema, the most important component would involve students working on data collection and processing (area b. on the model). The idea was for students working in pairs to work through a handful of films, producing complete shot lists (allowing cutting rates to be computed) and adding information about shot scale, camera movements, and scene boundaries. All the films would be Danish and from the same span of years. The hope was that the creation of a dataset of this sort would enable conclusions to be reached about the norms governing film production during the period in question.

While this would be the students’ most important contribution from the point of view of Tybjerg’s research interests within the overall plan of the course, it was only intended as a fairly small preparatory assignment – unless individual students specifically chose to do further work on the data. In practice, however, this proved to be unrealistically demanding. Part of the reason for planning this data-collection work as a side exercise was that the ambition for the course had been to enable students to conduct independent research – an ambition moti-

⁸ The key text is Mick Healey’s 2005 article “Linking Research and Teaching”; its central importance is stressed in an extensive report surveying the literature on research integration in teaching: “In our review, we have not been able to identify an alternative typology that would represent a substantial divergence from this model” (Mari Elken and Sabine Wollscheid, *The Relationship between Research and Education: Typologies and Indicators. A Literature Review*, Nordic Institute for Studies in Innovation, Research and Education (NIFU) (Oslo, 2016), 16, accessed February 6, 2023, <https://www.nifu.no/en/publications/1351162/>; citing Mick Healey, “Linking Research and Teaching: Exploring Disciplinary Spaces and the Role of Inquiry-Based Learning,” in *Reshaping the University: New Relationships Between Research, Scholarship and Teaching*, ed. Ronald Barnett (Maidenhead: McGraw-Hill Education, 2005), 67–78).

⁹ “Research integration in teaching,” project home page, accessed February 8, 2023, https://kunet.ku.dk/work-areas/teaching/teaching_development/funds-for-experiments-with-research-integration/Pages/default.aspx.

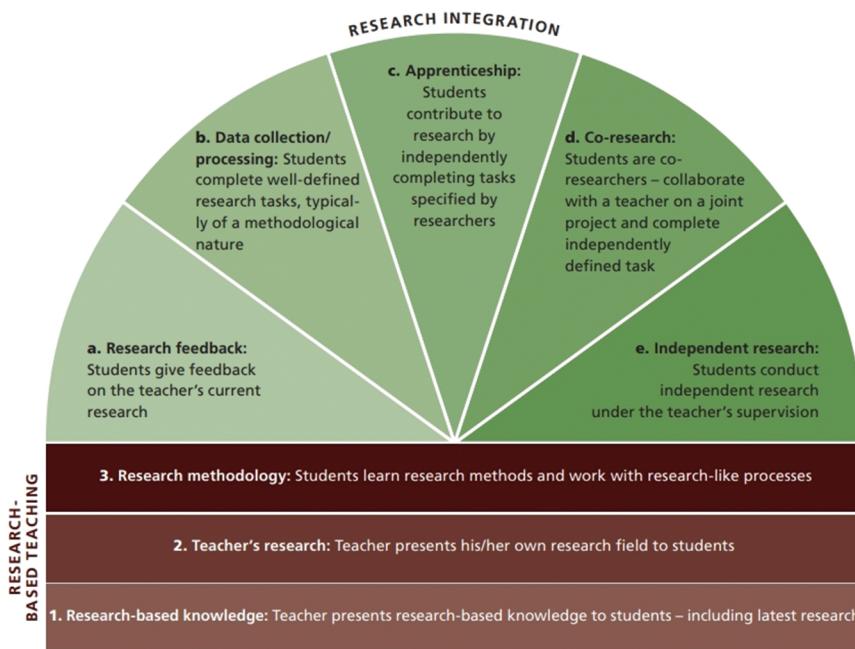


Figure 1: Research-teaching integration model. The darker the colors, the greater the degree of research independence and student autonomy. https://kunet.ku.dk/work-areas/teaching/teaching_development/funds-for-experiments-with-research-integration/Pages/default.aspx (accessed April 11, 2024).

vated to some extent by the way the models gave pride of place to independent research conducted by students. Another consideration was the hope that students could come up with innovative ways of using the material available on the website www.stumfilm.dk.

At [stumfilm.dk](http://www.stumfilm.dk), the Danish Film Institute (DFI) has made and continues to make available a rich trove of material on Danish silent cinema: all extant Danish silent fiction films are being digitized and presented, along with many related archival documents. The great majority of these films were produced by the dominant company Nordisk Film, which still exists and gave permission to make their films available, greatly reducing copyright concerns.

Students were given suggestions for possible ways of approaching the material but were also urged to come up with their own ideas. They were also given the opportunity to tour the DFI's archive facility, to give them an understanding of the physical character of the archival objects, the extensive infrastructure required to ensure their continued preservation, and the skilled labor involved in making digital versions available to users – an understanding, in other words, of

the efforts and support structures required to maintain the apparently seamless and effortless experience of using the *stumfilm.dk* website.

A complex archive or archive-like resource like *stumfilm.dk* may open up new avenues of investigation. It prompts us to ask: what can we do with it? What new questions will it enable us to answer? Thinking of the research process in this way may appear to turn the “proper” relation between research material and research questions on its head. Textbooks on empirical methodologies tend to insist on the primacy of research questions: only when researchers have formulated their research questions should they begin to gather data. While social science disciplines generally adhere to this model, scholarly work in the humanities, particularly historical scholarship, commonly begins with the archive. Researchers immerse themselves in documents, and only later formulate a research question or questions based on what they discover. Film historian Eric Schaefer disarmingly refers to this procedure as “critical mess historiography.”¹⁰ The historians’ immersion-first approach is a well-established research mode, although not always recognized as such in the methodological literature.

Even if our study ended up conforming to the research-question-driven model, the presence of this sort of alternative reveals the need for a framework for thinking about how research may be shaped by institutional and disciplinary contexts; in particular, for thinking more closely about how the archive is set up and what kinds of research procedures it does or does not facilitate – the archive’s affordances or *dispositif*, as it were. *Entangled film history* provides such a framework.

Entangled (Digital) Film History

The entangled approach emphasizes how film historians themselves are enmeshed in a particular context (scholarly traditions, archival access, technological resources, national institutions of learning).¹¹ In Film and Media Studies, the en-

¹⁰ Eric Schaefer, “The Problem with Sexploitation Movies,” *Iluminace* 24, no. 3 (2012): 151.

¹¹ See Malte Hagener, “Introduction: The Emergence of Film Culture,” in *The Emergence of Film Culture: Knowledge Production, Institution Building, and the Fate of the Avant-Garde in Europe, 1919–1945*, ed. Malte Hagener (New York: Berghahn Books, 2014); Daniël Biltreyest and Philippe Meers, “Comparative, Entangled, Parallel and ‘Other’ Cinema Histories. Another Reflection on the Comparative Mode within New Cinema History,” *TMG Journal for Media History* 23, nos. 1–2 (2020); Casper Tybjerg, “The European Principle: Art and Border Crossings in Carl Theodor Dreyer’s Career,” in *A History of Danish Cinema*, ed. Isak Thorsen, C. Claire Thomson, and Pei-Sze Chow (Edinburgh: Edinburgh University Press, 2021), 41–50; Casper Tybjerg, “Danish-German Cinematic Interconnections and the Prospects of an Entangled Film Historiography,” in *Danish and*

tangled approach has mainly been discussed as a framework for investigating the border-crossing and transnational character of many filmmaking careers and media industries. However, in its original formulation as *histoire croisée*,¹² it also encourages (film) historians to reflect on the context within which they do their work, how their work is constrained and facilitated by it, and how it makes some research questions salient while obscuring others.

To take a simple example: to do a quantitative style analysis of a film, or segment and annotate it in some specialized software, you really need to have the film as a file on your computer. The stumfilm.dk website, however, does not allow you to download the films directly. While we had initially hoped to work on all the 235 films available on the site, this was not possible in practice. We would have needed staff from the DFI to download the titles individually onto a hard drive for us, and that was just not feasible given the time available to carry out the investigation. We ended up with a much smaller sample of films; further, given the technical issues involved (of which more below), it would not have been realistic to attempt to analyze the full collection of films, even if it had been accessible.

This example is important to keep in mind when thinking about the way historical research is facilitated and shaped by the archive's affordances. The digital availability of large numbers of silent films enables researchers to view them systematically, but if they cannot be downloaded, certain types of investigation are very difficult to carry out. Another factor to consider is that many archives are organized as national (and nation-bound) institutions. This is apparent in the very names of both the website and the institution that runs it: for both stumfilm.dk and the Danish Film Institute, the nation of Denmark is evidently an important organizational principle. While this may seem unproblematic if you want to work on Danish silent film, the institutional framework makes it easy to think in only national terms and to overlook cross-border entanglements.

In particular, when trying to map out stylistic norms, film historians have long argued that national cinemas are not the most fruitful framework. David Bordwell has shown that the “standard story” of the history of film style has been resolutely internationalist from very early on.¹³ From the first, film has been a highly international medium. While the emergence and persistence of certain craft traditions may be best explained at the national level, this cannot be as-

German Silent Cinema, 1910–1930: Towards a Common Film Culture, ed. Lars-Martin Sørensen and Casper Tybjerg (Edinburgh: Edinburgh University Press, 2023), 24–50.

¹² Michael Werner and Bénédicte Zimmermann, “Beyond Comparison: *Histoire Croisée* and the Challenge of Reflexivity,” *History and Theory* 45, no. 1 (2006): 30–50.

¹³ David Bordwell, “The Power of a Research Tradition: Prospects for Progress in the Study of Film Style,” *Film History* 6, no. 1 (1994): 59–79.

sumed from the outset. Film workers have often travelled to work in other countries, transferring and absorbing skills and norms. Hollywood films have been shown in most places, making their stylistic and storytelling devices at least theoretically available as creative options for filmmakers elsewhere.

Even so, nationally focused works of film historiography predominate. Like archives, research institutions tend to be organized on national lines, and both films and written sources use a particular language, creating a strong countervailing pressure in the direction of methodological nationalism. It is not realistic to imagine that we can easily escape from the path dependencies created by these institutional facts. Instead, we should strive for compatibility and comparability, allowing scholars from different countries to work with each other's data and results.

The need to ensure compatibility was also present when thinking about how quantitative style analysis could be brought into the classroom in a way that would produce data useful for further research. Students would need to follow a clear, predefined procedure for gathering and organizing the data. As mentioned in the introduction, the original idea was to have students work in pairs or small groups¹⁴ to gather quantitative stylistic data on a handful of films, each group working with a different set for films. With enough students and a repetition of the exercise over several iterations of the course, a large and hopefully coherent dataset could be assembled.

However, the curriculum for the Film and Media Studies MA program of which the course was a part requires that the students work on more independent projects for their exams, allowing them to develop their own methodological reflections based on this work. Since this was a key learning outcome, the data gathering exercise had to be a preliminary one that would be practicable for the students to carry out; as a minimum, they should be able to easily access the films for analysis. The difficulties with access, outlined above, led to students being offered alternative options for their exercises. A few, however, were sufficiently intrigued by quantitative style analysis to focus on that approach.

¹⁴ András Kovács had three-person teams do the coding for his project: "Shot scale annotations were provided by three human coders (2 coders + 1 who made decision in case of disagreement)." See Savardi et al., "CineScale," 8.

Quantitative Style Analysis

The quantitative analysis of film style involves counting different types of stylistic features in a film, typically for the purpose of comparison with other films. The process is fairly demanding in terms of the resources it requires, which has been an obstacle to its widespread adoption. The earliest practitioners of the approach Barry Salt, Raymond Bellour, David Bordwell and Kristin Thompson¹⁵) would work on 35mm prints, but few researchers had the necessary access to extensive film archives and viewing tables. The advent of analog and digital video formats held out considerable promise, and the approach came within reach of anyone with an interest in the field (rather than just a privileged few) with the development of purpose-built and freely accessible digital tools, of which the most important has been Cinemetrics, launched in 2005.¹⁶

The film historian Barry Salt has been one of the biggest proponents of this quantitative method, from his pioneering articles in 1974 to his *magnum opus*, *Film Style and Technology* (1983, revised and expanded in 1992 and again in 2009), his findings based on years of collecting film style data from across the whole of film history.¹⁷ Salt argues that a central parameter of cinematic style is *average shot length* (ASL), which is the length of a film in seconds divided by the number of shots in it.¹⁸ The advantage of ASL is that it is easy to calculate: you just need to count each cut, which can be done with a clicker or by making marks on a piece of paper. Existing studies outside Denmark have mostly used this measurement. It has therefore also been used for the present study because it allows for easy comparison with these already existing studies, including Salt's.

Some researchers have pointed out that ASL can be misleading as a measurement, particularly if the film contains some very long or very brief shots. These outliers will distort the measurement, giving two films that are fairly similar in tempo quite different ASLs if one contains such outlier shots and the other does

15 Barry Salt, "Statistical Style Analysis of Motion Pictures," *Film Quarterly* 28, no. 1 (1974): 13–22; Raymond Bellour, "The Unattainable Text," *Screen* 16, no. 3 (1975): 19–27; David Bordwell, Janet Staiger, and Kristin Thompson, *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960* (London: Routledge, 1985).

16 For more historical detail, see Casper Tybjerg, "En introduktion til kvantitativ æstetisk filmanalyse i praksis," *Kosmorama* (2021).

17 Barry Salt, *Film Style and Technology: History and Analysis* (London: Starword, 1983), 2nd ed. (London: Starword, 1992), and 3rd ed. (London: Starword, 2009). See also Salt's compendium of his articles (with useful retrospective commentary), *Moving into Pictures: More on Film History, Style, and Analysis* (London: Starword, 2006).

18 Salt, *Film Style and Technology*, 3rd ed., 160.

not.¹⁹ To compensate for this distortion, these researchers recommend using *median shot length* (MSL) instead: recording the length of each shot, ordering them from the shortest to the longest, and then taking the one in the middle. MSL, according to these researchers, is a more reliable parameter when it comes to comparing the pace of individual films. However, it is somewhat more difficult to calculate since you need to know the length of each shot in the film.

Since the present study was initially conceived as one that would examine a very large corpus of films, the decision was made to focus on tempo. Tempo is an important stylistic characteristic of the film medium; it makes the medium's characteristic temporal dimension meaningful. Tempo has changed importantly over time and constitutes an essential part of the historically changing and developing stylistic conventions of cinema. For the purposes of this study, it was decided to treat cutting rates as indicative of tempo, using ASL and MSL as parameters. The quantification of stylistic parameters allows you to observe the development of different film techniques over time and to compare different groups of films (national cinemas, films by the same director, etc.) with each other. Here, Salt's data and research have laid the groundwork for subsequent studies.

In his contribution to the volume *Finnish Cinema: A Transnational Enterprise*, Jaakko Seppälä proceeds from Salt's work in an attempt to explain the peculiarity of Finnish silent fiction films and how foreign films influenced their stylistic development.²⁰ Like the present study, Seppälä looks at the development of tempo; Finnish silent films tend to be slow-paced. Seppälä has analyzed a substantial number of films from the period from 1920 to 1931, not only calculating their average shot length and median shot length, but also analyzing shot scale. As Seppälä points out, however, there is some disagreement among scholars working in this field on how best to code for shot scale, particularly regarding more distant shots.²¹

¹⁹ Nick Redfern, “The Average Shot Length as a Statistic of Film Style,” *Cinemetrics: Film Statistics: Give and Take* (n.d.), accessed April 11, 2024, https://web.archive.org/web/20230328210123/http://cinemetrics.lv/fsgt_q1b.php.

²⁰ Jaakko Seppälä, “Finnish Film Style in the Silent Era,” in *Finnish Cinema: A Transnational Enterprise*, ed. Henry Bacon (London: Palgrave Macmillan UK, 2016), 51.

²¹ Seppälä, “Finnish Film Style,” 60. Although there is a slight difference in nomenclature, both Barry Salt and András Bálint Kovács use seven shot scales, from Big or Extreme Close Up to Extreme or Very Long Shot; see Barry Salt, “[Data Method] Statistical Style Analysis” (n.d.), accessed October 23, 2023, http://www.starword.com/Data_Method/data_method.html; Kovács, “Shot Scale Distribution,” 50. However, Kovács observes that “Salt's definition of shot scales is slightly different from what I have used toward the long end of the scale” (54). Moreover, Kovács adds the category of “foreground shot,” which combines two different shot scales, with one significant pictorial element close to the camera and another fairly distant from it (51). Taken together, these differences were significant

It would not have been impossible to make an informed decision on how to code shot scale distributions in our course, but it would have required a fair amount of exploratory work and discussion of the advantages and disadvantages of different coding schemes. This would have meant that fewer movies could be analyzed within the time available, which is why we decided to concentrate on collecting data on tempo.

Practical Challenges and Solutions

One key challenge for carrying out the proposed research involved finding the right tools for the job. In order to gather the necessary data at scale, it seemed clear that it should be easy to gather the data (preferably through an automated process) and they should be output in a format that would be straightforward to share and would allow further data to be added fairly easily. For instance, if one already had a dataset containing a list of shots for a number of films with the length of each shot, it would be ideal if information about shot scale, the number and identity of characters, entrances and exits, camera movement, etc., could then be added. Cinemetrics is a very elegant piece of software, but it is not automated, and it does not really allow you to extract data. Regrettably, because of the untimely death of its developer Gunars Civians, it was not maintained for a long time, and since the web interface was a Flash program, it became unworkable for most users, although a new version (a Google Chrome extension) has just been released.²² The problem of software developed for research becoming obsolete because programs are not updated has also affected other specialized film studies tools, as they have often relied on funding for specific, time-limited research projects.

Tybjerg had hoped that the problem could be solved with software providing *automated scene* (or *shot-boundary*) *detection*, where the computer looks for changes in the pixel composition in the image that exceed a certain tolerance.²³ With the right software and settings, this method is surprisingly accurate, al-

enough that, when Kovács compared his analyses of the same films to Salt's data, "in all cases the [shot scale distribution] patterns of Salt's measurement and of mine were very different" (55).

22 Cinemetrics Measurement Tool, chrome web store, accessed April 11, 2024, <https://chromewebstore.google.com/detail/cinemetrics-measurement-t/bekhkeilpopjhchjdbhnfhcjkhljlgme?hl=en>.

23 "Fully automatic shot-boundary detection (SBD) has become the holy grail of video indexing," wrote Jeremy Butler in his article "Statistical Analysis of Television Style: What Can Numbers Tell Us about TV Editing?" *Cinema Journal* 54, no. 1 (2014): 30–31.

though some false positives and negatives do occur.²⁴ DaVinci Resolve, a powerful professional video editing program, seemed a promising candidate: it had the requisite automated scene detection capability; further, after running scene detection on a video file, the results can be viewed as an *editing decision list* that the program can export, allowing a shot list to be generated (almost) automatically; and, finally, it is available in a free version that does not restrict any features needed for the data-gathering process.

However, the DaVinci Resolve program is quite demanding in terms of the computer resources it needs to function optimally; it requires a lot of processing power to run scene detection on an entire film. Even powerful laptops struggle. Moreover, the results proved to be frequently unsatisfactory, with a lot of false positives or negatives (depending on the sensitivity chosen for the scene detection settings). Silent films look quite different from modern films with respect to the features the scene detection algorithms use to determine whether a frame is so different from its predecessor that a cut must have occurred. Silent films mostly lack color (many are of course tinted, but tints tend to be uniform within scenes, meaning that only the black-and-white patterns change from shot to shot), and prints often contain flaws (scratches, white blotches, uneven exposures) that are registered as shot changes by the scene detection algorithms. It is certainly possible to adjust the parameters of the scene detection feature of DaVinci Resolve to produce satisfactory scene detection results, but the processing power requirements meant that the necessary experimentation would have been very time-consuming and, worse, that it would be unlikely that students would be able to get the program to run satisfactorily on their laptops.

These difficulties had discouraged Tybjerg from making quantitative style analysis a central part of the course, but Borring and Vu were able to find a practicable alternative: PySceneDetect, a free program built with the Python programming language, designed specifically to detect shot boundaries in video files.²⁵

PySceneDetect requires users to have Python installed on their computers (along with a few supplementary pieces of software), but no skill in Python programming is needed to use it. The scene detect program runs off the command line. Once you have the program installed, you simply write `scenedetect` followed by the name of the video file you want to analyze at the command line prompt. You add further commands to control the parameters used by the scene detection

²⁴ A similar method is described in Brett Adams, Chitra Dorai, and Svetha Venkatesh, “Toward Automatic Extraction of Expressive Elements from Motion Pictures: Tempo,” *IEEE Transactions on Multimedia* 4, no. 4 (2002): 472–481.

²⁵ Available at <https://www.scenedetect.com/>, accessed April 11, 2024.

algorithm and to modify the output format (the [scenedetect.com](https://www.scenedetect.com) web page offers detailed instructions for using the program).

There are three algorithms for automatic scene detection in PySceneDetect: adaptive, content aware, and threshold.²⁶ After extensive experimentation between the different detection algorithms and thresholds on various parameters, we found a formula that works well on silent films. The command of the formula reads:

```
detect-adaptive --luma-only --threshold 4.5 --min-content-val 10 --frame-window 3
```

The adaptive algorithm is used; for each individual frame, it compares an average luma value (*luma-only* as the films are mostly grayscale) of the three previous and subsequent frames (*frame-window* value) to determine if a cut has happened. If the difference is above 4.5 (*threshold* value) and the difference between the frame in question and the previous frame is higher than 10 (*min-content-val* value) it is considered a cut. These values deviate significantly from the default values in the program. This is likely because of the way the visuals of silent films differ considerably from the modern movies from which the default values were created. Using the adaptive method (i.e., including previous and subsequent frames in the comparison) is especially useful when working with archival film material, as it avoids false positives if there are a few damaged frames. Precisely because the program includes the six surrounding frames in the assessment, it is not considered a cut if there is a sudden flash of light over a frame or two.

This formula was tested by holding the program's results up against a manual count. Three different films were tested. First, shots were counted by hand using a simple smartphone counting app. Next, the same film was run through the PySceneDetect program. When the shot counts were compared, the results were as shown in Table 1.

With deviations of 3 percent, 1 percent and 0.3 percent, the PySceneDetect algorithm has proven to be reliable.

PySceneDetect makes it easy to export the results to a spreadsheet or database. If the command `list-scenes` is added, the program outputs a shot list as a CSV file that can be imported into Microsoft Excel or other programs. The command `save-images` exports frame grabs of the first, middle, and last frames of each shot, allowing easy visual verification of the algorithm's segmentation. The frame grabs can also be used to code for other stylistic parameters.

²⁶ See <https://www.scenedetect.com/docs/latest/api/detectors.html>, accessed April 11, 2024.

Table 1: Comparison of PySceneDetect with manual counting.

DEN HVIDE SLAVEHANDEL (1910)			
Manual count: 65 shots	PySceneDetect: 67 shots	Variance: 2	Deviation: +3%
DER VAR ENGANG (1922)			
Manual count: 674 shots	PySceneDetect: 681 shots	Variance: 7	Deviation: +1%
NEDBRUDTE NERVER (1923)			
Manual count: 670 shots	PySceneDetect: 668 shots	Variance: 2	Deviation: -0.3%

Results

Armed with this effective tool, we proceeded to analyze a sample of 22 Danish silent feature films. We initially decided to focus on feature-length films (excluding short films and incomplete films from consideration) available on stumfilm.dk. The earliest available feature film was from 1910 and the latest from 1932. This fits with what we know about the history of Danish film production: the first feature films were made in 1910, and the 1932 film was the last silent film to be released (the other six features released that year were talkies).

We decided that, for the purposes of the study, it would make most sense to analyze a single random film for each year, ensuring that the whole period was covered. Since the number of feature films produced fluctuated a great deal over the period in question – from dozens every year in the mid-1910s to a handful or less at the end of the 1920s – and survival rates also differ a great deal, it is not possible to say whether the films chosen can be said to be a representative indicator of the tempo of Danish films in a given year.

The sampling of one film for each year in the period was possible until the years 1930 and 1931, where there were no surviving films available (two silent films were released in 1930 and just one in 1931, along with a total of three talkies). Each film from each year was chosen completely randomly from among those available. For all the years up to 1929, there was a minimum of three films to choose from. For the year 1932, there was only a single film. The sample selected in this way included 22 silent feature films from 1910–1932 (excluding 1930 and 1931), which was also a manageable quantity of films and data to work with. We were given access to file copies of the films in the sample by the Danish Film Institute.

An overview of the results of the analysis can be seen in Table 2.

Table 2: Cutting rates for the films included in the sample.

TITLE	YEAR	LENGTH	ASL (sec.)	MSL (sec.)
DEN HVIDE SLAVEHANDEL	1910	33 min.	29.6	19.9
VED FÆNGSLETS PORT	1911	41 min.	37.6	26.1
DØDENS BRUD	1912	41 min.	32.4	18.9
BALLETTENS DATTER	1913	56 min.	25.8	17.2
DET HEMMELIGHEDSFULDE X	1914	90 min.	16.3	12.3
NED MED VAABNENE	1915	65 min.	18.3	12.4
VERDEN'S UNDERGANG	1916	73 min.	15.5	11.5
KLOVNEN	1917	62 min.	18.1	13.2
HIMMELSKIBET	1918	81 min.	7.3	6
PRÆSIDENTEN	1919	71 min.	7.9	5.5
EN SKUESPILLERS KÆRLIGHED	1920	47 min.	14.5	8.6
VOR FÆLLES VEN	1921	139 min.	8.8	6.7
STORE FORVENTNINGER	1922	113 min.	8.6	6.4
NEDBRUDTE NERVER	1923	75 min.	6.7	5
HADDA PADDA	1924	83 min.	8	5.1
FRA PIAZZA DEL POPOLO	1925	107 min.	8.4	6.6
KLOVNEN	1926	139 min.	9.7	7.3
WESTER-VOV-VOV	1927	101 min.	6.2	4.3
JOKEREN	1928	102 min.	5.7	4
HØJT PAA EN KVIST	1929	97 min.	6.2	4.5
I KANTONNEMENT	1932	99 min.	6.6	4.4

The table includes information about length, since silent films did not have a set projection speed.²⁷ Individual projectionists could (and did) show a given film at very different speeds. Video versions or digitizations of the same film may also have been transferred at different speeds. Therefore, discrepancies may arise between different analyses of the same film. Barry Salt's dataset includes ASLs for 11 Danish silent films, of which two are also part of the sample examined here (see Table 3).

The ASL of KLOVNEN (1917) is given as 18.0 by Salt, while we have calculated it to be 18.1 – a negligible difference. However, Salt gives the ASL of DET HEMMELIGHEDSFULDE X as 12.0, much faster than the ASL of 16.3 recorded here. The most likely explanation is that Salt worked from a print run or transferred at a much faster speed, probably 24 frames per second, whereas the DFI's transfer was made at 16

²⁷ Kevin Brownlow, "Silent Films – What Was the Right Speed?" in *Early Cinema: Space – Frame – Narrative*, ed. Thomas Elsaesser and Adam Barker (London: BFI, 1990), 282–290; Casper Tybjerg, "About the Frame Rates," Visual essay, in Carl Th. Dreyer, *LA PASSION DE JEANNE D'ARC*, Blu-ray ed. (New York: Criterion Collection, 2017).

Table 3: From Barry Salt's dataset: cutting rates for Danish silent films.²⁸

TITLE	YEAR	ASL
EKSPEDITRICEN	1911	43.0
DE FIRE DJÆVLE	1911	21.0
DØDSSPRING TIL HEST FRA CIRKUSKUPLEN	1912	17.0
EKSPRESSENS MYSTERIUM	1914	21.0
DEN FREMMENDE	1914	16.0
DET HEMMELIGHEDSFULDE X	1914	12.0
DEN MYSTISKE FREMMENDE	1914	17.0
VERDEN'S UNDERGANG	1916	13.0
KLOVNEN	1917	18.0
BLADE AF SATANS BOG	1919	7.0
DER VAR ENGANG	1922	5.0

frames per second.²⁹ The transfer rates are unfortunately rarely given in open metadata, which is really regrettable. We would strongly recommend that archives digitizing films include transfer speeds in frames per second as part of the immediately available metadata for each title, preferably along with the length in meters of the source print. Until that happens, researchers doing this kind of research should provide the length in minutes of the video file used (as well as full information about its source), making it possible to see and compensate for differences in the speeds at which the film may have been viewed, transferred, or digitized.

As can be seen, the pace of Danish silent films increased significantly from 1910 to 1932. Both ASL and MSL decrease consistently over time. The slowest film is *VED FÆNGSLETS PORT* (*TEMPTATIONS OF THE BIG CITY*, 1911) with an ASL of 37.6 seconds and an MSL of 26.1 seconds. The film with the highest tempo in our sample is *JOKEREN* (*THE JOKER*, 1928) with an ASL of 5.7 seconds and an MSL of 4 seconds. A visualization of this data can be seen in Figure 2. As the graph makes clear, the pace increases sharply over the course of the 1910s, then levels off in the 1920s. This is clearly seen in the exponential trend line for both ASL and MSL.

²⁸ Source: Barry Salt's Average Shot Length data table, accessed October 23, 2023, http://starword.com/Data_Method/Data_Tables/data_tables.html. For *DER VAR ENGANG*, our manual test count produced an ASL of 6.7 rather than Salt's 5.0, but this is likely due to Salt working from an older print without the many rather long intertitles of the 2002 restoration.

²⁹ Claus Greffel, digital restoration and mastering technician, DFI, personal communication, June 15, 2023.

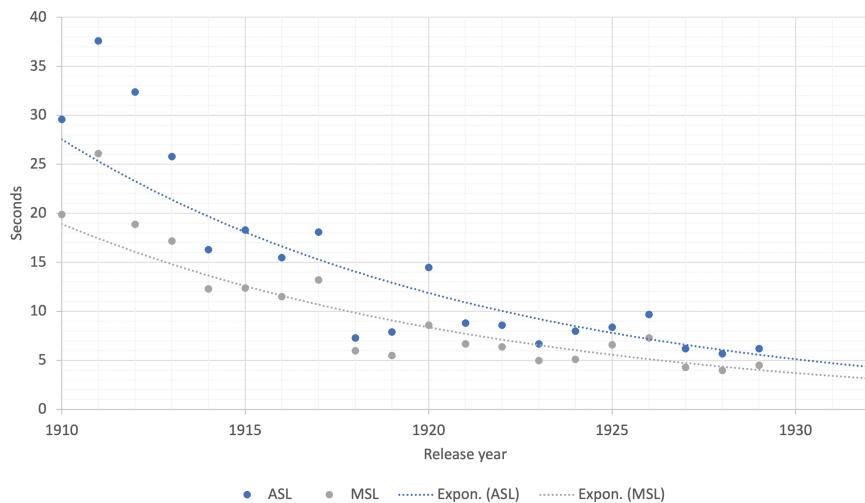


Figure 2: The pace of Danish silent films.

The overall development is clear and relatively homogeneous. There are not many deviations in the overall trend. A good example of this is A. W. Sandberg's *KLOVNEN* from 1917, which was remade in 1926 by the same director for the same production company. The 1917 version has an ASL of 18.1 while the 1926 version's ASL is 9.7. This strongly suggests that the drop in ASL between the two versions is due to the overall development towards a faster tempo in Danish film during the period. We also looked at whether any significant difference from the overall development could be observed if films made by the same directors or production companies were examined separately. No significant correlation was found, though the sample is really too small to draw conclusions. Still, the pattern seems to be an overall development, a change in stylistic norms, rather than something specific to particular filmmakers or organizations. One possible explanation could be the international adoption of scene dissection and the abandonment of the tableau style in the European cinema of the early 1920s.

We also sought to compare the data to international patterns. The lack of information about projection speeds makes it problematic to use Salt's data for silent films, but the patterns seem clear. Salt divides the silent feature period into three: 1911–1917, 1918–1923, and 1924–1929. A similar division of the sample yields the following results: from 1911 to 1917, the Danish films examined have an average ASL of 15.9 seconds (if we include the 1910 film, the average becomes 16.4). From 1918 to 1923, the ASL drops sharply to an average of 6.4. In the last period, from 1924 to 1932, the average ASL is 5.3. Table 4 compares the averages from the

sample with the average ASLs Salt calculates from his data on American and European films.³⁰

Table 4: Average ASLs in the United States, Europe, and Denmark.

	1911–1917	1918–1923	1924–1929
U.S.	9.6	6.5	5.8
Europe	15.0	8.5	6.5
Denmark	15.9	6.4	5.3

In the first period (1911–1917), the Danish films resemble the rest of Europe when it comes to tempo, with an ASL of 15.9 compared to the European average of 15.0. American films are significantly faster with an average ASL of 9.6 seconds. During the next period, the Danish films leave the European one behind and reach the American pace; while the rest of Europe has an average ASL of 8.5, Denmark and the U.S. are at 6.4 and 6.5. In the last period (1924–1929), the European average has fallen to 6.5, but the American average has fallen to 5.8 and the Danish even further to 5.3. On the face of it, it seems unlikely that Danish films of the late 1920s were actually faster paced than Hollywood pictures, and this is almost certainly an artifact of the small number of films in our sample. Still, the result is intriguing and suggests the need for further research.

Conclusion

The present study indicates that the cutting rates and thus the tempo of Danish films increased over the course of the silent period, rapidly at first, then more slowly. It suggests that Danish films caught up with the cutting rates of American ones in the late 1920s, a surprising result that calls for further research.

Perhaps more importantly, the present study can be seen as a proof of concept with respect to the automation of quantitative film style analysis. The reliability of the PySceneDetect algorithm when using our command parameters seems strong enough for generalizable conclusions. An obvious next step is to conduct the same

³⁰ Data for U.S. and Europe from Salt, *Film Style and Technology*, 3rd ed., 161, 192. The Danish titles individually listed by Salt (see n. 27) produce averages of 19.7 for 1911–1917 (9 films) and 6.0 for 1918–1923 (2 films). Salt lists no Danish films from 1924–1929.

study with a much larger sample. When DFI has finished digitizing all preserved Danish silent films, the study could be repeated with all available feature films. A large part of the work time lies in writing commands to the program and then processing the data for each individual movie, but this could easily be automated. It would not be overly difficult for someone with programming skills to write a program that automatically fed PySceneDetect with commands for each movie and then processed the data and compiled it into a spreadsheet. You could let the program run for a few days and then have all the data ready.

Of course, this would likely be a national database at first, and the entangled film history approach reminds us of the way methodological nationalism can limit our perspectives. Ideally, the database could gradually expand to include international data, perhaps allowing international patterns to emerge. Ahead of doing this kind of study, however, it would be important to secure the availability of metadata regarding the provenance of prints, the speeds at which the digital files have been transferred, and so on. To conduct a proper analysis of archival objects like the films presented on the *stumfilm.dk* site, researchers need to have a solid understanding of what these objects are and how they came to be. If our study shows anything, it shows the importance of understanding archival objects properly.

When planning courses that integrate digital methods, we think that it is important to recognize that it is very difficult to foresee all the technical and practical challenges that may arise, and that one should be ready to improvise and modify one's plans. Nevertheless, one outcome of this study has been to show that it would be feasible to do a quantitative style analysis exercise as part of the next iteration of the methods course. Video files for all the films from a particular year could be assembled and analyzed by two- or three-person teams.

Turning back to the research-integration model (Figure 1), we find that its categories do not wholly fit the work described in this article; they all make the process appear more structured and pre-planned than it tends to be in practice. Borring and Vu's work on cutting rates can fairly be described as category e.: "conduct[ing] independent research under the teacher's supervision." But arguably they made a more significant contribution by developing a set of commands for the PySceneDetect program that actually allows such research to be carried out at scale. In a sense, this is closer to c., as they "contribute[d] to research by independently completing tasks specified by researchers," even if "specified" in this case amounted to no more than Tybjerg saying, "it's cool that you want to do quantitative style analysis, but you'll have to figure out how to make it work in practice yourselves." While the model may suggest that e. is more "independent" (and therefore "better") than c., our experience indicates that the opposite was the case here.

Even as a retrospective categorization tool, the research-integration model (Figure 1) seems somewhat restrictive. It is useful when thinking about how to

integrate research with teaching, as long as the model and its categorization scheme is not applied too rigidly. From the point of view of students, the work certainly seems more rewarding if they are afforded at least some flexibility and have the opportunity to figure things out for themselves. From the instructor's point of view, having students carry out routine but time-consuming tasks like coding large datasets is certainly convenient and predictably useful. But even if giving students more loosely defined tasks and more freedom to grapple with how to resolve them may often fail to produce results that contribute to the instructor's own research projects, good students may come up with unforeseen solutions that advance projects in ways that the instructors could not have achieved on their own.

References

Adams, Brett, Chitra Dorai, and Svetha Venkatesh. "Toward Automatic Extraction of Expressive Elements from Motion Pictures: Tempo." *IEEE Transactions on Multimedia* 4, no. 4 (2002): 472–481.

Bellour, Raymond. "The Unattainable Text." *Screen* 16, no. 3 (1975): 19–27.

Biltreyest, Daniël, and Philippe Meers. "Comparative, Entangled, Parallel and 'Other' Cinema Histories. Another Reflection on the Comparative Mode within New Cinema History." *TMG Journal for Media History* 23, nos. 1–2 (2020): 1–9.

Bordwell, David. "The Power of a Research Tradition: Prospects for Progress in the Study of Film Style." *Film History* 6, no. 1 (1994): 59–79.

Bordwell, David, Janet Staiger, and Kristin Thompson. *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960*. London: Routledge, 1985.

Brownlow, Kevin. "Silent Films – What Was the Right Speed?" In *Early Cinema: Space – Frame – Narrative*, edited by Thomas Elsaesser and Adam Barker, 282–290. London: BFI, 1990.

Butler, Jeremy. "Statistical Analysis of Television Style: What Can Numbers Tell Us about TV Editing?" *Cinema Journal* 54, no. 1 (2014): 25–44.

Damsholt, Tine, and Marie Sandberg. *Af lyst eller nød: En etnologisk undersøgelse af integration mellem forskning og undervisning i praksis*. Copenhagen: University of Copenhagen, 2018. Accessed February 20, 2023. https://curis.ku.dk/ws/files/213594738/AF_LYST_ELLER_N_D_ELEKTRONISK_VERSION_2.pdf.

Elken, Mari, and Sabine Wollscheid. *The Relationship between Research and Education: Typologies and Indicators. A Literature Review*. Oslo: Nordic Institute for Studies in Innovation, Research and Education (NIFU), 2016. Accessed February 6, 2023. <https://www.nifu.no/en/publications/1351162/>.

Hagener, Malte. "Introduction: The Emergence of Film Culture." In *The Emergence of Film Culture: Knowledge Production, Institution Building, and the Fate of the Avant-Garde in Europe, 1919–1945*, edited by Malte Hagener, 1–18. New York: Berghahn Books, 2014.

Healey, Mick. "Linking Research and Teaching: Exploring Disciplinary Spaces and the Role of Inquiry-Based Learning." In *Reshaping the University: New Relationships Between Research, Scholarship and Teaching*, edited by Ronald Barnett, 67–78. Maidenhead: McGraw-Hill Education, 2005.

Kokeš, Radomír D. "Norms, Forms and Roles: Notes on the Concept of Norm (Not Just) in Neoformalist Poetics of Cinema." *Panoptikum*, no. 22 (2019): 52–78.

Kovács, András Bálint. "Shot Scale Distribution: An Authorial Fingerprint or a Cognitive Pattern?" *Projections* 8, no. 2 (2014): 50–70.

Redfern, Nick. "The Average Shot Length as a Statistic of Film Style." *Cinemetrics: Film Statistics: Give and Take*. n.d. Accessed April 11, 2024. https://web.archive.org/web/20230328210123/http://cine-metrics.lv/fsgt_q1b.php.

Salt, Barry. *Film Style and Technology: History and Analysis*. 3rd ed. London: Starword, 2009.

Salt, Barry. *Moving into Pictures: More on Film History, Style, and Analysis*. London: Starword, 2006.

Salt, Barry. *Film Style and Technology: History and Analysis*. 2nd ed. London: Starword, 1992.

Salt, Barry. *Film Style and Technology: History and Analysis*. London: Starword, 1983.

Salt, Barry. "Statistical Style Analysis of Motion Pictures." *Film Quarterly* 28, no. 1 (1974): 13–22.

Salt, Barry. "[Data Method] Statistical Style Analysis." n.d. Accessed October 23, 2023. http://www.star-word.com/Data_Method/data_method.html.

Savardi, Mattia, András Bálint Kovács, Alberto Signoroni, and Sergio Benini. "CineScale: A Dataset of Cinematic Shot Scale in Movies." *Data in Brief* 36 (2021): 107002. <https://doi.org/10.1016/j.dib.2021.107002>.

Schaefer, Eric. "The Problem with Sexploitation Movies." *Iluminace* 24, no. 3 (2012): 148–152.

Seppälä, Jaakko. "Finnish Film Style in the Silent Era." In *Finnish Cinema: A Transnational Enterprise*, edited by Henry Bacon, 51–80. London: Palgrave Macmillan UK, 2016.

Trowler, Paul, and Terry Wareham. "Tribes, Territories, Research and Teaching: Enhancing the Teaching-Research Nexus." January 1, 2008. Working Paper. Accessed October 23, 2023. https://www.researchgate.net/publication/252423791_Tribes_territories_research_and_teaching_Enhancing_the_teaching-research_nexus.

Tsivian, Yuri. "Cinemetrics: Part of the Humanities' Cyberinfrastructure." In *Digital Tools in Media Studies*, edited by Michael Ross, Manfred Grauer, and Bernd Freisleben, 93–100. Bielefeld: Transcript, 2009.

Tybjerg, Casper. "Danish-German Cinematic Interconnections and the Prospects of an Entangled Film Historiography." In *Danish and German Silent Cinema, 1910–1930: Towards a Common Film Culture*, edited by Lars-Martin Sørensen and Casper Tybjerg, 24–50. Edinburgh: Edinburgh University Press, 2023.

Tybjerg, Casper. "The European Principle: Art and Border Crossings in Carl Theodor Dreyer's Career." In *A History of Danish Cinema*, edited by Isak Thorsen, C. Claire Thomson, and Pei-Sze Chow, 41–50. Edinburgh: Edinburgh University Press, 2021.

Tybjerg, Casper. "En introduktion til kvantitativ æstetisk filmanalyse i praksis." *Kosmorama* (2021).

Tybjerg, Casper. "About the Frame Rates." Visual essay. In *Carl Th. Dreyer, LA PASSION DE JEANNE D'ARC*, Blu-ray ed. New York: Criterion Collection, 2017.

Werner, Michael, and Bénédicte Zimmermann. "Beyond Comparison: Histoire Croisée and the Challenge of Reflexivity." *History and Theory* 45, no. 1 (2006): 30–50.

