Check for updates

Article

Philosophy of the Social Sciences
2024, Vol. 54(5) 363–386
© The Author(s) 2024

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00483931241255229
journals.sagepub.com/home/pos


The Explanatory Power of Discourse Analysis

Corrado Matta[1]

Abstract

This paper discusses the explanatory power of discourse analysis, an approach to typically considered one of the main qualitative methods for data analysis. Although discourse analysis is typically not used to develop explanatory models, some have claimed that discourse models can causally explain social phenomena, such as institutional change. I analyze a case of institutional change to discuss and provide arguments for two claims. First, discourse analysis cannot generate causal explanations of social phenomena. Secondly, discourse analysis can generate non-causal explanations, specifically constitutive explanations of social phenomena. I conclude the paper by discussing possible implications for social research.

Keywords

discourse analysis, constitutive explanations, qualitative research methods, discursive institutionalism

1. Introduction

This paper discusses the qualitative method known as *discourse analysis* and its uses in social science. More specifically, the object of discussion in the paper is the claim that models resulting from the analysis of discursive

Received 19 March 2024

[1]Linnaeus University, Växjö, Sweden

Corresponding Author:
Corrado Matta, Department of Pedagogy and Learning, Linnaeus University, Universitetsplatsen 1, Växjö 352 52, Sweden.
Email: corrado.matta@lnu.se

practices can be used to explain social facts. Analyzing discursive practices has become a common approach in qualitative research, with a growing number of research contributions, handbooks, journals, and conferences dedicated to the area of *discourse analysis*. It is therefore peculiar that neither the philosophy of the social sciences nor the methodological literature on qualitative methods has devoted much attention to the issue of explanatory power.

The main rationale for discourse analysis is to investigate how groups of people use language as a means to provide understanding about social facts that are not necessarily linguistic. An example of this approach is the study of scientific knowledge and scientific disciplines, where the status of any set of statements as *scientific knowledge* or *scientific facts* is studied investigating the features of the linguistic practices within that scientific discipline. Scientific knowledge and scientific facts are, according to this approach, something that we do by using language in a certain way.

More recently, discourse analysis has been applied to fields other than the study of scientific or disciplinary practices and has become a consolidated subfield of qualitative research, characterized by a variety of methodological procedures and theoretical interpretations. Despite this diversity, the different concretizations of discourse analysis share a minimal common denominator: they all analyze language to answer questions about social facts.

The discussion of my paper, concerning the explanatory role of discourse analysis, will center on the analysis of a case. The case is illustrative of how discourses might play a role in explaining social facts and will help clarify the problem at the core of this paper. The introduction of this illustrative case will take the first part of my discussion. After that, I will focus on the general features of the models that result from applying discourse analysis, and on the concept of explanation. As for the latter, I will argue that scientific explanations involve the specification of a mechanism or a part of a mechanism, as well as the specification of one or several counterfactual dependencies. I will argue that if models of discourse analysis satisfy these demands, then it is justified to ascribe explanatory power to discourse analysis. These two discussions will work as basic assumptions for my argument, which consists of two parts. The first part of the argument is negative and concludes that models from discourse analysis cannot provide causal explanations. The second part is positive and concludes that discourse analytical models provide *constitutive* explanations of social facts, and therefore can play a non-causal role in broader causal explanations. I conclude the paper with methodological recommendations for researchers who plan to integrate discourse analysis in causal explanatory models.

2. Introducing the Case: The Swedish School System and the *Goal-and-Performance-Management*

In this section, I will consider an example of social fact that works as a main explanandum: the shift toward goals and performance-oriented management (GPM) in the Swedish school system that took place through the 1994 curriculum reform.

This reform constituted a paradigm shift in the Swedish school policy. Before its introduction, the school curriculum was only considered as an instrument used to determine what content was to be taught in schools. After the 1994 reform, the curriculum expressed what goals, in terms of pupils' knowledge and skills, the instruction must achieve, and this is the essential idea of goal and performance-oriented management in the school sector: school teaching is based on—and importantly—evaluated in terms of—the achievement of goals rather than on the necessity of covering certain subject matters. The GPM is not a management framework specific to the educational sector, but rather a general framework for the whole public sector. The Swedish Association of Local Authorities and Regions defines the framework as consisting of two basic elements: goals, which are set by decision-makers (both local and statal ones), and results, which must be measured regularly and reliably to test whether the goals have been achieved. In simple terms, the GPM entails that Swedish authorities must organize their work based on goals rather than on tasks and that they must use results to evaluate their performance in terms of goal satisfaction rather than on fixed benchmarks.

Below, I will summarize the chain of decisions and policies that led to this institutional change, as it is reconstructed in Wahlström's chapter (2009, chap. 3). Wahlström describes the GPM as the result of the introduction and transformation of another policy: *program budgeting*, which is an economic policy originated in the post-war US. To understand how Sweden adopted the GPM, Wahlström suggests, it is necessary to clarify how program budgeting was introduced in the Swedish public sector and successively transformed from an economic to a management policy.

In the third chapter of the monograph, Wahlström reconstructs the events and circumstances preceding the introduction of the GPM in Sweden. Using a data set consisting of official documents (policy propositions, government official investigations, pro-memorias from state departments and offices) and supported by further research about the history of the Swedish public sector (Amnå 1981; Sundström 2003). I summarize below the chain of events, grouped into four main phases. Some details are left out for the sake of space, but this summary should provide a sufficiently clear picture of the events. Each main set of events is summarized by a capital letter, which I will use whenever I need to avoid repetition.

2.1. First Phase, Introduction (I)

1. After the end of the Second World War, *program budgeting* was introduced in the US military sector. This is a significant change in the management of the military sector, consisting of connecting budgeted expenses to pre-determined goals. Budgeting is connected to performance analysis for the first time in military management.
2. During the 1950s, interest grew in program budgeting in Sweden. During this period, the Swedish Department of Defense Research Unit is interested in introducing the model into the Swedish military sector. Engineering universities started to teach about performance analysis as an approach to management.

2.2. Second Phase, Lobbying (L)

3. At the beginning of the 1960s, a group lobbying for introducing program budgeting in the Swedish public sector was formed within the Swedish Taxpayers' Association.
4. The group's lobbying was successful: in 1963, the Swedish Agency for Public Management nominated the group's founding member as an expert advisor for a work group assigned to produce a proposal draft for the introduction of program budgeting in Sweden. Later the same year, the Swedish Agency for Public Management appointed the same member as the project leader of an official investigation into the feasibility of introducing program budgeting.
5. The official investigation of the Swedish Agency for Public Management concluded that it is difficult to measure the effectiveness of program budgeting in the welfare sector, as it is difficult to construct performance indicators that are as reliable as those used to measure the performance of firms (such as net profit). The investigator concluded that each public authority should set its own goals and decide how to measure the satisfaction of these goals.

2.3. Third Phase, Testing (T)

6. The 1970s saw a phase of testing program budgeting in several public agencies. The Swedish National Audit Office (NAO, an authority closely related to the Agency for Public Management) was appointed as responsible for assessing the results of these tests. The tests showed that program budgeting was riddled with several difficulties and unanswered questions. In their report, the NAO described the negative results as *implementation problems*: program budgeting requires training and motivation, and goals and performances are very difficult

to define ex-ante, thus explaining why program budgeting might not work at first.

7. Another phase of testing started in the early years of the 1980s, structured in 3-year budgeting periods, and testing was made permanent in 1988. The test phase showed the same kind of problems as the first testing wave. The results of this new testing wave were also interpreted as implementation problems.

2.4. Fourth Phase, Translation (Tr)

8. A parliamentary proposition (Swedish Ministry of Finance 1987) was drafted. In this proposal, there is the elimination of the main mechanism of program budgeting, that is, tying the budget to goals and performances. The resulting policy proposes a system based on goals and performances that is supposed to inform the whole Swedish public sector. This is the Goal and Performance-oriented Management. In the proposition, the elimination of the budgeting part of the policy is motivated by the same technical implementation difficulties observed during the testing phase.

2.5. Final Phase, Implementation (F)

9. The same year, through a parliamentary vote, the GPM was adopted as the official management form for the Swedish public sector. A GPM-based reform of the Swedish national school curricula will follow in just a few years, namely, 1992.

Wahlström points out a visible contradiction in the chain of events leading to the introduction of the GPM: why did the government choose to move on and make the GPM a general policy even though all tests showed negative results? The results of the tests, Wahlström argues, seemed only to be used to determine how to implement the policy (i.e., as a budget policy or as a management policy), but the basic principles of the policy—that is, setting goals and measuring performances—seemed never to be genuinely under scrutiny.

Wahlström provides elements that might potentially explain this discrepancy and the fact the object of the evaluation was never the basic principles of both program budgeting and GPM. Wahlström refers to Sundström's study (2003), which explains that the basic principles of program budgeting and GPM were represented in the public and political debates as a matter of *sheer common sense*. Moreover, according to Sundström, the process leading to the adoption of the GPM is located in a context of growing critique in the public opinion against the Swedish public sector bureaucracy,

which was experienced as lacking transparency and accountability, being out of touch with the public, and restricting the people's possibility of choice.

Wahlström does not aim explicitly at providing an explanatory model but seems to intend that without reference to the broader cultural context of the period, the chain of events leading to the rise of the GPM might miss a link. In this way, Wahlström seems to agree with the *cultural turn in institutional studies* (Alasuutari 2015), according to which, explaining institutional change requires locating the decision-making process in a cultural context.

The two contextual elements that Wahlström mentions are not by themselves a full-blooded explanation, as the author does not discuss in detail the connection between these and the chain of decision-making events.[1] Hence, the issue is whether contextual elements concerning how the policy reform was represented in the public and political debate might be explanatory. This is the main issue of this paper.

Discursive Institutionalism (DI) (Alasuutari 2015; Schmidt 2008, 2015) is an example of cultural institutionalism that claims that such cultural facts are in fact explanatory. The advocates of DI propose to explain institutional change by integrating reconstructions of decision-making processes with analyses of the *discursive practices* related to the targeted institution. These analyses describe the discursive context of a decision-making event, that is, the discursive facts of a specific historical period. Advocates of DI argue that analyses of discursive practices can provide the kind of models that are necessary to explain institutional change, in that models of discursive practices can account for why political decisions can be taken in specific historical moments. In short, DI proposes a model for explaining institutions composed of a decision-making part and a discursive part.

Hence, if the advocates of DI are correct, then the analysis of discourse should contribute to the explanation of the institutional changes. In this paper, I will assess if this claim is justified. If the answer is positive, I will be able to conclude that discourse analysis is explanatory. In the next section, I provide

[1]An intuitive connection might be that the decision-makers who drafted and voted on the proposal had the attitude that "setting goals and evaluating performances is a matter of common sense," and that they held this attitude because of the influence of the public and political debates. However, neither Wahlström nor Nordström provide evidence of such attitudes. For instance, their attitude could have been that the proposition was a "high-stakes bet." Or, they might have thought that the proposition was divisive or controversial, or that it had support from only a part of the population, but that was necessary to solve some problems. Alternatively, they could even have thought the proposition was highly unpopular or risky, but the public sector required a "shock therapy." In all these cases, the decision-makers might have drafted the proposal and voted for it even if they did not have the specific attitude "the GPM is a matter of common sense."

the first clarification that is necessary for my discussion, which concerns what kind of models results from the application of discourse analysis.

3. Enter Discourse Analysis

Discourse analysis is the term used to group all approaches in qualitative research that focus on analyzing discursive practices, that is, the nexus of activities, typically linguistic or otherwise symbolic, that characterize social practices such as science, politics, religion, and so on. For instance, the discursive practice of medicine consists of all activities, such as argumentative styles, experts' claims, and the use of artifacts such as tables, diagrams, grids, and more, that contribute to making certain claims as medical facts. With some simplification, the literature on discourse analysis can be subdivided into foundational works on the theory of discourse, which includes scholars such as Foucault (1969 [1972]) and Laclau and Mouffe (1985), and empirical approaches, discussing the methodological procedures for the study of discursive practices, such as Fairclough (1995, 2003), Van Dijk (2006), Jørgensen and Phillips (2002), and Wodak and Meyer (2001). In this paper, I will rely heavily on Foucault's theory of discourse, trying as much as possible to isolate the parts of that theory that agree with the most common research practices in discourse analysis.

Discursive practices are often conceptualized as centering on networks of *central claims*, that is, the kind of linguistic activities that have privileged statuses within a linguistic practice, such as truths or facts. To describe these linguistic events, Foucault uses the term *statements* (Foucault 1969 [1972]).[2] Foucauldian statements are not considered in light of their semantic contents such as opinions, beliefs, or desires. Instead, discourse analysis typically operates by bracketing central claims from the intentions of individuals involved in that particular discursive practice and tries instead to characterize how language is used in a context, and this bracketing is usually described as what distinguishes discourse analysis from other more interpretive methods in the social sciences. Therefore, statements are descriptive facts about discursive practices, as Dreyfus and Rabinow—who authored a popular commentary on Foucault's theory of discourse—put it:

> [...] Foucault is interested in just those types of speech acts which are divorced from the local situation of assertion and from the shared everyday background so as to constitute a relatively autonomous realm [...]. (Dreyfus and Rabinow 1982 [1983], 48)

[2]Throughout the text, I will use the term statement referring to Foucault's theory, and not to its ordinary use.

Hence, discourse analysis is concerned with statements as types of performative events: how claims are collectively used. Dreyfus and Rabinow (1982 [1983]) use the term "serious speech acts" (1982 [1983], 48) to describe statements, highlighting the centrality of use. Statements are collective ways of using claims by virtue of which they can be described as "facts" or "truths." For instance, the fact that people in a scientific discipline typically do not use a claim as an object of scrutiny, but rather as a starting point for scrutinizing other claims is an indicator that claim is a statement. Consider the main equations of accepted physical theories (e.g., Newton's Law in classical physics), the idea of representative democracy in modern Western societies, or the idea of free trade in Western economic discourse. It is important here to reiterate that even if these claims and ideas might depend on peoples' beliefs and desires, discourse analysis is not concerned with this dimension, and only studies what characterizes statements at the level of discourse. Therefore, discourse analysis' concern with Newton's law, or the concept of representative democracy, is not the fact that they are accepted by the agents—i.e., that agents have certain attitudes—but rather what makes these claims appear as central, or facts, at the level of discourse.[3]

The articulation of which statements make up a discursive practice is neither the only task nor the main task of discourse analysis. Surely, it is customary for the practice of discourse analysis, especially in its more contemporary form, to include a descriptive part including the main statements of a discourse. However, statements are mostly considered a starting point, or an explanandum, rather than an explanans. Instead, the main task of discourse analysis is to articulate the *regularities*—which Foucault calls "rules of formation" or "conditions of possibility" (1969 [1972], 38)—that fix the status of statements as such. These regularities determine (according to discourse theory) which ideas are more central and primitive, and which are more peripheral and derivative. They fix ideational contents by establishing similarities and oppositions (i.e. defining different ideas as competing with one another, and similar ideas as allied ones), and constraining inferential practices by affording translations and anaphora (allowing steps from one statement to another). A model resulting from discourse analysis is therefore a

[3]Foucault alternates a form of strong constructivism and skepticism concerning this issue, suggesting sometimes that the level of the discourse exhausts attitudes (i.e., attitudes are not genuine phenomena), and sometimes that attitudes are too opaque to be studied properly. More modern approaches to discourse analysis as qualitative methods tend not to commit to these claims and maintain instead a more pragmatic attitude, according to which bracketing out the level of attitudes is only a methodological or modeling choice. To make this paper relevant to a broader class of methodological approaches in qualitative research, I will only focus on this more pragmatic approach to discourse analysis in the remainder of the paper.

structure consisting of a set of statements and a set of rules of formation. The status of these rules is a debated issue. For example, in the work of Foucault, rules are regularities immanent of linguistic practices, and, very much in the same way as statements are not modeled as representational objects, the rules of formation are not to be considered as the rules accepted by the participants of the targeted discursive practice. Both Foucault and the main commentators (Dreyfus and Rabinow 1982 [1983]; Kusch 1991; Tiisala 2015) use the terms "rules" and "regularities" to describe the central element of models of discourse, showing that rules of formation are generalized patterns immanent to the discursive practice and at the same time have a regulative role.[4] According to this theory, a discursive performance is a statement within a discursive practice that is made in accordance with the rule of formation of that discursive practice. The issue of the status of rules of formation is central to the discussion of this paper and concerns the issue of how regularities of discourse can have regulative force. As I argue in the central part of the paper (consisting of the penultimate and third to last sections, where I present my two arguments), the explanatory power of discourse analysis is related to a great extent on the issue of whether regularities can have regulative force. I set aside the issue for the moment and return to it later in the paper.

Model construction in discourse analysis employs different procedures, but this variety of procedures can be grouped into a small set of *analytical categories*. Each of these categories guides the analytical work by simply providing a magnifying lens with a specific focus, guiding the researcher into what to look for in a discursive practice. The analytical categories provide the core elements of the discourse model, that is, the rules of formation. Even if these categories are named in different ways depending on which source is used, their description typically fits well with Foucault's discussion in the *Archaeology of Knowledge*. The following description of the analytical categories is mostly based on Martin Kusch's critical reconstruction of Foucault's methodology (1991), which provides the clearest available systematization of this work:

- Discursive elements concerning *objects*, such as grids, tables, classification sheets, and so on. These elements describe regularities about the entities involved in the discursive practice.

[4]Other main methodological sources for discourse analysis, such as Van Dijk (2006), conceptualize the rules of discourses as cognitive phenomena. The "immanent" interpretation of rules of formation does not need to exclude that cognitive phenomena might be involved in the emergence of systems of statements. Therefore, this interpretation can be considered as a modeling choice: an immanent discourse analytical model describes certain facts about language use and does not incorporate the cognitive facts that might contribute to the practice in the model.

- Discursive elements concerning *subjects*, such as criteria of competence. These elements describe regularities concerning subject positions, such as who may speak and about what.
- Discursive elements concerning *concepts*, that is, regularities concerning relations between statements. These regularities describe which structural relations between different statements are possible, such as ordering and dependence relations between statements—which statement can succeed or precede another statement in a rational argument; or possible procedures of rewriting—such as translations, change of sign systems, rules for approximation, criteria for definitions, ellipses, anaphora, and metaphors.
- Discursive elements concerning *strategies*, that is, regularities concerning relations between discursive practices.

It is instructive to reconnect this methodological discussion to the case discussed in the previous section. The explanatory question is "Why did the Swedish school system shift to the GPM?" Part of the explanation is provided by Whalström's reconstruction of the chain of decision-making events, summarized in the diagram below (Figure 1): where the dashed arrow symbolizes the fact that the outcomes of the testing phase do not seem to explain the translation of the program budgeting into GPM, and the subsequent adoption of this management system. The events are happening in the order represented in the diagram, but the last two events do not seem to be explained by the first three. In order to connect these events into a coherent whole, a contextual claim is introduced, namely, the fact that GPM is represented as a matter of common sense in public debate. (I will use the letters CS to describe this claim, whenever I need to avoid repetition.) This fact about the context in which the decision-making chain is located helps make sense of the step between T and Tr: the GPM is an idea that was already included in the original policy (i.e., program budgeting), and this idea was never the object of testing *because* it was a matter of common sense. Therefore, when the test showed negative results, the choice was to eliminate the budgeting part of the policy and keep (i.e., *reaffirm*) the GPM part.

From a discourse theoretical point of view, CS is a statement. As Wahlström and Sundström explain, the basic principles of program budgeting and GPM were represented in the public and political debates as a matter of *sheer common sense*, which is a fact about the discursive context in which the

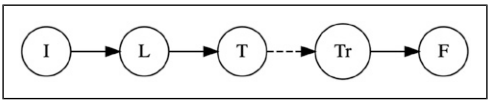


Figure 1. A linear representation of the chain of events discussed by Wahlström.

decision-making process is located. Tr and F describe two specific instances of performing this particular *serious* speech act. From the perspective of DI, it is by virtue of its statement status that this idea can play an explanatory role in the rise of the GPM. Therefore, it is plausible to expect that the explanatory model of F should involve an account of what confers CS (of which Tr and F are instances) its statement status. This justifies using discourse analysis, which leads to the identification of a set of rules of formation for that statement. These rules describe patterns in discourse that, because of their regularity, consolidate (a) the GPM as an *object* opposed to the content-based management, (b) certain *subject* positions as experts (as exemplified by the founder of the lobbying group for program budgeting being appointed as an expert advisor), (c) certain *concepts*, i.e., relations between statements, such as the common sense character of the GPM is related to the lack of transparency of the task-based management, and (d) certain relations between discursive practices, such as the transferability of elements from the military discourse to the public sector.

I will restrict my discussion to one possible example of a rule of formation that could be generated using discourse analysis. This could be a regularity concerning what conclusions are drawn in rational argumentations about the GPM. It could be that the analysis of the public and political discourse concerning managing the public sector shows that the idea of the GPM is resistant to critique based on empirical observations. This could be for instance if there is a tendency to explain away or rationalize observations that seem to contradict the GPM. This regularity could be analyzed as a rule of formation according to which empirical observations should accommodate the GPM, rather than put it at risk, and that contradiction between the GPM and empirical observations should not lead to concluding against the former. This regularity (which I denote with the letter R) belongs to the analytical category of *concepts* (c), as it concerns relations of dependence between statements. This regularity is also an example of what, from the perspective of discourse theory, could provide a claim with the status of a statement (especially if in conjunction with several other rules of formation developed using other analytical categories). The fact that the discursive practice of school politics exhibits such regularity is what (together with other regularities) makes CS a statement.

The resulting simplified explanatory model (which I will call *M* for short) is illustrated in the graph below. M consists of two parts: a decision model (denoted as M_{dec} and represented by circles connected by continuous arrows in Figure 2), and a discourse model (denoted as M_{dis} and represented by the dashed rectangles connected by dashed arrows).

Having clarified the details of how discursive practices are modeled using discourse analysis, I move on now to the second issue, that is, what theory of scientific explanation provides a sensible normative standard for scientific

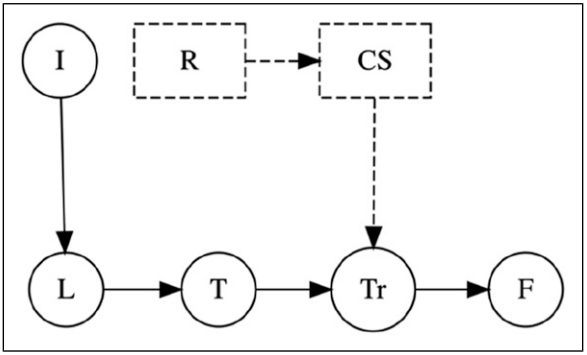


Figure 2. A representation of model M.

explanations of institutional change. As the main question of the paper is if discourse analysis is explanatory, it is necessary to define what an explanation should be. Depending on the standard that is accepted, the explanatory role of discourse analysis might change, one possibility being that discourse analysis is—vis-à-vis some suitable standard for scientific explanation—not explanatory at all.

4. Explanation

The advocates of DI, and especially Schmidt (2008), have been quite explicit that models of discourses explain. Moreover, it is plausible to assume that the kind of explanation that is assumed here is causal. This is because the main issue for DI is to account for what contributed to the occurrence of a specific event (institutional change), which is a clearly causal question.

This attitude is in clear contrast with the theoretical sources about discourse analysis (Foucault 1969 [1972]; Laclau and Mouffe 1985; Dreyfus and Rabinow 1982 [1983]), which have in most cases either denied or toned down the causal explanatory power of the theory, or simply tried to dodge the bullet. However, this early discussion about the causal role of discourses is affected by a major limitation, which entails that many critical claims should be taken with a proverbial grain of salt. Oftentimes, claims that qualitative methods are not explanatory, are made assuming the Hempelian conception of scientific explanation, and the Humean regularist approach to causality, both of which set too demanding standards for the social sciences (and for many natural sciences as well). These assumptions about causality and explanation motivated a general separatist attitude in certain social research contexts (Matta 2015), but lost most of their popularity among philosophers of science in recent decades, and have been replaced with perspectives on causal explanations that are more coherent with the epistemic demands of the special

sciences. In this section, I will argue for a theory of causal explanations in the social sciences that provides a more sensible standard and sets more realistic demands for discourse analysis.

Some popular contemporary approaches to causal explanations in the social sciences have attributed a pivotal role to *counterfactual claims* and *mechanisms*. The main idea behind the former term is that causal explanations, in contrast with descriptions, involve counterfactual claims. This means that if x explains y, then, if x had changed, y would have changed, too. Counterfactual claims are good tools for making an inventory of *what* contributed to a targeted outcome, that is, what ingredients we need in a causal account of that target. However, oftentimes social researchers are interested in *how* the ingredients together can generate an outcome. This requires a conception of how the different ingredients are related to one another, that is, a specification of what kind of concrete process or activity connects these ingredients. This leads us to the second demand that causal explanations should satisfy, that is, the specification of a *causal mechanism* (Glennan 2002; Hedström and Ylikoski 2010). For x to explain y, then x should specify both what contributes to y and how. In other words, and to use a popular definition (Illari and Williamson 2012), a mechanistic explanation specifies the *entities* and *activities* that together, and by virtue of their arrangements, contribute to a target outcome.

Under the assumption of these theories of causal explanations, the demands that discourse analysis must satisfy to play an explanatory role are the following:

- The model resulting from using discourse analysis should specify one or several counterfactual dependencies that are relevant to the targeted outcome.
- The resulting model should specify a mechanism, or a part of a mechanism, that is responsible for the targeted outcome.

Going back to the case of the Swedish school system reform, Smith's claim that discourses explain is justified if

- The discursive part of M, M_{dis}, describes one or several causal counterfactual dependencies, and
- M_{dis} contributes to the specification of a mechanism in M, that is, eliminating the discursive part of M results in a less specified mechanism.

5. Discourse Analysis is not Causally Explanatory

I have now enough background resources to move to the core part of the paper and discuss the possible explanatory role of discourse analysis. In the first step of my argument, I contend that the models generated using discourse analysis do not describe *causal relations*.

M_{dis} involves two arrows, and this two-arrow configuration is the general template for using discourse analysis for causal explanation: an arrow connecting rules of formation to statements, and an arrow connecting the statement (or statements) to a causal mechanism.

The most straightforward way to interpret the arrows in M_{dis} causally is by reference to rule-following. The decision-maker might have adhered to R, sanctioning the decision to ignore the negative result of the testing phase or to accommodate the negative result with the GPM. Moreover, the decision-maker might, because of her or his acceptance of rule R, have considered the GPM a matter of common sense. These two instances of adherence to rules might together with T work as causes for Tr, and eventually lead to F. The decision-maker had the results of the tests and some contextual factors about GPM as reasons for acting and acted based on these.

If discourse models explain by means of reference to this kind of causal relations, then the explanatory role of the discursive part of M_{dis} consists in the specification of two counterfactual dependencies that complete the gap in the mechanism described in M (which would satisfy the explanatory demands discussed in the previous section). More specifically, had the decision-maker not adhered to the rules R, she or he would not have considered the GPM a matter of common sense, but rather a policy as any other. Moreover, had the policy-maker not adhered to R, she or he would have interpreted the results of the testing phase as contradicting the GPM. As a result, the policy-maker would have weighted the test results differently in the decision of whether or not to decide in favor of the GPM.

This seems to be the essence of the argument for the causal role of discourses proposed by Dave Elder-Vass (2011). According to Elder-Vass, rules of formation explain in so far as they become socialized rules. The problem is that Elder-Vass is not providing an explication of how discourse analysis explains, but rather proposing a different model altogether. What explains in the model proposed by Elder-Vass is the fact that people have the attitude to adhere to the rules of formation, and not the rules of formations themselves. Hence, if Elder-Vass is right, attitudes are all we need to explain institutional change, and rules of formation are *interpreted*—i.e., modeled—as signs of these attitudes and commitments. However, the basic assumption of discourse analysis is that rules of formation are not interpretations, that is, models of attitudes. According to discourse analysis, discourses explain without reference to attitudes. Following Elder-Vass's proposal, rules of formation,

intended as *regularities* in linguistic practices, are not part of the explanatory model at all. Such regularities may mediate the process of socialization (Elder-Vass is not explicit about it, but his argument can be interpreted in this way), which might entail that they play a role in the explanation of people's attitudes, but the central issue is that Elder-Vass proposes that rules of formation are models of attitudes, which is simply what discourse analysis denies. Hence, discourse analysis—which is the analysis of regularities in linguistic practices that disregards the attitudes of the people involved in it—does not generate causal explanatory models in the sense described by Elder-Vass.

Whereas Elder-Vass proposes to interpret rules of formation as beliefs and attribute these beliefs a causal role in people's actions, others have discussed if rules of formation can be causally efficient when interpreted as regularities in discursive practices. A negative answer to this question was discussed by Dreyfus and Rabinow (1982 [1983]). The focus of their argument is the distinction between regularities and rules of discursive practices. As I mentioned earlier in the text, Foucault seems to argue that rules of formation have a regulative role; however, the methodology that he proposes—consisting of the use of the four analytical categories—seems to be only appropriate for articulating regularities. Therefore, once regularities in discursive practices are identified, it is unclear in what way these regularities can have any regulative or prescriptive force on the participants in a discursive practice.

Dreyfus and Rabinow take this issue as a starting point and argue that regularities in discursive practice, *qua* regularities, cannot have any regulative force on their own. To affect performances, rules of formations must entail prescriptions that are shared by participants, without the latter explicitly endorsing these prescriptions (to avoid the problem discussed concerning Elder-Vass's proposal above), or without involving other social practices such as socialization, education, or coercion.

Dreyfus and Rabinow show how this cannot be the case, by considering the case of grammar rules among natural language speakers. Competent speakers might be unaware of formalized grammar rules, but this does not preclude linguists from reconstructing such rules from speakers' practices. These rules are, however, approximations that are used to describe a complex practice. People act as a result of other cues (such as other, non-discursive, social practices, such as socialization or education), and the grammatic rules serve as a way of systematizing the complexity of the practice *at the linguistic level*. This is the key claim of Dreyfus and Rabinow's argument: if rules of formation are theoretical approximations used to systematize the complexity of a discursive practice (something i.e. more coherent with the methodology used to construe them), then they cannot describe a causal relation between rules and statements, because the rules are a systematization of the statements (i.e., of the performances that make up the discursive practice). As

approximations of the statements, rules of formation are theoretical redescriptions of discursive practices.

> Thus, although nondiscursive influences in the form of social and institutional practices, skills, pedagogical practices, and concrete models constantly intrude into Foucault's analysis [...] he must locate the productive power revealed by discursive practices in the regularity of these same practices. The result is the strange notion of regularities which regulate themselves. Since the regularity of discursive practices seems to be the result of their being governed, determined, and controlled, while they are assumed to be autonomous, the archaeologist must attribute causal efficiency to the very rules which describe these practices' systematicity. (Dreyfus and Rabinow 1982 [1983], 84)

According to Dreyfus and Rabinow, rules of formation and statements are two distinct descriptions of the same event. Because there is no ontological distinction between rules of formation and statements, there cannot be a causal relation between these two. In our case, if M_{dis} described causal relations, then R would cause CS and, in turn, Tr. This cannot be the case, because the rule of formation R is a descriptive model of the discursive practice of which both CS and Tr are parts. This is because, (a) the GPM being a matter of common sense means that it can resist contradicting facts, and (b) the translation of program budgeting into GPM is an instance of the same regularity summarized by R, that is, rational reasoning where GPM can withstand contradicting observations.

The two arguments show that discourse analysis does not satisfy the two demands on causal explanations that I assumed above. The arrows in M_{dis} do not describe causal counterfactual claims, as causal counterfactuals require two distinct events. Moreover, and for the same reason, M_{dis} does not contribute to the specification of the mechanism generating F, as it adds no new entities or activities to the model M. M_{dis} is neither a model of a causal mechanism nor a model of a part of a causal mechanism.

6. A Counterproposal: Discourse Analyses as Constitutive Explanations

In the second part of my argument, I propose that whereas it is correct to deny that discourse analysis generates causal explanations, the models developed with this method are explanatory in a different, non-causal, sense. More specifically, I argue that discourse analyses provide constitutive explanations.

As a first step, I define the concept of a constitutive explanation. As Ylikoski (2013) clarifies, a constitutive explanation is a non-causal explanation that "explain[s] how things have the causal capacities they have by appealing to their parts and their organization" (278). Constitutive

explanations differ from causal explanations in three respects. First, concerning their explanandum: whereas causal explanations have events as their explananda, constitutive ones have causal capacities. Causal capacities express the kind of causal effects that a system is *capable* of generating (glass can shatter, a fire can generate light, and a decision-making institution can implement a policy) in specific conditions. Importantly, as Ylikoski points out, causal capacities are properties of systems or entities, and as such, are not events. Secondly, they differ concerning the kind of counterfactual relation that is invoked: constitutive explanations explain by reference to composition whereas causal explanations explain by reference to *causal counterfactual dependency*. In a constitutive explanation, facts about the components of a system, and/or the arrangement of these components, must be able to account for the fact that the system has some causal capacity. Consider building a house with Lego bricks. The capacity of the house to afford changes in mental representations stimulating a child's playful enactment of certain scenarios (say, moving a small cat figurine inside the house) will depend on the form, material, quantity, and arrangements of the bricks. The house can—i.e., has the capacity to—cause the child's change of mental representation, and can therefore be played with, because of how many bricks it has, what they are made of (imagine building a house with bricks made of breadcrumbs), and how they are arranged together. It follows that similarly to causal explanations, constitutive explanations involve counterfactual dependencies. In the case of the Lego house, had the components differed in shape, quantity, or arrangement, the house's capacity to afford the playful enactment of the cat scenario would have changed. Thirdly, whereas causal explanations involve diachronic relations between explanandum and explanans, constitutive explanations are synchronic, meaning that the system maintains its causal capacity as long as the constitutive base (be that the property, quantity, or arrangement of components) is unchanged. A consequence of this characteristic of constitutive explanations is that explanans and explanandum are spatiotemporally coincident and thus ontologically dependent, or, as Ylikoski puts it: "the relata [of a constitutive explanation] are not independent existences" (2013, 284).

In light of this theory of constitutive explanations, I argue that discourse analyses may provide constitutive explanations of causal systems and that the

theory of constitutive explanations can clarify the regularity/rule ambiguity in the theory of discourses.

First of all, the proper object of discourse analysis is the discursive capacities of systems such as institutions, which are properties of such systems.[5] In the GPM case, the target system is an institution that I will call Swedish School Governance, that is, the complex of practices and regulations of decision-making bodies, experts, and opinion-makers that have the main task of regulating the Swedish school system. Consider again statement CS, that is, the fact that the GPM was a matter of common sense. In discourse analysis, statements are central elements of discursive practices and describe facts about the practice in general. They are therefore not circumscribed events, but rather descriptive features of the practice at large, and, in this case, features of the Swedish School Governance. Let us call these facts statement types, to distinguish them from statement tokens, which are instead circumscribed discursive events in which a statement is used by the practitioners participating in the practice. A system such as an institution affords its participants to perform serious speech acts, that is, its participants (*qua* participants to a practice) can perform statement tokens. Discourse analysis accounts for this affording capacity rather than for the occurrence of specific statement tokens.

Consequently, if the research goal is to explain the occurrence of an event, that is a causal explanation, then beliefs must be involved. A statement token can be interpreted as effect of an individual belief, and a statement type (the fact that a statement is central in a whole discursive practice) as a sign of a shared belief (as in Elder-Vass's proposal). This is a plausible approach, but it is not discourse analysis. Since explaining the *occurrence* of statement tokens requires invoking beliefs, and discourse analysis does not model beliefs, it can only explain the capacities of the system that performs statements. This concern is particularly relevant in the case discussed in this paper. The explanatory gap between the results of the testing phase (T) and the translation of program budgeting into the GPM raises the question of why this step was possible at all, that is, how the system could generate the *kind* of decision-making process described in the case.

This way of conceiving the main concern of discourse analysis is also recurrent in the literature on discourse analysis. As Dreyfus and Rabinow put it, rules of formation determine "what *can* be seriously said" (1982 [1983], 66;

[5]My discussion will require stretching the concept of a system/thing a bit. Constitutive explanations typically explain causal capacities of material things or systems, and it could be questioned if a social institution or a social entity can be a system in the relevant way. Institutions such as governments have capacities, such as the capacity to make decisions, and this is enough to qualify for an explanandum for constitutive explanations.

emphasis added). Similarly, Jäger describes the main goal of discourse analysis as

> The (dominating) discourses can be criticized and problematized; this is done by analysing them, by revealing their contradictions and nonexpression and/or the spectrum of *what can be said* and what can be done covered by them. (Jäger 2001, 34; emphasis added)

There is a major reward in considering statements as discursive capacities of institutions: such conceptualization allows for a purely descriptive and discursive—i.e., nonrepresentational—account of rules of formation. Let us consider a discursive practice as the set of all acts of moving pieces inside a checkerboard made by a community of users. Once we have observed what users do, we can describe what kind of moves occur. Without hypothesizing what users try to do, we can describe some moves as common, and others as rare. Rules of formation concerning objects and subjects in a discursive practice work as patterns of permutation in this example: they describe which particular performances are actualized, and therefore become *genuine* moves made by genuine players, and how distributed they are in the total space of possibilities of that practice. Moreover, looking at how players move the pieces, one might identify certain recurring patterns, such as moves that are preceded or followed by other moves. Rules of formation concerning concepts and strategies can be interpreted as analog to these more complex patterns, that is, describing what conditions or consequences for statements are possible, that is, included in the space of actualized acts. All four kinds of patterns/rules of formation are regularities, that is, descriptive features of the discursive practice.

This takes me to the issue of constitution, that is, the kind of explanatory account that rules of formation *afford* practitioners the possibility to make statements tokens. As I discussed in the previous section, statements types and rules of formation are regularities of recurring activities in a practice. These regularities do not *cause* agents to think "I am entitled to take this action" or "This action is sanctioned," but simply provide practitioners with an array or a repertoire of actions. This repertoire is a *constitutive fact* about a social system, that is, it makes up the social practice. In natural entities, constitutive facts will mainly concern the material constitution of the entities, that is their material parts. In contrast, social entities (social agents, social institutions, social collectives, etc.), consist of both material and immaterial constituents and parts. What makes up a collective such as a circle of friends or an editorial board is both the physical individuals and material resources in it and the social relations, shared feelings and beliefs, and very often a shared linguistic repertoire. This shared linguistic repertoire—i.e., these rules of formation—constitutes certain performances as statements, as it not only contains

examples of activities but also of series of activities (e.g., this act x is only performed after act y; act x is never performed after act y).

The capacities of the system that are explained by these constitutive facts are two. First, participants in a practice need only to conclude that "This possibility exists in this context; it is *on the table*," which is something that they can derive from simple observations of the system's regularities. Therefore, participation in a practice provides access to the shared discursive repertoires and allows the forming descriptive judgments of about the regularities of the system. The possibility of forming this individual belief is a causal capacity of the system, as the system affords that kind of causal effect on individual representation in the same way as constitutive facts about a stone confer it the capacity to shatter glass.

Secondly, a more complex consequence is that this constitutive fact of the system (once again, its shared repertoire of regularities) seems to afford *individuals* with certain discursive capacities, and maybe, even certain social powers. Possibly, participating in a practice and having access to a shared discursive repertoire makes it possible for individuals to enact certain discursive performances, such as the *capacity to perform the actions in that repertoire*, and *the capacity to perform actions with a certain status*. The former capacity, that is the capacity to perform a discursive activity part of the repertoire, is a simple indirect causal consequence of the first causal capacity of the system discussed above: if the system can afford a change of representation in individuals, then individuals can develop a capacity for certain performances. Hence, the explanatory value of R consists in describing which chains of statements (what Foucault calls *procedures of rewriting* 1969 [1972], 59) were available resources for the participant in that specific discursive practice. Discourse analysis might therefore explain the missing arrow in model M by stating that the chain of statements "negative test result → translation → GPM" was *on the table*, meaning that it was an available discursive resource. The decision-makers could reaffirm the GPM despite the negative results (i.e., the system had that capacity) because the rule of formation R provided them with the option of playing the decision-making game in that way (i.e., the constitution base of the game involved that move). In other words, the system had the capacity to change the decision-makers' repertoire in a way that made that performance appear like a possibility.[6]

An important consequence of interpreting discourse analysis from the perspective of constitutive explanations is that the dualism regularity/rule that

[6]However, the shared repertoire does not constitutively explain the individuals' capacities of enacting a performance, as constitutive explanations explain with reference to constitutive facts about the same system that has that capacity (in this case the individual). Hence, the constitution of the system cannot explain the capacities of individuals.

was understandably criticized by Dreyfus and Rabinow gets a simple solution. Rules of formation are regularities, in the sense that is discussed above, that is, they are descriptive patterns in discursive practice. Yet, they are rules, not in the sense of evaluative beliefs (beliefs about what is sanctioned or what is acceptable to do, and so on, which is how Dreyfus and Rabinow seem to understand rules), but rather in the original sense of the Latin term *regula*. This term denoted a straight stick from a tree that was used to perform measurements. The tree stick did not provide a rule for how to perform measurements correctly. Rather, it worked as a rule in the sense of an example that could be used for comparison. Rules of formation regulate practices by providing relevant examples.

As for the latter apparent individual capacity, i.e., the capacity of performing acts with a certain status (e.g., to speak as an expert, or to assert a fact), it can be argued that discourse analysis models these as facts about the system, rather than facts about individuals and their capacities. To qualify a performance as a statement (i.e., as an expert performance, as a fact, truth, or a matter of common sense) means to locate that performance in a particular position in the practice. Therefore, rules of formation can constitutively explain the fact that certain performances have certain statuses, but not the fact that an individual has the power to make that performance.

Finally, discourse analysis seems to fit the theory of constitutive explanations also from the point of view of the counterfactual dependencies that the method can afford. Rules of formation are counterfactually related to the system's capacity to afford certain changes in the participants' representations representation. If certain options (such as rewriting program budgeting into the GPM after the negative test results) were not on the table, that is, were not concretized parts of the discursive practice, then the participants would not have changed their representation of what was on the table in that way.

7. Conclusion (with a Suggestion for Integrative Approaches)

In concluding this paper, discourse analysis does not seem to satisfy the demands for causal explanations that I introduced earlier in the paper: the specification of one or several counterfactual dependencies and the specification of a mechanism or part of a mechanism. Rather, discourse analysis seems to satisfy a non-causal form of these demands. First, models generated using discourse analysis can track counterfactual dependencies, but not causal ones. Instead, they can track constitutive dependencies, describing how the discursive capacities of the targeted system depend on the available discursive resources (which are constitutive elements of the targeted social system). Secondly, models in discourse analysis do not describe mechanisms or parts of mechanisms—i.e., networks of entities and activities mediating or modulating

a more general causal relation. Therefore, discourse analysis cannot be used to further specify a mechanistic explanation. Instead, discourse analysis can contribute to a mechanistic explanation by providing an account of the constitutive constraints on an observed mechanism.

This conclusion entails an important methodological prescription for approaches—such as DI—which aims at integrating discourse analysis with other models to provide explanations of social phenomena such as institutional change. If, as I propose, discourse analysis produces constitutive explanations, then it can explain only a specific kind of explananda: causal capacities. This means that the integration of a discourse analytical model into a broader mechanistic model will play an explanatory role for institutional change (or any other social phenomenon) only in so far as there is a causal capacity to be explained. In the case of Wahlström's study, the explanatory gap concerned a causal capacity of the system, rather than a causal event. Therefore, using discourse analysis in this case would have contributed to answering the basic explanatory question motivating Wahlström's study. In cases when a model requires further specification, because of some gaps concerning the causal part of a mechanism (e.g., a missing mediator or modulator), then it could be that the inclusion of a discourse analytical model will not improve the explanatory power of the overall model. In simple terms, discourse analysis will contribute to a causal explanation whenever there is a genuine question about the causal capacity of a system. This conclusion can be helpful for approaches such as DI, that integrate discourse analysis with other models to improve the explanatory power of the model. These approaches might benefit from clarifying whether there is a genuine question about causal capacities that requires answering and what causal capacity needs an explanation, thereby making the contribution of discourse analysis clearer.

Acknowledgements

I would like to thank the participants at the ENPOSS 2023 conference at Radboud University in Nijmegen, for providing several helpful comments on this paper. I would also like to thank my colleagues at the SITE research group at Linnaeus University for commenting on an earlier version of this paper, and especially Ninni Wahlström for providing me with crucial insights into the case I discuss in this paper.

Declarations of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research for this paper was supported by

the Linnaeus-SITE (Studies in Curriculum, Teaching and Evaluation) research group, as a part of the research program *The New Meaning of Teaching—Curriculum, Knowledge, and Democracy in Changing Times.*

ORCID iD

Corrado Matta https://orcid.org/0000-0003-2282-8071

References

Alasuutari, Pertti. 2015. "The Discursive Side of New Institutionalism." *Cultural Sociology* 9 (2): 162-184. doi:10.1177/1749975514561805.

Amnå, Erik. 1981. *Planhushållning i Den Offentliga Sektorn? Budgetdialogen Mellan Regering Och Förvaltning under Efterkrigstiden [Planning in the Public Sector? The Budget Dialogue between the Cabinet and Bureaucracy in Sweden After 1945]*. Stockholm: Norstedt.

Dreyfus, Hubert L. and Paul Rabinow. 1982 (1983). *Michel Foucault: Beyond Structuralism and Hermeneutics*. 2nd ed. Chicago: University of Chicago Press.

Elder-Vass, Dave. 2011. "The Causal Power of Discourse." *Journal for the Theory of Social Behaviour* 41 (2): 143-60. doi:10.1111/j.1468-5914.2010.00449.x.

Fairclough, Norman. 1995. *Critical Discourse Analysis: The Critical Study of Language*. London: Longman.

Fairclough, Norman. 2003. *Analysing Discourse: Textual Analysis for Social Research*. New York: Routledge.

Foucault, Michel. 1969 (1972). *The Archaeology of Knowledge and the Discourse on Language*. Translated by A. M. Sheridan Smith. New York: Pantheon Books.

Glennan, Stuart. 2002. "Rethinking Mechanistic Explanation." *Philosophy of Science* 69 (S3): S342-53. doi:10.1086/341857.

Hedström, Peter and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36: 49-67.

Illari, Phyllis McKay and Jon Williamson. 2012. "What Is a Mechanism? Thinking about Mechanisms *across* the Sciences." *European Journal for Philosophy of Science* 2 (1): 119-35. doi:10.1007/s13194-011-0038-2.

Jäger, Siegfried. 2001. "Discourse and Knowledge: Theoretical and Methodological Aspects of a Critical Discourse and Dipositive Analysis." In *Methods of Critical Discourse Analysis*, edited by Ruth Wodak and Michael Meyer, 32-62. London: Sage.

Jørgensen, Marianne and Louise Phillips. 2002. *Discourse Analysis as Theory and Method*. London: Sage.

Kusch, Martin. 1991. *Foucault's Strata and Fields: An Investigation into Archaeological and Genealogical Science Studies*. Dordrecht, The Netherlands: Kluwer.

Laclau, Ernesto and Chantal Mouffe. 1985. *Hegemony & Socialist Strategy*. London: Verso.

Matta, Corrado. 2015. "Interpretivism and Causal Explanations: A Case from Educational Research." *Philosophy of the Social Sciences* 45 (6): 543-67.

Schmidt, Vivien A. 2008. "Discursive Institutionalism: The Explanatory Power of Ideas and Discourse." *Annual Review of Political Science* 11 (1): 303-26. doi:10.1146/annurev.polisci.11.060606.135342.

Schmidt, Vivien A. 2015. "Discursive Institutionalism: Understanding Policy in Context." In *Handbook of Critical Policy Studies*, edited by Frank Fischer, Douglas Torgerson, Anna Durnová, and Michael Orsini, 171-89. Cheltenham, UK: Edward Elgar.

Sundström, Göran. 2003. *Stat På Villovägar: Resultatstyrningens Framväxt i Ett Historisk-Institutionellt Perspektiv [State On Wild Roads: The Emergence of Performance Management in a Historical-Institutional Perspective]*. Stockholm: Dept. of Political Science [Statsvetenskapliga institutionen], Stockholm University.

Swedish Ministry of Finance. 1987. "Förslag till slutlig reglering av statsbudgeten för budgetåret 1987/1988 [Proposal for Final Regulation of the State Budget 1987/1988]." Proposition Number: 1987/88:150 https://data.riksdagen.se/dokument/GB03150

Tiisala, Tuomo. 2015. "Keeping it Implicit: A Defense of Foucault's Archaeology of Knowledge." *Journal of the American Philosophical Association* 1 (4): 653-73. doi:10.1017/apa.2015.29.

Van Dijk, Teun A. 2006. "Ideology and Discourse Analysis." *Journal of Political Ideologies* 11 (2): 115-40. doi:10.1080/13569310600687908.

Wahlström, Ninni. 2009. *Mellan Leverans Och Utbildning: Om Lärande I En Mål- Och Resultatstyrd Skola [Between Delivery and Education: About Learning in a Goal and Performance-Driven School]*. Göteborg: Daidalos.

Wodak, Ruth and Michael Meyer, eds. 2001. *Methods of Critical Discourse Analysis*. London: Sage.

Ylikoski, Petri. 2013. "Causal and Constitutive Explanation Compared." *Erkenntnis* 78 (2): 277-97. doi:10.1007/s10670-013-9513-9.

Author Biography

Corrado Matta is a senior lecturer at the Department of Pedagogy and Learning, Linnaeus University (Sweden). His main research areas are the philosophical foundations of social research methods, with a special focus on qualitative and mixed methods, and the philosophy of evidence-based practice in education. His latest research discussed issues such as the concept of evidential pluralism in education, the methodological rewards of data and theory integration, how neuroscience can (and cannot) support educational practice, and the abductive foundation of theory construction in qualitative research.