Article

Automation in Sports Reporting: Strategies of Data Providers, Software Providers, and Media Outlets

Jessica Kunert

Institute for Journalism and Communication Studies, University of Hamburg, 20146 Hamburg, Germany;
E-Mail: jessica.kunert@uni-hamburg.de

Abstract
This study examines how algorithmic processing affects structures and practices in sports journalism in Germany. A multi-level perspective is used to determine which strategies data providers, software providers, and media outlets use to develop automated reporting, which compiles perspectives across the entire line of news production. The results of 11 in-depth interviews show that non-journalistic actors are vital partners in the news production process, as all actors work together in data handling, training, and software development. Moreover, automation can generate additional content such as match and historical coverage to help address shortfalls in capacity. However, given the business case for automation, amateur football (soccer) is currently the only viable candidate for its use. Many actors involved in the process argue that automated content is an added value for their readers, but claim that content quality has to be put before quantity. This means that some media outlets edit automated articles to increase the quality of their sports journalism, but that this is done only on a small scale. Media outlets do not perceive their roles to be changing, but see automation as a helpful tool that complements their work; a few use automatically created articles as a baseline for in-depth reporting. Moreover, the so-called 'meta-writer' has not become a reality yet, as data-processing and news writing are still kept separate. This article sheds new light on the use of automation in the sports beat, highlighting the growing role of non-journalistic actors in the news production process.

1. Introduction

This article examines how algorithmic processing affects institutionalised structures and practices in sports journalism in Germany. It assesses the strategies that data providers, software providers, and media outlets have devised regarding the development of automated reporting. This study focuses on how German sports journalism manages the balancing act between being a data-rich (and thus easily 'automatable') beat on the one hand, and the impact of economic constraints and journalists' perceptions of 'good' sports coverage on the other hand.

The sports beat is especially well suited to automation due to the extensive structured data available and its routine events (Galily, 2018; Graefe, 2016). While simple automation processes have been used for weather and earnings reports for quite some time (Graefe, 2016, pp. 19–20), sports recaps (namely baseball) were the first application for automation by the major natural language generation providers Automated Insights and Narrative Science (Allen, Templon, McNally, Birnbaum, & Hammond, 2010; Graefe, 2016, p. 17). While academics and journalists often emphasize that sports reporting requires (human) creativity and emotions (Horky

& Stelzner, 2013), which sets it apart from other data-intensive beats such as finance, media outlets have successfully used automation, especially in amateur sports reporting (e.g., WashPostPR, 2017). In this article, I explore how German sports reporting uses automation, and identify lessons for the future.

I conducted guideline-based interviews to investigate the perceptions of sports data providers, software providers, and media outlets regarding automation. This approach examines views throughout the news production process, as the study participants include journalists as well as non-journalistic actors who may influence the (re-)definition of journalistic self-conceptions, and whose role in the news process might change with automation (see Carlson, 2015, p. 417). I demonstrate how journalists and these "journalistic strangers" (Holton & Belair-Gagnon, 2018) interact with each other, and how they define their relationship in automating amateur sports coverage.

This study focuses on the *production* of automated news rather than *audience perceptions*. The latter has been studied extensively in general (e.g., van der Kaa & Krahmer, 2014) as well as in the case of sports (Yao, Salmon, & Tandoc, 2018). This research has found that readers are ambivalent in their perceptions of automated news, ranging from being unable to differentiate between automatically and human-written material (e.g., Clerwall, 2014) to distinctly preferring either machine-written (e.g., Jung, Song, Kim, Im, & Oh, 2017) or human-written articles (e.g., Graefe, Haim, Haarmann, & Brosius, 2018). Although I did not survey the audience, it stills plays a major role for both journalists and journalistic strangers, as both anticipate audience reactions: The readers motivated initial decisions to automate, and their feedback sparks efforts to adjust the software settings for future articles.

The article begins with a literature review of automated journalism and its use in sports reporting. It then presents the methods and discusses the results. The article concludes with a discussion of the use of automation in sports reporting and areas for further research.

2. Literature Review

2.1. Automated Journalism in the Newsroom

Algorithmic processes are used to select, create and distribute news content. Automated content creation, using natural language generation, is defined as "algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming" (Carlson, 2015, p. 417). In principle, this definition also applies to the automatic creation of news videos and visualisations (e.g., Alhalaseh et al., 2015).

The automated composition of news texts proceeds in three steps (Dörr, 2016, pp. 703–704). The first step is inputting structured data. In sports, this is usually match day data, such as the number of goals, attendance, goal

scorers, or yellow cards. The second step is the "initial programming" (Carlson, 2015, p. 417): A journalist creates a template that is later filled with data (Graefe, 2016, p. 17). To add variability to the news texts, synonyms and 'if-then-else' branches, for example decision rules for the algorithms to decide whether the match was an 'easy win' or a 'bad loss' for a team, can be programmed in. The last step is the output, i.e., the articles generated from the data.

Automated journalism lends itself to data-intensive beats, such as crime, finance and sports (van Dalen, 2012). News agencies and individual news outlets increasingly use automated content creation (e.g., Fanta, 2017), such as the *Washington Post*'s Olympics coverage (Rojas Torrijos, 2019).

2.2. Changes in the Journalistic Profession through Automation and Journalistic Strangers

Automation processes have been found to heavily impact the journalistic profession, for example working routines and self-conceptions. Journalists generally feel optimistic about automated news reporting, as it frees them up from conducting grunt work and lets them report on in-depth issues (Wu, Tandoc, & Salmon, 2019, p. 1451). They are adapting to these changes by "re-examin(ing) their core skills" (van Dalen, 2012, p. 649); automation is not perceived as a threat, but rather as a form of hybrid work in which journalists maintain control over news generation (Graefe, 2016; Thurman, Dörr, & Kunert, 2017; Wu et al., 2019).

Carlson (2015, p. 423) finds that journalists either augment their core skills by becoming "meta-writers" who acquire computational skills and thinking (Gynnild, 2014; Lindén, 2017a), or outsource these skills by dividing tasks with programmers (Lindén, 2017b, p. 132). Carlson (2015, p. 417) classifies this negotiation of journalistic roles as a "technological drama," in which journalists have to redefine their roles while "technologists" find their way into newsrooms. Therefore, research on news automation must examine journalists as well as non-journalistic actors such as data and software providers.

Holton and Belair-Gagnon (2018) classify these non-journalistic actors as 'journalistic strangers,' who shape journalistic innovation and challenge the profession to evolve by introducing new perspectives and tools such as web analytics and programming. In the case of article automation, the journalistic strangers are mainly data and software providers, who are external to the news outlets, yet have vital tasks. They enter the news production process as "implicit interlopers" as they drive innovation by offering "potential contributions and improvements" (Holton & Belair-Gagnon, 2018, p. 74) to journalism, i.e., enhanced content and connection to audiences. Unlike "explicit interlopers," such as independently run blogs or the disclosure platform WikiLeaks, implicit interlopers do not challenge journalistic authority (Eldridge, 2019). They are generally welcomed to newsrooms, even

though this may be the result of a longer negotiation process of journalistic values and role perceptions (Chua & Duffy, 2019), which is also the case for automating content generation (Thurman et al., 2017).

For article automation, implicit interlopers are the main drivers of innovation, as they contribute the necessary technological aspects such as software platforms and data. These types of interlopers are not mere service providers working at the periphery of news production; they are powerful actors at the very heart of it (Ahva, 2019, p. 124; Tandoc, 2019, p. 141). This is surprising, because in the case of software providers, journalistic articles are not their main product, but rather "product descriptions, portfolio analyses, or patient summaries in hospitals" (Graefe, 2016, p. 19). Despite their central role as providers of the main infrastructures for automated sports content, readers are often not aware of their existence, as content may not be labelled as automated, but simply contain the brand of the news outlet (Braun, 2014, pp. 124–125).

Individual journalists' perceptions of these journalistic strangers seem to depend on the news beat. For instance, Thurman et al. (2017) found that finance journalists and editors felt positively about automation, whereas sports journalists vehemently opposed it. These discrepancies warrant a closer look at the sports beat.

2.3. Impact on the Sports Beat

The sports beat is a good candidate for article automation since it has routine events that can be covered instantaneously and at high volume using this technology (Galily, 2018; Graefe, 2016). Moreover, extensive structured datasets are collected and available, for example via OPTA (2020). For popular sports, such as football [soccer] in Germany (Allensbacher Markt und Werbeträgeranalyse, 2019), data are also collected for local events such as regional leagues. Without automation, these leagues would not be covered in the media due to a lack of personnel and resources (van Dalen, 2012, p. 652), which means that this type of coverage finds small, yet likely loyal, audiences (Carlson, 2015, p. 426). Moreover, automation allows articles to be sent out quickly, giving journalists time to work on in-depth reporting (Lindén, 2017a; Thurman et al., 2017, pp. 1249–1250), such as personal stories of coaches and players. The use of automation may thus be regarded as a service to loyal readers as well as a sound business investment for extending sports coverage (Boyle, 2006), achieved with the help of implicit interlopers, especially software providers. However, commercially used software templates only cover a limited number of sports (e.g., football in the case of Retresco's textengine), although non-commercial prototypes have been created, for example, for cricket (Gunasiri & Jayaratne, 2019) and ice hockey (Kanerva, Rönnqvist, Kekki, Salakoski, & Ginter, 2019).

Despite this potential, sports coverage has some characteristics that may hinder article automation and thus the involvement of implicit interlopers, as they profoundly change traditional processes of sports reporting. For instance, when reporting on a match, the context beyond the match statistics such as goals or yellow cards must be taken into account. However, such in-depth data that would help journalists who were not present adequately describe a match are often unavailable, especially for amateur leagues (see Lindén, 2017b). As a sports reporter said in Thurman et al.'s (2017, pp. 1247–1248) study, "the data might only present '10 percent of the story.'" He thus found a crucial part of reporting missing when asked to assess the quality of automatically created articles, and felt that these articles were "throw away, repetitive, not particularly interesting" (Thurman et al., 2017, p. 1248).

Another sports journalist in Thurman et al.'s (2017) study claimed that automation can help discover stories in the first place, especially when the data defy what he called "accepted wisdom" in sports (Thurman et al., 2017, p. 1248; Wu et al., 2019). This resembles the notion of a "hybrid collective" of an enhanced human–machine connection (Primo & Zago, 2015), yet shows that sports journalists in Thurman et al.'s (2017) study saw automation as merely a tool in the newsroom, but not one that writes publishable news stories.

Sports coverage is also associated with emotion and storytelling (Horky & Stelzner, 2013), which has not yet found its way into factual automated reports (Yao et al., 2018). However, it can be questioned whether results-led reporting requires an individual journalistic voice, as such coverage is usually done in fixed templates, and reads similarly to automatically created output (see Carlson, 2015, p. 425).

In addition to style issues, sports journalists have been found in early studies to be generally weary of automation technology. In describing journalists' reactions to StatSheet, a now defunct sports statistics website, van Dalen (2012) found that sports journalists felt they were competing with the underlying algorithm (p. 652), a sentiment echoed by Carlson (2015, pp. 422–423). The current dominant notion of a "hybrid collective" (Primo & Zago, 2015) between journalists and algorithms emphasises separating automatised routine tasks done by machines from skills that require humans (for the distinction between "low" and "high" journalism, see also Lindén, 2017a, p. 65). Instead of competing with the algorithm, van Dalen (2012, p. 653) advises sports journalists to "concentrate on their own strengths," and on "the human advantage of telling the story in the sporting world" (Galily, 2018, p. 50).

I aim to disentangle the relationship between the sports beat and automation, represented by the involvement of implicit interlopers, with a holistic view of the whole news production line, consisting of data providers, software providers, and news outlets. I focus on the case study of Germany, a big sports market with an emphasis on football:

RQ: How do data providers, software providers, and news outlets perceive and work with article automation in German sports reporting?

3. Methodology

I conducted interviews on the use and perceptions of article automation in German sports journalism and assessed them using qualitative content analysis (Schreier, 2012). I incorporated views from the entire production line of automated journalism, thus examining how journalists and implicit interlopers, namely data and software providers, assess their relationship and working processes (Ahva, 2019, p. 125; Holton & Belair-Gagnon, 2018, p. 76). Thus, I interviewed two providers of sports data, three software providers, five news outlets, and representatives of the German Football Association (DFB) between March 2018 and April 2020 (Table 1).

Starting with the interlopers, OPTA and Sportec Solutions are the most prominent data providers for sports data in Germany. For the software providers, AX Semantics and Retresco were interviewed. Retresco provides the 'textengine' used by many media outlets, including those interviewed for this study. ReportExpress is an app that provides software for reporting on amateur football matches. It is classified as a software provider since it focuses on producing content and is not a news outlet (its articles are not published in the media, but can be downloaded by its users). To account for a range of views from news outlets, I interviewed outlets with different strategies for automation in sports reporting: the online portal *OVB24*'s *Beinschuss* (in German: nutmeg or tunnel, a sports term), which covers amateur football in Upper Bavaria; *Der Spiegel*, which uses large volumes of data and some automation for sports news projects, but does not plan on implementing automation

to create text; *FussiFreunde* (in German: football friends), which covers amateur football in Hamburg; *Nordbayern Amateure* (in German: Northern Bavarian amateurs), which reports on amateur football in Middle Franconia in Bavaria; and *Sportbuzzer*, which previously used automation for their amateur football reporting in the Bremen area but has since stopped. *Fussball.de* (in German: football), run by the German Football Association, is the biggest provider of automated reporting on German amateur football. The German Football Association presents a special case, as it is an implicit interloper that covers amateur matches similarly to media outlets. However, the German Football Association claims that they have no journalistic aspirations and do not challenge journalistic authority, as their sole aim is to provide a service for their amateur leagues.

The first few interviews with data and software providers conducted for this research demonstrated that automated journalism is viable only for football in Germany, which is why the sample focuses on this sport. The interviews averaged one hour (range: 32 minutes to 2 hours). They were conducted in person (2), via telephone (6), via video chat (2), and via e-mail (1).

This sample offers an adequate overview of German news outlets' experiences with automation in sports coverage. Only a few software companies offer text automation for sports; Retresco is the main provider of this service. All five news outlets I interviewed use (or previously used) Retresco's textengine, and their accounts of the software differ only in details. Thus, other news outlets would likely report similar experiences. However, I did not interview outlets that use AX Semantic's platform. Moreover, as media outlets often do not label automated articles, identifying possible interviewees was not easy, which might mean that strategies regarding automation other than the ones presented were missed. Lastly, there

Table 1. Interviewees.

Type	Company/Outlet	Name	Position
Data provider	OPTA	Sven Tröster	Chief editor, Germany and Switzerland
Data provider	Sportec Solutions	Holger Rahlfs	Head of product
Software provider	AX Semantics	Frank Feulner	Chief visionary officer
Software provider	Retresco	Johannes Sommer	CEO
Software provider	ReportExpress	Gabriel Brass	CEO
Media outlet	Beinschuss	Martin Vodermair	Editor in chief
Media outlet	Der Spiegel	Patrick Stotz	Data journalist
Media outlet	FussiFreunde	Dennis Kormanjos	Editor
Media outlet	Nordbayern Amateure	Bastian Eberle	Team leader, amateur football
Media outlet	Sportbuzzer	Steffen von Deetzen	Product manager for the introduction of automation
German Football Association (DFB)	Fussball.de	Dr. Frank Biendara Anja Vianden	Managing director, DFB GmbH Project manager, robot journalism

is only one media outlet in the sample for which automation was not successful, meaning that there is not a balance of critical voices.

Interview questions fall into five categories: 1) workings of automation (viability for different sports, data collection systems, software systems, introduction of automation at news outlets, successes and obstacles associated with automation); 2) working with automated articles (article quality, types of customers of data and software providers, possible extension of the use of automation, editing practices at news outlets, reader reactions, practice of labelling articles); 3) working with the software (training of the software, journalists' technological skills, potential and limits of the software, working together across the journalistic production line); 4) data availability and maintenance (data depth and sources, maintenance, reaction to errors); 5) and change in sports reporting through automation (journalistic self-perceptions, changes in the profession, dangers and limits of automation for journalism, possible advantages of human journalists). The full interview guide is available in the Supplementary File.

The coding was structured by deductive categories derived from the literature on automation. These codes concerned the reception of automation in the respective newsroom and its impact on journalistic self-perceptions, the positive impacts of automation (e.g., saving time, exploring new story formats) as well as the negative aspects (e.g., no context, low quality), data availability, business considerations (e.g., cost), and user/reader feedback. New categories and codes were created inductively, which concerned mostly the interloper–media outlet relationships or were specific to sports (Schreier, 2012). These included working together across the production line, software development, possible business cases for sports other than football, the heightened importance of putting automated sports content into context, and, in the case of media outlets, creating a bond with local readers.

4. Results and Discussion

I identified four main themes: 'Actors: Working Together,' 'Content: Additional Coverage,' 'Money: Football Only,' and 'Quality: The Art of (Human) Sports Reporting.' These themes demonstrate the deeply intertwined relationship between journalists and implicit interlopers (Holton & Belair-Gagnon, 2018), showing their common working routines, the impact of the innovations that the interlopers contribute to newsrooms, how these innovations are affected by economic considerations, and how the innovations are negotiated through the lens of journalistic self-conceptions.

4.1. Actors: Working Together

The implicit interlopers are deeply involved in the working routines of automated journalism. Their work in data

handling, the training of journalists and newsrooms, and software development is a prerequisite for automated journalism. However, the interlopers' work is mostly separate from editorial work, and journalists are only involved in the technological processes to a limited degree.

4.1.1. Data Handling

As media outlets often have limited experience with data handling, data providers OPTA and Sportec Solutions assist their customers in selecting valid data for their reporting. Sportec Solutions highlight making data "consultable" for their customers, and emphasise that "the whole processes of refining match data according to the customer's wishes stays in our [Sportec Solutions'] hands." OPTA also advise customers on how to implement data. Even though both companies mainly work with professional football data, Sportec Solutions' assessment also applies to amateur data: "It's more economic if specialists do that [data handling], as it is not the customer's core business to work with data and operate a database." Thus, there is a clear distinction between data collection and preparation on the one hand, and data usage by media outlets on the other. Moreover, software providers often act as brokers between data providers and media outlets. However, there are exceptions, such as the Bavarian Football Association, which brokers its own data without intermediaries (*Beinschuss*, *Nordbayern Amateure*). This approach to data handling echoes Lindén's (2017b) call to divide tasks between programmers and journalists, and shows that data providers offer an innovative service that newsrooms cannot deliver themselves.

4.1.2. Training of Journalists and Newsrooms

As data handling is largely out of the media outlets' hands, so is tinkering with the automation software. Both AX Semantics and Retresco offer their customers basic training in using their software; the aim is for newsrooms to manage the software independently. However, this training does not include teaching customers how to fundamentally reprogram the software. According to Retresco, the underlying algorithm is complicated, which is why their software "looks much like [Microsoft] Excel." Thus, journalists may only input basic "rules for data interpretation," such as "synonyms for club names in singular and plural" (*Nordbayern Amateure*) or "team nicknames" (*Beinschuss*), so that the "texts don't always look the same" (*FussiFreunde*). *ReportExpress* also simplify working with their app by offering plug-and-play reporting, in which users "simply tap action and decision buttons" that, for example, describe the weather or the speed of the match. These limited options for working with technology contradict the prediction that journalists will acquire computational skills and thinking (Carlson, 2015; Gynnild, 2014; Lindén, 2017a). In fact, according to *Beinschuss* and *FussiFreunde*, to operate Retresco's tex-

tengine, one "doesn't need technological skills at all," as "just the index finger" (*Nordbayern Amateure*) suffices. These accounts differ heavily from the interviewee from *Der Spiegel*, who works with raw data and advocates having advanced software skills. He thus resembles the "meta-writer," as he distinguishes between "data journalists and traditional editors." However, journalists working with automation software do not program it themselves; they rely on a pre-set product.

4.1.3. Software Development

The software providers state that they continually improve the functionality and usability of their software. However, Retresco questions if more development was necessary for its language quality, asking "whether the football texts we now have are good enough." This sentiment echoes Carlson's (2015, pp. 424–425) finding that journalists generally claim that automated texts are of acceptable quality, as they mimic a "wire-service-style news story" that has no individual journalistic voice anyway. However, Retresco's assessment demonstrates that data and software providers introduce a new set of rules to sports journalism, which is guided by the opportunities the technologies provide as well as their limits, in this case the language and structural capabilities of the software (Tandoc, 2019, p. 141). Moreover, these capabilities have to be extensively defined. *Fussball.de* say that it took "a lot of manual work" to program the software and make it "talk sports" by, for example, adding "metaphors and idioms." For this programming, *fussball.de* employ a dedicated project team of "sports journalists, machine learning specialists, software specialists and linguists" that set up the software in the first place and continues to work on modifications. Thus, implicit interlopers are not only welcomed, but are invited to the news process in order to define the rules and semantics of automated sports coverage (Graefe, 2016, p. 18). However, in this case, the German Football Association is an implicit interloper itself, working with other interlopers to create automated news.

Nordbayern Amateure report that software providers respond to requests from media outlets to add further functionalities. ReportExpress also receive "wish lists" from their customers, mainly amateur clubs, indicating that users are involved to some degree in software development. Interviewees' descriptions of the audience feedback on the automatically created articles mirrors the ambivalent findings of previous research (e.g., Clerwall, 2014; Yao et al., 2018). For instance, *Beinschuss* state that a club was not happy to be repeatedly called the "shooting gallery of the league," and asked for a change in programming, which *Beinschuss* did by eliminating this phrase from their copy of the software. *Fussball.de* also reacted to amateur clubs' feedback and adapted the software accordingly. In addition, *fussball.de* also allows team managers on both sides to edit the automatically created articles. Thus,

there could be three reports of a single match, created by "the software, the home team, and the away team." Around 80% of teams do not use this option, which supports *fussball.de*'s claim that their article quality is "very good.''

Other reader feedback claims that the automatically written articles were "not exciting" (*Sportbuzzer*), and generated fewer clicks than usual. However, most interviewees report positive feedback from their readers (Retresco, *fussball.de*, *Nordbayern Amateure*). *FussiFreunde* say their readers perceive automation to be "normal," as they receive very little feedback anymore. Retresco and *Nordbayern Amateure* also state that readers might not even notice they are reading automated content since it is often not labelled as such (Braun, 2014). In summary, readers' assessments are ambivalent, and some interviewees actively work on the quality of the articles to address readers' feedback.

4.2. Content: Additional Coverage

The implicit interlopers have produced innovative tools that introduce new opportunities to cover previously underreported amateur football and for in-depth statistical and historical reporting. These types of reporting are not innovative *per se*, but would require extensive manual work to perform without automation.

4.2.1. Solving the Capacity Problem

All interviewees confirm that automation helps media outlets increase the quantity of news while working with limited resources (Graefe, 2016). They agree that automated journalism increases coverage in areas where there was hardly any before, in this case amateur football (see van Dalen, 2012). AX Semantics and Retresco emphasise the possibility of serving large regions with few journalists, as:

> Automated reporting is in some cases the only reason why there is a full report instead of a simple table or a summary of several matches….Fans appreciate that we report on their village team in a way that resembles the Bundesliga. (AX Semantics)

Sportbuzzer, *FussiFreunde* and *Nordbayern Amateure* agree, as they lack the time and personnel to cover every match manually, especially concerning pre-match reports and team statistics.

All interviewees mention that their readers benefit from automation. *Beinschuss* emphasise that automation offers "additional value" for their readers, even though the articles may be written based on "rudimentary data." *Nordbayern Amateure* add that amateur coverage is a "service for lower leagues" that they could not deliver without automation. The implicit interlopers, the data and software providers, are thus welcomed into newsrooms to help media outlets address economic chal-

lenges (Tandoc, 2019, p. 141). *Fussball.de* also emphasise the notion of offering a "service"; the representatives state that they want to give amateur football "more attention, as they are hardly featured in the media," and to do so "nationwide" for all matches, including youth teams. These examples show that implicit interlopers may help in "cultivating stronger press–public relationships" (Holton & Belair-Gagnon, 2018, p. 74), as the expanded coverage brings outlets and audiences closer together, especially where there was no coverage before. *Der Spiegel*, albeit without using text automation, add that they use structured data to create value for their audience with "timeless" stories, such as their "Fan Atlas," which gives a data-driven overview of football fan friendships and rivalries in Germany.

AX Semantics see automation as a starting point for in-depth reporting, as newsrooms "can deliberately select matches which have high news value, and then report on them again manually." *Beinschuss* and *FussiFreunde* agree on the benefits of utilising automation to find a "news angle" to report on manually (Allen et al., 2010; Thurman et al., 2017), and do not view automation as a threat to the journalistic profession, but rather as a helpful tool that summarises data in an efficient manner.

4.2.2. Statistics and Historical Reporting

Automation helps detect historical statistics, which is efficient and saves time. For instance, an algorithm can assess the team's performance in the latest match to determine "how teams and individual players performed [during the season]" (AX Semantics). Sportec Solutions add that historical data offer opportunities for deeper storytelling through "historical contexts, peculiarities, records" or "form curves" (*FussiFreunde*), which may help to verify "perceived truths" (*Der Spiegel*). The speed with which historical data can be generated is also valuable for professional sports (*Nordbayern Amateure*). Sportec Solutions argue that even though professional football is reported on anyway and thus automation does not fill the same gap as in amateur football, automation may provide "original texts," so that media outlets do not have to use "news agency texts which might not differ across customers." Thus, the innovative work of implicit interlopers allows media outlets to work with new formats and engage in forms of storytelling that would have taken extensive manual work otherwise (Lindén, 2017a). Through this coverage, media outlets can form deeper relationships with their (sometimes very small) audiences (Graefe, 2016, p. 26).

4.3. Money: Football Only

Although the interviewees repeatedly mention the notion of offering a service to their readers, automation in the sports beat hinges on business decisions. All but one interviewee claim that automation in Germany is eco-

nomically viable only for amateur football; it is less applicable to other sports.

Retresco state that the focus on football is a business decision: "Everything apart from football would be more of a hobby [to customers]," as other sports would not bring sufficient financial revenue. *Der Spiegel* does not work with text automation, but say that "readership interest" is crucial for "very complex and time-consuming data stories." Therefore, football is the subject of most sports reporting "even if data might be available [for other sports]."

OPTA and Sportec Solutions both emphasise that data collection is in principle possible for many sports, as is automated news creation, especially "for team sports similar to football, such as handball, basketball or hockey" (ReportExpress). Other sports in which "style" (AX Semantics) matters might be more difficult, as well as sports "deviating from shooting goals or hoops, such as curling, sailing or fencing" (ReportExpress).

Only *Beinschuss* aim to use automation for other sports, namely winter sports. The *Beinschuss* representative says that winter sports face similar capacity and personnel problems, but stresses that this endeavour depends on data availability. Nevertheless, all interviewees acknowledge that using basic data from all sports would in principle be of interest to the media, for example for tennis (*Nordbayern Amateure*). Therefore, even though structured data might be available, and it might be possible to develop templates (e.g., Allen et al., 2010; Gunasiri & Jayaratne, 2019), these developments hinge on business decisions. Such decisions also affect professional football in terms of historical reporting, as collecting more than the basics would be "too expensive" with "probably little gain" (OPTA). Thus, the challenge does not lie with the implicit interlopers here, as they can in principle offer more data and templates; it instead lies with the news outlets, which have to find economically viable business strategies and negotiate the trade-off between offering a service to their readers and sufficient financial revenue.

4.4. Quality: The Art of (Human) Sports Reporting

The trade-off between offering a service to readers and economic considerations also applies to the quality of automatically generated articles. While media outlets generally appreciate the increase in the *quantity* of articles, some have *quality* concerns regarding their formulaic structure and language, as well as incomplete content due to limited data availability. The outlets employ a range of strategies to deal with these issues, such as editing the automated articles.

4.4.1. Quality Before Quantity

Quality concerns especially relate to the language and structure of automatically created articles. For instance, *Sportbuzzer* was the only outlet in the sample to decide

not to use automated reporting after a test run. It claims that even though human-written match reports often "resemble automated reporting," as many of them are written using a template (van Dalen, 2012), the quality of the automated texts was "not what we had imagined." As the automatically created articles had to be heavily edited for language, the outlet claims that automation did not save its employees any time. *Sportbuzzer*'s representative emphasises that they do not want to offer their readers this type of coverage, as sports reporting is about "bringing in something "real," such as a "coach's comment."

FussiFreunde agree, and edit their articles, for example by adding quotes and further information, which teams even supply without being asked if a reporter cannot attend the match. *FussiFreunde* stress that they have a strong bond with their readers, and thus aim to add further information and "[bring] emotion to the facts" through exclusive reports, direct quotes, and storytelling (Horky & Stelzner, 2013). Thus, *FussiFreunde* still favours human reporting over automation, as they emphasise the "human advantage" (Galily, 2018; Graefe, 2016, p. 11), whereas automated articles serve as more of a baseline. Thus, these outlets recognise the advantages of automation, yet assert that journalists still oversee the process (Wu et al., 2019, p. 1454).

Nordbayern Amateure agree with putting "quality before quantity." They edit every automatically created text, for example when these are "too judgmental," and add further information such as "direct quotes." This kind of quality control—or, as Graefe (2016, p. 35) describes it, "man–machine marriage"—requires personnel, and thus *Nordbayern Amateure* contradict the notion that automation reduces jobs in journalism (Lindén, 2017a). The representative states the opposite: "Theoretically, we could even employ more people who operate the textengine."

Beinschuss also cover "special events" manually, which automation gives them the time to do. They report that automation is a way for their journalists to "concentrate on their own, good stories, which they most certainly would rather put on their CV than pre-match or post-match reports." However, they acknowledge that automation is the only way to offer "a short report on every match," which they could not achieve manually. Trying to achieve this goal before automation led to reports of "relatively low quality" due to personnel and time constraints.

While some quality considerations might be a programming issue—AX Semantics claim that a human may indeed feed "humour, irony, sarcasm, emotional value" into the software—automation also faces limits in specific forms of coverage. Retresco report that programming tournament modes would be too complex; however, these are a vital part of the sport. In addition to programming limitations, quality is also affected by limited data availability.

4.4.2. Data Limitations

The availability of structured data is the most important prerequisite for automated journalism (Graefe, 2016). However, even though Retresco claim that "data availability might be the most important thing," amateur football in particular suffers from limited data availability. As rarely more than the basics, such as "lineups and goal scorers" (Sportec Solutions), are available, automated articles rely on relatively few data points. *Beinschuss* concede that "the machine doesn't recognise special events," so the course of the match might not be adequately reflected in the data (*Sportbuzzer*). Sportec Solutions agree: "The data only show a piece of the whole football or Bundesliga experience. When a player announces the end of his career, receives flowers and has tears in his eyes, then we can't see that in the data." Again, emotions play a big role for the interviewees, echoing one of the sports journalists in Thurman et al.'s (2017) study who complains that automated texts do not offer relevant context. Moreover, aborted or cancelled matches are often not noted in the data, which may lead to false reports (*Beinschuss*), which is one reason why *Nordbayern Amateure* edit every automated text before publication. *Beinschuss* and *Nordbayern Amateure* mention that since limited data are available from the regional Bavarian football association, additional sources have to be brought in for Bavarian outlets.

Apart from the issues arising when aborted or cancelled matches are not reported, most interviewees say their data are errorless, even though they have to be examined thoroughly for "completeness, potential outliers, mistakes" (*Der Spiegel*). Thus, "databases have to be maintained" (Retresco) continuously so that readers do not find (and report) mistakes (*Sportbuzzer*).

Some of these problems might be cured with further software development. According to AX Semantics, their software offers to include context by adding a second data stream on weather or press releases about players. ReportExpress do not use third-party data streams at all, but ask their users to enter their assessments of the match's atmosphere and other surroundings along with factual data such as goals and yellow cards. In terms of mistakes in the data, ReportExpress do not scan for data mistakes, as users are responsible for their own data entry, and say that it is their sole responsibility to care for the "error-free running of the app." AX Semantics assure that the software can detect obvious mistakes, such as reporting "21 [goals] instead of two."

5. Conclusions and Outlook

This study shows how data providers, software providers, and news outlets perceive and work with article automation in German sports reporting. I demonstrate that the working processes of these actors are deeply intertwined, which can be seen, for instance, in how data and software providers deliver the prerequisites of article au-

tomation, without which this type of journalism cannot be done. Moreover, software providers and media outlets jointly adjust the software according to the individual outlet's needs based on ongoing feedback that includes the audience's reception.

These examples show that "journalistic strangers" (Holton & Belair-Gagnon, 2018) are vital in the automated news production process. They bring about an innovative form of reporting that helps media outlets scale up the quantity of their amateur football reporting, allowing outlets to offer a service to the lower leagues in their region. These implicit interlopers, namely the data and software providers, thus bring new opportunities to news outlets without challenging journalistic authority. Their central role implies a power shift, as journalists depend on them during the news creation process, and work with them to produce articles (Ahva, 2019; Primo & Zago, 2015; Tandoc, 2019). The ongoing negotiation between journalists and interlopers manifests itself in the tension between offering a service to readers while adhering to the manifests of sports journalism, which includes emotion and storytelling (Horky & Stelzner, 2013). This negotiation mainly applies to software providers, yet data providers are included in this conversation regarding data quality concerns.

Despite the vital role of the implicit interlopers, media outlets do not view their role as diminished or feel they are in competition with the automation software (van Dalen, 2012). As *FussiFreunde* explain, "nothing changes at all for us journalists," as they still have to add meaning and emotion to the automatically created articles, staying firmly in charge of the published product (Wu et al., 2019). But even if the automatically created articles are not edited, they still add extensive value for the news outlets' readers, including by granting the journalists more time for interviews and other exclusive stories. However, as the *Sportbuzzer* case shows, not everyone reaps the benefits of automation. Quality considerations demonstrate that the sports beat might be a unique case with regard to automation, as many interviewees, including *Sportbuzzer*, noted that sports stories require context and further editing. Nevertheless, *Beinschuss* concisely summarises the value of automation: "Automation is definitely an added value for our readers. So why shouldn't we make use of this?" Future research could examine whether this is different for other beats such as finance, where the raw numbers tell a story in themselves.

As power negotiations and shifts occur, an open question is whether data and software providers might turn from *implicit* to *explicit* interlopers, which feature their own media coverage and then do challenge journalistic authority, similar to blogs and citizen journalists (Eldridge, 2019). *Fussball.de*, run by the German Football Association, for instance, provide reporting on amateur matches, even though they claim that they "don't call [their coverage] journalism, but rather text generation according to data," as the software cannot "give context" or "judge" matches, and thus "the ma-

chine won't replace journalists." They regard automated content as a service to amateur clubs, which can use the articles on their websites and social media feeds. Despite these claims, their offering of extensive coverage opens up a new dimension of the automated news process. As in the examples above, implicit interlopers may become explicit interlopers even if that was not their original intention. In the case of *fussball.de*, the involvement of interlopers goes even further, as the German Football Association works with a team of software and machine-learning specialists as well as linguists, which demonstrates how an interloper works with other interlopers to provide automated news articles. Data and software providers may also become explicit interlopers. For example, Sportec Solutions and Retresco work on new media formats together, which are supposed to be used for both their customers and themselves. OPTA and Retresco also see other interlopers using automated sports content. Betting companies, for example, may employ historical data not only to calculate their odds, but also to generate stories around them. Thus, *fussball.de* and the other implicit interlopers take an active role in "defin(ing) the place of automated journalism within the larger context of news" (Carlson, 2015, p. 417), demonstrating how interlopers shape this field.

In all these cases, media outlets are cut out of the process. In this new environment, media outlets thus have to clearly establish the added value they give their readers with automation, and how their coverage differs from what interlopers deliver. *Nordbayern Amateure* could be a model for this avenue. They heighten the value of providing news content in large volume by editing every article with quotes and in-depth information, thus using automated articles as a baseline. Thus, they retain journalistic authority while making effective use of the technology. All in all, as data and software providers as well as other implicit interlopers continue to acquire journalistic authority, the power balance between interlopers as mere service providers on the one side and media outlets on the other is shifting. It is not yet known how these developments affect an already shrinking news market, and how interlopers will define their role in news production, even if they might not describe their content as 'journalism.'

As implicit interlopers may become explicit interlopers and take the journalistic stage, this expanded involvement of interlopers may have consequences for all actors in the journalistic news process. This includes the audience, which, as could be seen in this article, is perceived as almost a partner in developing automatically created output. Questions arise about how readers would perceive the interlopers' news-like narratives which take on journalistic authority over match odds and player profiles, especially when these narratives are indistinguishable from journalistic articles. Moreover, it remains to be seen whether audiences care about such a distinction in light of being offered a highly personalised product that news media are often not able to deliver due to person-

nel and budget constraints. I propose that researching this distinction between automatically created articles provided by news outlets and by interlopers is a route that research on audience perceptions of news automation could take. The present works mainly focus on the reception of new outlets' content, and examining interlopers as another message source could shed new light on audiences' perception of both source and message credibility of automated content.

All interviewees except *Sportbuzzer*, which stopped creating automated articles, either have plans for the enhanced use of automation or see the potential for further applications. *Fussball.de* is working on new features, such as putting player profiles into text (e.g., club history, minutes played), programming a skill for Amazon's Alexa, or adding articles for youth leagues. *Beinschuss* aim to experiment with winter sports and further areas such as weather and event calendars. Thus, it remains to be seen how automation will continue to be employed in sports newsrooms, and whether other sports coverage apart from amateur football will be automated; the future direction will largely depend on data availability. For professional football, OPTA and AX Semantics propose using more data sources, such as tracking data gathered from sensors on players' jerseys, to flesh out the reporting. While they acknowledge that privacy considerations may be a concern, AX Semantics say that this kind of data would bring about a "new space of meaning" for sports journalism. But even without innovative data collection methods, also professional football coverage is generally thought to benefit from automation. Unlike in amateur football, the mere coverage of professional football is not an issue, yet "being the first to cover it" (Retresco) could be. Automation could then be used to send out short snippets and direct audience attention to their outlet via push notifications. Internationalisation strategies might also play a role, since some software offers to simultaneously translate the generated articles into other languages. Both software providers mention the possibility of finding new markets, such as "fans who are located in other countries" (AX Semantics). In summary, automation holds considerable future potential, and it remains to be seen how media outlets and interlopers will exploit these opportunities.

Acknowledgments

Conflict of Interests

The author declares no conflict of interests.

Supplementary Material

Supplementary material (interview outline) for this article is available online in the format provided by the author (unedited).

References

Ahva, L. (2019). About actor positioning in journalism…slowly. *Media and Communication*, *7*(4), 123–126.

Alhalaseh, R., Munezero, M., Leinonen, M., Leppänen, L., Avikainen, J., & Toivonen, H. (2015). Towards data-driven generation of visualizations for automatically generated news articles. In M. Turunen, J. Paavilainen, F. Fagerholm, J. Linna, J. Okkonen, P. Huuskonen, & S. Kotilainen (Eds.), *Proceedings of the 22nd international academic Mindtrek conference* (pp. 100–109). New York, NY: Association for Computing Machinery.

Allen, N. D., Templon, J. R., McNally, P. S., Birnbaum, L., & Hammond, K. (2010). *StatsMonkey: A data-driven sports narrative writer*. Arlington, VA: AAAI Fall Symposium Series.

Allensbacher Markt und Werbeträgeranalyse. (2019). Beliebteste Sportarten in Deutschland nach Interesse der Bevölkerung an dem Sport in den Jahren 2017 bis 2019 [Most popular sports in Germany, interest in population, years 2017–2019]. Retrieved from https://de.statista.com/statistik/daten/studie/171072/umfrage/sportarten-fuer-die-besonderes-interesse-besteht

Boyle, R. (2006). *Sports journalism: Context and issues*. London: SAGE Publications Ltd.

Braun, J. (2014). Transparent intermediaries: Building the infrastructures of connected viewing. In J. Holt & K. Sanson (Eds.), *Connected viewing: Selling, streaming and sharing media in the digital era* (pp. 124–143). New York, NY: Routledge.

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, *3*(3), 416–431.

Chua, S., & Duffy, A. (2019). Friend, foe or frenemy? Traditional journalism actors' changing attitudes towards peripheral players and their innovations. *Media and Communication*, *7*(4), 112–122.

Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice*, *8*(5), 519–531.

Dörr, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, *4*(6), 700–722.

Eldridge, S. A., II. (2019). Where do we draw the line? Interlopers, (ant)agonists, and an unbounded journalistic field. *Media and Communication*, *7*(4), 8–18.

Fanta, A. (2017). *Putting Europe's robots on the map: Automated journalism in news agencies.* Oxford: Reuters Institute for the Study of Journalism. Re-

trieved from https://reutersinstitute.politics.ox.ac.uk/our-research/putting-europes-robots-map-automated-journalism-news-agencies

Galily, Y. (2018). Artificial intelligence and sports journalism: Is it a sweeping change? *Technology in Society*, *54*, 47–51.

Graefe, A. (2016). *Guide to automated journalism*. New York, NY: Tow Center for Digital Journalism.

Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, *19*(5), 595–610.

Gunasiri, D. Y., & Jayaratne, K. L. (2019). Automated cricket news generation in Sri Lankan style using natural language generation. *European Journal of Computer Science and Information Technology*, *7*(9), 42–56.

Gynnild, A. (2014). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism*, *15*(6), 713–730.

Holton, A. E., & Belair-Gagnon, V. (2018). Strangers to the game? Interlopers, intralopers, and shifting news production. *Media and Communication*, *6*(4), 70–78.

Horky, T., & Stelzner, B. (2013). Sports reporting and journalistic principles. In P. M. Pedersen (Ed.), *Routledge handbook of sport communication* (pp. 118–127). London: Routledge.

Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior*, *71*, 291–298.

Kanerva, J., Rönnqvist, S., Kekki, R., Salakoski, T., & Ginter, F. (2019). Template-free data-to-text generation of Finnish sports news. *Cornell University*. Retrieved from https://arxiv.org/abs/1910.01863

Lindén, C.-G. (2017a). Algorithms for journalism: The future of news work. *The Journal of Media Innovations*, *4*(1), 60–76.

Lindén, C.-G. (2017b). Decades of automation in the newsroom: Why are there still so many jobs in journalism? *Digital Journalism*, *5*(2), 123–140.

OPTA. (2020). Sports. *OPTA*. Retrieved from https://www.optasports.com/sports

Primo, A., & Zago, G. (2015). Who and what do journalism? An actor-network perspective. *Digital Journalism*, *3*(1), 38–52.

Rojas Torrijos, J. L. (2019). Automated sports coverage: Case study of bot released by The Washington Post during Rio 2016 and Pyeongchang 2018 Olympics. *Revista Latina De Comunicación Social*, 1729–1747.

Schreier, M. (2012). *Qualitative content analysis in practice*. London: SAGE.

Tandoc, E. C., Jr. (2019). Journalism at the periphery. *Media and Communication*, *7*(4), 138–143.

Thurman, N., Dörr, K., & Kunert, J. (2017). When reporters get hands-on with robo-writing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism*, *5*(10), 1240–1259.

van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice*, *6*(5/6), 648–658.

van der Kaa, H., & Krahmer, E. (2014). Journalist versus news consumer: The perceived credibility of machine written news. In *Proceedings of the computation + journalism conference New York* (pp. 1–4). Tilburg: Tilburg University.

WashPostPR. (2017, September 1). The Washington Post leverages automated storytelling to cover high school football. *The Washington Post*. Retrieved from https://www.washingtonpost.com/pr/wp/2017/09/01/the-washington-post-leverages-heliograf-to-cover-high-school-football

Wu, S., Tandoc, E. C., Jr., & Salmon, C. T. (2019). Journalism reconfigured: Assessing human–machine relations and the autonomous power of automation in news production. *Journalism Studies*, *20*(10), 1440–1457.

Yao, L. J., Salmon, C. T., & Tandoc, E. C., Jr. (2018). Automated journalism a home run for sports? On the effects of perceived authorship and text objectivity on receptivity toward machine-written sports news. In A. Sivakumar (Ed.), *Proceedings of the URECA@NTU 2017–18* (pp. 1–10). Singapore: Nanyang Technological University.

About the Author

Jessica Kunert is a Senior Research Associate at the Institute for Journalism and Communication Studies at the University of Hamburg. Her research focuses on the impact of algorithms and automation on news production and distribution, such as how algorithms are employed in newsrooms for automatic creation of text and video. She also studies communication patterns and networks in professional and amateur sports. She is on Twitter at @jessicakunert.