

“Automation *will* save journalism”

News automation from the service providers' point of view

Martin Kjellman

University of Helsinki

Faculty of Social Sciences

Magisterprogrammet i samhällsvetenskaper (Master's Programme in Social Sciences)

Journalistik och kommunikation (Journalism and Communication)

Master's thesis

Supervisor: Jaana Hujanen

January 2021



| | | | |
|--|--|---|--|
| Tiedekunta/Osasto Fakultet/Sektion – Faculty Faculty of Social Sciences | | Laitos/Institution– Department The Swedish School of Social Science | |
| Tekijä/Författare – Author Kjellman, Martin | | | |
| Työn nimi / Arbetets titel – Title “Automation <i>will</i> save journalism” – News automation from the service providers’ point of view | | | |
| Oppiaine /Läroämne – Subject Journalistik och kommunikation (Journalism and Communication) | | | |
| Työn laji/Arbetets art – Level Master’s thesis | Aika/Datum – Month and year January 2021 | Sivumäärä/ Sidoantal – Number of pages 75 | |
| Tiivistelmä/Referat – Abstract <p>This thesis examines how representatives of service providers for news automation perceive a) journalists and news organisations and b) the service providers’ relationship to these. By introducing new technology (natural language generation, i.e. the transformation of data into everyday language) that influences both the production and business models of news media, news automation represents a type of media innovation. The service providers represent actors peripheral to journalism. The theoretical framework takes hybrid media logics as its starting point, meaning that the power dynamics of news production are thought to be influenced by the field-specific logics of the actors involved. The hybridity metaphor is deepened by using a typology for journalistic strangers that takes into account the different roles peripheral actors adopt in relation to journalists and news organisations. Journalism is understood throughout as a professional ideology encountered by service providers who work with news organisations.</p> <p>Semi-structured interviews were conducted with representatives from companies that create natural language generation software used to produce journalistic text based on data. Participants were asked about their experiences working with news media and the interviews (N=6) were analysed phenomenologically.</p> <p>The findings form three distinct but interrelated dimensions of how the service providers perceive news media and journalism: an area that sorely needs innovators (potential) but lacks resources in terms of knowledge, money and will to innovate (obstacles), but one that they can ultimately learn from and collaborate with (solutions). Their own relationship to journalism and news media is not fixed to one single role. Instead, they alternate between challenging news media (explicit interloping) and inhabiting a supportive role (implicit interloping).</p> <p>This thesis serves as an exploration into how service providers for news automation affect the power dynamics of news production. It does so by unveiling how journalists and news organisations are perceived, and by adding further understanding to previous research on actors peripheral to journalism. In order to further untangle how service providers for news automation shift the balance of power shaping news production, future research should attempt to unify the way traditional news media actors and service providers perceive each other and their collaborations.</p> | | | |
| Avainsanat – Nyckelord – Keywords news automation, hybridity, journalistic strangers, natural language processing, computational journalism, service providers | | | |
| Säilytyspaikka – Förvaringställe – Where deposited | | | |
| Muita tietoja – Övriga uppgifter – Additional information | | | |

Contents

| | |
|---|----|
| 1. Introduction..... | 4 |
| 2. Theoretical framework..... | 7 |
| 2.1 Journalism and news organisations..... | 8 |
| 2.2 Hybridity..... | 10 |
| 2.3 Journalistic strangers..... | 13 |
| 3. Previous research..... | 15 |
| 3.1 News automation..... | 15 |
| 3.1.1 What it is and how it's used..... | 15 |
| 3.1.2 What it means..... | 20 |
| 3.2 Media innovation..... | 24 |
| 3.3 Peripheral actors..... | 27 |
| 4. Method..... | 31 |
| 4.1 Semi-structured research interviews..... | 32 |
| 4.2 Participants..... | 36 |
| 4.3 Ethical considerations..... | 37 |
| 4.4 Phenomenological analysis..... | 38 |
| 5. Results..... | 45 |
| 5.1 Potential: strategic field..... | 46 |
| 5.2 Obstacles: everything is lacking..... | 49 |
| 5.3 Solutions: learning to understand each other..... | 53 |
| 6. Discussion and Conclusions..... | 59 |
| 6.1 Summarisation and discussion of results..... | 59 |
| 6.2 Limitations and suggestions for further research..... | 62 |
| References..... | 66 |
| Appendix..... | 73 |

1. Introduction

On April 14th 2019, the people of Finland went out to vote in the parliamentary elections. Covering national elections is always busy for Yle, the Finnish public service broadcasting company, but during this particular election night the number of articles published exceeded 311. Why such a specific figure? 311 happens to be the number of municipalities in Finland in 2019, and in the parliamentary election, Yle published a unique article for each of the municipalities. This was possible thanks to a computer program called Voitto, which used election results as input data and produced news texts based on the results. 311 articles in a single night is quite a lot, but compared to what the news automation system Valtteri did in the municipal elections two years prior, it's nothing. Valtteri was able to produce 727,404 articles – in each of three languages, Finnish, Swedish and English. Valtteri was developed at the University of Helsinki and Voitto was created in-house at Yle (Leppänen et al., 2017; Vehkoo, 2017). Both are good examples of news automation, or actually a particular aspect of news automation: automatically generated news texts.

The idea behind news automation is to convert data into news texts. Structured data, often numbers or statistics, are used as input. Through the use of algorithms, the output is text written in everyday or “natural” language. This technology is called natural language generation (NLG). Historically NLG has been considered artificial intelligence (AI) (Reiter and Dale, 2000, p. 1) but due to high demands on accuracy, the technologies used in journalism are generally not on par with more complex modes of AI (Diakopoulos, 2019, p. 101). In journalism, NLG is used in domains where structured data is readily available and news texts follow recurring patterns. Besides political elections, prominent examples include sports reporting (Simola, 2019) and financial results (White, 2015). Recently, research exploring the possibility of automating more complex news stories, so-called “event-driven narratives”, has emerged as well (Caswell and Dörr, 2018).

News automation is already making its way into newsrooms and although it is happening slowly (Lindén et al., 2019, p. 14), the implications are profound. This is partly due to factors external to journalism – rapid technological developments and, most importantly, rapidly growing amounts of open digital data – and partly due to internal ones – diminished revenue

and a need to reduce costs while simultaneously meeting a growing demand for content (Graefe, 2016, pp. 30–31). News automation benefits from the former and potentially provides answers to the latter. In other words, news automation affects editorial processes, business models and organisational structures, all of which are features of media innovation in the broadest sense of the word (Westlund and Lewis, 2014, p. 11).

Media innovation is not limited to what goes on in news organisations. Innovation studies must take into account that “media, media practices, social practices, and society itself are inherently and inextricably entwined in contemporary mediatised society” (Bruns, 2014, p. 24). In this study, I use hybrid media theory as described by Chadwick (2017) to take this “entwinedness” into account. Furthermore, deep in the mix we find a number of different actors. These peripheral actors or “journalistic strangers” (Holton and Belair-Gagnon, 2018) work in tandem with news media, functioning as friends, foes or even “frenemies” (Chua and Duffy, 2019). In the case of news automation, the journalistic strangers are represented by software companies that develop NLG systems for news production. In order to achieve a more robust and domain-specific theoretical framework, the concept of strangers is combined with hybridity.

The field of research in which this and arguably other studies on news automation operates is digital media studies (cf. Eldridge et al., 2019). The body of research on news automation does not go that far back; the bulk of it coincides with the current era, which started some ten years ago with the emergence of the NLG company Narrative Science (Lindén et al., 2019, p. 9). Previous research has notably focused on the technological aspects of NLG and how it relates to journalism (cf. Diakopoulos, 2019; Leppänen et al., 2017), what it might mean for journalism as a profession (cf. Latar, 2015; Lindén, 2017), ethical aspects of autonomous systems producing news (cf. Diakopoulos and Koliska, 2017; Dörr and Hollnbuchner, 2017) and audience’s reception of automatically produced texts (cf. Clerwall, 2014; Melin et al., 2018).

When it comes to automation, news media is currently experiencing what Sirén-Heikel et al. (2019, p. 48) call an “in-between phase” in which the interaction between social practices and technology is shaping the news production to come. Two key findings by Sirén-Heikel et al. (2019, p. 59–60) highlight the need to examine the role of external NLG software providers

in the in-between phase: (1) most newsrooms are forced to rely on outside expertise in order to implement news automation systems and (2) news media representatives perceive automated content production as something to take into account regardless of whether it is actually used in traditional journalism. I believe that journalism research should do the same.

The aim of this study is to understand how the companies that develop the software that produces the news place themselves within the power dynamics of news production. Specifically, the goal is to examine how representatives of the companies perceive news organisations and journalists and their own relationship with them. More specifically, the thesis will answer the two following research questions:

RQ1: How do representatives for NLG service providers perceive news organisations and journalists?

RQ2: How do the service provider representatives experience their companies' relationship with news organisations and journalists?

As a reference point for understanding journalists, I use Deuze's (2005) definition of journalism as professional ideology (cf. section 2.1). A news organisation is defined as a unit composed of individuals who share the goal of producing and distributing news and clearly differentiates itself from other such units, for example in terms of ownership (Shoemaker and Reese, 2013, p. 130). Since the research questions explicitly concern the *experience* of a set of people (representing their companies), I rely on qualitative interviews with the service providers and phenomenological analysis of the interviews.

After the introduction, I lay out the study's theoretical framework. This is followed by a section on previous research on news automation, media innovation and peripheral actors. The next section describes my methodology and discusses the quality of the research design. The results are presented in a separate section. In the final section, I frame the results in relation to previous research and the theoretical framework and conclude by discussing the limitations of this study and suggesting further research.

2. Theoretical framework

This section starts out with a brief explanation of how I understand news organisations followed by an overview of Deuze's (2005) definition of journalism as professional ideology, which will be used as a reference point throughout the study. Next, I describe the main theoretical framework used in this study, which is a synthesis of Andrew Chadwick's (2017) concept of hybridity and Avery Holton and Valerie Belair-Gagnon's (2018) typology of strangers in journalism.

Since Chadwick (2017) builds on Pierre Bourdieu's field theory, the obvious choice would perhaps be to double down on this theory. For the purposes of this study, however, I believe that the strangers typology offers a more potent approach, simply because it was developed precisely for the purpose of studying non-traditional actors involved in journalism. While Holton and Belair-Gagnon (2018) do not reference hybridity in the context of strangers in journalism, it is not far-fetched to claim that the two concepts are related, an argument I will elaborate on at the end of this section. For example, in what is effectively this study's mirror image, Chua and Duffy (2019) used Holton and Belair-Gagnon's typology for studying journalists' attitude towards peripheral actors involved in innovation. When discussing future research, the authors suggest using hybridity for further exploration (p. 120). Here as well, this study operates in mirror image fashion by starting out with hybridity and then building on it.

Expanding the theoretical framework in this way is needed too, since the hybridity metaphor is not particularly useful unless the hybrids as well as their parent categories and the relationships between them are examined (Mast, Coesemans and Temmerman, 2017, p. 5). Holton and Belair-Gagnon's typology provides a structure for doing this when studying the interplay between traditional news media and external software providers. Finally, a key assumption of Deuze's (2005) definition of journalism as ideology is that "[t]he high modernism of journalistic professionalization has moved to a liquid modern state of affairs of feverish journalistic differentiation" (p. 450). That is an assumption prevalent in Chadwick's understanding of hybrid media, meaning that our definition of journalism also serves to deepen the explanatory potential of the hybridity metaphor.

2.1 Journalism and news organisations

According to Shoemaker and Reese (2013, p. 130), organisations can be viewed as units made up by individuals who share common goals and are distinct from other such units, for example in terms of ownership. A media organisation “creates, modifies, produces, and distributes content to many receivers” (ibid.) and in the case of news (media) organisations, that content is news. As a level of analysis, organisations can be placed between the level of overarching social structures and the level of routines as practiced by individual journalists. This allows us to examine content production from a perspective not explicitly tied to either of the other levels (Shoemaker and Reese, 2013, p. 134–135). Considering that in this study, media innovation (particularly news automation) is understood through organisational structures (Paulussen, 2016, p. 196; cf. section 3.2), the organisation as level of analysis is needed as a complement to the journalist level.

Following Deuze (2005), I understand journalism as a professional (or occupational) ideology. This means that I do not view journalism “as a profession, an industry, a literary genre, a culture or a complex social system” (Deuze, 2005, p. 444) although it certainly can be all those things. Instead, according to Deuze’s (2005) definition, journalism is understood “as a collection of values, strategies and formal codes characterizing professional journalism and shared most widely by its members” (p. 445). Depending on their cultural context and genre, journalists interpret these values, strategies and codes in various ways but they arguably remain constant (Deuze, 2005, p. 445). Since this is a non-comparative study with a global scope, a definition that emphasises presumably uniform aspects of journalism seems appropriate.

Deuze (2005, p. 446–450) identifies five values that constitute the ideology of journalism: public service, objectivity, autonomy, immediacy, and ethics. Providing a *public service* means monitoring those in power on behalf of citizens, or consumers depending on how the journalists view their audience. Sometimes disputed, *objectivity* refers to impartiality and neutrality as traits among journalists. It is a value traditionally claimed by journalists in order to strengthen their credibility. *Autonomy* describes the state of journalists enjoying freedom and independence in their work. It is a value used by journalists to fend off external influence on their work, and also a mechanism for preventing individual journalists from supporting

community engagements. Perceived as a virtue as well as a curse, *immediacy* refers to speed as an inherent factor of reporting news. Similar to objectivity, *ethics* – in the form of a self-proclaimed ethos rather than a universal code of conduct – serve to legitimise the position journalism holds in society.

In the context of automation and innovation, autonomy is the most relevant value (even though news automation can certainly be said to affect and be affected by all five of them). The transition to multimedia environments has introduced a need for multicompetent teams which transforms the autonomy concept into something more collaborative (Deuze, 2005, p. 456–457). When the service providers analysed in this study enter newsrooms, they drive innovation through collaborative efforts. Journalists do not always take kindly to such efforts though, and autonomy is a value typically held up by journalists resisting innovation (Deuze, 2005, p. 449).

More recently, Deuze (2019) himself has criticised his previous views on journalism, in particular the notion that journalism is somehow inextricable from how news media has traditionally been organised (p. 1). In the context of this study, this is relevant because media innovation, such as news automation, is all about the reorganising of the news industry (cf. section 3.2). Deuze does not reject journalism as ideology though. Instead he argues that it is upheld by individuals, not the news industry (p. 3). This further indicates that we specifically need to examine how software providers view journalists. So while ideology perhaps isn't such a definitive answer to the question of *what journalism is* as Deuze might have once believed, it arguably still provides a useful way of understanding journalism.

What journalism is does indeed not remain static (for an in-depth examination of this idea, cf. Deuze and Witschge, 2018), and the ideological values are constantly being challenged, mainly with regards to two aspects: control and transparency. Societal and technological changes shift the balance of who is in control of the news agenda and to what degree the process of news production is transparent (Deuze, 2005, p. 455). Within journalism, control and transparency are in a proverbial tug of war, with journalists giving up control as they make news production more transparent (ibid.). In this study, I view the issues of control and transparency through the lens of hybridity (Chadwick, 2017) and refer to the interplay between them as “the power dynamics of news production”.

In a time defined by rapid changes, what keeps journalists a cohesive professional group, Deuze (2005, p. 455) argues, is ideology. Because journalists constantly use the ideology of journalism to negotiate their own position in society (Deuze, 2005, p. 446), this definition of journalism is particularly useful when examining the relationship between news media and other (for example peripheral) actors. I further argue that what potentially sets the NLG providers' relationship with news organisations apart from their relationship with other types of media organisations, not to mention other industries, is the presence of the professional ideology of journalism. In other words, the ideology of journalism is what makes the interaction between NLG providers and news organisations unique and a big reason for it being worth examining. Finally, we should acknowledge the complex nature of the service providers' interaction with journalists and news organisations. In order to adequately examine this interaction, we need the concepts introduced in the following two sections: hybridity and journalistic strangers.

2.2 Hybridity

I want to start by following Chadwick (2017) back to the original Greek meaning of the word hybrid, where it refers to “something that questions conventional understandings and the accepted order” (p. 10). This highlights the main strengths of the concept: it helps us examine things without getting stuck in our usual, familiar conceptions and categories and allows us to examine the border areas (Chadwick, 2017, p. 10). One of the key ideas expressed by Chadwick (2017) is that hybridity helps replace “‘either/or’ patterns of thought” with more fruitful “not only but also” ones (p. 4). Even though the hybrid thinking that Chadwick introduces is broad and his own focus is explicitly on political communication, technological development is the driving force behind what the author calls “the hybrid media system” in the 2013 book (2nd ed. 2017) with the same title. Rapid technological developments is what creates the need for new ways of understanding communication where old dichotomies are replaced by border areas in which symbiotic relationships between actors become ever more important (Chadwick, 2017, pp. 3–4). Software providers for news automation operate in precisely this type of border area.

This brings us to the term media logic. The concept, introduced by Altheide and Snow (1979), is used to explain how “the assumptions, norms, and visible artifacts of media, such as templates, formats, genres, narratives, and tropes have come to penetrate other areas of social, economic, cultural, and political life” (Chadwick, 2017, p. 23). Acknowledging the diffusion described here represents a first step towards examining “the power of media and the power relations within media” in a manner that takes into account interactions between the media system and other areas of society instead of viewing them as separate from each other (ibid.). But to what extent does media logic shape the practices of other areas of society in the 21st century? Since the vast power originally ascribed to institutional media by Altheide and Snow has rapidly diminished, Chadwick (2017) argues, we need an update of the media logic concept (p. 24). The media landscape of today is more polyphone and fluid than what was the case pre-internet when the media logic was still dominated by mass media, whose logic is now facing competition from new types of media and by extension new media logics (Chadwick, 2017, pp. 24–25). By arguing that there are several parallel media logics, created through interactions between different fields, Chadwick (2017) introduces what he calls *hybrid media logics* (pp. 22–26). Here, Chadwick follows Deuze (2004), whose concept “multimedia logic” (p.140) is seen as a characteristic of 21st century journalism and marks a shift from a singular media logic towards a more flexible set of logics. With hybrid media logics, we no longer assume that media influence goes one-way. Instead, the assumption is that society is “being shaped by more complex interactions between competing and overlapping media logics” (Chadwick, 2017, p. 25). For the purposes of this study, I view the ideology of journalism as a key element of the hybrid media logics.

In this study, I am particularly interested in exploring what Chadwick (2011), describing the hybrid news system, calls “subtle but important shifts in the balance of power shaping news production” (p. 6). For instance, how is this balance of power influenced when news production is carried out using software created by external providers? Software from external providers is not a new phenomenon, with tools such as Photoshop and Indesign having been staples of news production for quite some time. What arguably sets news automation software apart is the automation part, the fact that “algorithms enter into the evaluative phase of cognitive labor, in effect judging and making decisions” (Diakopoulos, 2019, p. 17). This decision-making process particularly involves prioritisation and optimisation and has huge implications for news media because it will be shaped by the

values of the people who create the algorithm (Diakopoulos, 2019, p. 19). So, since the power balance of news production is now less than ever exclusively influenced by actors traditionally operating within the media system (Chadwick 2011, p. 19), it must be examined with this in mind. In order to come to terms with this, Chadwick (2017, p. 20) borrows from Manuel Castells' views on institutional power – in this context power wielded by journalistic institutions – facing counter-power. Castells originally had social movements in mind, but the hybrid media logics concept allows us to describe the relationships between media institutions and other (traditionally) peripheral actors in terms of power.

In digital journalism studies, hybridity has been applied to make sense of the power dynamics of news production. Hermida and Young (2017) used hybridity as defined by Chadwick to examine the influence data and computational journalism has on norms and practices in legacy media. Since news automation is closely related to computational journalism (see more detailed discussion in section 2.2.1), it is highly relevant that they rely on hybridity as a way of “understanding of technological change in media that supports a nuanced ability to interrogate past assumptions around journalistic agency and power” (Hermida and Young 2017, p. 162). More specifically, Hermida and Young (2017) were able to look at the power dynamics within newsrooms by framing it as encounters between different kinds of media logics (p. 160).

Hybridity has also been used to understand the relationship between journalism and external software providers, which of course is exactly what this study sets out to do. Ananny and Crawford (2015) applied it to study how the professionals who develop news applications view their relationship to journalism. Since these programmers and app designers are actors “who may not self-identify as journalists but nonetheless define the conditions under which news is created and circulates” (p. 193), the importance of the power-dynamics-of-news-production perspective becomes apparent.

Finally, we need to address some of the criticism associated with hybridity as a means to understand contemporary journalism. In an article accompanied by the suitable disclaimer “Dealing with the mess (we made)” Witschge et al. (2019) note that we do not rid ourselves of binary thinking by simply merging dualities such as “objective versus subjective” or “commercially driven versus publicly motivated” (p. 655). The authors call for “an

experientialist, practice-based approach” (p.652), one that moves beyond merely replacing the either-or with both and instead views the binaries as continuums (p. 656). This is essentially the approach used by Ananny and Crawford (2015) researching news app developers. The same approach also allowed Baack (2018) to examine the interplay between data journalists and civic technologists as something that happens on a continuum constituted by different practices (p. 688–689). In this study too, my aim is to “avoid a priori delineation of actors based on predefined categories” (Baack 2018, p. 675), thus sidestepping the pitfalls of an ungainly use of hybridity by opting for the experientialist approach favored by Witschge et al. (2019).

2.3 Journalistic strangers

Building on the work of innovation studies and organisational studies Holton and Belair-Gagnon (2018) introduce the concept of *journalistic strangers* to describe actors not traditionally thought of as journalistic ones. The strangers can identify as journalistic actors, or not, but they are involved in journalism and the shaping of its norms and practices (p. 73). The terminology owes to sociologist Georg Simmel’s metaphor of strangers, i.e. actors who did not originally belong to a group and bring something new to this group (Holton and Belair-Gagnon 2018, p. 72). The definition of strangers includes individuals as well as institutions, and in the case of journalism, the strangers are mainly technology-oriented (ibid.). In part due to the rise of the internet and social media, Holton and Belair-Gagnon (2018) note, the amount of types of journalistic strangers has been steadily increasing. Thus they identify a need for a typology that accurately reflects the actual nature of these strangers’ contribution to journalism (p. 73). Furthermore, the stranger typology represents an attempt to reckon with the journalism studies tendency to position journalism at the center and non-traditional actors at the periphery (Holton and Belair-Gagnon 2018, p. 71). Because as Tandoc (2019) puts it: “journalism being the centre of our investigation should not automatically assume that it is at the centre of social life” (p. 140).

In what is effectively an expansion of Eldridge’s (2017) concept *media interlopers*, Holton and Belair-Gagnon (2018) provide three categories of strangers: explicit interlopers, implicit interlopers and intralopers. *Explicit interlopers* are actors who operate outside traditional journalism but are nonetheless actively engaged in news production. They can be bloggers

working independently or organisations such as Wikileaks working partly in tandem with legacy media. Explicit interlopers typically challenge existing journalistic norms, e.g. by encouraging participation and claiming a place in journalism without labelling themselves journalists (pp. 73–74). *Implicit interlopers* are actors who are less directly involved with journalism. Their role is usually a supportive one vis-à-vis traditional journalism. Their focus being to help journalists produce better content or engage more effectively with the audience, they do not necessarily challenge traditional journalism head on and are generally more accepted by journalists. Examples of implicit interlopers include programmers and web analytics professionals (pp. 74–75). Finally, *intralopers* are non-journalistic actors working within traditional news organisations. They reshape news production by injecting it with new norms and practices of their own and driving newsroom innovation (p. 75).

These categories represent a more nuanced way of examining the new hybrid media logics that shape the power dynamics of news production. The journalistic strangers typology allows us to acknowledge that the relationship between the actors involved in news production is “not as dichotomously straightforward as insider/outsider or interloper/journalist” (Holton and Belair-Gagnon 2018, p. 71). By conceptualising the software providers and their relationship to journalists in this way, we move one step closer to the more nuanced hybrid thinking proposed by Mast, Coesemans and Temmerman (2017, p. 5) and Witschge et al. (2019, pp. 655–656). For the sake of narrowing the scope, the working assumption of this study is that software providers for news automation are best represented by the implicit interlopers category, i.e. actors who occupy a supportive role. But this is by no means a given. To cite just one example, the web analytics companies studied by Belair-Gagnon and Holton (2018) aimed to be “readily available to help with projects and dissect data alongside journalists” (p. 502), thus expressing clear intraloper tendencies.

The most obvious connection between hybridity and journalistic strangers is that both owe to Pierre Bourdieu’s field theory. In Chadwick’s (2017) case the connection to Bourdieu is explicit (p. 23). In Holton and Belair-Gagnon’s (2018) case it is implicit through Eldridge (2017), whose interloper media concept is built around field theory (p. 41). Hybridity and the journalistic strangers typology can be viewed as specialised variations of field theory that fit the aims of this study. For this reason, I will not dwell on the specifics of field theory, but note that this is the territory here.

3. Previous research

I have chosen to split this section into three parts. The first one (3.1) deals with research explicitly focused on news automation. Here, I aim to explain what news automation *is* (and perhaps isn't) as well as what it *means*. The second section (3.2) lays out the basics of media innovation and tries to situate news automation within this research tradition. Since this study puts focus on non-journalistic actors and their role in a particular type of media innovation, the third and final section (3.3) looks at the growing body of research on actors peripheral to journalism and news media.

3.1 News automation

The following two sections are meant to provide an overview of news automation. I start out by discussing the terminology and go on to describe news automation from a practical point of view. In the second section, my aim has been to situate news automation in a wider societal context.

3.1.1 What it is and how it's used

An algorithm can be defined “a series of steps that is undertaken in order to solve a particular problem or to accomplish a defined outcome” (Diakopoulos, 2019, p. 16). In news production, these steps are undertaken by an algorithm that runs on a computer (ibid.). The process of using algorithms to write news texts based on structured data is referred to using a wide array of terms (for an overview, cf. Lindén, 2017, p. 125), including the mostly appropriate *automated journalism* (e.g. Carlson, 2015; Graefe, 2016) and the popular but somewhat misleading metaphor *robot journalism* (e.g. Clerwall, 2014). Since this study is informed by the interdisciplinary project Immersive Automation (www.immersiveautomation.com), I stay with the term favoured in that project: *news automation* (Lindén et al., 2019). As the title of Diakopoulos' (2019) book on the subject, “Automating the News: How Algorithms Are Rewriting the Media”, would have it, what is primarily being automated is the news, not necessarily journalism. Given that I define journalism here as professional ideology (Deuze, 2005), I believe that in this context it is

appropriate not to view journalism as what is being automated (even though the idea of automating ideology certainly is interesting).

Essentially, news automation is a continuation of what Coddington (2015) calls journalism's "quantitative turn" (term initially used by Petre, 2013) and can be traced back to the sub-field of journalism previously known as computer-assisted reporting (CAR). Unsurprisingly, since reporting is seldom done without computer-assistance these days, what used to be called CAR has mutated into a variety of different categories. Stavelin (2013) provides a useful typology, listing the following categories: computer-assisted reporting & precision journalism, data journalism, database journalism, data-driven journalism, and computational journalism (pp. 28–35). On the industry side of things, the term data journalism seems to have replaced CAR as the umbrella term most commonly used to describe journalistic practices that rely on analysing and presenting data (Coddington, 2015, p. 334).

In 2012, Van Dalen noted that "with the introduction of machine-written news computational journalism entered a new phase" (p. 649), thus linking news automation to computational journalism. Indeed, of all the categories listed by Stavelin (2013), computational journalism appears to be the one that best fits news automation (Diakopoulos, 2019, p. 27; Lindén, 2017, p. 125). The most commonly used definition, by Hamilton and Turner (2009), describes computational journalism as "the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism" (p. 2). However, as Coddington (2015, p. 335) points out, there is significant overlap between this definition of computational journalism and data journalism. It is therefore useful to emphasise that what sets news automation apart from practices that fall under the industry-favoured umbrella term data journalism, is automation. The key difference, as pointed out by Carlson (2015), is that the process of analysing data and producing news texts is done "with limited to no human intervention beyond the initial programming choices" (p. 416).

The process referred to by Carlson, these initial programming choices, is called natural language generation (NLG). Together with natural language understanding, NLG forms the field of research called natural language processing or NLP (Reiter and Dale, 2000, pp. 2–3). In essence, NLG is the process of converting non-linguistic information (such as structured data) into human language (the kind that news texts are made of) (Reiter and Dale 2000, p.

1). According to Reiter (2013), one key aspect that sets NLG systems apart from other types of NLP is the high level of choice making required: what information to include, what word to choose etc. (p. 13e). The general NLP process consists of three stages: document planning, microplanning and realisation. At the document planning stage, decisions concern what information should be included in the final text and what the overall structure of the text should be. The microplanning stage is concerned with how to express the information linguistically. Finally, the realisation stage is where decisions are made about how to generate actual text (Reiter 2013, pp. 2e–3).

The parallels between NLG choice making and journalistic decision-making are arguably quite evident. Diakopoulos (2019, pp. 98–100) points out that, in news automation, the choices made at the document planning and microplanning stages can be more or less technologically sophisticated but do boil down to “editorial priority” (p. 99). The realisation stage is linguistically important but not as relevant from an editorial point of view (p. 101). Less sophisticated decision-making includes pre-written templates to which information is added based on rather straightforward rules. More sophisticated systems use sets of rules that include the ability to perform linguistically advanced tasks such as correctly conjugating words (p. 98). At present, most news automation utilise some variation of the traditional NLP process described here; more advanced technologies in the vein of artificial intelligence are yet to make their way into the news media (Lindén et al., 2019, p. 44). While for example various styles of machine learning make it possible to take a statistical approach to NLG, these systems are inhibited by their unpredictability (Diakopoulos, 2019, p. 101). Instead, the combination of templates and statistical approaches has yielded promising results and may prove useful (*ibid.*).

So, “How should an NLG system be designed to meet journalistic requirements?” This is the question that Leppänen et al. (2017, p. 188) set out to answer. Of course it is also the question that a growing number of news organisations have been asking themselves for the better part of the 2010s. Leppänen et al. (2017) describe “an architecture for a journalistic NLG system” (p. 190) that was later realised for the purpose of covering the 2017 Finnish municipal elections (p. 191). The NLG applications used by news organisations obviously vary a lot, but since the architecture laid out by Leppänen et al. provides a good overview of the general idea, it is summarised here.

Their original architecture was designed to be application independent, meaning that the components are usable in different domains, political elections being one among many potential areas of application. The first step was to create a “Generic representation of facts”. At this stage, data is structured as fact triplets containing information about some of the standard journalistic questions who (entity), where (location) and what (value). Each fact then becomes a sextuplet since every part is a pair: for example, location is coupled with a definition of its type. Thus, the first half of the pair would be location type, e.g. “municipality”, and the other half the actual location, e.g. “Jakobstad” (a small municipality on the western coast of Finland). Facts are represented by language templates that can be combined into sentences and paragraphs. In a “Data-driven, Application-Independent Fact Ranking Method”, newsworthiness is calculated based on two factors: outlieriness and importance. These are determined based on what weighting is given to each part (entity type, location type, value type) of the fact triplets.

These stages are preparations for the final document plan, in which the calculated newsworthiness determines the structure of the text. This follows the so-called inverted pyramid structure of journalistic text in which the most important (newsworthy) fact is presented followed by supporting facts. Measures were also taken to make the system language-independent, which means isolating language-specific aspects and creating templates for each desired language. Finally, in the microplanning stage, the small templates representing a fact are used to construct the story (Leppänen et al., 2017, pp. 190–191). This system was used to report on the 2017 Finnish municipal elections. Through a website launched on election night, the system attracted 398 unique users accessing 573 unique stories. About 60% of the stories were viewed only once, a clear indication of the personalisation potential of news automation (Leppänen et al., 2017, p. 193).

Broadly speaking, news organisations look to news automation for two main reasons: in order to reduce the amount of time journalists spend on repetitive tasks, and in order to produce higher volumes of content (Lindén et al., 2019, p. 5). Graefe (2016) lists six distinct potential advantages of news automation: speed, scale, accuracy, objectivity, personalisation and news on demand. Speed and scale both stem from the fact that a news automation system can instantly produce any number of articles that the data allows for and as soon as the data is

available. For example, using NLG software the Los Angeles Times were able to report on every earthquake instead of only the ones that reached a certain magnitude. Enhanced accuracy stems from the fact that a correctly written algorithm does not make mistakes. However, due to the inability of algorithms to adapt, unexpected events always risk compromising the accuracy of news automation systems. The rule-based logic of algorithms can be seen as a way to remove bias, thus achieving more objective reporting, but this assumes no bias was programmed into the algorithm and that the data do not contain any bias. Personalisation is achieved by algorithmically tailoring stories according to what information is relevant to a certain audience or even individuals. News on demand follows a similar logic: as long as the data is there, the content can be produced instantly (pp. 22–27).

From a business perspective, providing more content means expanding one's potential audience. The content can be monetised either by attracting advertisers through better audience metrics or by converting readers to subscribers (Diakopoulos, 2019, pp. 115–116). Business-wise, other benefits include reducing labour costs by outsourcing some of the news production to external tech actors, and producing content tailored to gain visibility on Google or other search engines (ibid.).

The limitations of news automation are closely associated with its potential. Access to data is what makes news automation possible but it is also the most critical limiting factor (Diakopoulos, 2019, p. 117; Graefe, 2016, p. 27). The algorithms themselves come with their own inherent limitations: they follow certain rules and are therefore inflexible (Graefe, 2016, p. 27). As Diakopoulos (2019) puts it: “Automation fundamentally lacks the flexibility to operate beyond the frontier of its own design” (p. 122). Moving beyond descriptive accounts of some event or phenomenon is difficult for a news automation system; since algorithms lack common sense and context, news automation systems are ill-equipped to provide interpretations and explanations of any kind (Diakopoulos, 2019, p. 124).

In the literature on news automation, the ground zero example of a journalistic NLG system would be the quarterly earnings reports that Associated Press (AP) started automating in 2014 using software from external NLG provider Automated Insights and data from Zacks Investment Research. Automating the writing of earnings reports allowed AP to increase the amount of stories tenfold while at the same time reducing the error rate (White, 2015). The

other major NLG software provider in the United States, Narrative Science, started out automating college baseball games recaps in 2009 (Levy, 2012). By 2012, Narrative Science software was being used by business magazine Forbes to produce earnings previews (Graefe, 2016, p. 20). In 2013, the investigative journalism organisation ProPublica's Opportunity Gap news application utilised the software to generate descriptions of more than 52,000 schools (ibid.).

In the United Kingdom, the project RADAR (Reporters and data and robots) evolved into a subscription-based local news agency in 2019. RADAR produces local news stories using software from external NLG provider Arria, human-authored text templates and public data from sources such as the NHS (UK's national health service). The operation now includes RADAR's own management system for storing and distributing data as well as personalised distribution according to the customer's areas of interest (Lindén et al., 2019, pp. 18–19). In the Nordic countries, Swedish software provider United Robots has developed a system for automatically writing Swedish language stories on a number of different team sports. Since 2017, the company has started automating stories on bankruptcies and the real-estate market (Lindén et al., 2019, p. 15). United Robots' sports software Rosalinda is also being used in Finland (ÖT, 2016). Some of the other notable automation efforts in Finnish newsrooms are in the sports reporting domain but not exclusively so. The public broadcaster Yle's system Voitto writes about domestic ice hockey games and was developed in-house (Vehkoo, 2017). During the 2019 Finnish parliamentary elections, Voitto was also used to produce stories for each municipality in Finland (Yle, 2019). In a joint project, University of Turku used the news agency Suomen Tietotoimisto's news archive and machine learning to develop a system for reporting on ice hockey and floorball called Scoopmatic (Simola, 2019).

3.1.2 What it means

So how did we get here? And by "here" I mean the point where automation and algorithms are ubiquitous in journalism. The precision journalism that Philip Meyer introduced in the 1970s is often cited as a progenitor for the different types of data-driven journalism (e.g. Gynnild, 2014, p 718). The idea Meyer (1973) championed was, in essence, to bring statistical analysis and empirical methods from the social sciences to journalism. The term Meyer used was computer-assisted reporting. However, this approach is in no way unique to

or even necessarily at the core of news automation. In fact, Hammond (2017, p. 410) proposes that data journalism in the age of big data might even be doing the opposite of what Meyer set out to do. Instead of hypothesising and interpreting, Hammond argues, modern day data journalism is preoccupied with letting the data speak for itself (p. 412). Hammond goes so far as to call this an “epistemological shift” (p. 412). This notion is echoed by Anderson (2018) claiming that, compared to CAR, the way data journalism works today is “reversed” (p. 3). Here, both Hammond and Anderson speak broadly enough to allow us to ignore the distinctions between data journalism and news automation laid out above. If anything, the industrial level that automation brings to news production means that it is even closer related to the concept of big data (Hammond 2017, p. 413).

I would argue that the term big data, in part since it escapes definition (Mayer-Schönberger and Cukier, 2013, p. 6), is not very useful in the context of news automation (for a more in-depth discussion on big data and journalism, cf. Lewis and Westlund 2015). What matters, however, is the datafication of society. Datafication refers to the process of quantifying any phenomenon imaginable (Mayer-Schönberger and Cukier, 2013, p.15) and it is what drives the rapidly growing amounts of data and the growing possibilities to do something with this data. Approaching news automation from a quite practical point of view, Leppänen et al. (2017, p. 195) cite data availability as one of the requirements. Listing the limitations of news automation, Graefe (2016, p. 27) also puts data availability up front and emphasises the importance of data quality and structure in any functional news automation process.

An important difference between news automation and other kinds of data-driven news production is that news automation demands a continuous flow of data. Whereas other kinds of data journalism might rely on multiple types of sources and methods (cf. Borges-Rey, 2016 for an overview), the content produced using news automation is often based solely on data (Diakopoulos, 2019, p. 120). With the datafication of society, the amounts of public data are larger than ever before. Automation offers a potent way of journalistically monitoring and analysing all this publicly available data on for instance unemployment or school results (Magnusson, Finnäs and Wallentin, 2016). Among European news agencies for instance, sports and finance are still the most prominent topics for news automation (Fanta, 2017, p. 13). But news automation is being explored in a number of different domains, as for example crime (Young and Hermida, 2015) and air quality (Dierickx, 2018), the latter of which also

represents so-called sensor journalism (cf. Koerth-Baker and Carlson, 2017). The fact that news automation depends on a continuous influx of new data means that so-called “[s]treams of data” (Diakopoulos, 2019, p. 118, emphasis in original) are particularly interesting. Pipelining public data or data measured using sensors are examples of such streams. Apart from using sensors of their own, news media can contribute to the datafication of society by quantifying existing reporting. This practice is referred to by Caswell and Dörr (2018) as “event-driven narratives”, with stories about car chases as an example of stories following patterns clear enough to be broken down into discrete quantities.

If we go back to Hamilton and Turners definition of computational journalism, the other key aspect besides data, is algorithms. In an early overview of research on computational and algorithmic journalism, Anderson (2013) highlights the possibilities of big data and goes on to express a desire to come to terms with what is considered an “algorithmic revolution in knowledge production” (p. 1006). Napoli (2013) sees an “algorithmic turn [...] taking place in the media sector” (p. 3) and emphasises that this algorithmic turn is felt throughout the media industry. Social media and streaming services are prominent examples, with algorithms being used in the consumption as well as the production of content (Napoli, 2013, pp. 10–15). News media and journalism is merely one area among many in which the algorithmic turn of the media sector is being felt but, as Napoli (2013, p. 16) argues, it might perhaps be the most controversial.

A good way to untangle the potential controversy here is to recall that journalism professionals aim to “self-legitimize their position in society” (Deuze, 2005, p. 446). One of the key values of the ideology of journalism discussed by Deuze is ethics (p. 449), which in turn is closely associated with the concept of transparency (Allen, 2008, pp. 323–324). Although transparency is not an unchallenged virtue in journalism, it may function as a way of strengthening legitimacy (ibid.). Inversely then, a lack of transparency might serve to weaken legitimacy. In an attempt to address this, Diakopoulos and Koliska (2017) emphasise the need for algorithmic transparency in news media. Simply put: journalistic decision making using algorithms is at odds with transparency, which means that the premise of the news product is blurred and this forces news organisations to think about how to disclose information about their algorithms (Diakopoulos and Koliska, 2017, p. 813). In the same vein, Dörr and Hollnbuchner (2017) identify a number of ethical challenges for news

automation: At the input level, questions concern data quality and origin. What are potential biases in the data? How are sources being handled? Are there privacy issues? At production level, the processing of data is merely the first concern; since algorithms are used to produce the content, the code itself becomes a potential source of bias. Here, the challenge for newsrooms is at least twofold: making sure the algorithms meet journalistic standards, and communicating the decision-making process to the audience (pp. 412–413). Thus, control and transparency, the two issues that Deuze (2005, p. 455) identifies when considering how societal and technological developments affect journalism, are both evident in news automation's use of algorithms.

In recent years, the body of research on news automation has notably expanded as digital journalism studies have approached the phenomenon from a number of different angles. The work done by Leppänen et al. (2017) represents a technological point of view rooted in computer science and concerns the technological prerequisites for guaranteeing journalistic standards in news automation (cf. section 3.1.1). Others, such as Latar (2018), explore what news automation and the larger developments surrounding it might mean for human journalists. Latar (2018) argues that, despite their limitations, “robot journalists” might tempt media organisations to replace human journalists (p. 38). This view is in stark contrast to the much less fatalistic one expressed by Lindén (2017), who concludes that any such risks are lessened by journalists' high adaptability, and that journalism as ideology will probably continue to soften any blows that technological advancements might deal to the profession (p. 136).

Then there is the audience perspective, where scholars examine audiences' attitudes towards automatically produced content. A 2014 study by Clerwall found that the only statistically significant difference between how audiences perceive automatically and human written news texts was that the human written ones were more pleasant to read (pp. 525–526). Subsequent experimental studies have also shown that audiences find human-written news texts more likeable (Melin et al., 2018, p. 43360) and readable (Graefe et al., 2018, p. 604). In other aspects, such as credibility (Wölker and Powell 2018, p.12) and expertise (Graefe et al., 2018, p. 604), human-written texts have not been found to outperform machine-written ones. Melin et al. (2018) on the other hand found that human-written texts statistically significantly outperformed machine-written ones in every aspect, but in all cases except likeability only

slightly. Potentially, there might be a cultural element involved: Jung et al. (2017) showed that in South Korea, readers preferred the texts they were told were machine-written. The texts used in all of these studies however, were written on completely fact-based (data-based) topics. This is where news automation systems perform well and as Graefe et al. (2018, p. 605) note, the results are not generalisable to other kinds of content.

Where does the research go from here then? First of all, despite all the hype around AI, the simple fact that automation is good for answering the “what” of journalism but poor when it comes to the “why” makes media organisations approach it with caution (Sirén-Heikel et al., 2019, pp. 61–62). What is happening when automation is introduced in the newsroom though, is that the roles are changing. Journalists turn to (a) calibrating the systems and monitoring news production as a whole and (b) doing in-depth reporting (Wu, Tandoc and Salmon, 2019a, pp. 1454–1455). The introduction of automation also has implications for the power dynamics of news production. As Wu, Tandoc and Salmon (2019a) conclude, autonomous systems do shape editorial decision making (pp. 1453–1454) and effectively “become the co-creators of news, capable of mediating and altering the final news product” (p. 1453). The power dynamics are further complicated by the fact that most newsrooms do not possess sufficient knowledge of data or NLG to handle news automation by themselves, forcing them to rely on outside expertise for these purposes (Sirén-Heikel et al., 2019, p. 60). In addition to the transparency issues discussed previously, what is potentially at stake here then is autonomy, a core value in the occupational ideology of journalism as described by Deuze (2005, p. 447).

3.2 Media innovation

Throughout the twenty-first century, legacy news media have been adapting to digitalisation (Paulussen 2016, p. 193), but due to factors such as difficulties reshaping old business models and lacking ability to understand technological developments, they have usually been reactive rather than proactive (pp. 194–195). In the last ten years, scholars have started to seriously address this. Pavlik (2013) arguably paints with a broad brush when stating that “innovation is the key to the viability of news media in the digital age” (p. 190), but goes on to identify three areas in which innovation has proven successful: “(1) creating, delivering and presenting quality news content; (2) engaging the public in an interactive news discourse both

in terms of citizen reporters and in concert with social media where a public dialog can easily occur online; and (3) employing new methods of reporting optimized for the digital, networked age” (Pavlik, 2013, p. 190). All of these areas are still quite broad but can be narrowed down, following Westlund and Lewis (2014), by adding that merely producing content is not enough. Westlund and Lewis (2014) argue that true media innovation extends to editorial processes, business models and organisational structures (p. 11). Then the third area listed by Pavlik becomes highly relevant from a news automation perspective.

In his seminal book on innovation in news media, *Digitizing the news: Innovation in Online Newspapers*, Boczkowski (2004) noted that newspapers’ transition to the web happened carefully, defensively even (p. 51). However, while finding that a lot of the early web endeavors happened reactively (p. 171), Boczkowski rejects a technologically deterministic explanation model and concludes that news organisations display a clear agency of their own in the innovation process (p. 177). This agency is related to journalism as an ideology and can, in part, function to slow down innovation by questioning its impact on editorial independence (Deuze, 2005, p. 449). More recent research has reinforced that “the primary brakes on media innovation, for better and for worse, are incorrigibly systemic” (Ess, 2014, p. 2).

From a news automation perspective, the most relevant technological developments are the ones even more recent than the transition to the web of the late nineties and early 00s, mainly those concerning datafication (described in section 3.1). But following Boczkowski’s (2004, p. 178) recommendation, I want to do a little historicising. It is useful to keep in mind that the digital disruption in journalism is merely the latest incarnation in a long line of communications revolutions dating back hundreds of years. Kovarik (2011, p. 1) calls digitalisation the fourth mass media revolution, the three previous ones being the printing revolution that set it all off in the middle of the 15th century, the visual revolution spurred by – among other things – the popularisation of photography in the 19th century, and the electronic revolution that brought us radio and television. It makes sense then to view digitalisation as part of a continuous development, and as something continuous in and of itself. When comparing the ongoing disruption brought on by the internet to the computerisation of the newsroom in the second half of the 20th century, Mari (2019, p. 94) notes that new ways of working might have been inevitable but that they did build on the old

ways nonetheless. Mari contrasts this McLuhanesque viewpoint by stating that journalists were in fact themselves driving and shaping the computerisation of the newsrooms (ibid.). While on the topic of computerisation of the newsroom, we should also note that news automation is not necessarily inherently different from the previous phases of the digital revolution. One end goal of even the earliest computerisation initiatives of the mid 1900s was to get rid of redundant steps of news production (Mari, 2019, p. 32).

Reviewing research on newsroom innovation, Paulussen (2016) identifies and divides into two categories a few key factors that influence the innovation process (p. 196). The first category, *organisational structures*, includes newsroom integration and resource reallocation. Newsroom integration refers to how newsrooms reconcile the old with the new, accommodating elements such as market strategy and work organisation. Resource reallocation includes strengthening the organisation's technological pedigree. This can be done structurally by adjusting workflows and investing in infrastructure, such as content management systems. It can also be done by enhancing the organisation's know-how by recruiting and training staff. In the reallocation of resources, the business view and the journalistic view have often been found to be at odds with one and other (p. 197). The second category, *professional cultures*, includes habits and role perceptions. Habits are simply some work-related routines that are more or less taken for granted, which for example may cause older media logics to retain more influence than newer ones even when innovation is introduced (p. 198). Role perceptions are built on the professional norms that govern how journalists perceive their professional role. Here – crucially – the ideologically rooted desire of journalists to control news production function to inhibit innovation (p. 199).

These are some of the factors we expect to be at play when news organisations venture into news automation. But the perspective laid out above is merely part of it, because neither newsrooms nor newsroom innovation exist in a vacuum. When external actors, such as software providers, are involved in the innovation process, the picture we set out to examine is a larger one. On a macro-level, Krumsvik et al. (2019) propose, there are three main categories for the influences on media innovation: “(1) media institutional factors; (2) technological developments; (3) sociocultural conditions and power relations” (p. 196). The first category essentially covers the same factors presented by Paulussen (2016) but notably adds media legislation, e.g. the state support model found in European countries.

Technological developments, while almost aggressively obvious a factor in the context of news automation, must be explicitly stated as something that forces or enables innovation. This does not mean that we get to ignore the hybrid media logics of Chadwick (2017, pp. 22–26) though. It is arguably power rather than technology that functions as the driving force (Trappel, 2015, p.19), hence the third category. The sociocultural conditions and power relations are also acknowledged by Bruns (2014) who argues that innovation studies must take into account that “media, media practices, social practices, and society itself are inherently and inextricably entwined in contemporary mediatised society” (p. 24). The collaboration between news organisations and software providers represents one such area in which the “entwinedness” is made manifest, and in order to understand this, we need to introduce the concept of peripheral actors.

3.3 Peripheral actors

A solid body of work exists on how innovation is driven by the newsrooms themselves (cf. Paulussen 2016). But as we have seen in the previous section, innovation also involves what can loosely be described as peripheral actors, the journalistic strangers conceptualised by Holton and Belair-Gagnon (2018). Moving forward, it is worthwhile to keep in mind that some of these actors, including the software providers for news production, are potentially directly involved in the news production through their innovations. This development needs to be taken seriously too, since in terms of responsibility it marks a “shift from the individual [journalist] level to the level of the organization or third parties” (Dörr and Hollnbuchner, 2017, p. 413).

In recent years, journalism scholars have started paying attention to peripheral actors in journalism. Some research perspectives include the impact of philanthropic foundation funding on journalism practice (Ferrucci and Nelson, 2019), “trading zones” such as the code-sharing platform GitHub as a pathway for peripheral actors (Haim and Zamith, 2019), and the inability of scholars to adequately examine the valorisation as well as suspicion encountered by peripheral actors in sub-Saharan Africa (Wahutu, 2019). In the context of this study, the most relevant perspective is the one of peripheral actors who directly influence news production through technological innovation.

Using Holton and Belair-Gagnon's typology of journalistic strangers, Chua and Duffy (2019) examined how journalists perceive peripheral actors' involvement in news production. They found that the interviewed journalists had gradually come to acknowledge the importance of the peripheral actors' input (p. 119). Experts in particular were welcomed by journalists as these represent so-called "respected implicit interlopers" (p. 116). The implicit interlopers were nonetheless perceived by journalists as professionally separate from themselves, being characterised as "frenemies" even (p. 116). The authors identified four forms of proximity through which innovation involving peripheral actors can be understood (pp. 119–120). Physical proximity refers to physical distance: journalists and coders sitting in the same room tend to view each other more favourably. Temporal proximity means that the more time spent working together, the better actors understand each other. Professional proximity refers to how much the overall goals associated with different professions overlap. For example, are commercial goals emphasised by one actor in conflict by editorial credibility emphasised by another? Here, the innovation at the core of technology and the consistency at the core of journalism are thought to be "a key barrier to normalisation of innovation" (p.120). Finally, control proximity is about the involved actors acknowledging authority. As one group, such as management, takes control over some process, proximity occurs as a result of "clear lines of engagement" (p. 120).

Up until recently, peripheral actors' influence has mostly been examined from the newsroom's point of view (Belair-Gagnon and Holton, 2018, p. 492). But in order to fully grasp "how journalistic strangers ... may be challenging the epistemologies of journalism" (Holton and Belair-Gagnon, 2018, p. 72), we also need a less newsroom-centric approach. Presenting such an approach, Domingo and Wiard (2016) argue that it is important to listen to how the actors involved in news production describe themselves since "[t]hey may not define their engagement with news as journalism, but their practices may be significant in their interaction with others to shape the news coverage" (p. 402). Ananny and Crawford's (2015) study on news app designers is a good example of an attempt to understand news production from the perspective of the peripheral actors. The interviewed app designers were found to work "in a space between technology design and journalism, influenced by both but not entirely beholden to either" (Ananny and Crawford, 2015, p. 204). The authors call this a "liminal press" (p. 204), operating through what Chadwick (2017) would call hybrid media

logics. Within this “liminal press”, older and newer media logics mix, resulting in a redefinition of how news production functions (Ananny and Crawford, 2015, p. 204).

Similar results are presented by Belair-Gagnon and Holton (2018) examining how web analytics companies position themselves in relation to news production. Belair-Gagnon and Holton (2018) found that these companies try to understand and shape news production but do not claim journalistic responsibilities of their own. Through the technology they introduce, these peripheral actors influence the norms and values of the newsroom, pushing them towards a more profit-oriented focus (p. 505). The managers of the web analytics companies were found to assume three main roles in their cooperation with journalists: disruptors, connectors and routinisers. As disruptors, they bring new insights and frame these as necessary to journalism (p. 499), thus reshaping news production, for example by disrupting editorial decision-making (p. 500). As connectors, their role is twofold. They connect newsrooms with audiences (via analytics), but they also connect newsroom professionals with external expertise (in analytics). This second type of connecting encourages experimentation, which means that journalistic practice becomes connected to the economic goals of the web analytics companies (p. 501). As routinisers, they work on the newsroom’s terms, according to its norms and values in order to make the use of their products more routine. This entails providing services such as training, support and regular check-ins, while still claiming the position of an outsider (pp. 501–502). These findings indicate that the same type of actor can function as different types of strangers, since these roles closely resemble explicit interlopers (disruptors), implicit interlopers (connectors) and intralopers (routinisers).

As far as service providers for news automation, this kind of peripheral actor is yet to be closely examined as such. Wu, Tandoc and Salmon (2019b) apply field theory to examine automation in journalism from the point of view of technologists. The study uses a broader definition of news automation that encompasses other steps of news production and distribution besides NLG, such as data gathering and analytics (p. 1241), but the research questions concern how technologists perceive the journalistic field and the influence of automation on said field (p.1243). The relationship between the tech field and the journalistic field can be said to revolve around the precarious economic circumstances many news organisations experience. The pressing need to find sustainable business models makes media organisations more easily influenced (pp. 1244–1245) and tech actors understand this

situation well (p. 1250). At the same time, technologists experienced that financial restraints make smaller organisations reluctant to adopt automation, while the larger ones are held back by inflexible organisational structures (p. 1249). Technologists also thought that media organisations' concerns regarding the reliability of the technology is connected to an inability to understand it, and that this was another limiting factor (*ibid.*), thus expressing the same sentiments as news media representatives themselves (Sirén-Heikel et al., 2019, p. 60).

The main study focusing particularly on service providers with NLG as their primary area of expertise is a 2016 overview by Konstantin Dörr. Dörr (2016) identified 13 service providers in Europe and the United States with NLG as their main business (p. 712). The companies employed between 5 and 80 people (Dörr, 2016, pp. 713–715). Based on size, clients and products launched, Dörr identified Automated Insights and Narrative Science as the most influential companies in the United States, Retresco and Aexea in Germany, Syllabs in France and Arria in the United Kingdom. One Chinese provider was also identified, but too late to be included in the main study. Out of the companies that were included, 10 created products with journalistic application.

These companies also provide services to a number of other markets besides journalism, such as e-commerce, finance and healthcare (Dörr, 2016, p. 712). According to Dörr (2016), the desire not to depend on one market segment likely stems from “limited availability of data, the time-consuming individualization of journalistic products due to high-quality standards in journalism, and the general view that journalistic products alone are hardly profitable” (p. 712). The unfavourable financial conditions cause even well-established software providers to experience uncertainty and have a hard time making headway in the media market; the products that were launched focused mainly on sports and finance due to the data availability and scalability of these domains (Dörr, 2016, p. 716). Looking again to the findings of Wu, Tandoc and Salmon (2019b), one reason actors from the tech sector nonetheless wish to work with news media is that media organisations can be of assistance in developing technology, both through analysis of the content they produce and as innovation collaborators (p. 1245). Perhaps more interestingly, tech actors view news media as a good arena for giving their products visibility. Tech actors showing that “they understand journalism’s influence over other fields” (Wu, Tandoc and Salmon, 2019b, p. 1250) again indicates that the relationships between journalists and strangers are closely associated with hybrid media logics.

4. Method

This section begins with a general discussion of the reasoning behind my methodological choices. I then highlight a few important aspects of the qualitative research interview. This is followed by a description of how the interviews conducted for this study were designed and executed. Finally, the method for data analysis is described and discussed.

Keeping in mind the less newsroom-centric approach described by Domingo and Wiard (2016), a good starting point seemed to be to “listen to the actants [or actors] and how they describe themselves and the world around them” (p. 402). For this, I use *semi-structured life-world interviews*, the purpose of which is “obtaining descriptions of the life world of the interviewee with respect to interpreting the meaning of the described phenomena” (Kvale, 2007, p. 8). Given that in the modern world, the prevalence of interviews has reached such levels that Gubrium and Holstein (2002, p. 9) argue that we now live in an interview society, the decision to conduct interviews in order to gather information might seem almost self-evident. Nonetheless, since appropriateness can be seen as a useful guideline for conducting qualitative research (Flick, 2007, p. 4–5), I will be discussing the appropriateness of this decision. Throughout this section, this discussion of appropriateness will, following Hennink, Hutter and Bailey (2011, p. 58), serve as a discussion of the *quality* of the study. In the spirit of (among others) Kvale (2007, p. 123), I use the quality concept here as a stand-in for validity. I return to issues of validity, reliability and generalisability in section 6.2.

While some aspects of news automation are relatively well-researched, the point-of-view of service providers is not. The unassuming nature of interviews as a research method make them effective for approaching a largely unexplored research area (Esaiasson et al., 2012, p. 253). Extending this logic, Lindlof and Taylor (2011) list as one of the basic reasons for interviewing the possibility of obtaining otherwise unavailable information (p. 173). And for the purpose of “understanding the social actor’s experience, knowledge, and worldviews” the interview is a particularly useful tool (Lindlof and Taylor, 2011, p. 173). Not to be underestimated either is a different kind of appropriateness of the research interview: for small-scale projects, the interview is all but unmatched as a time- and cost-effective way of examining “complex and subtle phenomena” (Denscombe, 2014, p. 186).

4.1 Semi-structured research interviews

Next, I will briefly discuss qualitative research and a few important theoretical and practical implications of the semi-structured research interview. I then go on to describe the setup for the interviews conducted in this study.

Denzin and Lincoln (2018) define qualitative research as the attempt to “study things in their natural settings, attempting to make sense of or interpret phenomena in terms of the meanings people bring to them” (p. 10). From an epistemological perspective, qualitative research generally adopts a constructionist viewpoint as opposed to a positivist one. As rough as this distinction might be, constructionist assumptions are at play when we use qualitative methods such as interviews (Flick, 2007, p. 11–13). Semi-structured interviews are concerned with *subjective viewpoints* and the theoretical positions for this approach are symbolic interactionism and phenomenology (Flick, 2007, p.10–11).

In this study, I adopt a phenomenological position. This supports an exploration of how things are experienced (Giorgi, 1997, p. 236–237) while simultaneously recognising that human knowledge is socially derived (Eberle, 2014, p. 186–187). The life-world of the semi-structured life-world interview does not refer to an entirely private realm – it is shared and situated in time and space (Justesen and Mik-Meyer, 2011, p. 18). This socially shared life-world can be considered the *natural setting* referred to in Denzin and Lincoln’s (2018) definition of qualitative research. For the social sciences, the collective nature of the life-world is key. In the most practical of terms, it means that the knowledge gained in interviews not only represents the point-of-view of a single person but reflects the collective knowledge of the social context described by the interview subject (Justesen and Mik-Meyer 2011, p. 55). In this study, that context is primarily the companies themselves and perhaps secondarily the social context in which news production takes place.

Moving on to describing the specifics of interviewing, Gillham’s (2005, p. 70) summary of the semi-structured interview provides a good overview: All interview subjects get asked the same questions and all the interviews are carried out in roughly the same amount of time, in order to allow for a purposeful comparative analysis. Open questions are used to allow open answers, and the interviewer intervenes if they determine that an answer needs to be

expanded upon (this is also done to make sure all topics are covered; Gillham calls these interventions *probes*). The questions used are developed through a process intended to “ensure their topic focus”.

Next, I will describe my research design via what Kvale (2007, p. 35–36) calls the “Seven stages of an interview inquiry”:

1) Thematizing

Here, we answer the questions *why* and *what* before moving on to the *how*, namely the method (different approaches to interviewing). The *why* refers to the purpose of the study. The *what* refers to previous knowledge of the research subject. The interviews used in this study were carried out by a group of people, including myself, a PhD student and a senior researcher. Due to a tight interviewing schedule, I initially had to rely on my colleagues for the *what* part and, under their tutelage, quickly catch up on previous research. These circumstances also meant that for this study, the *why* actually developed gradually. Little by little, what was initially a collective *why*, became a personal *why*. This *why* was obviously related to the initial one, but was informed by a steadily solidifying *what* on my part. Ultimately, I was able to transition from the more general research questions used to design the interviews to the specific research questions used in this study. Kvale (2007) understands this process as a part of “the temporal dimension of an interview design” (p.41) simply called “getting wiser” (p. 42) and points out that this is what an exploratory study is all about. These iterations also allowed me to continuously evaluate the study design thus ensuring the validity of the final outcome. The temporal dimension refers to practical aspects of the research process, such as maintaining an overview of the research design, ensuring the interdependence of its different parts and going back before moving forward.

2) Designing

Now that we have already touched upon the design part on a general level, it’s time to describe the process of preparing for the interviews. Since this was largely a group effort, I will change pronouns, meaning that the more abstract we occasionally used so far will temporarily be replaced by an actual we. A lot of care was put into the formulation of

questions. For these purposes, working as a group is a good way to enhance interview quality (Kvale, 2007, p. 139) and the iterative process of discussing and revising was indeed very beneficial.

Originally, the interview guide leaned heavily towards what Kvale (2007, p. 71) calls a *conceptual interview*, in which one attempts to get the subject's take on phenomena of interest. In this case, some of the phenomena were news automation, journalism, journalists and news organisations. It was also a lot closer to an unstructured interview than what ended up being the case. After some crucial feedback from senior researchers, we decided to take a different route. The guide was revised, keeping the initial ideas but restructuring the interview. The new structure was built around a specific project that the company had carried out together with a news organisation and in which the participant had been personally involved. The number of sections and questions remained the same, but instead of themes, the five main sections now concerned different phases of the project (which the subject was asked to choose in advance), the end-goal being to get more detailed descriptions. Many characteristics of the conceptual interview remained, but since the topic now concerned specific events experienced by the participants, the focus came to lean a bit more towards the *narrative interview* (Kvale 2007, p. 72–73). The final five main sections were the following: project background and start, designing the product for media organisations, implementing the product, looking back at the project, and self-reflection and future.

Having prepared the revised guide (cf. Appendix A), we also opted for a pilot interview, which is another way to strengthen interview quality (Kvale, 2007, p. 139). As the group's least experienced researcher, carrying out the pilot interview was a useful way for me to get practice and feedback. At the same time, the pilot interview provided crucial insights into the usefulness of the questions, another instance of ensuring validity (quality) by evaluating study design.

3) Interviewing

Between March and May 2019, nine interviews were conducted with representatives from NLG service providers working with news organisations (cf. section 4.2 for details on the participants), two in Swedish and seven in English. In this study, six interviews were

analysed. For logistical reasons – a tight time frame and wide geographical spread of the participants – all interviews were conducted and recorded via video calls. Since the participants are experts in their field, the interviews can be classified as elite interviews. This kind of interview poses certain challenges, the main one being that experts might be used to being interviewed and have the ability to steer the interview using pre-prepared *talk tracks* (Kvale, 2007, p. 70). I would argue that the open format of the semi-structured interview makes this particular challenge even bigger. In order to account for this, our planning also included carefully structuring the interview guide. During the interviews, however, our team only noticed a few instances of talk tracks being used. After the first few interviews we decided to let the talk tracks slide early on in the interviews in order for things to get going, and only later start directing the flow of the interview more carefully, a strategy that we found sufficient.

4) Transcribing

The English interviews were transcribed by Rev (a company that specialises in audio-to-text services) and the Swedish ones by Tutkimustie (a company that provides supporting services for research). In order to guarantee the accuracy of the transcriptions, they were compared to the original audio recordings and edited. In many instances this turned out to be essential since the domain knowledge possessed by our team helped sort out misunderstandings that inevitably occur in the transcription process. Having done most (five) of the interviews myself and listened to all the others more than once was important, since transcriptions alone are not in fact the be-all end-all elements for analysis although they are often considered to be (Kvale, 2007, p. 92–93).

5) Analyzing

The mode of analysis appropriate for each study should be based on what is being studied and what the aim of the study is as well as the material itself (Kvale, 2007, p. 36). In this study, the interviews were analysed using a phenomenological approach (cf. section 4.4).

6) Verifying

In this part of the process, questions concerning validity, reliability and generalisability are addressed (discussed in section 6).

7) Reporting

Findings are to be reported according to scientific standards (cf. section 5 for results). This includes adhering to the ethics of scientific research. The ethical considerations of this study are discussed in section 4.3.

4.2 Participants

The software providers that participated in this study were chosen from a shortlist of contacts ensembled in previous research projects (notably *Immersive Automation*, cf. Lindén et al., 2019). Interview requests were emailed to contacts from eleven different companies in Europe and the United States that fit the profile: NLG companies working with news organisations. The companies were informed about the purpose of the interview and asked to use this as guidance in choosing a suitable staff member to participate. Companies were explicitly asked to choose representatives that had been working directly with journalists. Striving for gender balance we also requested them to prioritise female participants and ended up with two female participants. In the final sample of six interviews analysed in this study, one participant was female. Representatives from nine companies agreed to participate, which is quite remarkable given that gaining access is a common issue in expert or elite interview studies (Kvale, 2007, p. 70). The high response rate can be ascribed to the fact that connections to the companies were already established through extensive networking and highlights some very practical benefits of participating in conferences and other industry gatherings.

All of the service providers had experience from working with news organisations. Most but not all of the companies also work with partners from other sectors. A review of the service providers' websites showed that they all emphasised data processing and NLG as parts of their services. Most of them mentioned providing AI, although the interviews showed that at

least in their work with news organisations, none of them used what Leppänen et al. (2017) refer to as “end-to-end architectures” (p. 189). In order to prepare for the interviews, we asked the individual participants to provide information on the product or service their company offers news organisations. The participants were also asked to answer three questions regarding their own educational background, previous experience and work with their current company. Three participants mentioned having experience from working in journalism. Their educational backgrounds were quite varied, ranging from computer science to social science, business and linguistics.

In order to get a sample that was of high quality and at the same time easily manageable for analysis, six interviews were chosen for the purpose of this study. This meant following Szklarski’s (2019) recommendation and discarding some interviews that did not entirely fit the scope of the investigation (such as the pilot interview). In practice, this process went the other way around: the interviews that best fit the scope of the study were chosen and analysed until saturation, i.e. the point when no new relevant information is found (Hennink, Hutter and Bailey, 2011, pp. 88–89), was reached, leaving out the remaining interviews at this point.

4.3 Ethical considerations

Social scientists have a duty to ensure that their research does not harm “individuals, communities and environments” (Israel and Hay, 2006, p. 3). While for-profit companies might not necessarily enjoy the same rights to be protected as, say, marginalised groups such as minorities, there are other reasons to conduct ethical research. One is trust, which is essential for upholding good relations both to the communities being studied and to the general public researchers (c)aim to serve (Israel and Hay, 2006, p. 3–4).

Since the participants of this study represent their companies, they do not necessarily need to be protected on a personal level. But the principle of informed consent, which states that participants need to know why research is being done as well as how it will be reported (Israel and Hay, 2006, p. 61) still applies, not least for transparency reasons. The purpose of this study and the interviews was clearly stated twice: first in the initial email to the service providers and once more before the interviews. Participants were informed that the interviews would be recorded, transcribed and used for the current study and possibly for future studies

on news automation. They were also informed that any direct quotes would be pseudonymised.

The decision not to report detailed information individually for any of the companies or participants is connected with the pseudonymisation. Due to the somewhat sensitive nature of the interviews (asking participants to speak about business partners) I did not want it to be possible to identify any specific person or company. Here, I have chosen to interpret in the strictest sense “The ethical principles of research with human participants and ethical review in the human sciences in Finland” provided by The Finnish National Board on Research Integrity TENK. According to these principles, the researcher’s responsibility to protect their subjects extends to any subject of research, including for example property (TENK, 2019, p. 50), which I take to mean anything from a car to a company. My hope is also that providing this kind of discretion will have allowed the participants to speak more freely.

In order to protect the integrity of the participants, our research team also needed to consider the General Data Protection Regulation (GDPR) of the European Union. We thoroughly discussed data sharing before starting out, the idea being to establish clear and secure practices. Here, we quickly realised that this side of things is largely uncharted territory, GDPR having come into effect less than a year before the start of the research project. We decided that the best practice was to use pseudonyms for all participants from the beginning. Each participant was given a code and a code table document was created as a backup reference, and this document was never uploaded to any third-party cloud service.

4.4 Phenomenological analysis

In the analysis of the interviews, I mainly rely on a phenomenological approach. However, since phenomenology is more than merely a tool for analysing qualitative data (Eberle, 2014, p. 184), it should be noted that I do not aim for my research to be phenomenological in the absolute sense of the word. Instead, I make use of a few key ideas typically associated with phenomenology that pair well with my methodology and theoretical framework. In the following section, I present the aspects of phenomenology that I will be applying in my analysis and discuss their usefulness. Finally, I summarise the analytical procedure by describing the steps used in the data analysis.

Interview analysis is generally made up of three major parts: data reduction, data reorganisation and data representation (Roulston, 2014, p. 301). Quite a few different research practices influence interview analysis; Roulston (2014, p. 301) cites hermeneutics, grounded theory, ethnographic and narrative methods, and phenomenology. The phenomenological approach also constitutes the core when Kvale and Brinkmann (2009, p. 205) describe the “meaning condensation” strategy of interview analysis. According to them, the phenomenologically informed analysis is useful when dealing with “extensive and often complex interview texts” (p. 205). On a data reorganising level, the aim of the phenomenological approach is to extract main themes from the texts. The themes then allow themselves to be further explored using theoretical frames (Kvale and Brinkmann, 2009, p. 207), thereby moving on to the data representation level. Explorations of this kind might in fact be necessary, since phenomenology can be criticised for favouring description over analysis (Denscombe, 2014, p. 103).

Phenomenology offers useful concepts for exploring the phenomena at hand. First we have the *eidōs*, or the essence of the phenomena. If we allow ourselves to indulge in some philosophy, we are, as Eberle (2014) puts it, looking for “the platonic *idea*, the invariant properties that are universal” (p. 185). This might be to push our ambitions a bit far though. A more pragmatic way of looking at things is to simply refer to the essence as “the most invariant meaning for a context” (Giorgi 1997, p. 242). For research purposes, it might also be useful to differentiate between philosophical and scientific essences, the scientific ones being determined by context and scientific discipline (Giorgi, 1997, p. 244–245).

The process of reducing a phenomenon to its essence is called eidetic reduction (Giorgi, 1997, p. 240), a key part of which is the “bracketing of the assumptions of the natural attitude that we regularly rely upon in everyday life” (Eberle, 2014, p. 186). Freeing the analytical process from assumptions means the approach is decidedly inductive (Szklarski, 2019, 151–152), a feature that pairs well with the methodological aim of this study, namely to explore an area of research of which we have little previous knowledge. To go all out Locke and become the *tabula rasa*, the clean slate, is of course impossible. Instead I note that for Moustakas (1994, p. 97) the bracketing process involves keeping the research topic and question in mind. The general idea is to not let previous knowledge blur the vision (Giorgi, 1997, p. 240).

Finally, eidetic reduction includes a related concept called horizontalisation: In the beginning of the analysis, every aspect of the phenomenon at hand is ascribed equal value. Only later do we get rid of irrelevant and overlapping findings (Moustakas, 1994, p. 97).

In my analysis, I use Giorgi's (1997, p. 245–248) general outline of the phenomenological method as a basis. This gives us a five-step procedure (cf. Table 1) in which the fifth step, following Szklarski's (2019, p. 155–159) more refined procedural description of Giorgi's method, can be divided into two parts.

| Step | Actions |
|---|---|
| 1) Collecting of verbal data | Document (e.g. interviews) the experiences of the participants in as concrete a way as possible. |
| 2) The reading of the data | Get a grasp of the material as a whole. |
| 3) The dividing of the data into parts | For each text (participant) separately, split their statements into parts (one or multiple sentences or just a phrase) that express a single thought/idea/experience. These parts are called “meaning units”. |
| 4) Organization and expression of raw data into disciplinary language | Assess what each meaning unit is expressing and rephrase them in a manner faithful to the scientific discipline and the context of the research. |
| 5a) Expressing the structure of the phenomenon | Compress the data in order to make the final expression of structure more manageable. Then we eliminate duplicates and finally, merge similar meaning units. |
| 5b) Extracting the essential themes | From the compressed material, identify central themes before filtering out the essential ones, i.e. those that are found in all the texts (participants). |

Table 1. A step-wise summarisation of the phenomenological analysis process.

1) Collecting of verbal data

The first step would be the interviews as described in the previous section. By asking the participants to talk about a specific project in which they were directly involved, the idea was to obtain “a concrete, detailed description of the subject’s experience and actions, as faithful as possible to what happened as experienced by the subject” (Giorgi, 1997, p. 245).

2) The reading of the data

The reading of the data means trying to get a sense of the data as a whole. The idea is not to draw specific conclusions but to better understand the individual parts when moving forward. Szklarski (2019, p. 155) adds that this is also the stage where possible unsuitable texts can be removed from the data set, e.g. interviews that are incomprehensible (to the researcher) or whose focus do not match the scope of the study. At this stage, the transcriptions were checked for errors while listening to the recordings. This process aided me in obtaining an overview of the material both as a whole and as individual parts. Before proceeding to the next step, all the interviews were read through at least once more while writing memos (an approach inspired by grounded theory).

3) The dividing of the data into parts

This step splits the text parts called *meaning units* (Giorgi, 1997, p. 246). A meaning unit is a piece of data – a phrase, sentence or sequence of sentences – that contains a meaning significant to the study. This division of texts into meaning units requires that our disciplinary perspective is balanced against the “discovery-oriented” nature of the phenomenological approach (Giorgi, 1997, p. 247). At the same time, crucially, the phenomenon of interest also needs to be present in the search for meaning. Thus, the meaning of the linguistic expression is determined by the purpose of the study, the lens through which the data is viewed. However, once this lens is applied, meaning must be allowed to manifest itself, to be discovered. This procedure is carried out by marking the places in the text where a transition in meaning happens, thus breaking the entire text down into meaning units. Below an excerpt from the transcript of the interview with Participant E split into meaning units.

(E1) Generally, the journalists in [the country], I don't know, elsewhere, but generally they are not ... They see the robots like a competitor. I think it's absurd, but they think it can be a competitor. (E2) In this case, this is a huge obstacle. It's really difficult to convince them that Terminator is not the way things are working right now. We don't know if we will have a strong artificial intelligence one day. This point is really difficult. (E3) The second point is the way they handle the information. For instance, we work with [national news organisation]. We thought it could be interesting for [the news organisation] and for the journalists to have a summary of a football match written for them, and then they could work on the specifics. We talk about the goals about the red flags, and all that, but then they can talk about how was the football match. Okay? But it's very difficult for them because there is the part of the text which is not written by them, and they want to own their text. (E4) This is also a very difficult point, because we think that the collaboration between the robot and the journalist is really the best way to tackle a lot of information with a journalistic point of view. But it's too difficult to have them do that.

At this point, the meaning units are expressed in the original language used by the interview subjects. Finally, Szklarski (2019, p. 156) recommends that the meaning units are summarised and first person expressions are modified to third person. In practice, I decided to conflate this third-person-substep with the next step.

4) Organization and expression of raw data into disciplinary language

Next, the meaning units are expressed using language more faithful to the scientific discipline, in our case digital journalism studies. Using a process called “free imaginative variation”, the researcher deploys their “disciplinary intuition” (Giorgi, 1997, p. 247). The process involves looking at the meaning units from different angles in search of possible meanings. This is a key step in order “to arrive at structural descriptions of an experience”, the end goal always being to find the essence of the phenomenon at hand (Moustakas 1994, p. 98). Szklarski (2019, p. 157) refers to this step as a way of interpreting both explicit and implicit expressions of meaning and compressing them at a more abstract level. Simply put, what we aim to do here, is determine what lies underneath the everyday language. To avoid

overextending at this stage my interpretations of what was being expressed, I decided to condense meaning but save further abstractions to the later phases of the analysis.

(E1) Based on interactions with journalists, Participant E believes that journalists in general view robots as competitors, a view not in the slightest shared by Participant E.

(E2) Participant E has had a hard time trying to convince journalists that news automation has little in common with stereotypes inspired by Terminator.

(E3) In Participant E's experience, journalists have a hard time putting their name on something they did not write entirely themselves.

(E4) In Participant E's opinion, a human-machine collaboration is a good solution in many instances, but this view is not shared by the journalists they encounter.

5) Expressing the structure of the phenomenon

As we were able to see in the description of free imaginative variation, the “structural descriptions” have already come into play in the previous step. Accordingly, Giorgi (1997, p. 247) notes that the process of the fifth step is basically the same as in step four. Now the units at play are the transformed ones, and the process of free imaginative variation is applied to these units in order to further locate structures in the data. Ultimately, we want to arrive at units “essential for the phenomenon” (Giorgi, 1997, p. 247).

Szklarski (2019, p.158–159) splits this step into two parts. The first part is meant to further compress the data in order to make the final expression of structure more manageable. First, all irrelevant meaning units are eliminated. Then, obvious duplicates are eliminated and finally, similar meaning units are merged. In the second part of step five, central themes are identified (Szklarski notes that this is the final sequence of a process that actually starts in step four). This is, effectively, the final step of the eidetic reduction. Free imaginative variation is now applied to the central themes, this time with the aim of extracting essential themes. Themes that are not essential, i.e. can not be found in all the texts, are said to

constitute the existence of the phenomenon. The essential themes, the ones that can be found in all the texts, constitute the essence of the phenomenon (Szkłarski, 2019. p. 159).

Analysing content in this way proved to be a bit slow-moving, but it quickly became clear that what is lost in speed is made up in quality. I found the procedure to be rather well-suited for an inexperienced researcher, because since the process involves keeping all the data until the final stages of analysis, it very nearly guarantees that no information is lost in the coding process.

5. Results

The aim of the phenomenological analysis was to arrive at the essence of the examined phenomena: news organisations and journalists and the service providers' relationship to these. The interviewees' accounts of their work with news organisations were categorised into a few central themes. Unsurprisingly due to the heterogeneous nature of the interview sample, the central themes were not uniform but rather formed by clustering similar themes. Together they form three overarching structures that run through the interview material as a whole. These structures constitute the essential themes, i.e. themes that are "essential for the phenomenon" (Giorgi, 1997, p. 247).

In this section, the three essential themes are presented as distinct but interrelated aspects of how the service providers a) perceive news organisations and journalists and b) perceive their relationship to them. The first aspect concerns their reasons for collaborating with news organisations (section 5.1). The second aspect is the perceived obstacles and limitations for them collaborating with news organisations (section 5.2). The third aspect is the perceived solutions for the encountered obstacles and limitations (section 5.3). The results are summarised in Table 2.

| | How service providers perceive news organisations/journalists (RQ1) | How service providers perceive their relationship to news organisations/journalists (RQ2) |
|---|--|--|
| Potential: strategic field | Needs: use data, more content, reach new audiences in new ways, monetise. | Training ground, proof of concept, high demands on accuracy of data and language. |
| Obstacles: everything is lacking | Resistance from journalists, weak understanding of data and tech, lack of financial resources. | Difficult to collaborate with, to understand and be understood by journalists. |
| Solutions: learning to understand each other | Journalistic competence beneficial and practical. | Inhabiting a supporting role, learning from each other. |

Table 2. The 3x2 matrix summarises the three different aspects (Potential, Obstacles and Solutions) of the views expressed by the service providers. Each aspect is made up of two interconnected dimensions: a) perception of news organisations and journalists, and b) relationship to news organisations and journalists.

5.1 Potential: strategic field

The participants view news media as an area in which they have a lot to offer. They firmly believe that news organisations are in need of renewed business models, and that the NLG technologies provided by their companies can be utilised to this end. It comes down to providing more content, personalising the content, reaching new audiences and thus increasing revenue. The software provider representatives see news media struggling to adapt their reporting to an online business logic. This is where the software providers come into the picture, providing their technology and their expertise to reshape how the news organisations use data for news production, in particular how it should be distributed and monetised in an online environment.

The conclusions we made was that there is a high demand for [a type of publicly available data] but also that what used to work well in print ... [describes how the content was previously packaged] ... does not work as well online because it generates very few page views. You are not making the most of the data.

- Participant D

The quote by Participant D also highlights the centrality of data in the software providers' relationship to journalists and news organisations. Data is perceived as an essential resource that is underutilised by news organisations. For an actor that understands how to utilise data, news organisations are potential customers. The financial difficulties of many news organisations is a related factor since automation can be used to increase profitability. What makes the service providers perceive news organisations as having a lot of potential as partners then, is their need to produce large quantities of content and the vast amounts of relevant data available to them, combined with their underdeveloped ways of utilising said data.

I think [the media organisation's] vision was, that there's a lot of publicly available data, that's out there ... I only have so many journalists, they're going to write a story for say one particular region, or one town, but they don't have coverage in the other towns, but they have the data to support that ... but if I automate this, I can write one story and it'll give me local news, which I'd never be able to touch from a manpower perspective, because we just don't have the number of journalists to cover things that are a really local or community, from a community perspective.

- Participant C

The service providers do not merely view automation as an opportunity for news media, but as a necessity. The way they see it, utilising automation better would help news organisations in their financial struggles, for example by becoming more resource efficient and reaching new audiences. As far as their relationship to the news organisations, the service providers generally believe the automation solutions they provide can help news organisations become more profitable or at least find business models better suited to an online environment. An awareness of the precarious financial positions of much of traditional news media is prevalent

among the participants, as is a belief that the services provided by their companies are indispensable.

Automation will [emphasis] save journalism, the same way that automation saved the auto industry or finance industry or ... because the economics has changed and the business models have changed and the way the customers consume content, fragmentation of the audiences and everything ... I think the benefits are clear, and the risks are basically for news organizations that are ... I mean at least in many countries, the news organizations are either in a very bad financial position, or they are on the way to being in a very bad financial position. So unless they adopt not only news automation, but other types of technology, let's say, I don't think they will survive. That's my opinion. The same way that happened in other industries. It's not like something new.

- Participant A

These views, of news media to be in need of new business models and of themselves to be in a position to help drive that change, mean that the service providers express explicit interloper tendencies. They might not aim to produce content themselves or “challenge journalistic norms” (Holton and Belair-Gagnon 2018, p. 73) head-on. However, at the very least they are making “products to the creation and distribution of news” and “calling for improved practices” (ibid.).

Although some participants explicitly stated that they consider journalism to be important for society as a whole, this was still not the main reason for working with news organisations. From the perspective of their own interests, software providers view news organisations as important strategic partners, or as Participant B puts it: journalism is “a highly significant field when it comes to public perception”. Except for this, the high demands on quality and autonomy prevalent in journalism makes it a good training ground for NLG companies looking to improve their technology. Participant F highlights how working with news media forced their company to enhance a) the fluency of language:

In news, we always have to aim at having a nicely written fluent article where at the beginning you talk about Chelsea Football Club, then you talk about Blues, and then you mention the club by the name of the coach; but when you're describing a dashboard with

some business intelligence, it doesn't have to be anything stylistically at the high level; it just has to be fully understandable. It doesn't matter that it's like pre-canned monotonous text.

and b) the flexibility of the service:

It should be fully under control of the newsroom, so the journalists should be able to change the wording of those articles if they don't like something. We will make sure that the grammar is right, and the choice of words, and so on, all the formal things, but they should be the ones who decide about the content; and if they don't like it, they should be able to change it.

Ultimately, using news media as a training ground might supply service providers with proof of concept when marketing to other business sectors. As Participant D puts it:

If it's good enough for [one of the largest national newspapers] and every local newspaper in [the country], it's good enough for other lines of business. So it's been a choice, partly by necessity, put also by choice, to focus on media.

5.2 Obstacles: everything is lacking

Even though the service providers see a lot of potential in news media and journalism and in associating themselves with them, none of them experience their work with news media as easy. The very same reasons – the financial struggles, the underutilisation of data, the high demands on accuracy – that make journalism an interesting endeavour for the service providers also function as major restraints. In the service providers' experience, these factors combined even seem to amplify each other. The requirements that news organisations have might not even be realistic, and even when they are, they might simply be too expensive. Ultimately, when news automation is introduced, the professional ideology of journalism, which demands high accuracy and a high degree of control over news production, clashes with the financial realities of the news organisations. This means that news organisations might have needs that they simply cannot afford to meet, forcing them to make compromises.

Sometimes they want us to change the way the engine writes the text, but in this case, we say "Yes, no problem. But that's the cost." Generally, they say, "Okay. No, it's perfect like that."

- Participant E

In innovation projects such as implementing news automation, the lack of money causes problems implicitly as well. Several service providers noticed that the media partners' small or non-existent IT-departments posed a considerable challenge. One way this manifests itself is that it slows down development, sometimes considerably.

I'm seeing that with another provider for public services we are working with. They are forced to buy everything that enables them from external sources [...] That is really a time-consuming process. We have been working with that provider for two years. And just the other day they said, "Okay, now we have ... the provider that does Python code for us in place and briefed." I was like dude what the [expletive]? It's [been] two years.

- Participant B

In essence, the service providers' experiences point to a severe lack of resources in news media, and the above mentioned lack of money is just one of the factors here. Another one is a lack of knowledge, a third is a lack of will to innovate. No single factor was necessarily prevalent all the time in the participants' experiences, but taken together they add up to a significant resource deficit. The most prevalent obstacle experienced by the software providers was a reluctance among journalists towards innovation. The participants' accounts suggest that some of the reluctance might be rooted in the ideology of journalism: when working with news organisations, the service providers are faced with reservations not encountered in other industries, and journalistic values were sometimes raised to negotiate how technology should be implemented without compromising for example autonomy. But the experiences recounted in the interviews also suggest that the opposite might be equally true, that professional ideology can be used to rationalise reluctance that is not necessarily ideologically rooted. The participants often interpreted the reluctance among journalists as fear.

Generally, the journalists in [European country] ... they see the robots like a competitor. I think it's absurd, but they think it can be a competitor. In this case, this is a huge obstacle.

- Participant E

Even in the one case where the participant had not perceived direct resistance from journalists, an implicit resistance could still be found:

So the editor-in-chief, the newsroom manager, the CEO, they need to give their support, make it clear to the organisation that “this is something that we are going to do”, that this is important. That’s a key to success. We’ve had some cases where we’ve worked with lower level management, such as a head of news or a head of sports, where the higher level management did not really want to have anything to do with the project. Then it’s much, much harder for the project to succeed. It gets poorly executed. You kinda don’t get any support. You get no traction or impact internally and it’s easy for the reporters or editors involved to ignore [the project].

- Participant D

This highlights the significance of control proximity, i.e. “when both groups recognise the authority of a third, most frequently management” (Chua and Duffy, 2019, p. 120). More importantly, it illustrates the complex interplay between internal and external motivations: even if journalists do not explicitly resist innovation, other factors might serve to stifle development projects. In a reversed fashion, tightly exercised control proximity can potentially serve to stifle other kinds of proximity – physical, temporal and professional (Chua and Duffy, 2019, pp. 119–120).

I always try to have access to the newsroom. I never have been able to talk to the journalists. Never. I don't know why. I always say, "Let me introduce the product. Let me talk to them. I know how to talk to them, because I do that for ... I've done that for seven years. So, I know how they feel about it." But no, I never have access to the newsroom.

- Participant E

However, not all participants experienced control proximity to be an issue in either direction (not too loose nor too tight), meaning that the service providers express a quite scattered view

on this particular aspect of the *organisational structures* (Paulussen, 2016, p. 196) of news media.

The final factor in the resource deficit of news media depicted by the participants is knowledge. A similar type of bottleneck for innovation as a non-existent IT-department, this lack of knowledge manifests itself as a poor understanding of both data and language technology. Ultimately, these types of misunderstandings might lead to resources being wasted. The different types of resource deficits (e.g. money and knowledge) actually appear to amplify each other, potentially creating a kind of negative synergy.

I was just talking to [online news site]. They wanted to start with us and they sent me a stream from [international news agency] and said, "Okay we did buy a data stream. Can you use it?" And the only thing that the data stream would give us was just written articles. And I said no we are not a rewriting tool. We need structured data. So this is the kind of story that repeats in different shapes and colors kind of every time I talk to somebody from publishing.

- Participant B

The participants also experienced a discrepancy between journalistic needs, such as accuracy and control, and the realities of the tech. This discrepancy is exacerbated by the publishers' poor understanding of the technical side of things. Some of the news organisations that the participants work with are quite pragmatic about it and do not even attempt to understand the systems on a technical level as long as the content is free from errors, thus relinquishing some control and – interestingly since the two are often at odds with each other in journalism – some transparency as well. In other cases however, the news organisation staff's poor understanding of the tech caused them to express demands that were often perceived as unrealistic.

The thing is that they want machine learning, they want artificial intelligence; but on the other hand, they also want to fully understand the system, which is not always possible with machine learning systems, and without a deep understanding of math, and so on. They also want this immediate fixing of errors, which again, in case of machine learning, is not always obvious how to do that.

- Participant F

The misunderstandings were not a one-way-streak. The service providers had experience of misguided attempts or missed opportunities stemming from misunderstandings on their part. They did not always realise what it is that journalists and news organisations do, what their day-to-day work and reasoning looks like. Their lack of understanding had both a principal dimension – how can journalistic autonomy be guaranteed – and a practical one – what kind of technology is of actual use and what degree of IT-skills can be expected among the end-users? On these issues, the news organisations deviated from the service providers’ customers in other industries and forced them to adapt, many describing a somewhat painful learning curve. In some instances, these failures to reach mutual understanding even caused the collaborators to lose sight of the strengths of the technology, thus undercutting the strength of the collaborative process.

The one big mistake that we made was, we really let ourselves be dragged into the discussion about quality and whether the tools that we were offering could do the same that journalists do to the same extent. We really blasted out a lot of money with basically no use. Just for exploring ... whether we could write celebrity news. Surprise you can't because celebrities tend to be original to stay in the discourse so they always come up with new [expletive] that they can do ... This was just not feasible and it came out of the notion that we really tried to ... that we really followed the journalists when they said "I want to replace this and that". And that replacing aspect instead offering something new, something that would be complementary to what they already had in place. This was a really ... I would call it an expensive mistake.

- Participant B

5.3 Solutions: learning to understand each other

The service providers partly paint a rather bleak picture of their news media partnership, but their experiences do involve overcoming obstacles. Reaching mutual understanding involves journalists understanding the technology and the data it builds upon. It also involves news organisations understanding how to successfully use the technology, mainly by integrating it into workflows and, crucially, business models. On the other hand, the process also involves the service providers understanding journalists. The journalism-as-training-ground and

journalism-as-proof of concept tropes have somewhat exploitative connotations, but the participants actually express a good deal of open-mindedness. While these might be the original reasons for collaborating with journalists, in practice they might not be the only benefits. Once inside, the service providers tend to view the journalistic expertise found in news media organisations as beneficial, if they did not already. Coming from a journalistic background themselves, Participant D is very much in touch with the logic of news media.

What we need, what I want, is to get in touch with someone, a reporter or journalist that is knowledgeable in the area that the text is about. That is the most important thing for us. So that the text is ... up to the standards of the publicist. We are not particularly dependent on the technical expertise of the media companies, but rather it is editorial competence that we're looking for.

For others the learning process was more challenging, but all of the participants eventually came to understand the value of listening to journalists. It allowed the service providers to design software that actually met the specific needs of news organisations (e.g. accuracy and autonomy). It also allowed them to focus more on good technological solutions instead of guessing how to best produce news, thus not losing sight of the strengths of the technology (cf. section 5.2). In fact, for all service providers, understanding the specific needs of news organisations and journalists was a key to success. In other words: succeeding meant adapting to the logics of news media, specifically the ideology of journalism (Deuze, 2005).

Participant A's software company even started employing journalists of their own and producing news independently, effectively becoming a news agency, distinctly inhabiting the role of explicit interloper. This is quite remarkable since it confirms the scenario imagined by elite media representatives about automation allowing news production to take place without a traditional newsroom (Sirén-Heikel et al., 2019, p. 59). However, Participant A's company does not actually wish to be an explicit interloper:

We don't want to be a news agency. We just want to be a tech provider, which is what we are. But because we are small and we are new, then we have to do everything.

In fact, the more important finding was that all the service providers aimed at being implicit interlopers. That means they wished to adopt the position of experts, inhabiting a supportive

role, focusing on tech. The collaborations they describe worked best when everyone involved understood each other as well as their own role in the collaboration. Or to use Chua and Duffy's (2019, pp. 119–120) terms, when the different types of proximity (physical, temporal and professional) are strong. In order to achieve this, mutual understanding was needed. Reaching this point involves the service providers understanding the needs of journalists. Contrary to what is the case with their customers in other sectors, the end users in news organisations usually did not have a solid IT background.

So we had to lower the bar. And actually in a very indirect way cooperating with journalists and people who come from a really editorial background gave us a new set of features which is some AI based features that help users annotate their sample texts into something that the machine understands. So we have a scripting language in the background ... and we actually have a component that reads a sample story and forecasts the code that somebody would have to code ... based on the user's input. And this is for journalists. In order for them to understand it, we just made it into a magic stick shaped button.

- Participant B

Reaching mutual understanding also involved journalists understanding the benefits of technology. Mostly this did not entail understanding the software on a deep technological level, but rather understanding how to take advantage of it. In the participants' experience, there are a few different ways in which news organisations and journalists have come to appreciate the news automation software. On an organisational level, taking advantage meant realising how to utilise data and how to distribute and monetise the content – for example personalisation and search engine optimisation, converting new subscribers and increasing ad revenue. On the level of the individual journalist it mainly meant getting rid of repetitive tasks, allowing the journalists to focus on other tasks such as in-depth reporting.

Then the journalists, in general, what they see as a positive thing is that they free up their time. They don't have to do, let's say, this very boring news that doesn't add any value. So that's what the journalists see.

- Participant A

Mutual understanding was reached in multiple different ways, and the participants' experiences seem to indicate that all the different types of proximity described by Chua and Duffy (2019, pp. 119–120) are important: Adopting or doubling down on the logics of news media, sometimes even employing journalists, can be seen as efforts to tighten professional proximity. Some participants emphasised the importance of physical proximity (getting access to the newsroom). The kind of proximity seen as most beneficial was control proximity, i.e. the need to acknowledge the authority of a specific group, such as management or business level.

They are able to explain to [the journalists] how this will impact the job of the journalists, because then they say “no one's going to lose a job and this is going to be complementary content, blah blah blah”, you know all these things, because our content is published automatically.

- Participant A

Once this mutual understanding is reached, the fruits of actual collaboration can be reaped. News organisations and journalists get to utilise their strengths, which in the service providers' estimation mainly include a combination of domain knowledge and journalistic competence (e.g. finance reporters being skilled at producing texts about the stock market that lives up to journalistic standards). Similarly, the service providers get to utilise their strengths, focusing on technical solutions and data handling. In this way, designing and implementing news automation systems becomes a joint endeavour. The participants describe meaningful collaboration as a back-and-forth between their companies and the news organisations.

Obviously, we know ... where our limits are. I'm listening to them in those areas that they are experts on ... Yeah, I think they're experts in certain areas, and there is no need for us trying to persuade them to do it differently. I would say that all of these things are decided and designed in common meetings where people from us, like the IT people, and the computational linguists, and so on, meet with the experts relevant for the given thing on their side, and together they design the solution, and then we design the technical insides.

- Participant F

As we can see, the news organisations that the participants have worked with are by no means perceived to be passive in the innovation process. News media might perhaps traditionally have been reactive rather than proactive (Boczkowski, 2004, p. 171) but this does not mean they lack agency and can be reduced to mere objects of change.

One of the most surprising things to me is that many people say that media outlets are not innovators. But I think it's not true. I think most of the last innovation we saw were hanged on by media outlets ... they know what they want and they ask us, "We want that. Can you do that or no." We also provide the list of what we already do. In most cases, it's interesting to them, but they are not only users. They propose different things.

- Participant E

News media as a training ground, which is one of the perceived benefits of doing business with news organisations, takes on new dimensions when the agency of the news organisations and journalists is added to the equation. News organisations and journalists become driving forces, shaping innovation through their particular needs, logics and expertise. In a way, once the collaborators understand each other, things come full circle, with the possibilities-turned-obstacles actually turning out to be solutions as well. The following quote, regarding how dates (year/month/day) should be reported, summarises how the specific needs of a news organisation – a need that potentially might have been considered a hurdle – through dialogue ended up leading to an improvement.

[The media partner] came back and said, your date handling maybe isn't as strong as we would like it to be. In our last release, and I would say that this is directly related to them, ... a lot of the date functions that we came up with were really driven from their need to translate one date format into a variety of formats that you might need to use when you're ... writing a news story. That would be an example of innovation that I think they really drove us to get better, with the date/time functions.

- Participant C

Finally, working with journalists might even affect the service providers in more fundamental ways. The relationship between service providers and news organisations bear the stamp of hybrid media logics (Chadwick, 2017, p. 22) with both parties adapting to each other. In

perhaps its most notable form, hybrid media logics manifest themselves as the ideology of journalism flowing over into the tech companies. It is easy to see that this is inevitable in cases where the service providers employ journalists, but it can happen in more subtle ways as well.

For instance, we are creating an ethical committee. I'm sure that if we weren't working with the media industry and journalists, we wouldn't have asked this type of question. The responsibility we have is for all of our clients, but the way we think [about] this responsibility is induced by our work with journalists and the media outlets.

- Participant E

Ethics being a core value of the ideology of journalism, this statement is noteworthy since it illustrates yet another dimension to the ways in which news organisations and service providers influence each other, the other main ones being journalists working directly for the service providers and news organisations influencing the software through joint innovation. This third dimension represents the sharing of values, and not just on the individual level but on an organisational one. Again, and unsurprisingly since the service providers are quite different from one another, while none of these dimensions were apparent in all of the interviews, they indicate that the logics of news media strongly inform the relationship between news organisations and service providers.

6. Discussion and Conclusions

I start this section out by briefly summarising the results. I then examine them in relation to previous research and the theoretical framework. Finally, I discuss the limitations of this study and make suggestions for further research.

6.1 Summarisation and discussion of results

The service providers consider news organisations to have a lot of potential as customers. In their experience, journalism puts high demands on the tech they provide – in terms of control over the production process and in terms of data and language quality. This makes news media a potential training ground in which the service providers are pushed to develop their software and eventually develop a proof of concept. News media's struggle to develop profitable business models combined with an underutilisation of data is perceived as a possibility to introduce innovation in the form of news automation. The same factors that make news media such an interesting prospect for the service providers also add up to the most prominent obstacles for fruitful cooperation. In their work with news media, the service providers describe a resource deficit – lack of money, lack of knowledge and lack of will to innovate – that is perceived to stifle innovation. Ultimately, the service providers do not merely perceive news media and journalism as a set of possibilities turned obstacles. In an example of hybrid media logics (Chadwick, 2017, pp. 22–26) at work, the service providers acknowledge the mutual nature of influence in the collaborative process. Taken together, these aspects, i.e. potential, obstacles and solutions, provide three distinct but interrelated dimensions of how representatives for service providers perceive news organisations and journalists.

The results undoubtedly paint the picture of an ambivalent relationship between service providers and news organisations. They also confirm many of the findings made in previous research on news automation and media innovation. In an age of information overflow created by technological advancements, news media organisations tend to fight fire with fire by looking to more technology for a solution (Örnebring and Ferrer Conill, 2016, p. 214). However, since news organisations often lack the knowledge, money and infrastructure

needed to put data and automation to good use, they are forced to look for outside help (Dörr, 2016, p. 712; Sirén-Heikel et al., 2019, p. 60). Furthermore, as Wu, Tandoc and Salmon (2019b) have shown, being forced to renew business models puts news media in an even more vulnerable position (pp. 1244–1245). This position is well-understood by tech actors (p. 1250) who also see news media as a display window of sorts (p. 1245). These factors add up to the *organisational structures* (Paulussen, 2016, p. 196) of media innovation. Media innovation is also shaped by *professional cultures* (Paulussen, 2016, p. 198), or as I have chosen to frame it here, the ideology of journalism (Deuze, 2005). It is the structures of journalism that cause the accuracy requirements on news automation software to be quite high (Dörr, 2016, p. 712; Leppänen et al., 2017, p. 189). It is not surprising then, that the software providers interviewed in this study should view news media as an area where their competence will a) be sorely needed, b) be proven and c) potentially be refined.

Furthermore, the collaboration clearly revolves around media innovation. But as b) and c) show, organisational structures and professional cultures (ideologies) are not enough to explain the collaboration. Media innovation is also inevitably “a form of social and societal innovation” (Bruns, 2014, p. 24). When viewing news media as an area to prove and refine their technology, the service providers are being influenced by larger societal circumstances: public perception, financial interests and so forth. In fact, this goes both ways since journalists aim to “self-legitimize their position in society” (Deuze, 2005, p. 446). This is in part why, for all their promise, the organisational structures and professional cultures (ideologies) of news media organisations do not solely serve to pave the way towards success for the service providers.

The service providers experience their work with news media to be anything but straightforward. Here, my findings confirm and deepen those made by Dörr (2016), who concluded that the high quality-standards of journalism (p. 712) coupled with the uncertain economic conditions of news media (p.716) cause difficulties for NLG software providers looking to make headway in news automation. To this we can add the ideologically motivated resistance from journalists, explicit as well as implicit, which is typically associated with media innovation (Deuze, 2005, p. 449). We can not however make any conclusions regarding the control proximity that Chua and Duffy (2019, pp. 119–120) found to be influencing the collaboration between journalists and peripheral actors: the service

providers encountered varying levels of managerial involvement (control) and experienced it in quite varied ways too. This is further discussed in the section on limitations (6.2).

In their relationship to journalism the service providers clearly see themselves as implicit interlopers: desiring the role of experts aiding journalism and news production, they position themselves as supporting actors rather than disruptors challenging journalistic norms. Here my findings would at first appear to be at odds with those made by Chua and Duffy (2019), who found that “respected implicit interlopers” (p. 116) were welcomed by journalists. The service providers studied here clearly fit Chua and Duffy’s description of respected implicit interlopers, but nonetheless experienced aversion more often than not. What this suggests is a discrepancy between how the service providers view themselves in their relationship to journalism and how they are viewed by the journalists. Because even though they see themselves as implicit interlopers, a type of actor not associated with disrupting journalistic practices (Holton and Belair-Gagnon, 2018, p. 74), their role might be a bit more complicated. As we have seen, being an explicit interloper – questioning the practices and business models of news media – is at least to some extent their *raison d’être*.

Instead of adopting a fixed position, the service providers seem to be oscillating between different roles, interloping explicitly or implicitly depending on what serves their purposes. None of the service providers can be said to inhabit or aim to inhabit the role of intralooper, but there is no reason to believe that the oscillation between roles found in this study could not extend itself to the intralooper role as well. For example, among analytics companies, the so-called “routinisers” (Belair-Gagnon and Holton, 2018, p. 501) lean towards the intralooper role.

Hybrid media logics provide a useful way of looking at this oscillation: the collaboration between the service providers and news organisations involves a mutual shaping of each other’s views and practices. The practices of challenging and empowering are some the most relevant ways in which the service providers affect the power dynamics of news production. Interestingly, in order to achieve each of them, the service providers need to understand the logics of news media and journalism. So while the service providers see a need to map their own logics onto the news organisations (explicit interloping), they also embrace the logics of journalism, not only in the collaboration but in their organisation as a whole. The

participants' accounts of their collaborations with news organisations do point to them acknowledging the importance of journalistic expertise. Insofar as the collaborations are successful, they involve the service providers being able to reconcile their view of the relationship with the journalists' view.

How the power dynamics of news production is influenced depends on how the news organisations and the service providers handle the combination of the other party's perceived expertise and lack thereof. The results of this study point towards service providers on the one hand attempting to make news organisations acknowledge their own resource deficit, and on the other hand allowing them to outsource the work of overcoming the resource deficit. This entails bridging knowledge gaps to some extent and perhaps dispersing with some doubts in the process: the connection between reluctance (or indeed even fear) and lack of knowledge is easy to make. For the service providers, the course of action has a lot to do with role perception: depending on how they perceive their own role, the impact they have on news production will differ.

6.2 Limitations and suggestions for further research

The goal of this study was to examine how service providers for news automation view journalism and news media and their own relationship to them. In the positivist tradition, the criteria by which to evaluate research are validity – whether we measure what we set out to measure – and reliability – the trustworthiness and reproducibility of the study (Kvale, 2007, p. 122). However, the usefulness of these concepts to qualitative research rooted in the constructionist tradition has been called into question (cf. Flick, 2007, p. 65; Kvale, 2007, p. 122–123). Instead, scholars tend to propose evaluating qualitative research in terms of quality. In the remainder of this section, I evaluate my study by discussing the appropriateness (Hennink, Hutter and Bailey, 2011, p. 58) of my overall research design. Specifically, I look at the limitations of my theoretical and methodological framework.

First a note on the limits of generalisation. Thanks to previous mapping of the field (notably Dörr, 2016, and Lindén et al., 2019) and to the high response rate of our interview requests (9 out of 11 agreed to be interviewed) we were able to cover a large share of the relevant actors. Thus, looking strictly at the numbers, the generalisability of the results is quite high, and

saturation being reached in the analysis stage further indicates high generalisability. The remaining issues concern the material fitting the scope of the study. Having conducted a majority of the interviews myself, and proofread and reread the transcripts of all of them, I was in a good position to assess which were the ones that best fit the scope. In this sense, the interviews analysed do match the research questions quite well. That said, the results are not 100% percent generalisable to the rest of the material (and by extension the population). For example, the pilot interview (which was not included in the analysis) was educational and important for fine-tuning the interview guide but did not correspond as well to the research questions as the subsequent interviews. In another discarded interview there was the issue of the participant not having actually worked directly with news media companies. In sum, the results provide a good foundation for further researching the relationships between service providers and news media but are not applicable to the entire field of news automation.

By using hybridity as defined by Chadwick (2017) as my theoretical framework, I wanted to frame the results in terms of hybrid media logics (*ibid.*) and the power dynamics of news production (Chadwick, 2011). In order to do that and to mitigate the weaknesses of the hybridity metaphor (Mast, Coesemans and Temmerman, 2017), I also applied the journalistic strangers typology (Holton and Belair-Gagnon 2018). This did allow me to look at the hybrid media logics in more detail. But hybridity's relative vagueness and inability to come to terms with false dualisms coupled with the highly subjective nature of the material meant that I had to stop short of actually examining the power dynamics in-depth.

Throughout this study, I have defined journalism as professional ideology. This definition proved well-suited for understanding how such standards as editorial independence manifest themselves in the power dynamics of news production. But as the rather inconclusive findings on certain aspects of the organisational structures that inform media innovation show, it leaves us with a few blind spots. Specifically, to understand the impact of control proximity (Chua and Duffy, 2019, p. 120) we might need other definitions. For example, a comparative studies approach (cf. Hanitzsch, 2007 or Örnebring, 2012) that takes into account varying cultures of journalism (as opposed to Deuze's rather uniform ideology concept) could help account for differences between countries. It seems to me that the best way to actually come to terms with how the different kinds of proximity between actors influence the power dynamics is a case-study approach. Journalism as ideology has also

functioned as the main representation of the logics of news media. But different types of media logics might be relevant as well, as might different interpretations of journalism. Further research into the power dynamics of news automation also needs to approach the subject with a clear understanding of the logics that the service providers bring as tech actors.

A second blind spot is methodological. The service providers shift between explicit and implicit interloping, and there might be a temporal dimension to these shifts (for example starting out as explicit interlopers and becoming implicit as time passes). But the design of this study does not allow for conclusions about this. A longitudinal approach might shine some more light on the shifting of roles found in this study.

Ultimately, the contribution of this study is not to untangle the power dynamics, but hopefully to point out some of the ways in which service providers for news automation are taking part in the power dynamics of news production. A solid body of work is emerging on many other aspects of how news automation shapes these power dynamics (cf. section 3.1.2). This study adds a missing piece, but it couples qualitative interviews with a phenomenological analytical approach, which means that the results are limited to a single point of view. In the same manner as the vast majority of studies in the past has focused solely on the perspective of news media, this study has focused solely on the perspective of the service providers. Future research should attempt to go deeper into the power structures by merging the two perspectives. Methodologically, the aforementioned case-study approach or some similar kind of field work (cf. Goffman, 2010) could provide a nuanced view of the collaborations between news organisations and news automation service providers. Theoretically, it might involve a more comprehensive conceptualisation of power in the vein of Foucault.

Finally, the contradictory findings regarding the perceived resistance faced by journalists and the desired role of implicit interlopers (i.e. an actor that should not necessarily be met with aversion) is another aspect that can not fully be accommodated by our theoretical framework. These contradictions imply that it might not be possible to divorce the service providers from the technology they provide. I argue that journalists might perceive the technology itself as the explicit interloper, not least since the discourse surrounding news automation is so heavily influenced by the robot metaphor, which depicts NLG technology as an almost

human-like competitor (Lindén and Dierickx, 2019, p.154). Because of this, some variation of actor-network theory (Latour, 2005), which views technology as an actor in and of itself, might be one way to more fully comprehend the power dynamics of news automation.

References

- Allen, D.S. 2008, "The trouble with transparency: The challenge of doing journalism ethics in a surveillance society", *Journalism Studies*, vol. 9, no. 3, pp. 323-340.
- Altheide, D.L. & Snow, R.P. 1979, *Media Logic*, 2nd edn, Sage, Beverly Hills, Calif.
- Ananny, M. & Crawford, K. 2015, "A liminal press: Situating news app designers within a field of networked news production", *Digital Journalism*, vol. 3, no. 2, pp. 192-208.
- Anderson, C.W. 2018, *Apostles of certainty : data journalism and the politics of doubt*, Oxford University Press, New York, NY.
- Anderson, C.W. 2013, "Towards a sociology of computational and algorithmic journalism", *New media & society*, vol. 15, no. 7, pp. 1005-1021.
- Baack, S. 2018, "Practically Engaged: The entanglements between data journalism and civic tech", *Digital Journalism*, vol. 6, no. 6, pp. 673-692.
- Belair-Gagnon, V. & Holton, A.E. 2018, "Boundary work, interloper media, and analytics in newsrooms: An analysis of the roles of web analytics companies in news production", *Digital Journalism*, vol. 6, no. 4, pp. 492-508.
- Boczkowski, P.J. 2004, *Digitizing the news : innovation in online newspapers*, MIT Press, Cambridge, Mass.
- Borges-Rey, E. 2016, "Unravelling data journalism: A study of data journalism practice in British newsrooms", *Journalism Practice*, vol. 10, no. 7, pp. 833-843.
- Bruns, A. 2014, "Media innovations, user innovations, societal innovations", *The Journal of Media Innovations*, vol. 1, no. 1, pp. 13-27.
- Carlson, M. 2015, "The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", *Digital journalism*, vol. 3, no. 3, pp. 416-431.
- Caswell, D. & Dörr, K. 2018, "Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories", *Journalism practice*, vol. 12, no. 4, pp. 477-496.
- Chadwick, A. 2017, *The hybrid media system : politics and power*, Second Edition edn, Oxford University Press, New York, NY.
- Chadwick, A. 2011, "The political information cycle in a hybrid news system: The British prime minister and the "Bullygate" affair", *The International Journal of Press/Politics*, vol. 16, no. 1, pp. 3-29.
- Chua, S. & Duffy, A. 2019, "Friend, Foe or Frenemy? Traditional Journalism Actors' Changing Attitudes towards Peripheral Players and Their Innovations", *Media and Communication*, vol. 7, no. 4, pp. 112-122.
- Clerwall, C. 2014, "Enter the robot journalist: Users' perceptions of automated content", *Journalism practice*, vol. 8, no. 5, pp. 519-531.
- Coddington, M. 2015, "Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting", *Digital journalism*, vol. 3, no. 3, pp. 331-348.

- Denscombe, M. 2014, *The good research guide : for small-scale research projects*, Fifth edition edn, Open University Press, Maidenhead, Berkshire.
- Denzin, N.K. & Lincoln, Y.S. 2018, *The SAGE handbook of qualitative research*, Fifth edition edn, SAGE, Los Angeles.
- Deuze, M. 2019, "What journalism is (not)", *Social Media Society*, vol. 5, no. 3, pp. 2056305119857202.
- Deuze, M. 2005, "What is journalism? Professional identity and ideology of journalists reconsidered", *Journalism*, vol. 6, no. 4, pp. 442-464.
- Deuze, M. 2004, "What is multimedia journalism?", *Journalism studies*, vol. 5, no. 2, pp. 139-152.
- Deuze, M. & Witschge, T. 2018, "Beyond journalism: Theorizing the transformation of journalism", *Journalism*, vol. 19, no. 2, pp. 165-181.
- Diakopoulos, N. 2019, *Automating the news : how algorithms are rewriting the media*, Harvard University Press, Cambridge, Massachusetts.
- Diakopoulos, N. & Koliska, M. 2017, "Algorithmic transparency in the news media", *Digital Journalism*, vol. 5, no. 7, pp. 809-828.
- Dierickx, L. 2018, "News bot for the newsroom: how building data quality indicators can support journalistic projects relying on real-time open data", *Global Investigative Journalism Conference 2017 Academic Track*. Available: <https://ijec.org/2018/02/02/research-news-bot-for-the-newsroom-how-building-data-quality-indicators-can-support-journalistic-projects-relying-on-real-time-open-data/> [2021, January 25].
- Domingo, D. & Wiard, V. 2016, "News Networks" in *The SAGE handbook of digital journalism*, eds. T. Witschge, C.W. Anderson, D. Domingo & A. Hermida, Sage Publications, Ltd, London ; Thousand Oaks, California, pp. 397-409.
- Dörr, K.N. 2016, "Mapping the field of algorithmic journalism", *Digital journalism*, .
- Dörr, K.N. & Hollnbuchner, K. 2017, "Ethical challenges of algorithmic journalism", *Digital journalism*, vol. 5, no. 4, pp. 404-419.
- Eberle, T.S. 2014, "Phenomenology as a Research Method" in *The SAGE handbook of qualitative data analysis*, ed. U. Flick, SAGE, Los Angeles, pp. 184-202.
- Eldridge, S.A., Hess, K., Tandoc Jr, E.C. & Westlund, O. 2019, "Navigating the scholarly terrain: introducing the digital journalism studies compass", *Digital Journalism*, vol. 7, no. 3, pp. 386-403.
- Eldridge, S.A., II 2017, *Online journalism from the periphery : interloper media and the journalistic field*, 1st edition edn, Routledge, Taylor & Francis Group, London.
- Esaiasson, P., Esaiasson, P., Gilljam, M., Oscarsson, H. & Wägnerud, L. 2012, *Metodpraktikan : konsten att studera samhälle, individ och marknad*, 4 [rev.] uppl edn, Norstedts juridik, Stockholm.
- Ess, C.M. 2014, "Editor's Introduction: Innovations in the newsroom—and beyond", *The Journal of Media Innovations*, vol. 1, no. 2, pp. 1-9.

- Fanta, A. 2017, "Putting Europe's robots on the map: automated journalism in news agencies", *Reuters Institute Fellowship Paper*, , pp. 2017-2009.
- Ferrucci, P. & Nelson, J.L. 2019, "The new advertisers: How foundation funding impacts journalism", *Media and Communication*, vol. 7, no. 4, pp. 45-55.
- Flick, U. 2007, *Designing qualitative research*, SAGE, Los Angeles, Calif. ; London.
- Gillham, B. 2005, *Research interviewing : the range of techniques*, Open University Press, Maidenhead.
- Giorgi, A. 1997, "The theory, practice, and evaluation of the phenomenological method as a qualitative research procedure", *Journal of phenomenological psychology*, vol. 28, no. 2, pp. 235-260.
- Goffman, E. 2010, "On fieldwork" in *SAGE qualitative research methods*, eds. P. Atkinson & S. Delamont, Thousand Oaks, CA: SAGE Publications, , pp. 124-132.
- Graefe, A. 2016, "Guide to automated journalism", .
- Graefe, A., Haim, M., Haarmann, B. & Brosius, H. 2018, "Readers' perception of computer-generated news: Credibility, expertise, and readability", *Journalism*, vol. 19, no. 5, pp. 595-610.
- Gubrium, J.F., Gubrium, J.F. & Holstein, J.A. 2002, *Handbook of interview research*, SAGE, Thousand Oaks, California ; London.
- Gynnild, A. 2014, "Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets", *Journalism*, vol. 15, no. 6, pp. 713-730.
- Haim, M. & Zamith, R. 2019, "Open-Source Trading Zones and Boundary Objects: Examining GitHub as a Space for Collaborating on "News"", *Media and Communication*, vol. 7, no. 4, pp. 80-91.
- Hamilton, J.T. & Turner, F. 2009, *Accountability through algorithm: Developing the field of computational journalism*. Available: <https://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf> [2021, January 25].
- Hammond, P. 2017, "From computer-assisted to data-driven: Journalism and Big Data", *Journalism*, vol. 18, no. 4, pp. 408-424.
- Hanitzsch, T. 2007, "Deconstructing journalism culture: Toward a universal theory", *Communication theory*, vol. 17, no. 4, pp. 367-385.
- Hennink, M., Hutter, I. & Bailey, A. 2011, *Qualitative research methods*, 1st ed edn, SAGE, London.
- Hermida, A. & Young, M.L. 2017, "Finding the data unicorn: A hierarchy of hybridity in data and computational journalism", *Digital Journalism*, vol. 5, no. 2, pp. 159-176.
- Holton, A.E. & Belair-Gagnon, V. 2018, "Strangers to the game? Interlopers, intralopers, and shifting news production", *Media and Communication*, vol. 6, no. 4, pp. 70-78.

- Israel, M. & Hay, I. 2006, *Research Ethics for Social Scientists: Between Ethical Conduct and Regulatory Compliance*, SAGE Publications, London.
- 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*, vol. 71, pp. 291-298.
- Justesen, L. & Mik-Meyer, N. 2011, *Kvalitativa metoder : från vetenskapsteori till praktik*, Studentlitteratur, Lund.
- Koerth-Baker, M. & Carlson, E. 2017, , *Sensor Journalism: Reporting Opportunities and Ethical Concerns*. Available: <https://niemanreports.org/articles/mining-the-data/> [2020, Oct 30].
- Kovarik, B. 2011, *Revolutions in communication: Media history from Gutenberg to the digital age*, 1st edn, Bloomsbury Publishing USA.
- Krumsvik, A.H., Milan, S., Bhroin, N.N. & Storsul, T. 2019, "Making (Sense of) Media Innovations" in *Making Media: production, Practices, and Professions*, eds. M. Deuze & M. Prenger, Amsterdam University Press, , pp. 193-206.
- Kvale, S. 2007, *Doing interviews*, Sage, Los Angeles.
- Kvale, S. & Brinkmann, S. 2009, *InterViews : learning the craft of qualitative research interviewing*, 2nd ed edn, Sage, Los Angeles.
- Latar, N.L. 2018, *Robot journalism : can human journalism survive?* World Scientific Publishing Co. Pte Ltd, Singapore.
- Latar, N.L. 2015, "The robot journalist in the age of social physics: The end of human journalism?" in *researchgate.net*, ed. G. Einav, Springer, , pp. 65-80.
- Latour, B. 2005, *Reassembling the social : an introduction to actor-network-theory*, Oxford University Press, Oxford.
- Leppänen, L., Munezero, M., Granroth-Wilding, M. & Toivonen, H. 2017, "Data-driven news generation for automated journalism", *Proceedings of the 10th International Conference on Natural Language Generation*, pp. 188.
- Levy, S. 2012, , *Can an Algorithm Write a Better News Story Than a Human Reporter?*. Available: <https://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/> [2020, April 6].
- Lewis, S.C. & Westlund, O. 2015, "Big data and journalism: Epistemology, expertise, economics, and ethics", *Digital journalism*, vol. 3, no. 3, pp. 447-466.
- Lindén, C. 2017, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", *Digital journalism*, vol. 5, no. 2, pp. 123-140.
- Lindén, C., Tuulonen, H., Bäck, A., Diakopoulos, N., Granroth-Wilding, M., Haapanen, L., Leppänen, L., Melin, M., Moring, T. & Munezero, M. 2019, *News automation: The rewards, risks and realities of 'machine journalism'*, WAN-IFRA.
- Lindén, T.C. & Dierickx, L. 2019, "Robot journalism: The damage done by a metaphor", *Unmediated: Journal of Politics and Communication*, vol. 2, pp. 152-155.

- Lindlof, T.R. & Taylor, B.C. 2011, *Qualitative communication research methods*, 3rd ed edn, SAGE, Thousand Oaks, Calif.
- Magnusson, M., Finnäs, J. & Wallentin, L. 2016, "Finding the news lead in the data haystack: Automated local data journalism using crime data", *Computation Journalism Symposium*.
- Mari, W. 2019, *A Short History of Disruptive Journalism Technologies: 1960-1990*, Routledge.
- Mast, J., Coesemans, R. & Temmerman, M. 2017, "Hybridity and the news: Blending genres and interaction patterns in new forms of journalism", *Journalism*, vol. 18, no. 1, pp. 3-10.
- Mayer-Schönberger, V. & Cukier, K. 2013, *Big data : a revolution that will transform how we live, work, and think*, Houghton Mifflin Harcourt, Boston.
- Melin, M., Bäck, A., Södergård, C., Munezero, M.D., Leppänen, L.J. & Toivonen, H. 2018, "No landslide for the human journalist-an empirical study of computer-generated election news in finland", *IEEE Access*, vol. 6, pp. 43356-43367.
- Meyer, P. 1973, *Precision journalism : a reporter's introduction to social science methods*, Indiana University Press, Bloomington.
- Moustakas, C.E. 1994, *Phenomenological research methods*, SAGE, Thousand Oaks, Calif. ; London.
- Napoli, P.M. 2013, "The algorithm as institution: Toward a theoretical framework for automated media production and consumption", *Fordham University Schools of Business Research Paper*, .
- Örnebring, H. 2012, "Comparative journalism research—an overview", *Sociology Compass*, vol. 6, no. 10, pp. 769-780.
- Örnebring, H. & Ferrer Conill, R. 2016, "Outsourcing Newswork" in *The SAGE handbook of digital journalism*, eds. T. Witschge, C.W. Anderson, D. Domingo & A. Hermida, Sage Publications, Ltd, London ; Thousand Oaks, California, pp. 1 online resource (xviii, 605 pages).
- ÖT 2016, , *Textrobot skriver lokal sport för HSS Media*.
Available: <https://www.osterbottenstidning.fi/Artikel/Visa/122240> [2020, April 6].
- Paulussen, S. 2016, "Innovation in the Newsroom" in *The SAGE handbook of digital journalism*, eds. T. Witschge, C.W. Anderson, D. Domingo & A. Hermida, Sage Publications, Ltd, London ; Thousand Oaks, California, pp. 1 online resource (xviii, 605 pages).
- Pavlik, J.V. 2013, "Innovation and the future of journalism", *Digital journalism*, vol. 1, no. 2, pp. 181-193.
- Petre, C. 2013, , *A Quantitative Turn in Journalism?*.
Available: <https://blog.chartbeat.com/2013/10/31/quantitative-turn-journalism/> [2020, Feb 17].

- Reiter, E. 2013, "Natural Language Generation" in *The handbook of computational linguistics and natural language processing*, eds. A. Clark, C. Fox & S. Lappin, Wiley-Blackwell, West Sussex, England, pp. 13e-6a.
- Reiter, E. & Dale, R. 2000, *Building natural language generation systems*, Cambridge University Press, Cambridge.
- Roulston, K. 2014, "Analysing Interviews" in *The SAGE handbook of qualitative data analysis*, ed. U. Flick, SAGE, Los Angeles, pp. 297–312.
- Shoemaker, P.J. & Reese, S.D. 2013, *Mediating the message in the 21st century: A media sociology perspective*, Routledge.
- Simola, K. 2019, , *STT ja Yle tiivistävät yhteistyötä uutisrobotiikassa*. Available: <https://suomenlehdisto.fi/stt-ja-yle-tiivistavat-yhteistyota-uutisrobotiikassa/> [2020, April 6].
- Sirén-Heikel, S., Leppänen, L., Lindén, C. & Bäck, A. 2019, "Unboxing news automation: Exploring imagined affordances of automation in news journalism", *Nordic Journal of Media Studies*, vol. 1, no. 1, pp. 47-66.
- Stavelin, E. 2013, *Computational Journalism. When journalism meets programming*, PhD Thesis, The University of Bergen. Available: http://stavelin.com/uib/ComputationalJournalism_EirikStavelin.pdf [2021, January 25].
- Szklarski, A. 2019, "Fenomenologi" in *Handbok i kvalitativ analys*, eds. A. Fejes & R. Thornberg, Upplaga 3. edn, Liber, Stockholm, pp. 148–164.
- Tandoc Jr, E.C. 2019, "Journalism at the Periphery", *Media and Communication*, vol. 7, no. 4, pp. 138-143.
- TENK 2019, *The ethical principles of research with human participants and ethical review in the human sciences in Finland*. Available: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf [2021, January 25].
- Trappel, J. 2015, "What to study when studying media and communication innovation? Research design for the digital age", *The Journal of Media Innovations*, vol. 2, no. 1, pp. 7-22.
- Van Dalen, A. 2012, "The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists", *Journalism Practice*, vol. 6, no. 5-6, pp. 648-658.
- Vehkoo, J. 2017, , *Uutisrobotit tulivat vihdoin Suomeen*. Available: <https://www.journalisti.fi/artikkelit/2017/4/uutisrobotit-tulivat-vihdoin-suomeen/> [2020, April 6].
- Wahutu, j.S. 2019, "Prophets without Honor: Peripheral Actors in Kenyan Journalism", *Media and Communication*, vol. 7, no. 4.
- Westlund, O. & Lewis, S.C. 2014, "Agents of media innovations: Actors, actants, and audiences", *The Journal of Media Innovations*, vol. 1, no. 2, pp. 10-35.
- White, E.M. 2015, , *Automated earnings stories multiply*. Available: <https://blog.ap.org/announcements/automated-earnings-stories-multiply> [2020, April 6].

- Witschge, T., Anderson, C.W., Domingo, D. & Hermida, A. 2019, "Dealing with the mess (we made): Unraveling hybridity, normativity, and complexity in journalism studies", *Journalism*, vol. 20, no. 5, pp. 651-659.
- Wölker, A. & Powell, T.E. 2018, "Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism", *Journalism*, , pp. 1464884918757072.
- Wu, S., Tandoc Jr, E.C. & Salmon, C.T. 2019a, "Journalism reconfigured: Assessing human-machine relations and the autonomous power of automation in news production", *Journalism Studies*, vol. 20, no. 10, pp. 1440-1457.
- Wu, S., Tandoc Jr, E.C. & Salmon, C.T. 2019b, "When journalism and automation intersect: Assessing the influence of the technological field on contemporary newsrooms", *Journalism Practice*, vol. 13, no. 10, pp. 1238-1254.
- Yle 2019, , *Ketkä olivat kotikuntasi suurimmat ääniharavat? Katso Voitto-robotin tulokset*. Available: <https://yle.fi/uutiset/3-10717811> [2020, April 7].
- Young, M.L. & Hermida, A. 2015, "From Mr. and Mrs. outlier to central tendencies: Computational journalism and crime reporting at the Los Angeles Times", *Digital Journalism*, vol. 3, no. 3, pp. 381-397.

Appendix

Appendix A – interview guide

1. How did the project start – background (NO NEED TO STATE NEWS ORG)

- a. How would you define news? (in your mind, what is it?)
- b. How did the project start?
- c. Who reached out to whom?
- d. What did the media organisation want/ need?
- e. What was your main selling point, how did you present your service?
- f. Have you been involved in several similar projects?
- i. WHY DID YOU CHOOSE THIS PARTICULAR PROJECT?

2. Designing the product for media organisations

- a. To what extent did you collaborate with the news organisation in designing the automation service/ software system?
 - i. How long did you work on the system before they could put it into use?
 - ii. EXTRA: have them explain what changes they needed to make.
- b. What people were you in contact with, what positions in the news organisations did they represent?
- c. In your experience, are these people the right ones to be in charge of these issues?
- d. Were there specific issues or (journalistic) requirements that are related to how journalists work / the work processes in the newsroom?
- e. Considering that this system is supposed to produce news texts, are there differences in how it is designed compared to a system intended for non-journalistic texts?
 - i. Were there obstacles, bottlenecks that impacted on implementation, or other

issues that you encountered in the beginning of the project?

3. Implementing the product

a. And the implementation process, how did that start?

i. Was there anything concerning the production of news texts that caused problems?

b. What kind of reception did you get by the news organisation/newsrooms? How were you received by the news organisation?

i. Details on how it was done (e.g. workshops)

c. How did you open up how the system works for the news organisation?

d. Did you notice differences in the way you talked depending on what person in the organisation you were talking to?

i. If so, how? What type of language, words, terms etc. do you use when talking about your system?

ii. What did you do when somebody didn't understand?

iii. Have you noticed something in particular that can be difficult for them to grasp?

iv. Do you recall any situation when you thought "oh, they didn't understand" / "this person understands exactly what I am talking about"?

e. Did you discuss responsibility in case something would go wrong?

i. In case you have a written agreement regarding responsibility, what's in it?

ii. Ownership – who owns what?

f. How do you follow up how the system works, have you had to make any changes?

g. Was there any documentation of the process, if so, who did it?

h. Did you encounter any kind of resistance in the newsroom?

4. Looking back at the project

a. How well did you feel that newsrooms were able to adopt your systems?

b. Who were the end users of the system?

- c. How did the project change your view on news journalism?
- d. How do you feel that your product has impacted, or will impact, on the news organisation?
- e. Is there anything you would have done differently?
- i. Did you learn anything from the project?
- f. EXTRA (in case they did not answer already): What factors do you feel are crucial when considering that the newsrooms actually implement the system? What has to work/fall into place?

5. Self-reflection & future

- a. What added value, or pain-killers, do you feel that your software can give to journalists and news organisations?
 - i. Specify their arguments (e.g. quality, what does that mean?)
- b. How do you see the role of your company, and your product, in news journalism? (What are you and what do you want to be?)
 - i. Do you think that this project, or other projects with news organisations, has impacted the way you work and think as a company?
 - ii. When designing systems, to what extent do you consider the final users, the readers and consumers?
 - iii. How are you planning on further developing your news-oriented services?
- c. From a commercial perspective, how profitable is it to do business with news organisations?
- d. How do you think that automation will change journalism? Potential/challenges?
- d.e. Is there anything that you fear with your system, considering the current state of AI and how people are using media?
- e.f. Are there any technical innovations currently not being used that you think have potential?
 - i. What is keeping them from being put into use?
 - i.ii. What is the next big thing?