Meaningful Lines

Social Semiotic Investigations of the Graphical Line, Used as a Connector in Digital Data Visualizations

Verena Elisabeth Lechner

UiA University of Agder

Doctoral Dissertations at the University of Agder 302

Verena Elisabeth Lechner

Meaningful Lines

Social Semiotic Investigations of the Graphical Line, Used as a Connector in Digital Data Visualizations

Dissertation for the degree philosophiae doctor

University of Agder
Faculty of Humanities and Education
2020

Doctoral dissertations at the University of Agder 302

ISSN: 1504-9272
ISBN: 978-82-8427-004-3

Print: 07 Media, Kristiansand

Acknowledgements

If I were asked to visualize the journey of writing this dissertation by drawing a line between two dots, this line would probably end up as a curvy line rather than a straight one. There have been many ups and downs, and today I stand in great dept to everybody who has supported me during this time. I would like to thank the following people:

My supervisor, Martin Engebretsen, for his excellent supervision throughout this project and for welcoming me into the INDVIL research group. Thank you, Martin, for never getting tired of listening to my thoughts about lines, whether in English, Norwegian or a mixture of both. I am deeply grateful for all your comments and advice, which have had an immense impact on the quality of this dissertation.

My co-supervisor, Jana Holsanova. Thank you for all your supportive feedback, and for keeping an eye on the big picture of the project.

The Research Council of Norway, for funding the INDVIL research project (grant number 259536) and thus my fellowship (see indvil.org).

The Faculty of Humanities and Pedagogy and the Department of Nordic and Media Studies at the University of Agder, for supporting the project.

Hubert Matt and Karl-Heinz Weidmann from the University of Applied Sciences, Vorarlberg, Austria, for supporting me since long before I started writing this dissertation.

My colleagues and friends at the University of Agder and all INDVILers. Thank you for your support both on professional and personal levels. I am extremely grateful to my colleagues at Knausen and the coffee corners on both levels of the E-building for their encouragement and for contributing to making these years a great time in my life. Special thanks to Malin Toften Mangersnes, with whom I spent almost three years and shared all the ups and downs of our journeys. Thank you for all your language feedback and for always having your door open for me.

My family and friends for their continuous support. This thesis would not have been possible without you believing in me and keeping me grounded. Special thanks to my brother, Eric Lechner, for contributing his technical expertise. My deepest appreciation goes to Bernhard Fäßler, for going through this adventure with me and doing everything necessary to make me feel happy.

Verena Elisabeth Lechner
Kristiansand, September 2020

Summary in Norwegian

Avhandlingen undersøker den grafiske linjen brukt som bindeelement i digitale datavisualiseringer (DV) fra et sosialsemiotisk perspektiv. Den utforsker de semiotiske funksjonene til bindelinjer i digitale DV-er med empirisk bakgrunn i to korpusanalyser. Disse analysene handler om hvordan to viktige semiotiske funksjoner realiseres grafisk i de valgte linjene, og hvorvidt det finnes grafiske konvensjoner som regulerer disse realiseringene.

Avhandlingen undersøker bindelinjen som et sentralt element i flere ulike DV-typer: linjegrafer, tidslinjer, rutekart, forbindelseskart, nettverksdiagrammer og liknende. I dag brukes DV-er i en rekke ulike sosiale kontekster og spiller derfor en viktig rolle i samfunnet. Nye digitale teknologier gjør det mulig å gi grafiske linjer i DV-er andre visuelle egenskaper enn det som tidligere var mulig, for eksempel da de ble tegnet for hånd eller trykket. Denne utviklingen kan ha påvirket DV-designpraksisen, og denne situasjonen trenger oppdatert forskning.

Avhandlingen er artikkelbasert og består av tre artikler og et utvidet sammendrag, den såkalte *kappa*. Det utvidede sammendraget kontekstualiserer de tre artiklene og omfatter en introduksjon, definisjoner av sentrale termer, mer omfattende deler om det teoretiske rammeverket, tidligere forskning og metodologi samt etiske hensyn, refleksjoner rundt metodiske beslutninger og en konklusjon. De to overordnede forskningsspørsmålene som avhandlingen forsøker å besvare, er disse:

(1) Hva karakteriserer de grafiske formene og semiotiske funksjonene til bindelinjer i våre dagers offentlig tilgjengelige, digitale DV-er?

(2) Hvilke konvensjoner, som regulerer forholdene mellom disse formene og funksjonene, kan observeres i slike DV-er?

Artikkel I legger frem det teoretiske grunnlaget for hele avhandlingen. Den demonstrerer hvordan sentrale konsepter fra sosialsemiotisk teori kan bli brukt til å analysere det semiotiske potensialet til bindelinjer i DV-er. Artikkelen

identifiserer fire semiotiske funksjoner i linjer (i tillegg til å representere en forbindelse): (1) indikere graden av sikkerhet som en modalitetsmarkør, (2) indikere en narrativ eller konseptuell påstand, (3) indikere kohesjonsmønstre og (4) regulere leserens posisjon til objektene som representeres. Artikkelen inneholder eksempler på DV-er og drøfter disse for å demonstrere hvordan bindelinjer gjennom sin visuelle utforming kan realisere hver av disse semiotiske funksjonene. Den konkluderer med en oppfordring til videre empirisk og korpusbasert forskning for å undersøke hvorvidt og hvordan disse funksjonene realiseres i samtidens DV-er, og for å undersøke konvensjonene etter som de potensielt utvikler seg over tid.

Denne tråden tas opp i Artikkel II og III, som hver fokuserer på én av de semiotiske funksjonene som ble identifisert i Artikkel I. Undersøkelsene som presenteres i disse to artiklene, er basert på et korpus spesifikt utviklet for avhandlingen. Det inneholder 163 prisbelønte DV-er, der bindelinjene har en ledende rolle i å kommunisere det intenderte budskapet til DV-en. Alle DV-ene retter seg mot lesere i den vestlige verden og ble tildelt priser mellom 2015 og 2018.

Artikkel II konsentrerer seg om hvorvidt og hvordan bindelinjenes spesifikke visuelle egenskaper indikerer *modalitet* og *usikkerhet* i DV-ene i korpuset. I den teoretiske delen av denne artikkelen drøfter jeg forholdet mellom modalitet (som drøftet av lingvister og semiotikere) og usikkerhet (som drøftet i DV-praksisfeltet). Usikkerhet og modalitet er tett forbundet, og førstnevnte blir identifisert som et bredere konsept enn sistnevnte. Korpusanalysen som presenteres i denne artikkelen, fokuserer bare på visuelle representasjoner av modalitet som handler om lavere sannsynlighet eller pålitelighet. Den stiller spørsmål om (1) hvordan bindelinjer i DV-ene i korpuset signaliserer lavere sannsynlighet og pålitelighet, og (2) hvorvidt det kan observeres grafiske konvensjoner knyttet til slike representasjoner. Det ble funnet at flere visuelle egenskaper av de undersøkte linjene ble brukt til å signalisere lavere sannsynlighet og lavere pålitelighet. Lavere sannsynlighet ble oftest signalisert med visse mønstre som brøt linjen. Det ble også observert at slike linjeavbrudd sjelden ble brukt til andre formål. Dette indikerer en grafisk konvensjon om å bruke avbrudd på bindelinjer for å signalisere modalitet, av typen lavere sannsynlighet. Undersøkelsen fant ikke liknende konvensjoner for signalisering av lavere pålitelighet. Det ble observert at denne funksjonen har et større spekter når det gjelder bindelinjens visuelle utforminger.

Artikkel III utforsker hvilke visuelle egenskaper som brukes i korpuset for å la bindelinjer indikere en narrativ påstand. Det sosialsemiotiske konseptet *representerende strukturer*, som grunnleggende skiller mellom *narrative* og *konseptuelle representerende strukturer*, står sentralt i denne undersøkelsen. Denne artikkelen presenterer en korpusundersøkelse som utforsker hvilke grafiske variasjoner i bindelinjen som uttrykker narrativitet, og hvor ofte og i hvilke kombinasjoner de brukes i det undersøkte korpuset. På samme måte som i Artikkel II blir visse grafiske konvensjoner identifisert på grunnlag av den empiriske undersøkelsen. I det undersøkte korpuset blir opptil fire visuelle egenskaper ble brukt i kombinasjon for å signalisere narrativitet. De fleste av de analyserte linjene har imidlertid bare én visuell egenskap som uttrykker retning og dermed narrativitet. Når de sees enkeltvis, forekommer de følgende to gruppene med visuelle linje-egenskaper oftest for å signalisere narrativitet: dynamiske egenskaper (først og fremst endringer i størrelse, men også i posisjon og mønster) og pilspisser eller liknende grafiske elementer som indikerer retning (f.eks. et ikon av et fly). Dette indikerer at det finnes en konvensjon om å signalisere narrativitet med linjer gjennom dynamikk i størrelser (dvs. linjer som vokser i én retning) og pilspisser eller liknende (først og fremst pilspisser eller liknende i én av endene). Man kan konkludere med at konvensjonene som ble observert i denne studien, er påvirket av inntoget av digitale produksjonsteknikker og skjermbaserte visningsflater, men de bygger også på en lang tradisjon med analog visualisering.

For å oppsummere: Avhandlingen ser på de semiotiske funksjonene til bindelinjer i nåtidige offentlig tilgjengelige digitale DV-er fra et sosialsemiotisk perspektiv. Den presenterer to korpusanalyser som bygger på et korpus med 163 prisbelønte DV-er. De to analysene undersøker hvordan de fokuserte bindelinjene signaliserer henholdsvis modalitet og narrativitet. I tillegg til at avhandlingen gir et originalt bidrag til forskningslitteraturen innen datavisualisering og visuell kommunikasjon, er resultatene særlig verdifulle for DV-designere, programvareutviklere, brukere som setter opp retningslinjer for DV-design samt instruktører og lærere innen visuell kommunikasjon.

Summary in English

The dissertation investigates the graphical line, used as a connector in digital data visualizations (DVs), from a social semiotic perspective. It explores the semiotic functions of connecting lines in digital DVs, building empirically on two corpus analyses. These analyses are concerned with how two key semiotic functions are graphically realized in the selected lines and whether graphical conventions regulating these realizations can be identified.

In the dissertation the connecting line is researched as a central element of several different DV types – line graphs, timelines, route maps, connection maps, network diagrams, etc. Today, DVs are used in a broad range of social contexts and thus play an essential role in society. New digital technologies make it possible to apply other visual characteristics to graphical lines in DVs than was previously possible, for example, when they were hand-drawn or printed. This change may have influenced the DV design practice, a situation calling for updated research.

The dissertation is article-based, comprising three articles and an extended abstract. The extended abstract frames the three articles and includes an introduction, definitions of key terms, more comprehensive sections on the theoretical framework, earlier research and methodology, as well as ethical considerations, reflections on methodological decisions and a conclusion. The two overall research questions addressed in the dissertation are:

(1) What characterizes the graphical forms and semiotic functions of connecting lines in current, publicly available digital DVs?
(2) What conventions, regulating the relations between these forms and functions, can be observed in such DVs?

Article I lays the theoretical basis for the whole dissertation. It demonstrates how key concepts from social semiotic theory can be used to analyse the semiotic potential of connecting lines in DVs. Four semiotic functions of lines (additional

to representing a connection) are identified: (1) indicating certainty as a modality marker, (2) indicating a narrative or conceptual claim, (3) indicating patterns of cohesion and (4) regulating the reader's position to the represented objects. Exemplary DVs are discussed to demonstrate how connecting lines, through their visual design, can realize each of these semiotic functions. It concludes with a call for further empirical, corpus-based research to gauge if and how these functions are realised in current DVs, and to investigate the potentially evolving conventions.

This thread is taken up by Articles II and III, which each focus on one of the semiotic functions identified in Article I. The research presented in these two articles is based on a corpus, specifically developed for this dissertation. It contains 163 award-winning DVs, where connecting lines are in the leading role of communicating the intended meaning of the DV. They are all targeted at audiences in the Western world and were awarded in the years between 2015 and 2018.

Article II concentrates on whether and how *modality* and *uncertainty* are indicated through certain visual characteristics of the connecting lines in the DVs of the corpus. In the theoretical part of this article, I discuss the relation between modality (as discussed by linguists and semioticians) and uncertainty (as it is discussed in the practice field of DV). Uncertainty and modality are intertwined, and the former is identified as a wider concept than the latter. The corpus analysis presented in this article focuses only on visual representations of modality that relate to lowered probability or reliability. It asks: (1) how connecting lines in the DVs of the corpus signal lowered probability and reliability, and (2) whether any graphical conventions concerning the same issue can be observed. The findings show that several visual characteristics of the studied lines were used to signal lowered probability and lowered reliability. For signalling lowered probability, patterns which lead to interruptions of the line are most used. It was also observed that such line interruptions were rarely used for other purposes. This indicates a graphical convention for using interruptions of connecting lines for signalling modality in terms of lowered probability. For signalling lowered reliability, the study did not reveal any similar conventions. For this purpose, a broader range of visual line characteristics was found.

Article III explores what visual characteristics are used in the corpus to let connecting lines indicate a narrative claim. The social semiotic concept of *representational structures*, distinguishing at its broadest level between *narrative* and *conceptual representation structures*, is central to this investigation. The corpus study presented in this article investigates which graphical variations of the connecting line, expressing narrativity, can be observed in the corpus and with what frequency and in what combinations they are used. Based on the results, evolving graphical conventions were identified, as in the study presented in Article II. Up to four visual characteristics were used in combination to signal narrativity. However, most of the analysed lines have only one visual characteristic that expresses direction, and thus narrativity. Viewed individually, the following two groups of visual line characteristics are most prevalent for signalling narrativity: dynamic features (above all changes in size, but also in position or pattern) and arrowheads or similar graphic elements indicating direction (like, e.g., an icon of a plane). This indicates a convention for signalling narrativity by lines through dynamics in size (i.e., lines growing in one direction) and arrowheads or similar (above all arrowheads or similar at one end). One can conclude that the conventions observed in this study are influenced by the advent of digital production techniques and screen-based display devices, but they also build on the long tradition of analogue visualization.

Summing up, the dissertation takes a social semiotic view on the semiotic functions of connecting lines in current, publicly available digital DVs. It presents two corpus analyses, building on a corpus of 163 award-winning DVs. The two analyses investigate how the focused connecting lines signal modality and narrativity, respectively. In addition to the dissertation making an original contribution to the research literature within data visualization and visual communication, the results are particularly valuable for DV designers, DV tool developers and practitioners setting up DV design guidelines as well as educators and teachers in visual communication.

Contents

List of Tables

List of Figures

1 Introduction

The way we see our world depends heavily on the information we have at hand. As a basis for creating a picture for ourselves of what is going on, we can draw on data – a resource in constant growth. Even when looking only at the years from 2010 to 2017, the volume of digital data created globally increased by a factor of 13 (from 2 to 26 zettabytes), and a steep further growth curve is forecasted (Statista, 2018; based on data from Reinsel et al., 2018). This development makes explicit that the number of quantified phenomena rises - a process sometimes called "*datafication*" (Mayer-Schönberger & Cukier, 2013, p. 78). The collected data cover all aspects of our society, such as our words, locations, and interactions, as Mayer-Schönberger and Cukier further explains (p. 83-94).

However, having a substantial amount of data available does not automatically mean that we can make use of them. The size of the global datasphere leads to the challenge of making meaning from all these data. Data visualizations, further referred to as *DVs*, can be used as a tool to make data more accessible and usable. They may be valuable in data analysis processes and can also be used for communication purposes. It is, therefore, hardly surprising that DVs are used in a broad range of social contexts and play an important and growing role in contemporary mediated communication (Kennedy & Engebretsen, 2020, p. 19; Riche et al., 2018, p. 2).

1.1 Motivation

For DVs to work according to their potential, their users must have a shared understanding of their meanings. Communication based on graphical depictions, such as DVs, relies on visual language[1] – a language where visual expression and meaning are linked systematically.

1 The choice of the term *language* in this context needs an explanation. The verbal and visual modes of expression each have their own set of semiotic affordances as pointed out, for example, by Kress and Van Leeuwen (1996/2006, pp. 4–5) and Saint-Martin (1987/1990, pp. x–xi). Thus, one may question why using linguistic

To be able to apply this visual language effectively in communication, visual literacy skills are needed (Kędra, 2018, p. 70). Visual literacy skills include, amongst other things, knowledge about the basic components of visual language, and about what universal visual codes or conventions exist (Kędra, 2018, p. 77). Both the DV designers (used as a collective term for everybody involved in the process on deciding for the visual appearance of a DV) and the readers (in this dissertation also referred to as *viewers*) need such knowledge to communicate with each other.

From a DV designer point of view, Brinton (1914) observed that "unless the facts are presented in a clear and interesting manner, they are about as effective as a phonograph record with the phonograph missing" (p. 2). He regarded it as challenging to know what rules need to be applied when designing visualizations, as can be read in his seminal work *Graphic Methods for Presenting Facts* (1914):

> The trouble at present is that there are no standards by which graphic presentations can be prepared in accordance with definite rules so that their interpretation by the reader may be both rapid and accurate. It is certain that there will evolve for methods of graphic presentation a few useful and definite rules which will correspond with the rules of grammar for the spoken and written language. (p. 3)

Now, more than a hundred years later, we can ask: Has this situation of not having any definite rules changed? Iliinsky and Steele (2011) made an observation that relates precisely to this issue:

> The second challenge for designing more complex visualizations [besides having to find ways to visually encode more than just a few dimensions] is that there are relatively few well-known conventions, meta-

terms and frameworks for the analysis of graphical elements in DVs is fruitful – an issue further problematized in Section 4.5. One reason for using the term visual *language* is that I apply a broad understanding of the term *language*, building on the definition of Lotman (1971/1977), who stated: "We understand language to mean any communication system employing signs which are ordered in a particular manner" (p. 8). Moreover, several scholars of areas central to this dissertation (social semiotics, visual literacy and design studies) use the term *visual language* (Avgerinou, 2009, p. 29; Horn, 1998; Kędra, 2018, p. 77; Kress & Van Leeuwen, 1996/2006, pp. 4, 5; Poulin, 2018, p. 7).

> phors, defaults, and best practices to rely on. Because the safety net of convention may not exist, there is more of a burden on the designer to make good choices that can be easily understood by the reader. (p. 4)

Following on from this, definite rules have not been developed, and there is a lack of clear and publicly shared conventions in the field of DV. Pointing to the same issue, Munzner (2014) offers future DV designers some “rules of thumb” (p. 117) – advice and guidelines for how to design DVs (p. 117–144). She calls them *rules of thumb* because “they are not set in stone; indeed, they are deeply incomplete” (p. 117). She further comments that “the characterization of what idioms are appropriate for which task and data abstractions is still an ongoing research frontier, and there are many open questions” (p. 117).

However, it is reasonable to assume that the community of DV users has built up (and still is building up) certain graphical conventions. It is an underlying hypothesis of this dissertation that such conventions do exist, or are in the process of being shaped, particularly in the domain of digital DVs. This can be expected because large and heterogeneous audiences can access, and are targeted by, digital DVs. However, as Iliinsky and Steele (2011) and Munzner (2014) point out, the available knowledge about such conventions is limited, leaving the designers with wide freedom in their design decisions, and the readers with uncertain interpretations.

The visual language of graphics as applied in DVs is still only partly described (see, e.g., Bertin (1967/2011) and Kress & Van Leeuwen (1996/2006) – Chapter 5 presents the current state of knowledge concerning the graphical line within DVs). For developing a more detailed description, the issue of evolving conventions is a core challenge and it is important to develop descriptions of visual language that are more nuanced and more widely shared than is currently the case. Existing conventions must be detected and described, and research findings must find their way into textbooks and education (see, e.g., Holsanova, in press; Kress & Van Leeuwen, 1996/2006, pp. 16–17; Unsworth, 1997, p. 35).

1.2 Research Focus

The objective of this dissertation is to investigate graphical conventions in the domain of digital, publicly accessible DVs. As for the term *convention*, I adopt the definition suggested by Kennedy et al. (2016), who define it as “a symbolic or social practice that is shared, readily understood and widely accepted by members of a cultural group" (p. 717).

The study of graphical conventions in digital DVs represents a vast field of research. In this dissertation, I focus on only one graphical element, which forms the basis of many different DV types: the line, used as a connector (further called *connecting line*). Timelines, line graphs, route maps, connection maps, flow maps, network diagrams and slope charts are just a few examples of visualization types that have connecting lines as their central graphical element. Lines in such DVs do not only connect entities; they can also have additional semiotic functions. Which ones they are, and what the relations are between the visual appearance of the lines and their perceived meaning potentials, is highly valuable information – not least for DV designers. This kind of knowledge informs the design decisions made for every single graphical element of a DV. Knowing what is "readily understood and widely accepted by members of a cultural group" (Kennedy et al., 2016, p. 717) is the basis for being able to decide whether to follow these conventions or not[2]. This dissertation explores the graphical conventions of the visual language within DVs, focusing on the semiotic functions of connecting lines, complementing the basic function of connecting.

The focus of this dissertation differs from many other studies that investigate only the visual appearance and the meaning potential of connecting lines within one specific domain (like cartography) or concrete DV application (see Chapter 5). Instead, I research connecting lines as an element appearing across several different DV types. In this sense, the dissertation draws on the work of Tim Ingold. In *Lines: A brief history* (2007) he also focuses on the line as a phenomenon in various places of appearance, and covers a wide spectrum of examples of lines, as they occur in genealogical trees, drawings, writing, maps, textile techniques and so on. Nevertheless, his anthropological view on lines consciously does not cover any impact of the advent of computers (2007, p. 166). This again points to the need for more research within this area.

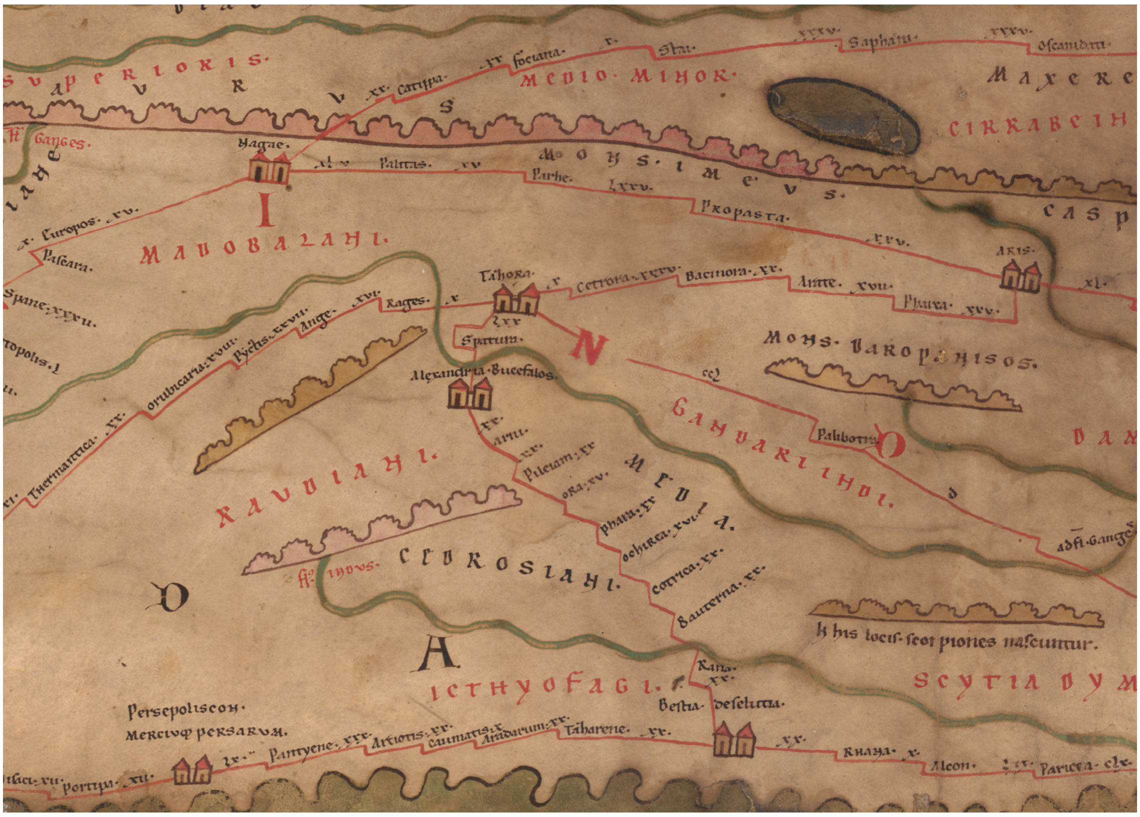
2 Obviously, the designer's decisions are not based only on the desired semiotic functions. Other reasons like, for example, the circumstances of the developer's tools or aesthetic reasons, may also influence how graphic elements in DVs are designed. This has to be borne in mind, but due to the fact that the exact reasons for those decisions are seldom published, it will often not be possible to discover the exact reasons for the visual appearance of the final result. In this dissertation I expect the desired semiotic function to be the most important reason for the designer's decisions on the analysed DVs.

1.3 Historical Perspective on the Use of Connecting Lines

Using lines to represent connections between elements is an old technique. Brinton, for example, shows many examples of "relationship charts [emphasis deleted]" (1939, p. 68) basing on lines as the visual representation of connections (p. 68–72). Bertin also suggests in *Semiology of Graphics: Diagrams, Networks, Maps* (1967/2011, p. 270), several different variants of network visualizations using the line as the connecting element between two other elements. A much older, historical example of a DV with connecting lines as the central graphical element is the Peutinger Map (Unknown et al., ca. 1200), known as the first route map (see Figure 1). Red lines visualize the connections (roads) of the road network of the Roman Empire. It was drawn by hand on parchment (Unknown et al., ca. 1200). Therefore, elements such as the drawing tool, the surface of this parchment and the drawing abilities of the artist influenced the graphical form of the red lines (and all the other graphical elements).

Figure 1

Detail of the Peutinger Map (ca. 1200, original: 4th century BC according to Friendly & Denis, 2001)



Note. From "Tabula Peutingeriana." by Unknown et al., ca. 1200 (http://data.onb.ac.at/rep/10002029). Retrieved May 25, 2020. Provided by Österreichische Nationalbibliothek (Austrian National Library). Reprinted with permission.

In current DVs, using lines for representing connections between elements is still a ubiquitous technique. The number of DV types using lines to represent connections (see Section 7.3.6) and the number of published DVs that have this characteristic (see Section 7.3.3) document the need to visualize connections. Several tools, developed to create visualizations digitally (e.g., D3.js, Tableau, R, Plotly), make it easy to produce such visualizations. However, the design options for connecting lines (and all other graphical elements of DVs) have changed, not least through the advent of digital technologies. In contrast to early, printed DVs, where the designers could rely only on static black and white representations, it is now possible to use colour, transparency, animation, and interaction, to name a few additional methods. It is therefore of interest to investigate what techniques DV designers use in current digital DVs, how they do it, and for what purpose.

1.4 Research Questions, Approach and Research Material

The following overall research questions are addressed in this dissertation:

(1) What characterizes the graphical forms and semiotic functions of connecting lines in current, publicly available digital DVs?

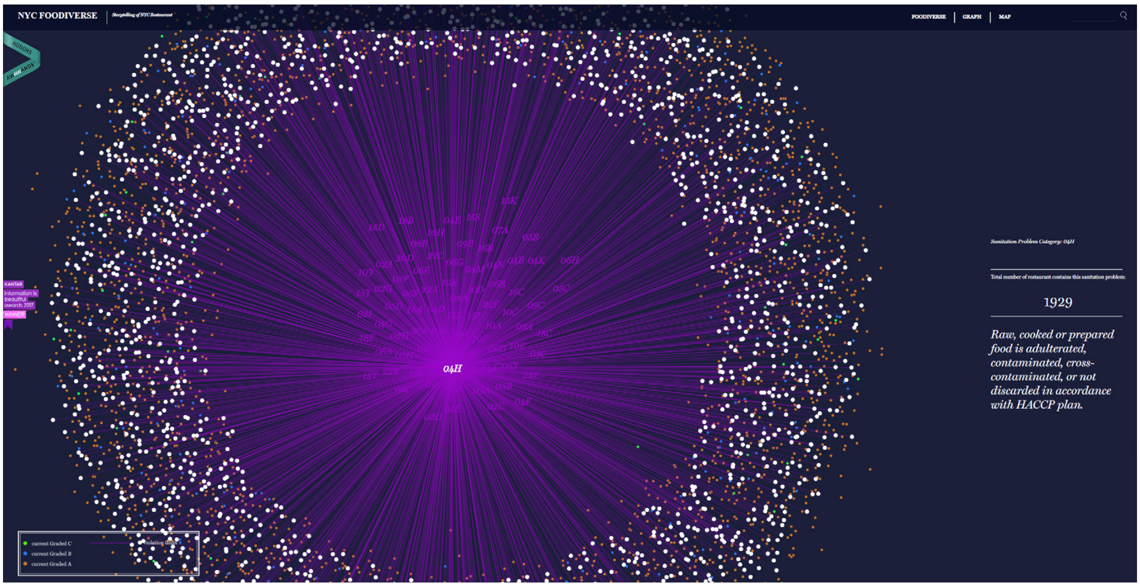
(2) What conventions, regulating the relations between these forms and functions, can be observed in such DVs?

I approach these research questions from a social semiotic perspective. This approach refers to the use of social semiotics as a theoretical basis, as well as to the use of social semiotic concepts as methodological tools. As Jewitt and Oyama (2001) put it, a social semiotic approach involves "the description of semiotic resources, what can be said and done with images (and other visual means of communication) and how the things people say and do with images can be interpreted" (p. 134). The social semiotic investigation in this dissertation is of a more descriptive type than social semiotics applied as a "tool for use in critical research" (Jewitt & Oyama, 2001, p. 136). The starting point for this investigation is a corpus that was built up specifically for this dissertation (see Section 7.3). This approach answers to the knowledge needs described earlier, which in short is a deeper, empirically based understanding of the graphical conventions governing the visual language of a domain gaining importance (the DV), with focus on a widely used graphic element (the connecting line).

I regard the DVs in the corpus (which constitute my research material) as a type of *visual text* that follows certain conventions constituting a visual language which is continually evolving. At the same time, DVs are also regarded as *multimodal texts*, since the graphical elements always interact with other semiotic modes. Thus, I use both concepts, *multimodal text* and *visual text* for referring to different aspects of DVs. This practice reflects my main interest in the graphical form and the meaning potential signalled by that form (as similarly reasoned by Björkvall, 2009, pp. 7–8). Figure 2 shows an example of such a DV, as found in my research material (Su, 2017). It shows which restaurant in New York has what kind of sanitation problem and includes both graphical elements (including many connecting lines) and verbal language.

Figure 2

Screenshot of a DV Serving as an Example of the Research Material for This Dissertation



Note. From “NYC FOODIVERSE” by W. Su, 2017 (http://nycfoodiverse.com). Retrieved April 25, 2018. Copyright 2017 by W. Su. Reprinted with permission.

1.5 Summary and Objectives

Summing up, this dissertation aims to explore, theoretically and empirically, the relationships between the graphical forms and the semiotic functions of graphical lines serving as connectors in DVs. The research is relevant to several fields, such as design and media studies, social semiotics and multimodality studies, visual literacy and visual culture studies. However, I directly address mainly the fields of design studies and social semiotics, which is reflected in the conceptual paradigms and the literature reviews presented. My knowledge interest is chiefly to understand current semiotic practices and to create awareness among DV designers as well as design and visual literacy educators. The objective is not to define a set of rules about how lines need to be designed, serving as a cookbook for practitioners. Instead, this work presents an exploration of a single graphical element within current visual text material that could be useful to future research and education in relevant fields, as well as to designers to help them make informed choices within future DV designs.

1.6 Thesis Structure

The dissertation comprises three articles and an extended abstract. The extended abstract frames the three articles, which are embedded in the extended abstract as Chapter 10.

The chapters of the extended abstract include the following contents (numbered according to the chapter numbers): (1) an introduction, (2) definitions of key terms, (3) the positioning of the research in the traditions of science, (4) the theoretical framework, (5) a literature review, (6) an overview of the articles, (7) the methodology, (8) ethical considerations, (9) summaries of the articles, (10) the articles, (11) reflections on methodological decisions, and finally (12) the conclusion.

2 Definitions of Key Terms

In this chapter, the key terms of this dissertation are defined.
To begin with, I clarify the overarching term of the chosen research material itself – *data visualization* (DV) – and contrast and compare it with alternative and similar terms. Being specifically interested in *connections*, visualized through *lines*, I subsequently present my understanding of these two terms and the combination of both (the *connecting line*). Finally, I explain which types of lines are included in this dissertation.

2.1 Data Visualization

DVs are the text type which forms the research material of this dissertation. Instead of using this term, I could have used *information visualization*, or more casually *data viz* or *info viz*, as they refer to the same phenomenon (Iliinsky & Steele, 2011, p. 7). Both *data visualization* and *information visualization* are used in various fields like cartography, journalism, statistics and graphic design, although in varying frequency (Engebretsen & Weber, 2018, p. 279). I have chosen to mainly use the term *data visualization* since it underlines that data are being visualized.

Seen from a functional point of view, a DV fulfils two main functions, as Engebretsen and Weber (2018) point out: "(1) it visualizes non-visible data (quantitative or qualitative data) … , and (2) it provides a cognitive tool for discovery and exploration and thus enables the recipients to gain new insights" (p. 279). In the words of Kirk (2016), DV is "the representation and presentation of data to facilitate understanding" (p. 19).

More specifically and in dissociation from an *infographic*, a DV is according to Illinsky and Steele (2011), drawn by algorithms and therefore it is easy to simply regenerate the same DV with new or updated data. What follows is that DVs are "relatively data-rich" while "often aesthetically barren" (Iliinsky & Steele, 2011, p. 7). That last aspect could lead to discussions within the DV community, since defining what is data-rich and aesthetically barren is challenging. What Iliinsky and Steele mean by the term *data visualization* becomes clearer when one regards what they consider to be an *infographic*. Other than the afore-mentioned alternative terms (information visualization, info viz, data viz) *infographic* cannot simply be used instead of *data visualization* (Iliinsky & Steele, 2011, p. 5). An infographic is different in form and origin, as they specify (p. 5): When creating infographics, manual tools are used instead of algorithms, which means that the data get custom treatment and cannot simply be replaced. This is only possible with relatively little data as a basis. With relatively little data and manual production techniques, as Iliinsky and Steele describe further, it is possible to develop infographics with an aesthetical richness that are very "specific to the data at hand" (Iliinsky & Steele, 2011, p. 5). However, in practice, it is not always easy to label such visual material as either an infographic or a data visualization since the boundaries blur and are not always so clearly visible. That is the case because what constitutes data-richness and aesthetical richness, and our understanding of what is considered as *manually treated*, changes over time. At the time of Brinton, one may have thought of everything produced with digital tools as not having been hand treated, whereas nowadays, digital drawing tools like Adobe Illustrator are seen as manual tools in contrast to DV production tools like D3.js, Tableau, R or Plotly.

In this dissertation, I define DVs as multimodal expressions having the function of visualizing non-visible data to facilitate understanding. They are simple to regenerate with other data, and thus non-manual tools (i.e., not only Adobe Illustrator or hand-drawing, for example) are assumed to have been used in at least parts of the DV production process. I take into consideration that in many DV production workflows, several tools are used for the different stages. I acknowledge the blurring boundaries between DVs and infographics but, at the same time, define the opposing poles.

In the following, I explain the broad types of DVs and the terms used for them. Although DVs can be categorized according to several criteria, on the formal and on the content side (Engebretsen & Weber, 2018, p. 278), I will focus only on explaining the differences between *graphs*, *charts*, *maps* and *diagrams*. Some of these terms are used interchangeably by other scholars or without any clear distinction between DVs. For instance, Few uses the term *graph* in his book *Show Me the Numbers: Designing Tables and Graphs to Enlighten* (2012) simply as a term to refer to a "visual display of quantitative information [emphasis deleted]" (p. 45), as an option for showing data in tables. According to Few, graphs are only characterized by displaying their values as visual objects "within an area delineated by one or more axes" (p. 45), while these axes provide scales. Similarly, Kirk uses the term *chart* in his explanation of what DV is, as another word for *data visualization* (2016, p. 21). Even when it comes to different types, the heading is called "Chart Types" (2016, p. 157). However, the single types that he describes contain different terms, as in "bar chart" (p. 161), "chord diagram" (p. 189) "flow map" (p. 206), and so on.

As this, I assume, could cause confusion, I decided to use only the term *data visualization* as the general term, and other terms like *graphs*, *charts*, *maps* and *diagrams* only when I refer to specific types (subgroups) of DVs (see Section 7.3.6). As a basis for this, I use the specifications of Kosslyn (1989), who explained the four different "types of visual displays [emphasis deleted]" (p. 185–186) as follows:

Graphs are "the most constrained form, with at least two scales" (p. 186). The values presented along these scales have a symmetrical paired-with relation[3]. Within *charts*, not only (symmetrical) paired-with relations are possible but so are other relations[4], and they show "discrete relations among discrete entities" (p. 186). The entities shown need to be visually linked to other entities, for example, by lines. *Maps*, as Kosslyn (1989) goes on to explain, correspond "non-arbitrarily to a part of the pictured territory" (p. 186). *Diagrams*, as the fourth type of visual display, are "schematic pictures of objects or entities", which means that "unlike charts and graphs, the parts of a diagram correspond to parts of some actual object or entity; and unlike maps, parts of diagrams do not represent locations of a territory" (p. 186).

2.2 Connection

As mentioned in the introduction, one function of lines is to represent connections (Poulin, 2018, p. 32); thus, the line itself has the function of a *connector* (Engelhardt, 2002, p. 40). The phenomenon of connections occurs in various fields and disciplines, and the terminology differs since many synonyms exist for the term *connection*. To mention a few examples, the words *relation*, *relationship*, *link*, and *tie* can be applied with similar meanings (Fergusson, 1992, p. 88).

3 Like, for example, *Five rolls cost 3 euros*, which could also be formulated as *3 euros are worth five rolls*.
4 To name a simple example: *At 7 and at 10 am the temperature is 7° Celsius*. The inversed association is ambiguous, since it is possible that the temperature can be 7° Celsius at other times.

Nevertheless, it is necessary to describe the fine differences between the terms (if they exist) and point out in which fields, touched on in this dissertation, the different terms are more likely to be used.

Brinton uses the terms *relation* and *relationship* in his 1939 published book *Graphic Presentation* and even names a category of charts "relationship charts [emphasis deleted]" (1939, p. 68). This category contains visualization examples not only using lines as the relation-forming elements, but also other possibilities, like physical proximity or grouping within the same framing, to show relations. In addition, other categories he describes represent different forms of relations too, like classification charts which, according to his definition, differ only in that they show the represented relations with classifications (p. 43–72). Bertin (1967/2011) explicitly mentions the possibility that a line can represent a relationship between single elements (p. 271). Having a closer look at the terms used in the original French version instead of the English word *relationship*, the French terms *correspondance* and *liaison* are used (Bertin, 1967, p. 271). As an English translation of the former, *correspondence* or *connection* could have been used, as the translation of the latter (*liaison*), *liaison*, *association*, *link*, or *relation* could have been applied (Chalmers & Pierquin, 2003). The translator has decided on the term *relationship*, and thus follows the approach of Brinton.

In his dissertation aiming to providing a theoretical framework for studying diagrams, Richards (1984) also uses the terms *relationship* and *relation* in his proposed method of analysing diagrams (p. 3/21). However, in the verbal translation of a figure showing two letters with a line between them he states, "A is connected to B", and further on describes the line as being a "connection" (p. 3/21). Thus, it appears that Richards uses *relation* and *connection* interchangeably in certain contexts. Engelhardt (2002), who was strongly influenced by the dissertation of Richards, also had a varied use of terms. In his chapter about object-to-object relations Engelhardt states:

> **Linking** is a basic type of object-to-object relation that involves graphic objects with two *syntactic roles*: *nodes* and *connectors*. A **connector** is a graphic object in the shape of an arrow, band or line that is anchored to two other graphic objects (*nodes*), connecting them. (p. 40)

More recently, Ware (2013) speaks of *links* that represent *relationships* in his explanation of the visual grammar of node-link diagrams. His educational background in the field of psychology of perception and computer science (2013, p. xxi) may have been an influencing factor. That becomes obvious in the examples he mentions (like software structure diagrams, data-flow diagrams and organization charts) (pp. 221–226). The fact that links are such a central unit within the field of computer science (e.g. in hypertexts (Nielsen, 1995)), could be one reason for him to use the term *link* rather than others.

Looking back on the terms used by these authors, all with a strong position in the field of information visualization, I conclude that the terms *relationship* and *relation* are widely used. Nevertheless, other terms that also describe the same matter appear in their publications. Engelhardt uses the word *connector* for the object that performs the action of linking. In terms of what phenomenon is represented by these links or connectors Brinton, Richards, Ware and Engelhardt mention either *relationships* or *relations*, which shows at least one similarity.

Viewed from an etymological perspective, *relate* was formed on the stem of the Latin *relātus* (“Relate”, 1996/2003). It meant to “give an account of” and to “refer back” in the 16th century, and to “bring into connection or comparison” (“Relate”, 1996/2003) in the 17th century. In comparison, the term *connect* derives from the Latin *connectere*, which simply means to bind or fasten together (“Connect”, 1996/2003), whereas *link* derives from old Norse, standing for a “loop of chain” (“Link”, 1996/2003). To close the circle, the term *line* has two senses, either “flax”(“Line”, 1996/2003a), going back to the Latin *līnum* or “cord, string” (“Line”, 1996/2003b), deriving from the Latin *līnea*. Looking at these etymological origins, it becomes apparent that *link*, *line* and *connect* have a background as physical entities or actions that can be performed with them, all of which having certain common qualities. Because of their equal level of physical solidness, the word groups of the etymological roots of *line* (flax, cord, string) and *connect* (to bind or fasten together) fit together much better than the one of *line* does with those of *link* or *relate*.

To come back to authors that draw parallels from visual material to linguistic structures, the terminology used by Kress and Van Leeuwen will be investigated next. In *Reading Images. The Grammar of Visual Design* (1996/2006) they investigate patterns of contemporary “Western visual design” (p. 4) within various examples like advertisements, art pieces, photographs or diagrams. They use the terms *connecting lines* and *vectors* to explain the relations between objects and processes represented in visual material (1996/2006, pp. 42, 97). Without going into detailed explanation of the term *vector* (which is extensively done in Section 4.4.3 and Article III), Kress and Van Leeuwen use that term for a subcategory of connecting lines indicating directionality. In one of their more detailed descriptions of network diagrams, they describe the participants in a network with the term *node* and the vectors or lines between the participants with the term *link*, which together "show the multiple interconnections between participants" (Kress & Van Leeuwen, 1996/2006, p. 84). This again shows how great a variety of terms can be used even within one publication.

To recapitulate, the terminology used to conceptualize connections varies a lot both between and within the fields of research, and sometimes even within individual published works. Nevertheless, it can be assumed that the term *link* has a special association with the field of computer science, and the term *vector* has been strongly characterized and defined by scholars working in the field of social semiotics. For this dissertation, I have chosen the term *connection* as the central one (rather than *link*, *relationship*, *relation*, *tie* or others) because it has been used in several central fields and also explicitly in combination with the line (as *connecting line* by Kress and Van Leeuwen (1996/2006)). Moreover, it fits well with the etymological roots of the term *line*. Nevertheless, I also use the other synonyms mentioned earlier, for example, when referring directly to an author's work. I pay special attention to the fact that Kress and Van Leeuwen used the term *connecting line* as the umbrella term for both vectors and connecting lines without an indicator of directionality. Therefore, every use of the term *vector* refers to this strong differentiation.

2.3 Line

Until now, only the etymological background of the term *line* and its synonyms have been discussed. This section presents the variations of understandings of the term *line* and looks into the question of what can form a line. Starting with the question what a line is, an often-quoted statement by the Russian painter Kandinsky in *Point and line to plain* (1926/1947) shall first be at the centre of attention:

> The geometric line is an invisible thing. It is the track made by the moving point; that is, its product. It is created by movement—specifically through the destruction of the intense self-contained repose of the

> point. Here, the leap out of the static into the dynamic occurs.[5] (p. 57)

Within this quote, he emphasizes the invisibility of the line itself, being the product of the moving point. The Greek mathematician Euclid (Euclid & Heath, n.d./1956) has even defined the line as a "breadthless length" (p. 153).

Wong (1993) doesn't take such an extreme position, but the basic idea is similar, as can be observed in his description of the two reasons why a line is recognized as such:

> A form is recognized as a line because of two reasons: (a) its breadth is extremely narrow, and (b) its length is quite prominent. A line generally conveys the feeling of thinness. Thinness, like smallness, is relative. The extreme ratio between length and breadth of a shape makes it a line, but there is no absolute criterion for this. (p. 45)

Looking at the graphical examples given by Kandinsky (1926/1947), they all show the line as a continuous form. According to Kandinsky's explanation of the development of the line, this means that the moving point itself constantly touches the surface during its movement. But as the gestalt laws of proximity[6] and good continuation[7] state, single elements grouped in a certain way or incomplete lines, can also be perceived as lines (Lauesen, 2005, pp. 68–69). A sequence of single elements that can be perceived as a linear entity can therefore also be considered as a line, as for example dashed lines or a sequence of written characters or words (Hammer, 2008, p. 74).

Moreover, imaginary reference lines may occur between single form elements, which is especially relevant for composition (Hammer, 2008, p. 74). Kress and Van Leeuwen (1996/2006) mention literally and figuratively drawn lines that are formed by elements in the image (p. 175–176), but also vectors,

5 Or in the German original: "Die geometrische Linie ist ein unsichtbares Wesen. Sie ist die Spur des sich bewegenden Punktes, also sein Erzeugnis. Sie ist aus der Bewegung entstanden – und zwar durch Vernichtung der höchsten in sich geschlossenen Ruhe des Punktes. Hier wird der Sprung aus dem Statischen in das Dynamische gemacht." (Kandinsky, 1926/1928, p. 51)

6 In German "Faktor der Nähe [emphasis deleted]" (Wertheimer, 1923, p. 308).

7 In German "'gute' Fortsetzung" (Wertheimer, 1923, p. 324).

formed, for example, by people's gestures that are directed towards a particular place (p. 63–64). These are all examples of such imaginary reference lines.

Within this dissertation, I include continuous lines and incomplete lines as well as sequences of visual objects that form a line as potential connecting lines. By contrast, I exclude imaginary lines, because this dissertation investigates the *graphical* forms of the lines.

2.4 The Connecting Line and Which Lines Are Included

To be included in the research material of this thesis, the lines need to have a connecting function. Looking back on what has been discussed in Section 2.2, it should, therefore, be possible to detect the two parts of a connection: the single components and the relationship as Bertin (1967/2011) describes it (p. 271), or the nodes and the connector, to mention the terms used by Engelhardt (2002, p. 40). When reflecting on these remarks about the two parts of a connection and the necessity of detecting both parts so that the line has a connecting function, one may ask whether it is not the case that all lines automatically connect a starting point and an endpoint.

Krämer (2016) presents two production methods of lines, which lead to the distinction between what she calls a connecting and a motion line: Either two points on a plane are connected with a line or, starting from a point, a line is produced through a fluid movement with a writing or drawing instrument (p. 102)[8]. In my interpretation, this says something about what was first – both endpoints, or only a starting point – where the line explores the way and then finds an endpoint. The former type is what Krämer would call a *connecting line*[9], whereas the second type she would call a *motion line*[10] (p. 102). At the same time, for Krämer, a line can embody and realize both aspects within one inscription and oscillate between them (p. 109)[11]. The Swiss painter Paul Klee (1921–1922/1991) and the British anthropologist Tim Ingold (2007) follow a similar approach to Krämer to describe different types of lines. For Klee, the *free line*[12] evolves when a point is put in motion, which means that it is completely

---

8 My translation. Original quote: "Entweder werden zwei Punkte auf einer Fläche durch einen Strich verbunden; oder vom Punkt ausgehend wird mit dem Schreib-/Zeicheninstrument in einer flüssigen Bewegung ein Strich produziert." (Krämer, 2016, p. 102)

9 My translation. Original term: "Verbindungslinie" (Krämer, 2016, p. 102)

10 My translation. Original term: "Bewegungslinie" (Krämer, 2016, p. 102)

11 My translation. Original quote: ". . .verhält es sich bei der Bewegungs- und Verbindunglinie so, dass eine Linie innerhalb einer Inskription den Wechsel beider Aspekte verkörpern und realisieren kann und zwischen diesen oszilliert." (Krämer, 2016, p. 109)

12 My translation. Original term: "freie Linie" (1921–1922/1991, p. 99)

unbound in terms of its movements, comparable to a walk with no aim other than the walk itself[13] (1921–1922/1991, pp. 97–101). In contrast to the free line, the *restricted line*[14] more closely resembles a course of business, because the target (the endpoint) is to be reached as fast as possible (Klee, 1921–1922/1991, p. 101). Similarly Ingold (2007) distinguishes between lines that are "the trace of a gesture [emphasis deleted]" (p. 74) comparable to "wayfaring" (p. 75) and lines that are "an assembly of point-to-point connectors [emphasis deleted]" (p. 74–75) called "threads" (p. 41), rather comparable to a "transport" (p. 75).

As stated earlier, Kandinsky (1926/1947), by contrast, describes only one way of producing a line, namely through taking the static presence away from the point (p. 57). The time aspect, which leads to the line type distinctions of Krämer, Klee and Ingold, is brought in by Kandinsky by introducing the term *tension*[15] as "the force living within the element" (p. 57), which, together with the *direction*, forms the *movement*. In other words, if the line has a high tension, it is comparable to Krämer's *connecting line*, Klee's *restricted line* and Ingold's *connector*.

Looking at the research material of this dissertation – DVs – it becomes clear that the message the designer wants to convey, often motivating the type of DV used, is possible to classify with the help of all four of the former concepts (Krämer, Klee, Ingold, Kandinsky). Having Krämer's classification in mind, it becomes clear that lines in some DV types are easy to classify as connection lines and motion lines. For instance, lines in network visualizations, slope charts, dumbbell charts and parallel coordinates can clearly be classified as connection lines, while the lines in route maps or flow maps, for example, are motion lines. By contrast, lines in some other DV types, like line graphs, timelines or connection maps, are not so clearly assignable to one or the other line type suggested by Krämer. For Krämer, this could be explained by the fact that lines can also embody both aspects, as explained earlier. For this reason, but also because of the continuous development of hybrids and new DV types, it seems questionable whether this distinction is helpful for the investigations of this dissertation. Therefore, I adopt the approach of Kandinsky, seeing all lines as an outcome of a movement. For me, the term *connecting line* can include lines of both of Krämer's line types: motion and connecting lines.

---

13 My translation. Original quote: "Es ist sozusagen ein Spaziergang um seiner selbst willen." (1921–1922/1991, pp. 100–101)

14 My translation. Original term: "befristete Linie" (1921–1922/1991, p. 101)

15 By tension is meant what could also be called power, not the friction towards the surface on which the line is to be produced.

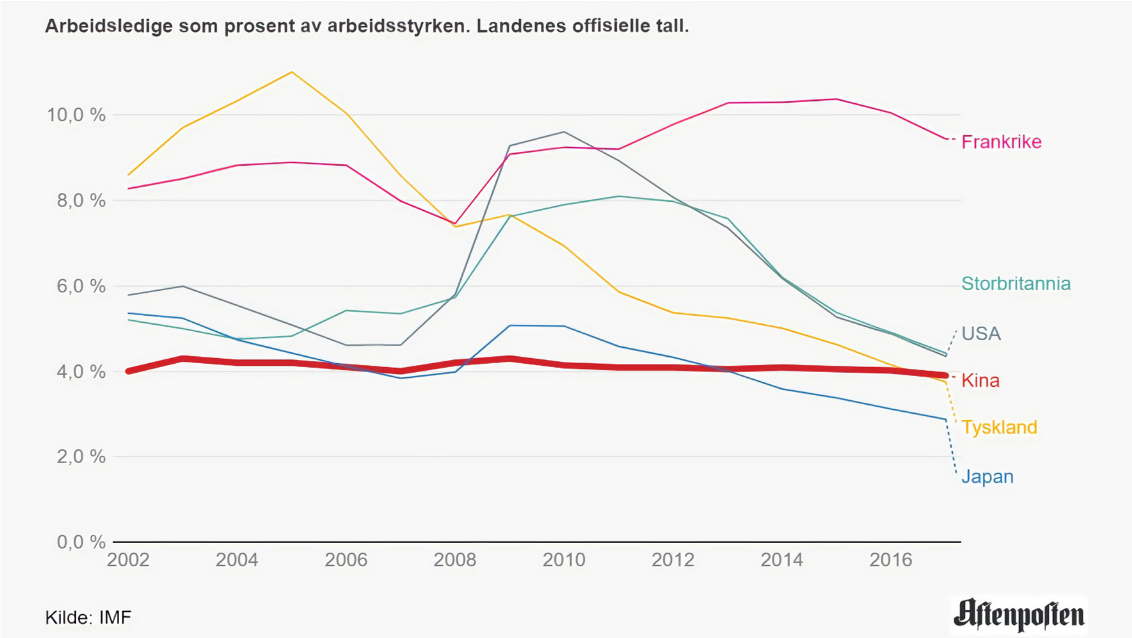
For the current study, the classification into different functions of lines is much more important than distinguishing between different production processes of lines. To make such a classification, Krämer (2016) names the three prominent types *object lines*, *auxiliary lines* and *border lines*[16] (p. 100–102). She describes the difference between the first two types with the example of a coordinate cross. The lines building up the coordinate cross itself, labelled with values, she calls auxiliary lines, while the lines shaping the graphs that are drawn into the evolved space of the coordinate system, she calls object lines (p. 102). Border lines, which she compares to auxiliary lines, are lines that only define two different territories, the identity of these territories only evolves through the border line. Krämer's mention of border lines also points to the fact that, within the group of object lines, the lines have various purposes. Additional to the function of separating elements (in contrast to what is done with connecting lines), lines can, for example, also define contours or be the base element for textures and patterns (Poulin, 2018, pp. 32, 37).

For this dissertation, this functional difference is essential because it makes the difference between (1) DVs in which one or more connecting lines communicate the main message, and (2) DVs where other graphical elements, like dots or shapes, are used to imply and communicate the main message, while lines are only used for compositional purposes, to help the reader to make meaning of the other visual material. This is the case in bar charts or scatter plots, for example. In this dissertation, I consider only lines that communicate the main message of the DV (i.e., object lines) as potential objects of study. Figure 3 shows an example of a DV published by the Norwegian newspaper *Aftenposten* (Kløvstad Langberg, 2019), where connecting lines are used to signal the main message (and therefore constitute object lines), whereas auxiliary lines and border lines have a more subordinate function.

16 My translation. Original terms: "Objektlinien", "Hilfslinien" and "Grenzlinien" (2016, pp. 100–102)

Figure 3

Line Graph Showing Unemployment as Percentage of the Labour Force



Note. Using Krämer's (2016) terms, the coloured continuous lines showing the development of the unemployment in each country are object lines. All other lines are auxiliary lines (the grey thick line that constitutes the time axis and the coloured, dashed lines that connect the object lines with the country names) or border lines (the grey thin lines in the background).
From "Kinas ledighet har knapt rikket seg på 15 år. At noe ikke stemmer, er åpenbart." by Ø. Kløvstad Langberg, 2019 (https://www.aftenposten.no/meninger/kommentar/i/gPqk6a/Kinas-ledighet-har-knapt-rikket-seg-pa-15-ar-At-noe-ik-ke-stemmer_-er-apenbart). Retrieved March 5, 2019. Copyright 2019 by Aftenposten. Reprinted with permission.

To recapitulate, in terms of the visual appearance of lines, I include as objects of study continuous and incomplete lines as well as sequences of visual objects that are meant to form a line. In terms of relevance for the message, they must be object lines; I exclude auxiliary lines. In addition, these object lines need to have the function of connecting two elements at their ends, which can be either two different elements or two different locations of the same element, which means that location A is connected to location B by the path of the line. How explicitly visible these two elements are made, plays a subordinate role. Outside the focus of this dissertation are border lines, contour lines, lines that are used as the base element for textures and patterns as well as lines with any other function that has not been described in this chapter. Table 1 presents an overview of all included and excluded line types.

Table 1

Overview of Included and Excluded Line Types

Included line types	**Excluded line types**
Overall function:	
connecting lines • in the words of Krämer: connecting and motion lines (connecting either two different elements or two different locations of the same element) • in the words of Ingold: traces and threads • in the words of Klee: free and restricted lines	border lines contour lines lines that are used as the base element for textures and patterns
Visual appearance:	
continuous lines incomplete lines sequences of visual objects that are perceived as lines	imaginary lines
Relevance for the message:	
object lines	auxiliary lines

3 Positioning the Work in the Traditions of Science

As the central element of this dissertation is a sign (the line) which is investigated concerning the relations between its physical forms and its potential meanings, this project belongs to the research tradition of semiotics. Anchored in the tradition of hermeneutics, semiotics offers an approach to studying the meaning of visual text (Stokes, 2003, p. 71), among which DVs are a particular kind. DVs in this dissertation are seen as visual text, and with semiotics it is possible to unpack the meaning potentials inherent in these textual forms.

Designers decide on the specific visual appearance of DVs. They have specialized knowledge about how visual text needs to be designed so that it best carries the intended meaning potential. This is why not only theoretical concepts of the field of linguistics and semiotics are used in this thesis, but also knowledge from the field of graphic design (see Chapter 7). This shows the interdisciplinary orientation of this research.

Viewed from a wider perspective, I see DVs from a social constructivist position. DVs are viewed as a part of the collective human discourse and thought, which generate the social and human world (Collin, 2013). Thus, when people interact with others through DVs, these DVs are part of the constructing process of the "fluid and fragile" (Neuman, 2006, p. 89) social life. The constructivist view on the relation between language and reality is that language can never objectively reflect the reality (Neuman, 2006, p. 89). Instead, as Neuman continues, people in a society need to learn language, because it contains "built-in social constructions" (p. 89). If they have learned these social constructions, they see the world in a certain way. Therefore, the way of seeing is coloured by the social constructions. Consequently, my choice of theory affects my social constructivist perspective on the research material. Putting social semiotics at the centre of the theoretical framework, I follow M.A.K. Halliday (the founder of that discipline), who stated: "Reality is a social construct, it can be constructed only through an exchange of meanings. Hence meanings are seen as constitutive of reality" (Halliday, 1978, p. 191).

By placing the visual text or, more specifically, the connecting line within DVs, at the centre of the investigation, I consider the communication process as a whole, the relations between the participants (the DV designer, the viewer etc.) and the immediate production and reception process only indirectly. This means, for example, that I do not directly investigate the designer's intentions or thoughts (which could, e.g., be done by interviewing them or using thinking aloud protocols) or the viewer's reactions or reading behaviour (for which one could, e.g., perform user tests using eye-tracking). Rather, I see these actors as members of a communication culture that together build up conventions that are reflected in the visual text.

4 Theoretical Framework

This chapter is devoted to the theoretical basis for the research presented in Articles I, II and III. Each article already contains a theoretical section, the length of which was dependent on the overall article length restrictions. Some parts of the theoretical sections of the articles turned out to be shorter than ideal. Therefore, this chapter outlines the core concepts used in this dissertation in more detail. It first provides a historical context for the study of signs in general and more specifically for the field of social semiotics and its application to multimodal text. Further, I present the social semiotic concepts most central to this dissertation. These are the (1) metafunctions, (2) modality, (3) representational structures, (4) cohesion and (5) viewer position. Finally, I discuss the aspects of social semiotics that are criticized by the research community and related to the investigations presented in this dissertation.

4.1 The Study of Signs

The study of signs is often associated with two scholars, who are considered the founders of two different branches. The Swiss linguist Ferdinand de Saussure (1857-1913) saw the sign as a two-part construction, consisting of a signifier and a signified (Skaggs, 2017, p. 39). This “science which studies the role of signs as part of social life [emphasis deleted]” (Saussure, 1916/1983, p. 15) he called “semiology [emphasis deleted]” (Saussure, 1916/1983, p. 15). In contrast to Saussure, the American philosopher Charles Sanders Peirce (1839–1914) has seen the sign as one part of a three-part relation and named this theory “semiotics” (Skaggs, 2017, p. 39). The “sign, its object and its interpretant” (Peirce, 1935, p. 332 (CP 5.484)) are the three parts defined by Peirce, which are all included in the sign process. The interpretant is the effect which is produced by the sign in the mind of the receiving person, whereas the object is what the sign stands for (Peirce, 1960, p. 135 (CP 2.228)).

Semiology, which was developed by Saussure, has a linguistic structuralist background, while semiotics, developed by Peirce comes from the pragmatic tradition in philosophy (Schrøder, 2009, p. 481). Saussure focuses only on verbal language (in opposition to Peirce), and for him, the meaning of signs can only be explained based on their position in the language system, through its contrast to the other signs in the system (Schrøder, 2009, pp. 481–482). While focusing on the language system (*langue*), he discarded language use (*parole*) from his research field (Hodge & Kress, 1988, p. 16). Furthermore, he categorized the study of *langue* into the *synchronic* and the *diachronic*, whereby he focused only on the former, not studying the changes in the system over time, but only its stages at a certain point in time (Hodge & Kress, 1988, p. 16). This approach is considered by some to be a weakness of the theory. Hodge and Kress (1988), for example, expressed their concerns as follows: "The repressed is the dynamic principle of his system. That repressed is invariably energy, movement, process: whatever changes, or causes or describes change." (p. 17)

In contrast, in Peirce's sign theory (including the interpretant) the sign does not exist only in a closed system, but he stresses the "process in the study of signs" (Hodge & Kress, 1988, p. 20), because what characterizes pragmatism is "its emphasis on human beings as agents and their practical relations to the world" (Delanty & Strydom, 2003, pp. 277–278). This enables Peirce to include the cultural context around the use of signs in his theory, so the sign gets its meaning through its use in a social context (Schrøder, 2009, pp. 481–482).

Peirce (1960) introduced the categorization of signs into "icon", "index" and "symbol" (p. 143 (CP 2.247)). This categorization shows the different kinds of reference a sign can have to the object. Saussure (1916/1983) didn't use such a subdivision of the sign, but instead said that "the linguistic sign is arbitrary [emphasis deleted]" (p. 67). In Peirce's model, only the symbol is arbitrary. For this dissertation, Peirce's theory, including the different references of signs to their objects, is valuable because the line, when serving as a connecting element in DVs, can have several forms of references to the represented object.

The visual appearance of a line can point to objects of the physical world with a similar function of connecting one element to another, like ropes, chains, or paths. In such cases, the sign is an icon, because "it is like that thing and used as a sign of it" (Peirce, 1960, p. 143 (CP 2.247)). If the line is referring to a connection by indicating a particular aspect or quality of that connection, in the same way that a weak relationship between two persons could be shown by a very thin line, it is seen as an index. If a line represents a relation between two persons as an abstract connection in a way regulated by a cultural convention, the sign is a symbol, because "it denotes by virtue of a law" (Peirce, 1960, p. 143 (CP 2.249)), which the recipient has learned to read. But a line does not only have an iconic, indexic or symbolic relation to the object it represents. Most often, the meaning of a line is constituted through a combination of two or three relations working simultaneously (Skaggs, 2017, p. 60).

Summing up the difference between Saussure's semiology and Peirce's semiotics most relevant to this dissertation, Peirce considers the cultural context around the use of signs. In contrast to (Peircean) semiotics, (Saussurean) semiology implies "a tendency to give much attention to signs as such, less to society, and hardly any to the 'life' of signs in social practices" (Jensen, 1995, p. 3). This, however, is exactly what is central in social semiotics, which I have chosen as the main theoretical framework of this dissertation.

4.2 Text as a Sociosemiotic Process

The theoretical basis for investigating the "relationship between language and social structure" (Threadgold, 1986, p. 108) was laid by the linguist M.A.K. Halliday, who is regarded the founder of the discipline called *social semiotics* (see Halliday, 1978; Hodge & Kress, 1988). He had a background in European structural linguistics, having been a student of Raymond Firth and relying substantially on Louis Hjemslev, who relied on Saussure (Andersen et al., 2015, p. 147; Friman, 2009; Threadgold, 1986, p. 108). However, he (just as V. N. Vološinov and Umberto Eco) was "critical of the basic Saussurean dichotomies of *langue/parole*, *synchronic/diachronic* and *arbitrary/motivated* [emphasis added]" (Threadgold, 1986, p. 108). The reason is, as Threadgold continues, that these dichotomies do not fit together with their view of language being a social phenomenon and constantly in change. Vološinov, Eco and Halliday all had in common that they focused "on the text as sociosemiotic process" (Threadgold, 1986, p. 109) – they saw language as "interacting with the other semiotic systems within the culture" (Threadgold, 1986, p. 109).

Halliday developed a systemic functional theory of language (Systemic Functional Linguistics), which "looks at how language functions to make meaning in [the] context of situation" (Webster, 2019, p. 35). Within Systemic Functional Linguistics, language is not described as a rigid system, but as a flexible one, where its social use is considered as a key factor (Ledin & Machin, 2018, p. 16). In other words, meaning has, from a social semiotic perspective, a dynamic nature. To highlight this dynamic nature, Halliday (1978) uses the term *meaning potential* rather than *meaning* (pp. 21, 28, 139). Consequently, if meaning is unstable, the rules for making meaning change as well: “The ongoing text-creating process continually modifies the system that engenders it” (Halliday, 1978, p. 139). This means that the understanding of how the rules for interpreting texts evolve in a social semiotic perspective differs fundamentally from the Saussurean view, for instance. In the latter case, rules are seen as predefined and more or less consistent (Hodge & Kress, 1988, p. 12). In contrast, in social semiotics, people are seen as the ones who constantly generate and modify these rules through their involvement in social interaction (Van Leeuwen, 2005, pp. 47–48).

For this dissertation, Halliday’s general social semiotic view on language is fundamental as, in my understanding, research on conventions needs to consider the social and its relation to signs. It also motivates the dissertation’s focus on the development of conventions in the socially situated design and use of DVs. Several concepts described in Halliday’s *An Introduction to Functional Grammar* (1985/2004) and *Language as social semiotic* (1978) are used for this dissertation. The most central ones are presented in Section 4.4.

4.3 A Social Semiotic View on Multimodal Texts

Several scholars have developed Halliday’s legacy further, resulting in three main directions, as Andersen et al. (2015) describe: (1) Christian Matthiessen and Jim Martin are two scholars that worked further on the “*systems* for describing language and meaning making” (p. 142). (2) Gunther Kress and Theo van Leeuwen – and later Jay Lemke – are key figures for the development of “*multimodality research*, i.e. taking semiotics into modes other than the verbal” (p. 142). (3) Theo van Leeuwen and Jim Martin, for example, are concerned with “*discourse as social practice*” (p. 142). The second direction (multimodal research) is the most relevant one for this dissertation, as DVs most often combine verbal language with visual elements and numbers.

Photos, advertisements, artwork and cartoons, for instance, are included as text examples in Robert Hodge’s and Gunther Kress’s seminal work *Social Semiotics* (1988). Together with Van Leeuwen, Kress later developed a more elaborate grammar of visual design, published in *Reading Images: The Grammar of Visual Design* (1996/2006) (further referred to as *Reading Images*).

This seminal work is one of the main sources used for developing the concrete analysis methods for Articles II and III. The aim of *Reading Images* was, as they explain, to present a descriptive framework as a useful tool for visual analysis (p. 14). The framework they present provides insights into "contemporary visual design in 'Western' cultures" (1996/2006, p. 3). They do not define what exactly they mean by Western cultures. But, with this statement, they make clear that visual design is culturally specific, and that their framework may not be useful for investigations of visual design of all cultures. The boundaries of the language of visual design are, according to them, "not those of nation-states, … . Rather, this visual resource has spread, always interacting with the specificities of locality, wherever global Western culture is the dominant culture" (1996/2006, p. 4).

Before explaining the core concepts applied to my analyses, I will briefly comment on Kress and Van Leeuwen's understanding of the sign in relation to those of Saussure and Peirce. Kress and Van Leeuwen are both dismissive of Saussure's thinking that the sign is arbitrary (Kress and Van Leeuwen, interviewed in Andersen et al., 2015, pp. 74, 100). Instead, Kress (interviewed in Andersen et al., 2015, p. 73) was inspired by Peirce's focus on the ongoing process of semiosis and the introduction of the interpretant. However, he does acknowledge elements from Saussure's thinking as well, especially "the notion of convention, and the notion of reference" (interviewed in Andersen et al., 2015, p. 73).

In *Reading Images* (1996/2006), Kress and Van Leeuwen point out that visual structures, just like linguistic structures, indicate a particular interpretation that is formed by experience and social interaction (p. 2)[17]. This assumption is adopted by this dissertation, where I see the visual structure of DVs as leading the viewers to a certain interpretation, based on certain social circumstances as well as conventions. A sign-maker (such as a DV designer), "'chooses' a semiotic resource from an available system of resources" (Jewitt, 2009, p. 23) in order to signal a certain intended meaning. Such semiotic resources are, according to Van Leeuwen (2005) "the actions, materials and artefacts we use for communicative purposes, whether produced physiologically – for example, with our vocal apparatus …, or technologically – for example with pen and ink, or computer hardware and software" (p. 285). As he continues, the semiotic resources have a meaning potential (see Section 4.2) "based on their past uses, and a set of

17 They also critically reflected on this comparison between visual and verbal structures and pointed out that this similarity can easily be overemphasized (see also Kress & Van Leeuwen, 1996/2006, p. 76)

affordances, based on their possible uses, and these will be actualized in concrete social contexts where their use is subject to some form of semiotic regime" (p. 285). Applied to this dissertation's research focus, all graphic elements that can be used in a DV, as well as all visual characteristics that can contribute to shaping the graphic element, are semiotic resources, from which the DV designer can choose. What Van Leeuwen (2005) calls a *semiotic regime* points to the existence of some (modifiable and instable) rules, that have been already touched on in Section 4.2. These rules are informal and function in the form of social conventions. Conventions are defined by Kennedy et al. (2016) as "symbolic or social practice[s]" that are "shared, readily understood and widely accepted by members of a cultural group" (p. 717).

4.4 Social Semiotic Concepts Central to This Dissertation

Five social semiotic concepts central to this dissertation are presented in the following: (1) metafunctions, (2) modality, (3) representational structures, (4) cohesion and (5) viewer position. After the presentation of each of the concepts, I point out what I used them for in this dissertation.

4.4.1 Metafunctions

The three "functional components" (Halliday, 1978, p. 130), later called "metafunction[s]" (Halliday, 1985/2004, p. 29) describe three different aspects of meaning potentials of any clause. When talking about the "ideational metafunction" (Halliday, 1985/2004, p. 29) of a clause, one talks about the clause as a representation of the world or, more specifically, of a process going on in the world. However, a clause does not say only something about a process, but also about interpersonal relationships. Therefore, the "interpersonal metafunction" (Halliday, 1985/2004, p. 30) indicates that the clause can also be seen as an "exchange" (Halliday, 1985/2004, p. 59) within a social space involving a speaker or writer and a listener or reader. Third, the clause also represents a message with a certain internal structure, thus it also has a "textual metafunction" (Halliday, 1985/2004, p. 30). Any clause, any verbal utterance, carries all three metafunctions at the same time. In this dissertation, I use the concept of metafunctions in the first article, to contextualize the identified semiotic functions of the connecting lines.

4.4.2 Modality

Modality (as an aspect connected to the interpersonal metafunction) is another concept described by Halliday (1985/2004), that is central to this dissertation. However, he was not the first one to write about modality. Scholars since Greek antiquity have been concerned with the concepts of necessity and possibility, although without using the term *modality* (Van der Auwera & Aguilar, 2015). In the last decades, the phenomenon of modality has also been dealt with through approaches other than the functional linguistic one. Scholars within the fields of cognitive linguistics, formal syntactics and formal semantics, for example, have worked on the topic (Axel-Tober & Gergel, 2015; Boogaart & Fortuin, 2015; Kaufmann & Kaufmann, 2015). I will, however, not go any deeper into these other approaches, as the only relevant perspective for this dissertation is the one of Halliday and the scholars who later used and adapted his thoughts for the analysis of multimodal texts. The next paragraphs provide an overview of Halliday's view on modality, supplementing what in short is presented in Article II.

For Halliday (1985/2004), modality is "the speaker's judgement, or request of the judgement of the listener, on the status of what is being said" (p. 143); it "construe[s] the region of uncertainty that lies between 'yes' and 'no'" (p. 147) and is, therefore, an "expression of indeterminacy" (p. 148). His system of modality – on a clause level – includes four variables: the "modality type", "value", "polarity" and "orientation" (1985/2004, p. 150).

Regarding the first variable (modality type), Halliday (1985/2004) identifies two major types: "modalization" and "modulation" (p. 618). The first type (modalization) counts clauses that indicate some degree of either probability or usuality, related to a proposition; the second type (modulation) counts clauses that indicate some degree of obligation or inclination (Halliday, 1985/2004, p. 618). To give some examples of how a degree of probability can be expressed in verbal language, the use of adverbs (modal disjuncts) like *certainly*, *probably*, or *possibly* can be mentioned.

These examples of modal adverbs bring us to the next variable of modality, which is value. They express high, median and low modality value, respectively, which are the three modality value options suggested by Halliday (1985/2004, p. 620). The modality value options lie in different positions between the positive and the negative pole (e.g. *It is* or *It isn't* (Halliday, 1985/2004, p. 143)).

Expressions with a high, median or low modality value can all be realized both with a negative or a positive polarity, which is the third variable in the modality system (Halliday, 1985/2004, pp. 148, 619–622). For example, one may say *It possibly is* or *It possibly isn't* so as to express a low modality value with a positive respectively negative polarity (Halliday, 1985/2004, p. 623).

Finally, each type of modality (e.g., modalization) can be realized with two different orientations: objective or subjective orientation. Following Halliday (1985/2004), orientation is about whether the speaker states the source of conviction or not, and about how explicitly this is done (p. 149). He offers the following examples: "*It is certain (that) that is true* [objective orientation] and *I'm certain (that) that is true* [subjective orientation]" (p. 149). Within verbal language, there are many ways to formulate a proposition, which includes showing or not showing that the proposition is influenced by the subject expressing the proposition (Halliday, 1985/2004, p. 616). In the sentence *I'm certain that that is true* for instance, the source of conviction is stated very clearly (explicit subjective orientation), whereas in *That must be true* it is not made so clear (implicit subjective orientation) (Halliday, 1985/2004, p. 149).

Halliday's view on modality is used in this dissertation as a key position for clarifying the relation between modality and uncertainty in the field of DVs, and for developing the corpus analysis method applied in Article II. Besides Halliday's approach to modality, which was concerned with verbal language, in Article II I also include the perspective on modality of scholars concerned with semiotic systems other than verbal language. Hodge and Kress (1988) have noticed that modality is not indicated only in verbal language; "in non-verbal media the same kind of thing happens as in verbal language" (p. 121). Readers of visual text, for example, also assess its modality "in a reasonably predictable fashion" (Hodge & Kress, 1988, p. 128). Visual text, such as a cartoon or a DV, therefore also contains visual modality markers as the pendant to verbal modality markers such as *certainly*, *probably*, or *possibly*. For them (1988), using modality markers is about indicating reliability, truth and reality (p. 121). Kress and Van Leeuwen (1996/2006) also talk about "questions of truth and reality [that] remain insecure, subject to doubt and uncertainty" (p. 154) when introducing modality. We can thus observe that *modality* (concerned with *probability* – as in verbal language indicated by modal disjuncts, e.g., – and *reliability*) and *uncertainty* are intertwined. How this is the case is further problematized in Article II, where I use these terms as the basis for setting up the corpus analysis structure.

With the concept of modality in mind, one can investigate "whether a given 'proposition' (visual, verbal, or otherwise) is represented as true or not" (p. 154), as Kress and Van Leeuwen (1996/2006) explain. They borrow the basic concept of modality from Halliday, introducing several modality markers in visual text like colour saturation, contextualization (level of context shown), or representation (level of abstraction) (1996/2006, pp. 160–162). For each modality marker, they introduce a scale, running from the lowest to the highest modality value. Taking the example of representation as a modality marker, a photo has a very low modality value if no details (like creases in clothes, etc.) are shown – much lower than if it looks more like the situation in reality. Likewise, in other visual depictions, like the graphics used in a bar graph, showing pictorial detail is not so important to make the DV look reliable. This means that the modality value of different forms of visual depictions cannot all be assessed in the same manner.

To explain this difference, Kress and Van Leeuwen (1996/2006) introduce the concept of "coding orientation" (p. 163), a term originally deriving from Bernstein (1981). Coding orientation refers to what counts as real and reliable in a certain social group or institutional context. Kress and Van Leeuwen (1996/2006) distinguish between four coding orientations: the "technological", "sensory", "abstract" and "naturalistic coding orientation [emphasis deleted]" (p. 165). In each of these coding orientations a different goal is pursued and, thus, the visual style is adapted to reach this goal if the sign-maker wants to signal maximum reliability. As Kress and Van Leeuwen explain, in the technological coding orientation, effectiveness is key. Thus, the unnecessary application of colour, for instance, signals low modality. In the sensory coding orientation, one seeks to evocate pleasure, through using vibrant colours, for instance. In the naturalistic coding orientation, the more the depiction resembles what we would normally see with our own eyes in a real life situation, the more real and the more reliable it is assessed as. In the abstract coding orientation, semiotic reduction (showing only the essential qualities of what is depicted) is used to represent the "deeper 'essence' of what it depicts" or a "general pattern" (Van Leeuwen, 2005, p. 168). Accordingly, the reader (if visually literate enough) takes the coding orientation into account when assessing the modality of a visual text. Likewise, the producer of the visual text can consciously use certain visual characteristics to signal a certain modality value.

Since the time Kress and Van Leeuwen published *Reading Images* (1996/2006) and described the four coding orientations, digital technologies have developed further and thus, our practices have changed. This requires us to think about whether a new understanding and description of modality is necessary, as Ravelli and Van Leeuwen (2018) notice. The digital age has, as they continue, an impact on the "reality criteria offered by coding orientations" (2018, p. 15). Therefore, they write, there may be new norms for the coding orientations, and they do not even preclude the possibility that new coding orientations may exist. The "framework needs adaptation and expansion" (2018, p. 15), they claim, but which one, they don't describe in detail. Rather, they leave that open for further empirical research, pointing out that the original model of Kress and Van Leeuwen (1996/2006) "needs to be applied to the texts of our time" (Ravelli & Van Leeuwen, 2018, p. 18). In Article II, I use these four coding orientations for the assessment of modality in DVs. The possibility that new, so far unknown coding orientations may have evolved is reflected in the analysis procedure, offering an open coding orientation category that first needs to be observed.

4.4.3 Representational Structures

As pointed out in Section 4.4.1, one of the three metafunctions active in any clause is the ideational metafunction; a clause is always a representation of an aspect of a world (Halliday, 1985/2004, p. 29). In every semiotic mode, there is the possibility of choosing between different options on how to represent the "aspect of the experiential world" (Kress & Van Leeuwen, 1996/2006, p. 46). Kress and Van Leeuwen (1996/2006) developed a system of different kinds of visual representational structures, using art pieces, advertisements, abstract diagrams and illustrations, for example, as exemplary visual texts (p. 45–113).

The system has two main representational structure types and several subtypes. Drawing on Arnheim (1982), who distinguishes between "volumes and vectors, between being and acting" (p. 154), Kress and Van Leeuwen (1996/2006) distinguish between *narrative* and *conceptual* representational structures (p. 59). The difference between the two is, according to them, that in a visual text, the participants of a process can be either "represented as *doing* something to or for each other" (p. 59) (narrative structure) or rather "in terms of their more generalized and more or less stable and timeless essence" (p. 79) (conceptual structure). Exemplary for a process represented as having a narrative structure, they show a route map with an arrow pointing to a city. The arrow indicates direction and stands for a movement towards the city; it is what they call a "vector" (1996/2006, p. 59). Other examples of what could form vectors include triangles in artpieces, a hand pointing at something in illustrations and an eyeline (the direction of the glance) (Kress & Van Leeuwen, 1996/2006, pp. 56, 64, 67). If such a vector is present in a visual representation, they specify, then this process is represented as a narrative one. By contrast, if there is no vector used, it is automatically represented as a conceptual one. A family tree, showing the relations of the family members, is an example of a conceptual representation (Kress & Van Leeuwen, 1996/2006, p. 80).

Although they are aware that there is no one-to-one correspondence between visual structures and verbal structures, they draw several comparisons between them. Vectors, for example, are the visual equivalent of *action verbs* in verbal language (1996/2006, p. 46). For both narrative and conceptual structures, Kress and Van Leeuwen (1996/2006) list several subtypes. For this dissertation, only some of the narrative representational structural subtypes are relevant, and these are presented in detail in Article III. I will therefore not repeat what is written there, but will rather continue with describing two other concepts that were central in Article I.

4.4.4 Cohesion

Within any semiotic mode, complexes of signs can be formed into a text. In the words of Kress and Van Leeuwen (1996/2006), the signs are then made to "cohere both internally with each other and externally with the context in and for which they were produced" (p. 43). If a text has few coherent characteristics, it is difficult to understand as a text, because it doesn't have a common thread (Taboada, 2019, p. 311). To refer back to Section 4.4.1, this refers to the textual metafunction (Halliday, 1985/2004, p. 30).

Linguists such as Halliday and Hasan (1976/1994), have worked on characterizing the connectedness that makes a set of utterances a text. In their pioneering work *Cohesion in English* (1976/1994) they describe single text units as needing to have meaning relations, which they call "cohesive properties" (p. 4). There are several different ways of forming cohesion within verbal text. Halliday and Hasan (1976/1994, p. 4) identify five: (1) reference, (2) substitution, (3) ellipsis, (4) conjunction, (5) lexical cohesion. The following examples given by Sanders and Pander Maat (2006, p. 591) demonstrate how the cohesion is formed:

(1) reference: "*Jan* lives near *the park*. *He* often goes *there*."
(2) substitution: "Daan loves strawberry *ice-creams*. He has *one* every day."
(3) ellipsis: "All the children had an *ice-cream* today. Eva chose strawberry. Arthur had orange and Willem too."
(4) conjunction: "Eva walked into town, *because* she wanted an ice-cream."
(5) lexical cohesion: "It was *hot*. Daan was lining up for an *ice-cream*."

Without going deeper into cohesion in verbal language, I will now turn to how single multimodal text units (comparable to a clause in verbal language) can be tied together to cohesive entities. According to Van Leeuwen (2005, pp. 179–267), there are four ways to create cohesion in a multimodal text:

(1) In spatial arrangements (such as DVs), the "composition" (Van Leeuwen, 2005, p. 179) can help to form coherence. Here, composition refers to the way the elements of a multimodal text (like several connecting lines) are arranged in space (e.g., on a page or a screen). From Van Leeuwen's point of view, whether these elements are placed on the left or the right, on the top or the bottom, in the centre or at the margin, for instance, makes a difference to how we read them. Composition can thus be deliberately used to guide the reader and to give each element a certain role within a textual whole.
(2) In time-based text types, like music, the equivalent of spatial composition is "rhythm" (Van Leeuwen, 2005, p. 182). It is formed by alternation between two different stages. Examples are the alteration between night and day and the beat in the case of music, in opposition to the notes that are not stressed.
(3) "Information linking" (Van Leeuwen, 2005, p. 219) is the third way of forming coherence in a multimodal text. To explain this category, Van Leeuwen borrows the linguistic concept of cohesion, mainly by using conjunctions of verbal language as examples. If there are two separate items of information that have a relation, this relation can be either temporal, logical or additive (Van Leeuwen, 2005, pp. 222–221). In verbal language, a temporal link may be expressed by conjunctions like *then*, *next* or *finally*; a logical link by conjunctions as *because*, *if* or *likewise*; an additive link by such as *and*, *or*, or *but* (Van Leeuwen, 2005, pp. 222–225). Similar to using conjunctions in verbal language, it is possible to apply certain techniques for marking relations between images, and images and words, as Van Leeuwen further ex-

plains. In films, for example, a cut from one event to the next can indicate a temporal link, without having to show the whole sequence of events in detail.

(4) The fourth way to form cohesion in a multimodal text that Van Leeuwen (2005) mentions is through "dialogue" (p. 248). By *dialogue* he means not only the spoken dialogue between people such as where one person may ask something and the other one replies. A dialogic structure "exists whenever several 'voices' are seen or heard, whether sequentially or simultaneously – and this, given the multimodality of all communication, is perhaps always" (Van Leeuwen, 2005, p. 267).

In this dissertation I apply the cohesion forming techniques of composition and information linking to the analyses of DVs, as can be read in Article I. Rhythm is not considered any further because the spatial equivalent (composition) is more relevant for this dissertation. I also do not consider dialogue any further, as this dissertation's focus is only on lines and not on the interplay between lines and other elements of the DVs.

4.4.5 Viewer Position

Within visual communication, it is possible to design the viewer's position to what is represented. That is another aspect of the interpersonal metafunction (besides modality) that Kress and Van Leeuwen (1996/2006) have investigated.

In film production, frame size refers to the distance from which something (like a human body) is filmed. If a person is filmed from a very close position, the person is "portrayed *as though* they [the filmed person and the viewer] are friends" (Kress & Van Leeuwen, 1996/2006, p. 126). Likewise, a long distance between the camera and the filmed person evokes a feeling of relational distance between the filmed person and the viewer of the film. This effect can, as Kress and Van Leeuwen (1996/2006) emphasize, also be used in representations of non-human participants, and not only in film production. In this way, the producer of a visual text can design the distance to the represented objects and topics – in other words, how close or how distant they feel to the viewer.

Like the frame size, the perspective on what is represented also influences the relation between the represented and the viewer (Kress & Van Leeuwen, 1996/2006, p. 129). The angle from which the represented object is viewed influences how subjectively or objectively the represented object is perceived. Kress and Van Leeuwen present various examples, including photographs, advertisements and DVs. In Western cultures, for instance, objectivity can be visually indicated by a direct frontal or perpendicular top-down angle (Kress & Van Leeuwen, 1996/2006, pp. 143–144). If one wants to signal subjectivity, an oblique angle may be used.

In this dissertation, I apply the concepts of frame size and perspective on DVs to investigate how the design of the DV affects the viewer's position to what is represented (see Article I).

4.5 Criticism of the Social Semiotic Approach to Multimodal Texts

As demonstrated in Sections 4.3 and 4.4, the authors of *Reading Images* (Kress & Van Leeuwen, 1996/2006) and *Introducing Social Semiotics* (Van Leeuwen, 2005) – both central sources for this dissertation – use frameworks developed for understanding verbal language for the investigation of, among other things, visual texts. This move has evoked criticism from several scholars. The first part of this section presents their main arguments; the second part is devoted to concrete points of critique of the methodology and of some of the concepts applied by Kress and Van Leeuwen.

I will first account for an argument *for* taking a linguistic framework as a starting point for developments in the multimodal field. Forceville (2007) points out that "linguistics is not at all a bad discipline to start working from", as "language is the most detailed and refined mode for the communication of complex information" (p. 1236). Machin (2013) brings up a related argument, saying that the strength of "a social semiotic approach to different forms of communication … allowed not only deeper analysis, but as in linguistics, a more systematic level of description" (p. 348). Put another way, an advantage of using a social semiotic approach is that it allows for detailed descriptions and profound analyses. However, several scholars, such as Forceville (2007, 2010), Machin (2013) and Ledin and Machin (2018), put forward the risk of having tunnel vision and trying to reinvent the wheel. From their perspective, this risk arises from using concepts of only a single theory of language, while ignoring what is offered by other analysis methods, that have been developed in other academic fields.

Despite several thousand years of debate, it is still unclear in what respects language is similar or different to visual text (Machin, 2009, p. 181). This makes it hard to estimate how fruitful it is to transfer a linguistic framework to analyses of texts that do not, or do not only, consist of verbal language. Ledin and Machin (2018, p. 25) point to a problem related to this aspect of similarity/difference between language and visual text. They argue that if the same metafunctions and same concepts are used to describe various kinds of semiotic materials, the qualities and affordances that the specific semiotic materials represent tend to get lost. "Such analysis tends to seek to show how instances of communication can be described through Hallidayan terminology" (Ledin & Machin, 2018, p. 25). In other words, one aspect of critique is that the actual focus of investigation (namely the uses of multimodal text) is lost sight of, while the application of Hallidayan terminology becomes self-perpetuating. According to Machin (2013) the relevant questions are therefore: "What is the advantage of the tools we use – [and] what do we see or hear now that we did not so do beforehand using the tools from other scholarly fields?" (p. 349).

I will now turn to the critique that concerns mainly the methodology applied by Kress and Van Leeuwen and some of their concepts. In a review of the 1996 version of *Reading Images: The Grammar of Visual Design*, Forceville (1999) questions the representativeness of the multimodal examples. He is critical of taking examples from such a wide variety of sources (like diagrams, schoolbook illustrations, film frames and newspaper photographs) without accounting for the differences between them. Another aspect of critique concerns the intersubjective validity of the interpretations made in the book, and thus on some of the resulting concepts. Both Forceville (1999) and Bateman et al. (2004) demand much more empirical work to be done when applying these concepts to concrete analyses of multimodal text. According to Forceville (1999), "systematic testing of the concepts against new pictorial material" (p. 175) is badly needed. Similarly, Bateman et al. (2004) consider work on multimodal corpora of organized data as a promising road to follow.

This points to another, more detailed point of critique relevant to this dissertation. Bateman et al. (2004) are highly critical of the meaning potential of composition, as Kress and Van Leeuwen describe it (see Section 4.4.4.). According to Bateman et al. (2004), there is, for example, little support for the claim that elements placed on the left of a spatial field are to be considered as the "Given" (Kress & Van Leeuwen, 1996/2006, p. 180), while such placed on the right are rather considered to represent the "New" (Kress & Van Leeuwen, 1996/2006, p. 180). The same critique is also expressed by Forceville (1999).

The last point of critique that I will present in this section deals with the general problem of categorization, put forward by Forceville (1999). Bringing this forward is particularly relevant since the process of categorization has taken a great deal of time in the PhD project behind this dissertation. For describing the representational structures, for example, Kress and Van Leeuwen (1996/2006) suggest several hierarchical categories. Forceville (1999) notes that this taxonomy is complex and questions its practical applicability to new pictures. Likewise, he is critical of the way Kress and Van Leeuwen suggest assessing modality, by considering various kinds of modality markers and coding orientations. Forceville's general concern is that "categories are seldom clear-cut; many categories are fuzzy, and describe a continuum" (1999, p. 168). This, however, is a fact of which Kress and Van Leeuwen are aware (see Kress & Van Leeuwen, 1996/2006, p. 86).

Summing up, taking a linguistic framework as the starting point for the analysis of multimodal text, as I have done in this dissertation, is not free of critique and problems. Several scholars have expressed their concerns, not all of which I was aware of before conducting the empirical work for this dissertation. The way I reflect on these points of critique, and how I experienced or tackled them, is presented in Chapter 11.

5 Literature Review

This dissertation contains three articles (see Chapter 10), all of which point to already existing literature. The extent to which a literature review found space in these articles, however, varies due to article length restrictions. This literature review intends to combine and extend these sections of the articles, and is structured as follows:

It first provides a general view of the semiotic potential of graphical lines, presenting sources from the field of visual communication and design, as well as social semiotics. The second, bigger part of this literature review centres around uncertainty[18] and narrativity – the two core aspects investigated in Articles II and III. As for *uncertainty*, I first describe the way it is discussed in the DV domain, highlighting the wide range of what may be understood by the term *uncertainty*. Then I present existing literature reflecting on techniques for signalling uncertainty in DVs in general. This constitutes an essential separate part because many of these techniques have later been applied on lines as well. Last, I outline relevant studies investigating various techniques for signalling uncertainty *by lines*. Then, similarly to the aspect of *uncertainty*, this literature review provides insights into how *narrativity* is traditionally discussed in the DV domain. This part also includes a brief glimpse into techniques for signalling narrativity in DVs (other than through lines). Lastly, I present previous studies investigating lines signalling narrativity in DVs. Figure 4 provides an overview of this structure.

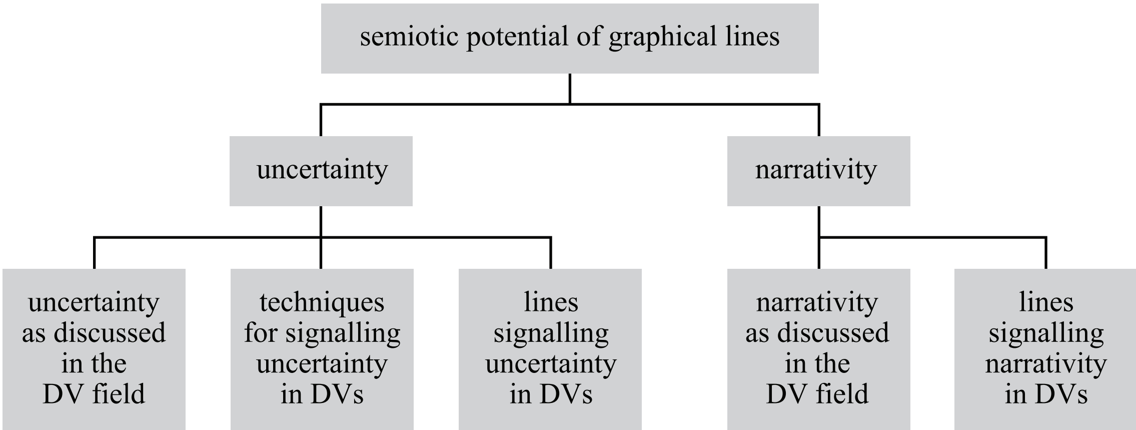
This literature review does not aim to be comprehensive but to give a good overview of the most relevant literature (from 1990 to today) touching on the investigated aspects of *graphical* lines. Literature illuminating the subject of the line in general (like, e.g., Ingold, 2007; Krämer, 2016; Mainberger & Ramharter,

18 As the literature discussed in this part of the literature review focuses on how *uncertainty* can be visualized, here I use the term *uncertainty* instead of *modality*, keeping in mind that these concepts are related, yet not identical.

2017) is therefore not focused on in this chapter. Instead, I assign more weight to presenting existing literature from a DV practice point of view.

Figure 4

Structure of the Literature Review



5.1 The Semiotic Potential of Graphical Lines in General

As the line is one of the most basic graphic elements (Bertin, 1967/2011, p. 44), it is not surprising that many handbooks contain sections about the semiotic potential of graphical lines. Poulin (2018, pp. 32–36), for example, points out that the visual characteristics of lines influence their meaning potential. If a line is straight, curvilinear, thin, thick, or whether it looks like it was painted with a brush, for instance, implies different kinds of meaning potentials[19]. Similar lists in other books present what the authors call the symbolic meaning[20] (Habermann, 2015) and "associative interpretation of lines" (Horn, 1998, p. 147). In all these three cases, the lines are looked at out of any application context.

To give two examples, Poulin (2018, p. 34) ascribes strength and power to a thick line, and Horn (1998, p. 148) claims that if a line is irregularly waved, it has the visual effect of uncertainty and weakness. However, the sources have limited empirical support. In Poulin (2018), there is no justification mentioned for how the author got to the meaning potentials of the line characteristics mentioned. Horn's (1998) explanations of the visual effects and emotions that he attributes to the different visual characteristics of lines are based only on "sugges-

19 An extensive discussion around the description of the visual appearance of lines (based on Bertin's visual variables (1967/2011)) is placed in Section 7.4.1, since this issue primarily concerns methodological decisions.

20 My translation. Original term: "symbolhafte Bedeutung" (Habermann, 2015, p. 649)

tions of three observers, a filmmaker …, an architect …, and a cartoonist" (1998, p. 147).

From a social semiotic perspective, the graphical line was investigated earlier as well. Cameron (2011) focused on the line used as a boundary line and explored its semiotic status. Fuller (2002) studied the arrow, as a line with a specific ending, in the context of wayfinding at airports. Neither Cameron (2011) nor Fuller (2002), however, systematically researched how certain kinds of visual characteristics of lines (as boundary lines or arrows) inform their semiotic potential. This was, by contrast, done by Johannessen and Van Leeuwen (Johannessen, 2017, 2018; Johannessen & Van Leeuwen, 2018). Johannessen (2017, 2018) presented a fine granular system to analyse the shape (or in other words, the outline of a stroke) of logos. He sees great potential in such fine-grained analyses for finding out about the meaning potential of a logo because the "qualities of the very shapes of the logo … can contain information about its articulatory history" (2018, p. 180). Whether a shape has "dips and bumps on the edge of the stroke" (2018, p. 175) or not, influences, as he explains, its experiential meaning potential because it reminds us of the circumstances of the activity of drawing the line. The concept of *experiential meaning potential* was coined by Van Leeuwen (1999), who, referring to metrical patterns of speech, described the meaning potential as "given by what we are actually *doing* when we produce the pattern" (p. 46). He continues that we have the "ability to turn action into knowledge, to extend our practical experience metaphorically, and to grasp similar extensions made by others" (p. 140). Therefore, he understands the experiential meaning potential as a source of meaning that can be projected to a semiotic expression, but which is based on bodily or social experience. Drawing on the concept of experiential meaning potential, Johannessen and Van Leeuwen (2018) investigate the line as a graphic trace, looking at a broader spectrum of texts (e.g., a handwritten sign, a childish computer font, a magazine page). They argue that:

> traces make meaning in an embodied and material way, through the bodily gestures and the physical tools and materials that articulate them, and that our understanding of traces is informed by our accumulated experience of the effect of these gestures, tools and materials on the traces they produce. Even when traces are no longer handmade, they will still be

> understood on the basis of this kind of experience, although the fact that they are not handmade will also contribute to their meaning …. (Johannessen & Van Leeuwen, 2018, p. 191)

5.2 Uncertainty

One aspect of *graphical* lines that is investigated in depth in this dissertation is their potential to signal modality and uncertainty (see mainly Article II). As described in Section 4.4.2, linguists and social semioticians have studied expressions of modality in verbal language as well as multimodal text, such as photographs, advertisements or DVs. Modality and uncertainty are related concepts, but the way practitioners and researchers in the DV domain approach the topic of uncertainty has not yet been covered. The following three sections provide an overview on (1) how uncertainty is discussed in the DV domain, (2) what techniques exist for signalling uncertainty in DVs in general, and (3) what former studies have investigated line characteristics for signalling uncertainty in DVs.

5.2.1 Uncertainty as Discussed in the Field of DV Design

The term *uncertainty* in the field of DV is associated with a wide range of meanings, as Hullman (2020) observes. She surveyed 90 professional visualization authors and interviewed 13 influential professional visualization designers and journalists. The associations with uncertainty these participants mentioned reflect how broad the understanding of uncertainty is in this field: "Interval, range, or region … error bars …. possible outcomes …. probability, confidence, variance, error, missing data and sampling and modelling artifacts", "qualitative expression of a gap in knowledge" "resolution", "visual inexactitude or encodings intentionally chosen to imply imprecision" (Hullman, 2020, p. 132) were just some terms and formulations mentioned. Instead of defining *uncertainty*, Wilke (2019, pp. 181–182) gets around this task by giving examples of situations where uncertainty could arise. Future events with an uncertain outcome and past events with missing data are two of those he identifies in *Fundamentals of Data Visualization* (2019, pp. 181–182).

Dasgupta et al. (2012) distinguish between *data uncertainty* and *visual uncertainty*, referring to uncertainty, the source of which lies in the data a DV designer gets at the very beginning of the process, and uncertainty introduced through the visualization process. A long list of reasons may cause data uncertainty. Measurement precision, incompleteness of data, inference, disagreement and data incredibility are those reasons named by Skeels et al. (2008). Bonneau et al. (2014) also notice that "uncertainty can arise in all stages of the analysis pipeline" (p. 5).

The approach Dasgupta et al. (2012) have on uncertainty in DVs is particularly relevant to this dissertation (mainly to the study presented in Article II). This is the case because they mainly focus on *visual uncertainty*, the "uncertainty that is associated with a visualization during encoding … and decoding" (Dasgupta et al., 2012, p. 1017). The taxonomy of visual uncertainty they propose (p. 1017) shows when and how visual uncertainty can be introduced during the encoding and decoding phase. Reflecting on what is important for this dissertation, I further explain only the steps of the encoding process. In the first step of the encoding process, they argue, the DV designer has to decide which kinds of data shall be included and excluded (based on the available data material, which could include data uncertainty) and whether the chosen data need to be configurated (e.g., introducing a different axis order or excluding outliers). In the second step, which they call the "visual mapping" (p. 1018), the data are transformed from numbers to a visualization. The designer first decides on how precisely to display the data (how well the different values are identifiable as distinct values). Then the DV designer chooses how much granularity to make visible (e.g., whether to show every single data point or to summarize some and show them as an area). Summing up, both incompleteness and configuration of data, as well as an imprecise and low granular visual representation, can cause uncertainty (Dasgupta et al., 2012). Said differently, the DV designer decides what level of certainty is intended to be signalled and then selects certain ways of graphical representation that may be useful for that purpose. In Section 5.2.2, I discuss the techniques used and researched for that purpose.

That DV designers sometimes do not try to visualize uncertainty at all and why this is the case, is discussed by Hullman (2020). She also states that "a formal science of visual communication stands to improve visualization-practice writ broadly, while integrating uncertainty as part of the definition of a visualization" (2020, p. 138).

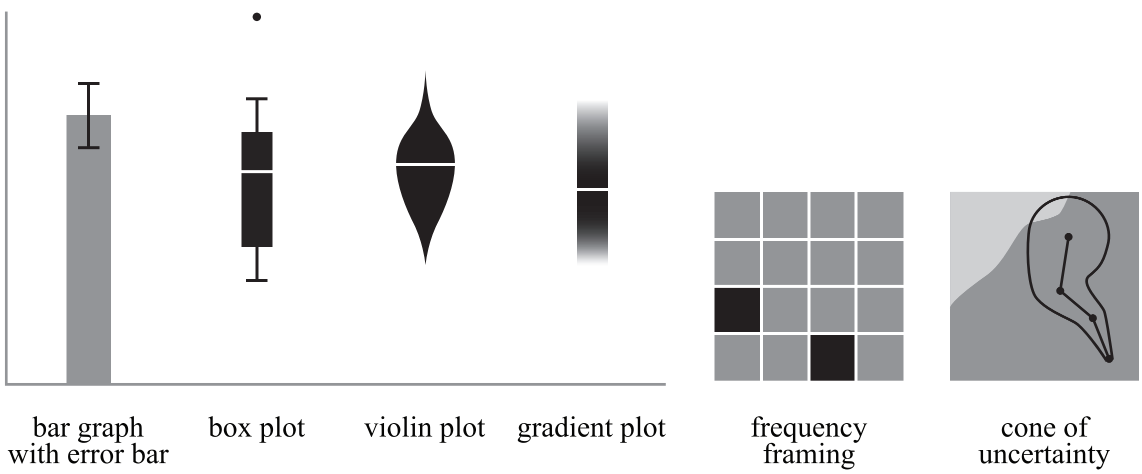
5.2.2 Techniques for Signalling Uncertainty in DVs

Various DV types and features have been developed to specifically represent *data uncertainty* (see Figure 5 for some examples). Two commonly used features of

DVs in scientific publications are *error bars* and *confidence bands* (Wilke, 2019, p. 181). Error bars can be combined with several kinds of DVs, like *bar graphs* or *scatterplots*. Yet, one cannot assume that a lay audience can correctly read such DVs, initially developed for the scientific context (Hullman, 2019; Wilke, 2019). As another technique applicable to scatterplots, Lu et al. (2020) suggest equipping the dots with winglets, the length of which indicates uncertainty. Bonneau et al. (2014) mention another ubiquitous DV type used for showing uncertainty information: the *boxplot*. There are several variations of this type (like the *violin plot* and the *gradient plot*) (Bonneau et al., 2014; Cairo, 2016), which may not make their interpretation easier. According to Hullman (2019) and Wilke (2019, p. 182) the technique of *frequency framing* is intuitive to read. A probability of 10%, for example, may be shown by ten graphical entities (like dots or icons) that have other visual characteristics than 90 others that are shown side by side. Uncertainty may also be shown on maps, a scenario that gains importance when it comes to the depiction of hurricane paths, for example. The *cone of uncertainty* is one kind of DV for showing the potential future path of a hurricane (Cairo, 2016; Hullman, 2019). In this case, the projected path is shown as a centreline with an area around it, which gets “wider the more the line moves away from the latest current location” (Cairo, 2016, p. 314).

Figure 5

Schematic Illustration of Some Visualization Types Developed Specifically to Represent Data Uncertainty



So far, I have presented some DV types and features used for signalling data uncertainty. This list is not aimed to be comprehensive, which may be impossible anyway, considering the number of possible hybrids and variations. Apart from using such DV types and features specifically made for visualizing uncertainty, visual properties of DV elements can also signal uncertainty (Hullman, 2019). In the following, I will give an overview of these more implicit techniques, that can be applied to different kinds of graphic elements (like, e.g., the line).

Generally, any kind of visual variable (such as *size*, *position*, *colour hue*; based on Bertin (1967/2011) – see Section 7.4.1 for details) can be used to signal uncertainty (whether based on uncertain data or not), as Kinkeldey and Senaratne (2018) observe. However, “some visual variables more intuitively connote certainty versus uncertainty than others” (Kinkeldey & Senaratne, 2018).

Cartographers have been concerned with this issue of representing uncertainty in an intuitive, subtle way on maps for several decades. MacEachren (1992) identifies that “an important representation issue for visualization of uncertainty …, is how Bertin’s graphic variables (with possible additions or modifications) might be logically matched with different kinds of data uncertainty” (p. 13). Without naming any empirical support, he states that modifying saturation is “arguably the most logical” (p. 13) way to depict uncertainty. As seeming “quite promising” (p. 14) he rates manipulating the “focus”, which can be done by adjusting “contour crispness”, “fill clarity”, “fog” and “resolution” (p. 14). To give two examples: an uncertain border on a map could be visualized with a fuzzy, fading line; an area of a map could be covered with a shape resembling fog, if the exact location is uncertain (“the thicker the fog, the more uncertain that part of the map” (MacEachren, 1992, p. 14)). Still in the context of cartography, McGranaghan (1993) notes that “in some cases, it may be desirable to give no clear indication of a data value to a user, but rather to graphically imply that the data quality is such that the data should be used with caution” (p.13). Graphic ambiguity may be used as an implicit strategy “to create visual and cognitive ambiguity related to uncertainty in the data” (McGranaghan, 1993, p. 11). For producing graphic ambiguity, he suggests various methods: showing multiple positions of graphical elements, using fuzzy, indistinct symbols or animation and mixing colours or patterns over blending zones. But as already mentioned, these are only suggestions, and McGranagham (1993) expresses the need for empirical research to demonstrate their validity.

Since the 1990s, several empirical studies have been conducted, not only in the cartography domain, about how uncertainty can be signalled. The systematic review presented by Kinkeldey et al. (2014) summarizes past user studies from the early 1990s to 2014, concerned with uncertainty visualization in the fields of scientific and information visualization, cartography and geovisualization. The studies compare the user performance of modifying *colour hue*, *value*, *sketchiness*, *saturation*, *whitening*, *transparency*, *texture* and *resolution* as intrinsic ways of indicating uncertainty (Kinkeldey et al., 2014, pp. 378–379). Summarizing their results, depicting uncertainty by manipulating *colour hue* and/or *value*, *whitening*, *transparency*, *texture* and *resolution* achieved good results.

Only a few studies have aimed to assess the intuitiveness of such techniques (Kinkeldey et al., 2014, p. 381). Most relevant to this dissertation (apart from those studies investigating lines, which are presented in Section 5.2.3), MacEachren et al. (2012) investigate the intuitiveness of representing uncertainty of point symbols by modifying 11 visual variables. The highest ranking was achieved through *fuzziness* (more fuzzy = less certain), *location* (further from centre = less certain) and *colour value* (lighter = less certain). *Arrangement* (poorer arrangement = less certain), *size* (smaller = less certain) and *transparency* (more obscured = less certain) were less intuitive than the first three. However, MacEachren et al. (2012) still deem them as acceptable for use. *Saturation*, which MacEachren earlier thought would be "arguably the most logical" (1992, p. 13), scored lower. Also, in the case of this visual variable, there was almost no difference in intuitiveness to signal uncertainty between a point with lower or higher saturation.

As can be concluded so far, the manipulation of the signs along several visual variables is deemed useful for signalling uncertainty. The link to visual metaphors plays an essential role for the intuitiveness (Kinkeldey & Senaratne, 2018). A fuzzy (i.e, blurry) border of a circle, for example, makes it harder to see, resembling how it would look in fog or with poor eyesight when not using any visual aids. That such metaphors and analogies to the "experiential world" (to use Kress and Van Leeuwen's words: 1996/2006, p. 46) are not only tested but also implemented into tools, can be shown by the following example: in 2017, Meeks added sketchy rendering as a feature into his DV framework *Semiotic* (Meeks, n.d.). Through this, users of that tool can profit from the "capacity [of the hand-drawn style] for communicating that the shape is unfinished or imprecise or otherwise less 'scientific'" (Meeks, 2017). Backed by the research published by Wood et al. (2012), he considers sketchiness as a characteristic of DVs useful to signal uncertainty or significance values. He further notices that "existing solutions for deploying sketchy rendering make it time-consuming to implement"

(2017). Therefore, practitioners could not experiment with sketchiness as much as they could with other features easier to apply, like animation.

So far in this section I have reported on techniques used for signalling uncertainty in general – without looking specifically at how this can be done with lines. Summing up, many approaches to visualize uncertainty are based on the principle of signalling it consistently with "our cognitive models of which perceptual elements contain variability or uncertainty" (Bonneau et al., 2014, p. 7).

5.2.3 Lines Signalling Uncertainty in DVs

As with any kind of sign, lines can be manipulated along their visual variables. Relevant to this dissertation, such manipulations have been evaluated in earlier studies for their ability to indicate uncertainty.

Boukhelifa et al. (2012) investigate how lines may be used to visualize qualitative uncertainty data, which they oppose to uncertainty data that "comes in numerical format" (p. 3). They distinguish between three main categories of methods for depicting uncertainty that are applicable to lines, that match the visual variable-based techniques mentioned in Section 5.2.2: "(1) color-based techniques that manipulate hue, saturation, or brightness dimensions; (2) focus-based techniques that modify contour crispness, transparency, or resolution; and (3) geometry-based techniques that distort line marks by applying a rendering style as in sketchiness" (p. 4). However, as they further state, no formal comparative evaluations existed of these methods applied to lines. Therefore, they conducted such a study, comparing four visual variables (*sketchiness*, *value* (i.e., greyscale), *blur* and *dashing*). Figure 6 shows examples of each of these characteristics. The criteria they were interested in are intuitiveness (for associating the line characteristics to uncertainty), accuracy (of assessing uncertainty values and generally, how many perceivable uncertainty levels can be expressed) and subjective user preference (which technique the participants prefer to represent uncertainty). As stimuli, they used image variations of a bar chart, a social network graph, a family tree, a Venn diagram, a rail network and a utility map. Of each of these stimuli, they manipulated one or more lines along the visual variables *value*, *blur*, *dashing* or *sketchiness*. The findings show that sketchiness of lines is only to 11.7% attributed to uncertainty. In terms of accuracy, none of the four techniques were very accurately perceived, having error rates of more than 10%. When asking about the subjective preference for visualizing uncertainty, dashing was the preferred technique (selected by 68.3%), before blur (selected by 15.1%), grayscale (selected by 12.5%) and sketchiness (selected by 3.12%). Some participants made comments on the application of sketchiness of lines to signal uncertainty, giving negative feedback on the "'informal' and 'unprofessional' look of sketchy lines" (Boukhelifa et al., 2012, p. 20)[21].

Figure 6

Examples of the Four Characteristics of Lines Tested by Boukhelifa et al. (2012)

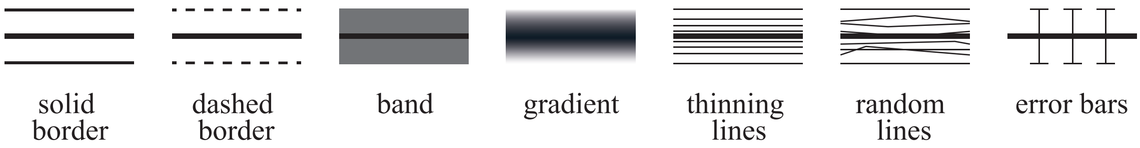
As explained earlier, the previous study focused on line characteristics for signalling uncertainty in the context of various DV types simultaneously. Of that kind, this one is, to my knowledge, still the only existing one. The following three studies present similar, comparative investigations in only one use-case (and thus DV type) each.

Tak et al. (2014) researched how various characteristics of lines representing the boundary between two earth layers (abstractly visualized as a cross-section of the earth) influence the test participant's judgement of probability that the boundary is at a certain position. Seven techniques were tested (see Figure 7): a solid border, a dashed border, a band, a gradient (blurring from the centre outward), several parallel thinning lines (decreasing their breadth from the centre outward), random lines and error bars (attached to the boundary). The results make clear that the technique used influences the perceived certainty. Which technique is most suitable for that purpose, is, however, dependent on "the definition of 'most suitable'" (Tak et al., 2014, p. 942). All techniques are suitable if the intended interpretation is that the certainty of the boundary is higher towards the centre and lower the bigger the distance to the centre is, or if it is that the boundary may also be outside the uncertainty range. The dashed border, random lines and the gradient stood out in the results because their perceived probability was near to normal distribution, which makes them "suitable for a wide range of data sources with an underlying normal distribution." (Tak et al., 2014, p. 942).

21 This kind of feedback may also point to a special semiotic function of sketchy lines – namely the ability to signal informality and unprofessionalism.

Figure 7

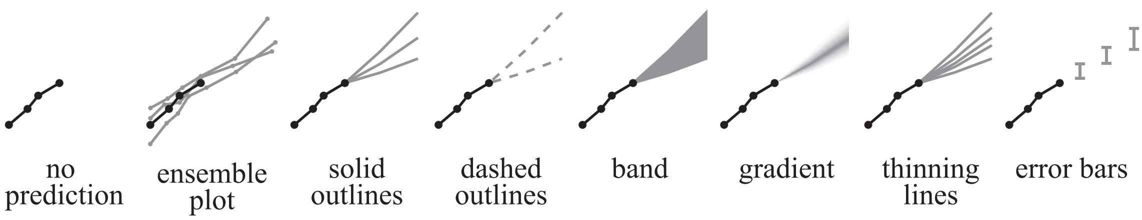
Schematic Illustration of the Seven Techniques for Visualizing Uncertain Boundaries of Two Earth Layers Tested by Tak et al. (2014)



Both Toet et al. (2019) and Liu et al. (2019) published studies concerned with visualizations of predictions through lines, although in different kinds of visualizations. The former looked at line graphs showing the uncertain future development of child growth. They tested eight different graphical representations of the predictions (see Figure 8). The techniques overlap with those tested by Tak et al. (2014) and are complemented by not visualizing the prediction at all (thus cutting the line at the last known data point). According to the test results, the best choice for non-expert users is using the ensemble plot (comparable to the random lines of Tak et al. (2014)) for indicating uncertainty. Thus, using a line with a gradient from a non-transparent centre to a transparent outer border, or dashed outlines on a broader line, for instance, are rated less useful. The authors point out that "it seems that this representation [the ensemble plot] effectively communicates essential unpredictability through the metaphor of 'multiple possible outcomes'" (2019, p. 10).

Figure 8

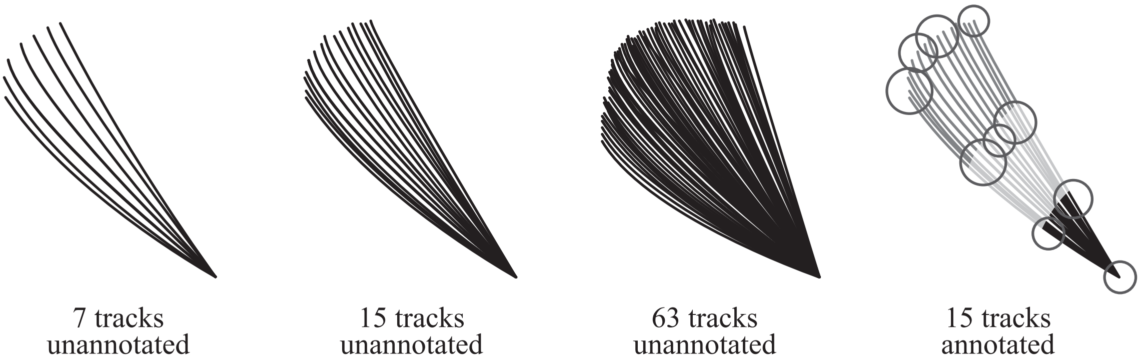
Schematic Illustration of the Eight Techniques for Visualizing Children's Growth Predictions Tested by Toet et al. (2019)



Focusing on maps showing the uncertain predictions of tropical cyclones, Liu et al. (2019) experimented with varying numbers of lines in ensemble plots and annotations to signal uncertainty (see Figure 9). The results of the experiment evaluating the effectiveness of such visualizations show that annotations make it possible to use fewer lines representing possible tracks.

Figure 9

Schematic Illustration of Exemplary Predictions of Tropical Cyclones on Maps, Tested by Liu et al. (2019)



Note. The annotations of the example on the right show different storm intensities (greytones), and storm sizes (radii of the circles).

Summing up the Sections 5.2.1. to 5.2.3, uncertainty in the DV domain is associated with a wide range of meanings. The reason for that may be that people with a wide variety of backgrounds, such as statisticians, computer scientists, designers and cartographers, work on the topic of visualizing uncertainty. Several empirical studies provide detailed insights into how uncertainty can be signalled in DVs *in general* by modifying the single graphic entities along visual variables. Visual metaphors and analogies to the experiential world play an important role in how intuitively uncertainty is associated with the visual sign. This knowledge is also applied in the studies presented that focus on how *lines* can signal uncertainty in DVs. At this point, one can also draw a connection to the concept of experiential meaning potential (see Section 5.1). Sketchy lines, for example, may indicate uncertainty because we know from our experience of drawing lines while being unsure of what the object to be drawn looks like, that our insecurity sometimes results in sketchy lines.

Although several studies focusing on how lines can signal uncertainty in DVs have been published, it is hard to compare their results. The reason is that they look at different techniques, with varying contexts of application and diverse research interests in mind. For this dissertation, a ranking of techniques is not essential. Most relevant is the knowledge that line characteristics can, through metaphors or analogies, indicate uncertainty in DVs. Thus, when encountering lines with such characteristics in the analysis, this knowledge can be applied to explore their meaning potential (see Section 3.4 in Article II).

5.3 Narrativity

The second aspect of *graphical* lines that is investigated in depth in this dissertation is their potential to signal narrativity (see mainly Article III). In social semiotic theory, narrativity is visually represented through vectors, indicating an action (Kress & Van Leeuwen, 1996/2006, p. 59). Lines that show a direction can serve as a vector and thus signal narrativity (see Section 4.4.3). In the following, I will first present how narrativity is discussed in the DV domain in general. Then I will provide an insight into previous research on how lines in DVs signal narrativity by indicating direction.

5.3.1 Narrativity as Discussed in the Field of DV Design

Narrativity in the DV domain is typically mentioned alongside storytelling. Research in that area includes investigations on how multimodal elements on websites containing DVs build a story. Weber (2020) defines a narrative as a "textual, visual, or multimodal representation that presents a story" (p. 297). A story, in turn, reflects a transformation of a situation, and thus consists of at least one unfolding event or happening (Weber, 2020, p. 297). In many cases, the focus of research in this area is not on single graphical elements within a DV (like the line), but rather on features and techniques that appear on these websites as groups of graphical elements. This applies, for instance, to the research presented by Bach et al. (2018). They present a set of narrative design patterns that they believe "can help storytellers design data-driven narratives that rely heavily, but not exclusively, on data visualization" (p. 109). This set comprises 18 such patterns, such as "call to action", "exploration" or "repetition" (p. 112), and their concrete implementation is demonstrated with exemplary visualizations.

Several other scholars have done comparable investigations by using a corpus of recent (at the time of corpus setup) DVs or data stories. Segel and Heer (2010) performed a systematic review of the design space of narrative visualizations, using a corpus of 58 narrative visualizations stemming primarily from online journalism websites. As a result, they present a design space organization with three divisions of features: "(1) genre, (2) visual narrative tactics, and (3) narrative structure tactics" (p. 1145). To mention examples of the second – and to this dissertation – most relevant division, one may visually structure a visualization through a progress bar or highlight elements through zooming (Segel & Heer, 2010, p. 1144). As these examples show, Segel and Heer's aim was to identify tactics, not to find out how to use single graphical elements to fulfil these tactics.

Building on Segel and Heer (2010), McKenna et al. (2017) researched the factors that shape visual narrative flow (flow-factors) and further use these factors for three studies looking at the reader's reactions and preferences related to these flow-factors. They identify seven flow-factors by analysing "a corpus of 80 recent [data] stories on the web" (p. 379). Examples of such flow-factors are *navigation input* ("how a reader interacts to progress through a narrative visualization" (p. 379)) and *level of control* ("how much control a reader has over the motion or animated transitions of story components" (p. 379)). These flow-factors can be approached by various techniques like providing buttons or a scroll bar to allow navigation input (Segel & Heer, 2010, p. 379).

Weber (2020) has looked at the same issues, but from a slightly different angle: she performed a qualitative corpus study on what visual elements may constitute narrativity in DVs. The corpus consists of 73 DVs collected from journalistic works that were shortlisted or won an award at the Malofiej Awards 2018, the Data Journalism Awards 2018 or the Kantar Information is Beautiful Awards 2017. Weber identified four broad techniques (the instance of a narrator, sequentiality, the temporal dimension and tellability) and listed how these broad techniques were realized in the corpus (2020, p. 307). She found that scrolling, animation and dynamic transition effects, for example, can express sequentiality; and that some DV types – such as timelines or flow maps – can be used to express a temporal dimension (Weber, 2020, p. 307).

Stolper et al. (2016, 2018) looked for visual, data-driven storytelling techniques in a corpus containing 45 data stories found by surveying blogs and following recommendations of the Visual.ly community. They found 20 such techniques, some of which overlap with the realizations observed by Weber (2020). Stolper et al. (2016, 2018) also mention – among other techniques – scrolling and flowchart arrows.

Summing up, the research presented in this section takes a broad perspective on how narrativity – in the sense of storytelling – is construed with multimodal elements on entire websites containing DVs. This means that the graphical elements central to presenting the data within the DVs (like connecting lines) play only a minor role in these investigations.

5.3.2 Lines Signalling Narrativity in DVs

Probably the most frequently examined feature of *lines* to indicate direction, and thus signal narrativity, is the arrowhead. In this section, I will, therefore, first present studies dealing with only the arrow and its function of signalling narrativity. Thereafter, I will describe relevant studies dealing with other visual characteristics of lines that may signal narrativity.

Of the six kinds of meaning potentials of arrows that Horn (1998) presents, several point to the function of signalling narrativity. According to him, they can (not only in DVs but in general) metaphorically indicate transformation, direction, flow, motion, force, time, and linkage (p. 98). More related to the research material of this dissertation, Van der Waarde and Westendorp (2000) identify seven functions of arrows used in user instructions, many of which overlap with those presented by Horn (1998). According to them, in user instructions arrows are used to (1) direct a movement, (2) show a physical change or transformation, (3) indicate dimension (e.g., by a line with arrowheads at both ends, complemented with a verbal length description) and (4) a connection (e.g., between an object and a label), (5) focus the attention by pointing to a specific location, (6) indicate sequence and (7) signal specific meanings as parts of icons or other signs[22]. However, neither Horn (1998) nor Van der Waarde and Westendorp (2000) investigate the relation between the arrow's visual appearance and its semiotic function or discuss any conventions regulating this relation.

More recently, Wong (2011) writes that arrows "are one of the most commonly used graphical devices in scientific figures" (p. 701). He notes that in the field of molecular biology, there are conventions for how to use arrows in diagrams. Different line lengths and line forms of arrows are used in that field as symbols, standing for what he refers to as "promoter[s]" or "primer[s]" (p. 701), for example.

That arrows are intuitively understood as indicating active processes has been shown by two experiments within the field of cognitive science, performed by Heiser and Tversky (2006). They investigate the function of arrows in the production and interpretation of mechanical diagrams. In the first experiment, the test participants looked at a diagrammatic depiction of a mechanical system

22 This list also includes arrows that I would not rate as vectors, such as the arrows with function number 3, 4 and 5. These arrows rather represent an aspect of the world "in terms of their more generalized and more or less stable and timeless essence" (Kress & Van Leeuwen, 1996/2006, p. 79, see Section 4.4.3) and thus represent a conceptual structure.

(a car brake, a bicycle pump, or a pulley system) either with or without arrows. They then had to write a description of the system. The results showed that "participants who described diagrams with arrows produced significantly more functional units … [i.e., verbs like *enter*, *open*, *close*, *push*] than participants who described diagrams without arrows" (Heiser & Tversky, 2006, p. 585). For the diagrams without arrows, the test participants used significantly more "structural predicates, specifically, more forms of the verb 'to be'" (Heiser & Tversky, 2006, p. 585). In the second experiment, the test participants had to sketch diagrams based on given verbal descriptions of the same mechanical systems. When the participants had functional descriptions, they used arrows more often than when they had structural descriptions (Heiser & Tversky, 2006, p. 588).

We will now look at two studies that deal with alternatives to arrowheads for indicating direction and thus signalling narrativity. Tversky et al. (2008) investigate how animated and static ways of telling stories in diagrams (like through using arrows) compare. Animations have, according to Tversky et al., "not proved to be superior to informationally equivalent static graphics in teaching structural or conceptual content" (2008, p. 265). Adding extra-pictorial information to static diagrams (such as arrows), however, is one strategy they consider useful for telling stories (p. 270). They suggest developing further and improving the use of animations by learning from the techniques that have proved to be useful in static diagrams (p. 281–282).

In the field of human-computer interaction, Koylu and Guo (2017) conducted a relevant study on direction-signifying, connecting lines in DVs. They evaluate the perception and task performance of users reading flow maps, by testing lines with five different visual line characteristics for signalling flow (i.e., direction). The test participants had to perform four reading tasks by looking at static flow maps using the five kinds of asymmetric lines, differing in shape, colour, and transparency (p. 313–316). The results, based on correctness and response time data, suggest that all five line types are potentially useful. However, the "influence of the design on performance and perception depends on the type of the task" (Koylu & Guo, 2017, p. 328). Therefore, they suggest that designers make their design decisions task-dependent.

To conclude Sections 5.3.1 to 5.3.2, we can state that previous research has investigated narrativity in the context of DVs both on the level of single graphical elements (such as the line) and on a much broader, multimodal level. Studies on the detail level show that lines with arrowheads may have several functions, and that signalling "unfolding actions and events, processes of change, transitory spatial arrangements" (Kress & Van Leeuwen, 1996/2006, p. 59) – and thus narrativity – is a central one. When it comes to other visual characteristics of lines used to signal narrativity, like animation or static techniques other than adding arrowheads, these have been researched only within specific use-cases of concrete visualizations.

5.4 Summary and Description of the Knowledge Gap

Summing up the previous work presented in this chapter, we can conclude that the graphical line has been investigated from several perspectives, including graphic design, cognitive science, cartography, human-computer interaction and social semiotics. More specifically, the same applies to the graphical line used to signal modality/uncertainty and narrativity. However, the existing literature either considers the meaning potentials of lines without any application context (see Section 5.1), within one specific domain, or related to one or a few concrete DV applications (see Sections 5.2.3 and 5.3.2). What has not been done before is a broad empirical investigation of the connecting line, as applied in DVs, based on a corpus that includes DVs of various types.

Corpus studies on visual texts, in general, are still rare (see Section 7.2). These are some relevant examples of studies that have a stable dataset[23], but contain other kinds of visual texts: Kembhavi et al. (2016b) and Hiippala et al. (2019) work with corpora containing diagrams from the school context (about 5000 items in the case of Kembhavi and 1000 in the case of Hiippala). Wilson and Landon-Hays (2016) investigate a corpus containing 1132 instructional images used by middle-school teachers. Thomas (2014) investigates a corpus of 24 product-packages.

The corpus-based studies presented in Section 5.3.1 (McKenna et al., 2017; Segel & Heer, 2010; Stolper et al., 2016, 2018; Weber, 2020) include 45 to 80 data stories or DVs. They contain various kinds of DVs, not only those where the graphical element used to signal the main message is one or more connecting lines. Also, the corpora vary in how systematically the dataset was selected. Moreover, the authors of these studies do not make transparent whether they set up a stable dataset, which could potentially be reused. To my knowledge, there is

23 By "a stable dataset" I mean that the analyst can access the identical data at any time. This is not the case, for instance, when accessing a visualization via the original URL, because the website's content may be modified or made inaccessible, or the access can become restricted.

no corpus yet that contains current, publicly available digital DVs as a stable and accessible dataset. Nor is there a corpus of such DVs where the graphical element used to signal the main message of all DVs is the connecting line. Consequently, nobody has so far performed a corpus study on such text, focusing on a single graphical element within the DVs using social semiotic concepts as a theoretical and methodological basis for the analysis.

To conclude, the research gap consists of the missing corpus-based study on what semiotic functions connecting lines realize in current, publicly available, digital DVs, and by which visual characteristics. This gap represents a small, yet important gap in our knowledge on visual texts and applied visual language. In this dissertation, I seek to fill this knowledge gap, providing a social semiotic, corpus-based approach. I set up a corpus containing such systematically collected units of texts, develop and test a systematic analysis method, and perform two corpus analyses. With these corpus analyses, I investigate whether and how connecting lines signal modality/uncertainty and narrativity, respectively, and whether any graphical conventions regarding these functions can be observed.

6 Overview of the Articles

The three articles presented in Chapter 10 constitute the core of this dissertation. This chapter provides a short description of each article, to provide an overview about the content and the relation between the articles. Summaries of all three articles, including the research results, can be found in Chapter 9.

The first article, *What a Line Can Say: Investigating the Semiotic Potentials of the Connecting Line in Data Visualizations*, lays the theoretical basis for the whole dissertation. It demonstrates how key terms from social semiotic theory can be used to analyse DVs, and especially single graphical elements within them. Focusing on connecting lines, this article highlights four types of semiotic potentials related to this graphical element and opens up for further empirical research how these semiotic potentials can be signalled. The second and third article presented in Chapter 10 deal with one semiotic potential each, as listed in the first article. Figure 10 shows an overview of how the three articles relate to each other.

The second article, *Modality and Uncertainty in Data Visualizations: A Corpus Approach to the Use of Connecting Lines*, concentrates only on the first of the four semiotic potentials identified in the first article, namely the indication of modality and uncertainty. Thereby it includes the social semiotic view on modality as well as the DV design view, where uncertainty is a matter of ongoing discourse, and clarifies the relation between modality and uncertainty. Furthermore, the article introduces a systematic method to analyse modality and uncertainty on different levels of DVs. This method can be used not only for corpus analyses focusing on connecting lines, but also for a broader range of visual elements. Finally, the corpus that was specifically developed for this dissertation comes into use. It contains 163 DVs, where connecting lines are in the leading role of communicating the intended meaning of the DV. This article presents an analysis of this corpus, focusing on whether and how modality and uncertainty (with all their semiotic varieties) are signalled by these connecting lines.

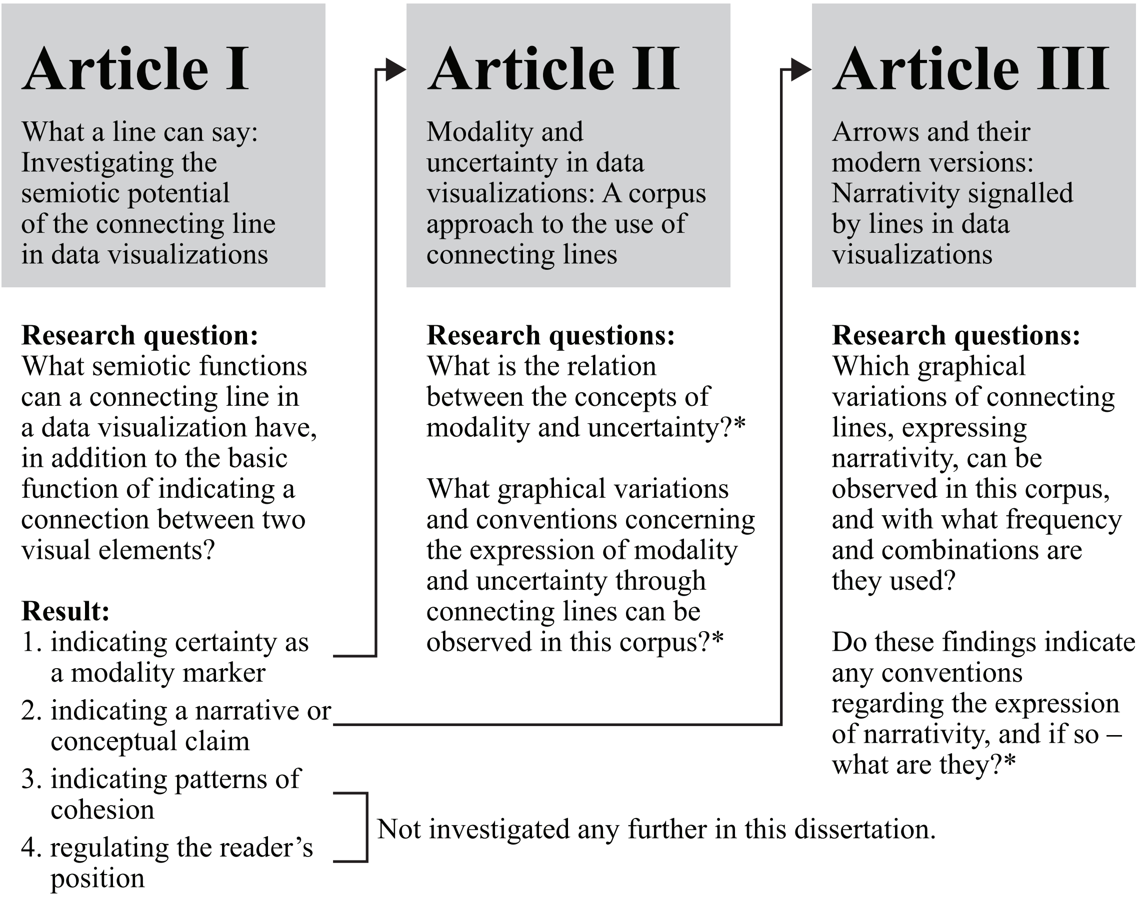
Following the corpus analysis, I draw conclusions on what graphical conventions the results point to.

The third article, *Arrows and Their Modern Versions: Narrativity Signalled by Lines in Data Visualizations*, is devoted to the second of the four semiotic potentials identified in the first article. It explores how visual characteristics of connecting lines are used to indicate a narrative claim. Based on the social semiotic concept of *representational structure*, the article presents an analysis method that I then apply in the corpus analysis building the core of this third article. The same corpus that was used for the second article is analysed; this time focusing on how the connecting lines indicate narrativity. (How conceptual claims are signalled by lines is not looked at, since that is characterized by the absence of visual characteristics that indicate narrativity.) Based on the results of this corpus analysis, I discuss whether any evolving conventions concerning the signalling of narrativity can be observed. As a side result of this empirical investigation based on the concept of representational structure, I point out the impracticability of using further subcategories of narrative representational structures, as the underlying literature defines them. This is another link to the first article, where I mention that this practicability could be investigated through analysing a larger corpus than just the few examples explored in Article I.

Recapitulating, Articles II and III concentrate on the first and second semiotic potential identified in the first article. The other two semiotic potentials named in the first article are not investigated any further in this dissertation, for reasons that shall now be explained. The first two semiotic potentials (indicating certainty as a modality marker and indicating a narrative or conceptual claim) can be realized through the connecting line itself or the line in combination with what it connects. For signalling the third and fourth semiotic potential, many elements of larger text units are involved – the connecting line is only one of them. In other words: in the first article I compare connecting lines to verbs – following Richards (1984, p. 3/21). Consequently, a line with its connected elements may be comparable to a sentence. The investigations of the first two semiotic potentials can be done staying on that sentence level. By contrast, for the third and fourth semiotic potentials to be realized, elements of the wholeness of the DV (thus several sentences if compared to verbal language again) are at interplay. Investigating them would focus not so much on the visual characteristics of the single lines, but rather on their interplay with other DV elements. Thus, I do not investigate the third and fourth semiotic potentials any further, since that would go beyond the focus of this dissertation.

Figure 10

Article Overview



Note. Research questions marked with * are reformulated statements written in the articles.

7 Methodology

This chapter includes a detailed description of the methodology used for the two corpus studies presented in the second and third articles. Although both articles contain a short description of the methodological approach, some aspects were assigned only a little space due to the word limits in the articles. Therefore, this chapter aims to complement the method sections in the articles. It offers more detailed information about the general considerations, the corpus setup and the practical steps in the analyses. As the first article was developed based on an in-depth reading of the underlying literature (from the field of social semiotics), no further explanations about the used methodology are needed for that article.

7.1 General Considerations for the Methodological Approach

Two corpus analyses were conducted to investigate the graphical variations of connecting lines that can be observed in the corpus, expressing modality and narrativity and what conventions potentially exist in this regard (cf., Articles II and III). These corpus analyses are based on two key concepts in social semiotic theory: *modality* and *representational structures*.

The analyses mainly cover the first step of what a social semiotic analysis most often includes, following Aiello (2020). Staying on a rather descriptive stage, I begin with building "a systematic inventory of data visualization resources" (Aiello, 2020, p. 56). Through the descriptions of visual characteristics used for the two semiotic functions I then contribute to the "attempt to outline a 'grammar' of data visualization design" (Aiello, 2020, p. 56). I don't follow Aiello's next step, which is to "*situate* data visualization resources in their social and cultural contexts" (p. 56) (apart from presenting a historical DV example in Article III). Neither do I conduct any user testing to find out how the participants use the visualizations. Moreover, the analyses and their outcomes only marginally touch the political dimension, which would be necessary for covering the third, transforming stage that Aiello (2020) identifies. Taking into consideration these differences to a full social semiotic analysis, I do not use this label for my study, even though the two analyses are based on a social semiotic theoretical framework. Instead, I use the broader term *corpus analyses* for the two analyses presented in Articles II and III. The argumentation for using the term *corpus* is presented in Section 7.2.

The method used for the second and third articles includes aspects of both quantitative and qualitative research. As described in Chapter 3, I see DVs from a constructivist position, which is reflected in the choice of social semiotics as a theoretical framework. This means that I consider DVs as "cultural objects, or media that communicate social meaning" (Neuman, 2006, p. 323), which is typical in a qualitative approach. However, the hypothesis that conventions exist for the application of certain visual appearances of connecting lines to signify modality and narrativity, is what motivates the corpus analyses in Articles II and III. Yet the aim of hypothesis testing, typical for a quantitative approach (White & Marsh, 2006, p. 30), is not central to this dissertation. Instead, the objective is to contribute to the development of a general, "big picture"(White & Marsh, 2006, p. 39), mapping both patterns and variations in the relationship between visual appearance and potential meanings.

Also, rather atypically for the quantitative approach (White & Marsh, 2006, p. 31), is the fact that the generalizability from the chosen sample (the DVs of the corpus) to the whole population of DVs, is limited. More important is the role model function that can be attached to the chosen corpus of award winners.

Because I used the theory described in the former chapters as a basis for setting up the analysis method, I follow a deductive approach – often associated with quantitative content studies (White & Marsh, 2006, p. 30). This also implies that the coding schemes (see Appendices G and J) were developed before the coding took place, based on those theories. However, since the first step of the corpus analyses presented in Articles II and III is done manually and individually for each DV in the corpus (further described in 7.2), this part can be seen as qualitative (Bell, 2001, p. 15).

Finally, I turn to the method of analysing the coded data. Due to the limited representativeness of the data material (see Section 7.3.2), no inferential statistical analysis was performed to describe the results, which, following White and Marsh (2006, p. 37), would be typical in a quantitative approach. Instead, frequency counts were generated and presented in combination with more qualitative descriptions of single examples. Table 2 shows a summary of all the methodological aspects mentioned in this section, ordering them according to where they belong on the continuous spectrum between the purely qualitative and purely quantitative traditions of research.

Table 2
Summary of the Qualitative and Quantitative Methodological Aspects of This Dissertation

← Qualitative		**Quantitative →**
• constructivist perspective on the research material	• underlying hypothesis, but hypothesis testing is not in focus – big picture shall be developed • limited generalizability, but the chosen DVs have the function of being role models • no inferential statistics, only frequency counts, in combination with in-depth descriptions of single examples	• deductive approach for developing the analysis method and the coding schemes

Recapitulating, the basis of the second and third articles are corpus analyses using key concepts from social semiotic theory to build up the analysis method that includes both qualitative and quantitative aspects.

7.2 What Is a Corpus, and Why Build One?

To investigate the potential conventions regulating the relations between graphical forms and semiotic functions of connecting lines in current, publicly available digital DVs, a collection of DVs is necessary. As Giorgia Aiello (2020) writes, identifying and cataloguing “resources that are representative of data visualization as a whole” is “a particularly challenging task, both because uses of data visualization cut across a vast range of social spheres, and because the existing empirical base to systematically describe key data visualization resources is still thin” (p. 54). When thinking of the vast number of DVs accessible on the internet, which may include connecting lines, it becomes clear that it is impossible to keep an overview of all of them. In particular, the fact that websites can be continuously modified, made inaccessible or that access to them can be restricted, makes it impossible to see this whole population of DVs as fixed and in its wholeness researchable. Or as Luc Pauwels (2011b) puts it:

> The Internet is, for the most part, and despite sophisticated search engines and efforts to add some structure and control to it, an uncharted domain: we don't conclusively know who or what is 'out there.' …. The Internet is a huge data repository, but not necessarily a very permanent or predictable one. (p. 572)

What was needed for this dissertation was a systematically selected, stable sample of DVs. This called for a closer look at how other fields working with corpora approach this issue. In research on verbal language, grammarians and lexicologists have a long tradition of working with text corpora (Halliday, 1985/2004, p. 48). Huge corpora for different languages have evolved, like The Oxford English Corpus, currently containing over 2.5 billion words (Oxford University Press, n.d.). Such a text corpus, as Halliday (1985/2004) describes it, "represents a systematic sample of text collected according to clearly stated criteria, whereas a text archive is assembled in a more opportunistic fashion; thus given such criteria, a corpus can be extracted from a text archive" (p. 49–50). However, he claims, that "the difference between a corpus and a text archive is not a sharp one" (p. 49). The size of a collection is, according to Sinclair (2004), not the only characteristic to define a corpus. Talking from a corpus linguistics perspective, he clarifies that corpora do not necessarily have to be big, but big enough, meaning that with more text, there is not more to find out (p. 188–189). Irrespective of its size, as Sinclair (2004) describes further, a corpus is much easier to define through the method used for its investigation: "You do not observe it directly; instead you use tools of indirect observation, like query languages, concordances, collocators, parsers and aligners" (p. 189). Summing up, the following characteristics define a corpus: It (1) contains a systematic sample, that (2) is investigated through indirect observation. I will now turn to why I had to develop a new corpus, and how it corresponds to the characteristics identified above.

An initial approach would usually be to see whether a ready-to-use corpus, containing the kind of material one is interested in, already exists or not. Several recent publications are based on research on a corpus, specifically built by the authors to investigate a specific aspect within this visual material (see Section 5.4).

Looking at corpora containing DVs, the ones presented by Hiippala et al. (2019; dataset: Hiippala, 2019/2020) and Kembhavi et al. (2016b, dataset: 2016a) are probably the ones which come closest to what I would have wished to have as a basis for this dissertation. However, they contain a very different kind of visual text than the one I was looking for: static diagrams representing science topics common in the primary school context. To my knowledge, there exists no generally accessible corpus containing current, publicly available digital DVs as a stable dataset (see Section 5.4). It was, therefore, necessary to build up a corpus of DVs specifically for this research.

The size of the corpus that I built up for this dissertation contains a much smaller amount of data than what is sometimes the case in corpora used for research on verbal language. The reason for this is the time required to build up a stable corpus of multimodal text (see Section 7.3). As explained earlier, the usefulness of a corpus and its status as such is not necessarily dependent on its size. More important is that it contains a systematically collected sample, and that an indirect observation method is used. Section 7.3 contains a description of the systematic corpus setup process. Indirect observation, as is typical for work on corpora of verbal language, means that tools are applied that allow for generating queries that research the whole corpus at once, without looking at every single text item.

To observe the data material in this corpus indirectly, it was first necessary to develop an annotation system for all the aspects to be queried. The data collected for the corpus used in this dissertation are non-linear data – data which do not "unfold along a single dimension of actualisation" (Bateman et al., 2017, p. 155). The data are not only accessible through one dimension, such as a temporal or spatial sequence, but, in the case of DVs, also include spatially distributed information. The tools for researching those kinds of data are less developed than those aimed at linear data (Bateman et al., 2017, p. 155). Computational methods, including machine learning and deep learning, are promising, as Bateman et al. continue. However, to my knowledge, no automatic tool exists for describing single graphic elements of DVs and interpreting their meanings. Thus, for the corpus analyses presented in Articles II and III, human annotation was needed first. The solution was to first annotate all the connecting lines in the DVs by hand and then integrate the resulting data into a database, which could be queried (see Section 7.4). In that sense, the observation method applied in the two corpus analyses presented in Articles II and III is only partly indirect, a fact that also limits the corpus size. Nevertheless, since the collected DVs were selected following strict criteria, and the investigation methods include at least partly indirect observation, I will call the collection a *corpus*, and the investigation of it, a *corpus analysis*.

7.3 Corpus Setup

In the following, I will present the applied criteria for building up the corpus, how the stable data basis was achieved, and which steps were taken before performing the corpus analyses.

7.3.1 Finding Sources for a Basic Collection of Relevant DVs

As mentioned in the first overarching research question (see Section 1.4), the research materials of this dissertation are current, publicly available digital DVs. As publicly available digital DVs, I defined those that are accessible via the internet during the time of corpus collection and made for being displayed on a digital platform (as opposed to print media). That includes not only DVs made for being viewed on websites, but also DVs made for being displayed on screens in exhibitions, for example. However, they needed to be presented online by their producers to be considered. Thus, the potentially usable research material was to be found on the internet. This implied challenges for the collection of the research material:

(1) It is hard to gain an overview of all DVs available on the internet. Thus a way to determine a systematic selection of DVs needs to be found.

(2) For defining DVs as current, it needs to be clear when they were produced or published.

The decision to use award-winning DVs tackles both issues. First, award-winning DVs presumably serve as role models for future developments in the design practice field, since they gain much publicity. When researching conventions in DV design, I thus follow the path of investigating such DVs, that were on the one hand awarded and therefore have high quality standards, and on the other hand are leading the way for practitioners who may follow such awards. Second, determining the production or publishing date of digital content is a challenging task, mainly because such metadata are not always available on the webpages, and directly contacting the producers is not feasible in terms of time resources. For the chosen DV awards, however, the regulations require that the visualizations have been produced or published in a certain period of time. Therefore, being able to decide on specific award years enables me to define what I consider as *current* DVs and makes a time restriction feasible. Another benefit of using only award-winning DVs is that, since the designers creating them are experts in their craft, random design decisions appear unlikely. One can assume that the designers base every design decision on an intention to express a certain meaning potential. This increases the meaningfulness of corpus analyses focusing on the relation between form and function of graphical elements.

After having decided on using only award-winning DVs for the corpus setup, I had to decide which awards to include. Focusing on *data visualizations*, it becomes clear that the chosen awards need to have a strong focus on DVs and not only on the broader field of journalism or infographics. Therefore, I refrained from looking at the winner lists of some competitions which might generally be interesting to look at from a broader perspective, but not for this dissertation. Table 3 shows which awards I excluded in the process of setting up the corpus and why.

Table 3
List of Excluded Awards and the Related Exclusion Criteria

Award	**Exclusion criteria**
Online Journalism Awards (Online News Association, n.d.)	too broad a focus (journalism)
NODA Awards (NxtMedia, n.d.)	too broad a focus (journalism)
Canadian Association of Journalists Awards (Canadian Association of Journalists, 2018)	too broad a focus (journalism)
Best of News Design Competition (SND, n.d.-b)	too broad a focus (news)
Best of Digital Design Competition (SND, n.d.-a)	too broad a focus (digital journalism)
Innovation by Design Awards (Fast Company & Inc, 2018)	too broad a focus (design in business)
World Digital Media Awards including the regional awards (WAN-IFRA, 2018)	too broad a focus (digital media)
Data Journalism Awards (Global Editors Network, 2018)	focus on data journalism, only very few DVs per year are awarded

The two awards selected for this study are the Kantar Information is Beautiful Awards (The Information is Beautiful Awards Ltd, 2018a) (further abbreviated as *Kantar Awards*) and the Malofiej Awards (Malofiej Infographic World Summit, n.d.-e). To work with only the latest DVs, the winner lists of the last three years (at the point of data collection) constituted the basis for the corpus setup.

Looking closer at the type of awarded entities, the Malofiej Awards, at least according to the award title, focus on infographics. Nonetheless, when looking at the winner lists by using my definition of DVs (see Section 2.1), many of the winners can be categorized as DVs rather than infographics. Consequently, I included the Malofiej Awards. The categorization of whether a winner is a DV or not was done manually in a separate step (see Appendix A). The same applies to the Kantar Awards, which focus on DVs, infographics, interactives and information art. The way the winner categories are defined for this award is different from those of the Malofiej Awards. The categories even differ from year to year. Therefore, I went through the winner list sources and defined which winners to exclude, either because they are not DVs (according to my definition, see Section 2.1), or because they were not made for the digital context (see Section 7.3.3 and Appendix A).

7.3.2 Limitations of Using Award Winners as the Basis for the Corpus Setup

Limitations need to be pointed out concerning two aspects when taking awards as a basis for such a corpus setup. First, what potentially can end up in the corpus is dependent on the awards' regulations on who is permitted to submit pieces of work. Second (and also linked to the awards' regulations), the countries where the submitted pieces of work are produced, and what audience they are targeted at, may differ from award to award. Both limitations have an impact on the spectrum of final corpus items.

Regarding the issue of who is invited to submit their work, the two awards follow different practices. The Malofiej Awards are "open to all general circulation newspapers – daily or non-daily, broadsheet or tabloid, printed or online – and magazines published anywhere in the world, as well as syndicates and agencies providing graphics" (Malofiej Infographic World Summit, n.d.-a). However, when looking at the winner lists from Malofiej 24, 25 and 26, it becomes clear that submissions from the United States and Europe – many from big newspapers like *The New York Times* and *The Guardian* – dominate these lists. In comparison, the Kantar Award is open for submissions from a much broader range of institutions and people. Besides the submissions that can be made by "practitioners, dataviz creators and teams from any kind of organization" (The Information is Beautiful Awards Ltd, 2016b) it is also possible to nominate DVs and infographics by others (not the producers themselves). The winner lists of the years 2015, 2016 and 2017 reflect that the range of producers is wider than that of the Malofiej Awards. Yet, Kantar winners also seem to be targeted mainly at audiences in the Western world (The Information is Beautiful Awards Ltd, 2018b).

The most pertinent question is about which expectations one can set for the results of the corpus analyses presented in Articles II and III concerning generalization. It is possible to claim that a corpus with 163 award-winning DVs from these two awards represents a broad enough spectrum of high-quality DVs to offer insights regarding general graphical conventions. But this is only possible with the following restrictions: the final list of award-winning DVs chosen for building up the corpus is by no means powerful enough to generate valid statements about the whole population of digital DVs in the world in the chosen timeframe. It is only a slice of all DVs produced in that time frame, and the fact that more weight is assigned to DVs from the USA and Europe needs to be kept in mind. In fact, all verbal text of the award winners finally used for the corpus was written in English, Spanish, Portuguese, German, French or Norwegian. Therefore, I assume that they were all targeted at audiences in the Western world. Thus, I presume my cultural background and language skills make it possible to analyse these award winners from a relevant cultural perspective. Consequently, the results can offer relevant and plausible information concerning how the acknowledged DVs targeted at the Western world have been developed during this time period. The corpus built up for this dissertation is the result of an attempt to build the broadest possible, systematically collected repository of highly acknowledged DVs, analyzable by one person within a restricted timeframe.

7.3.3 The Selection of Award Winners

Based on the main research questions, five criteria were essential for the selection of the award winners for the corpus. They needed to be (1) current, (2) publicly available, (3) digital and (4) contain at least one DV with (5) one or more connecting lines in the leading role of communicating the DV's meaning. The issue of selecting only current DVs was solved by looking only at the award winner lists of the three last years (at point of data collection). Thus, the six chosen award winner lists belong to the Malofiej Infographic World Summit 24, 25 and 26 (Malofiej Infographic World Summit, n.d.-b, n.d.-c, n.d.-d) and the Kantar Award 2015, 2016 and 2017 (The Information is Beautiful Awards Ltd, 2015, 2016a, 2017)). I considered all the other selection criteria in a manual categorization process:

The criterion *publicly available* means that the chosen award winners have to be generally possible to access (whether free of charge or by paying a subscription fee). This excludes those candidates that were not accessible, either because they are commercial products of the private industry or because the URLs were no longer available at the time of the review. Turning to the criterion *digital*, the difference between the Malofiej and the Kantar Awards has to be underlined. On the winner lists of the Malofiej Awards, two categories divide printed graphics from digital (respectively online) graphics. That makes it easy to exclude the printed award winners. Such a categorization is not available on the Kantar Award showcase website (The Information is Beautiful Awards Ltd, 2018b). For this reason, this categorization had to be done manually, excluding all award winners that were not produced primarily for presentation on a digital system (like posters). In the next step, those award winners that don't contain a DV (but an infographic, e.g.) were filtered out. This was done based on the definition of the term DV presented in Section 2.1. The final selection step was to select only award winners that contain at least one DV with one or more connecting lines in the leading role of communicating the main message of the DV.

Appendix A contains the six winner lists (three years per award), showing in detail how this filtering process was executed. A summary of the filtering steps can be found in Tables 4 and 5. In brief, of all 205 (70+135) publicly available, digital DV projects that were awarded in these years, 105 (36+69) contain one or more central DVs with one or more connecting lines in the leading role of communicating the DV's meaning.

Table 4

Summary of the Filtering Process of the Award Winners of the Kantar Information Is Beautiful Awards 2015, 2016 and 2017

Award[1], year	**How many single DV & infographic projects were awarded, that were publicly available at the time of data collection?[2]**	**How many of those (in the second column) were DVs?**	**How many of those (in the third column) were digital?**	**How many of those (in the fourth column) contained one or more central DVs with one or more connecting lines in the main role?**
K, 2015	30	27[3]	24	8
K, 2016	25	20[3]	18	13
K, 2017	39	31	28	15
Total	94	78	70	36

Note. [1] K = Kantar Information is Beautiful Awards.
[2] Winners that were awarded several times by the Kantar Awards were counted only once. Winners that were awarded both by the Malofiej Awards as well as by the Kantar Awards were only counted within the latter. In this step, all awards were filtered out that that were given to people (e.g., as teams or students), to DV tools or to websites with collections of many different projects.
[3] In these two years there was an award category called *infographic*. However, the definition used by the Kantar Awards and the one used in this dissertation (as described in 2.1), differed. Therefore, this categorization was redone for this dissertation.

Table 5

Summary of the Filtering Process of the Award Winners of the Malofiej Awards 24, 25 and 26

Award[1], year	**How many single online DV & infographic projects were awarded, that were publicly available at the time of data collection?[2]**	**How many of those (in the second column) were DVs?**	**How many of those (in the third column) contained one or more central DVs with one or more connecting lines in the main role?**
M, 24	43	34	21
M, 25	72	55	25
M, 26	55	46	23
Total	170	135	69

Note. [1] M = Malofiej Awards
[2] Winners that were awarded several times by the Malofiej Awards were counted only once. Winners that were awarded both by the Malofiej Awards as well as by the Kantar Awards were counted only within the latter.

Through this filtering process, 105 different award winners (award-winning websites) were identified that meet the criteria mentioned above. Table 6 shows how the number of 105 award winners was reached.

Table 6
Number of Selected Award Winners, Based on the Tables in Appendix A

Award, year	Number of selected winners
K, 2015	8
K, 2016	13
K, 2017	15
M, 24	21
M, 25	25
M, 26	23
Total	105

Of each of these 105 selected award winners, full-size screenshots, screencasts and PDF prints were taken for four reasons: Firstly, it was necessary to document the state of these webpages at the time of investigation. Secondly, it was important to have the material stored independently, so as not to have to access the winning pages with the URL each time, sometimes even requiring a subscription to the newspapers or publishing institution. Thirdly, these screenshots, screencasts and PDF documents show how the single DVs are embedded into the broader multimodal context, which was needed for the further analysis. Finally, the screencasts reveal all feedback to interactions with the websites, respectively DVs presented on these websites.

In the final stage of this award winner selection process, an Excel table was set up, listing these 105 award winners, their URL, the award name and year, the available metadata (such as authors, institution, country of origin, publishing date and date of last update) as well as the file names of the full-size screenshots, screencasts and PDF prints.

7.3.4 The Selection of Single DVs for the Corpus

Because the 105 selected award-winners could contain either one or more DVs of interest, a list of all single DVs was developed in the next step. In practice, the 105 award-winning websites were searched through for DVs that contained one or more connecting lines in the leading role of communicating the DV's meaning. Those found were given a title based on the surrounding verbal text. A screenshot of the specific website part including each DV was taken. These screenshots show only a static state of the visualization. This suffices, because the animations and interactive elements of each DV can be investigated in the 105 screencasts, covering the entirety of the whole websites, as explained earlier.

Because of the selection criteria for the 105 award-winners, at least one such DV was found per website, sometimes more. If DVs of various types were found on a website, or at least the kind of connecting lines used in the DVs varied, then all those DVs were selected. This way, 163 single DVs were identified. A separate Excel list was set up, containing the visualizations' titles, the file names of the screenshots and the corresponding award-winner. These 163 DVs build up the systematically and strategically collected corpus for the investigations presented in Articles II and III. All visualizations (with one exception, which was made specifically for internal company use) are targeted at the general public and were published in online news media or other channels of public information. A full list of these 163 DVs, their corresponding award-winner, their URLs and the award name and year can be found in Appendix B.

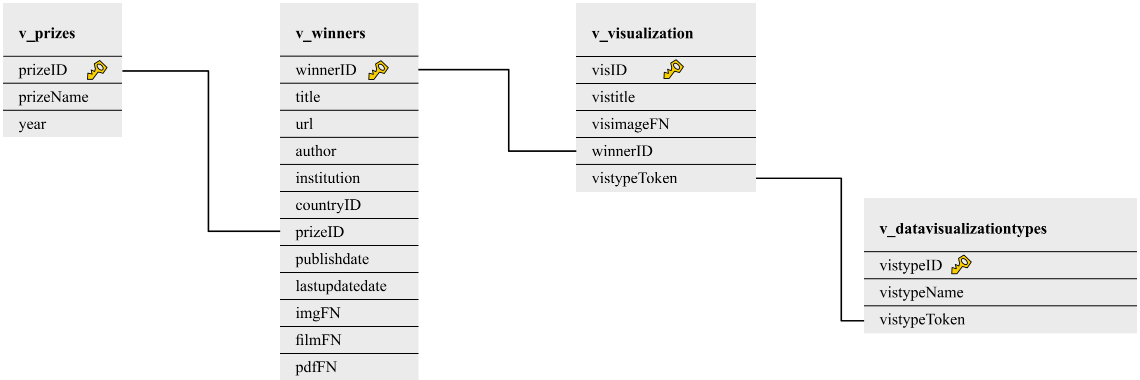
7.3.5 Database Setup

In the next stage of the process, a relational database (MySQL) was built up. It first contained only the two prepared lists containing the 105 award-winners and the one with the 163 single selected DVs. The prizes were split into a separate list to make the award-winner list more lucid. A fourth list was prepared for saving the DV types of each visualization (see Section 7.3.6). Figure 11 shows a structural overview of the database before the analysis step.

This relational database approach makes it possible to create queries across multiple tables, which is especially useful when it comes to investigating relationships between visual appearance and semiotic function of the connecting lines of the visualizations. Moreover, it makes it highly scalable and therefore it is possible to easily expand the database in the future (adding for instance new awards, new winners, new visualizations, etc.), without losing control due to the growing amount of data. As a third advantage, normalizing the database structure avoids redundancy.

Figure 11

Structural Overview of the Relational Database, Before the Analysis Step



Note. Each box represents a table. In each box, the table title (bold) and the column names are listed. The relations between the different tables are visualized through lines, and the primary key of each table is marked with a key icon. The filenames (here abbreviated as *FN – imgFN*, *filmFN*, *pdfFN*, *visimageFN* etc.) of the screenshots, screencasts and PDFs of the award-winners and their visualizations were saved in this database, so that these files could be easily accessed for the corpus analysis.

7.3.6 Categorization into DV Types

As can be seen in Figure 11, the DVs in the corpus were also categorized according to their DV types. This categorization offers an overview of the kinds of DVs contained in the corpus. It is also helpful when investigating patterns of form-function relations, which again can be related to specific DV types.

Such a categorization presupposes a selected repertoire of DV types. Table 7 shows the list of 23 DV types into which the DVs were categorized. How this repertoire of DV types was developed is described in Appendix C. After having categorized all the DVs into one of the DV types listed in Table 7, those data were imported into the prepared table of the database (see Figure 11).

Table 8 summarizes the result of the categorization process by showing the distribution of DV types in the corpus. The three DV types most frequently found in the corpus are the route map (59 instances), the line graph (37 instances) and the connection map (18 instances). Of all other DV types between one and six instances were found in this corpus.

Table 7

List of DV Types Used for Categorizing the DVs in the Corpus

DV type no.	DV type name
1	route map
2	flow map
3	connection map
4	line graph
5	network diagram
6	stream graph
7	lollipop chart
8	timeline
9	alluvial diagram
10	sociogram
11	dumb-bell chart
12	bump chart
13	parallel coordinates
14	slope chart
15	radar chart
16	direct trace visualization
17	hybrid of radial network diagram and network diagram
18	hybrid of timeline and linegraph (splinegraph)
19	hybrid of linear process diagram and dendogram
20	hybrid of line graph and dumb-bell chart
21	hybrid of dumb-bell chart and bubble chart
22	hybrid of connection map and line graph
23	hybrid of arc diagram and bubble chart

Table 8

Distribution of DV Types Within the Corpus of 163 DVs, Ordered According to Frequency of Occurrence

DV type	Instances of DVs
route map	59
line graph	37
connection map	18
flow map	6
slope chart	4

dumb-bell chart	4
sociogram	4
hybrid of dumb-bell chart and bubble chart	4
timeline	4
lollipop chart	3
stream graph	3
network diagram	3
hybrid of radial network diagram and network diagram	2
direct trace visualization	2
parallel coordinates	2
radar chart	1
hybrid of arc diagram and bubble chart	1
hybrid of timeline and line graph (splinegraph)	1
bump chart	1
hybrid of connection map and line graph	1
hybrid of line graph and dumb-bell chart	1
hybrid of linear process diagram and dendogram	1
alluvial diagram	1
total	163

7.4 Corpus Analyses

For both corpus analyses, each DV went through three steps after having defined which line the analysis would focus on:

(1) description of the visual appearance of the focused line
(2) semiotic analysis of the focused line, investigating the semiotic potentials regarding modality (Article II) and narrativity (Article III)
(3) description of the visual characteristics that are used to signal the semiotic potential analysed in step 2.

The final coding instructions (Appendices G and J) used for both analyses explain these steps in detail. Supplementary to the methodology presented in both papers and the appendices, the following three sections contain:

(1) detailed information about how the analysis method of the visual appearance of the lines was developed
(2) detailed information about the inter-rater reliability studies
(3) insights into how the single-coded corpus analyses were conducted from a practice point of view

7.4.1 Description of the Visual Appearance of the Focused Lines

To be able to analyse the effects of different visual appearances of lines concerning their meaning potential, it is necessary to define a consistent way to describe those visual appearances. Although Articles II and III both contain the description scheme that was finally used, they offer only a brief explanation of how this scheme was set up. Therefore, this section offers a detailed overview of the underlying theory about the visual description of graphic marks in general, and then more specifically of lines.

The basis for developing the visual description system of lines for this dissertation was Bertin's (1967/2011) concept of "visual variables" (p. 42). As his concept was further developed and adjusted by several scholars over time, I considered the developments relevant for this dissertation too. Moreover, some scholars contributed specifically to a nuanced visual description of lines. These nuanced ways of describing lines also had an impact on the categorization scheme that was ultimately used. In the following, I explain what Bertin's basic concept involved and what further developments were considered relevant for this dissertation.

Bertin (1967/2011) developed a system of eight visual variables, in which "a visible mark can vary" (p. 42):

> A visible mark can vary in position on a sheet of paper. … A mark can thus express a correspondence between the two series constituted by the TWO PLANAR DIMENSIONS. Fixed at a given point on the plane, the mark, provided it has a certain dimension, can be drawn in different modes. It can vary in SIZE VALUE TEXTURE COLOR ORIENTATION SHAPE and can also express a correspondence between its planar position and its position in the series constituting each variable. The designer thus has eight variables to work with. They are the components of the graphic system and will be called the "*visual variables*". (p. 42)

When developing this list of visual variables, Bertin (1967/2011) only thought of what was printable on a sheet of white paper with graphic means already available at that time (p. 42). Digital DVs – the research material used in this dissertation – by contrast, offer designers creative possibilities other than static print media. For this reason, I will further describe the changes that I applied to Bertin's system of visual variables.

Since in digital DVs, three-dimensional space can also be used, a way to integrate the third dimension into Bertin's system had to be found. Coming from the field of animation, Hayward (1984) mentions two visual variables, pointing to a third dimension: "change in angle" and "change in viewpoint" (p. 26). Another scholar thinking of the third dimension as a missing aspect in Bertin's visual variables is Marc Green. He (1991) specifically calls for the additional visual variable "binocular disparity" (p. 21). As for the use in this dissertation, the third dimension is included in the final description scheme by simply extending Bertin's variable *position* with the third planar dimension. This also affects the visual variable *orientation*, which now also includes orientation possibilities using the third planar dimension.

Morrison (1974) made a change to Bertin's visual variables of colour and value: he suggested using colour as the main dimension, with the three constituent dimensions hue, value and intensity (also called saturation) (p. 123). In this dissertation, the formerly two visual variables *colour* and *value* have thus been merged into one main dimension *colour* with the three subdimensions as described before.

I will continue with visual variables that have to do with how the surfaces of the displayed units appear. MacEachren notes that the visual variable *clarity* (earlier called *focus*) had to be added to Bertin's list of visual variables (MacEachren, 1992, p. 14, 1995, p. 275). He presents three subdimensions of clarity: crispness, resolution and transparency (1995, p. 276), which I have used for the final visual description scheme in this dissertation. That makes it possible to describe blurring, pixilated and transparent lines better.

Moreover, Bertin's visual variable *texture* in this dissertation is renamed *pattern*, following MacEachren (1995), who points out that pattern is "a higher level visual variable consisting of units that have shape, size, orientation …" (p. 273), which also makes sense if one thinks of the wide variety of line patterns.

Lastly, the way in which Bertin's visual variables are complemented with animation and interaction shall be explained. Both aspects were irrelevant in a print environment, which was Bertin's starting point. When investigating DVs in a digital environment, however, animation and interaction are widely used and therefore need to be considered. Hayward (1984, pp. 26–27), Green (1991, pp. 20–21), DiBiase et al. (1992, pp. 205–206), Weger (1999, p. 46) Saulnier (2005) and Saulnier et al. (2006) were concerned with aspects of change in visualizations. However, only the contributions of Saulnier (and the team around her) and Weger have been used in this dissertation for integrating the aspect of animation into Bertin's visual variable system.

Using the French term "le mouvement" (2005, p. 190) Saulnier describes this term as a process of change, and therefore has a very broad understanding of it. With its four subvariables trajectory[24], speed/frequency[25], phase[26] and duration[27] such a process of change can, as she understands it, be applied not only to the variable position (2005, p. 190). Rather, she mentions examples of effects on other variables suggested by Bertin (Saulnier, 2005, p. 190; Saulnier et al., 2006, pp. 59, 75): the effect on the size could, for example, be perceived as an enlargement, the one on colour, for example, as a blinking, the one on orientation could be perceived as a rotation, and so on. Thus, it is not the dynamics themselves that carry the information, but always their effects on the different variables (Saulnier, 2005, p. 190; Saulnier et al., 2006, pp. 59, 75). Although Saulnier used the term *movement*, I decided to apply the term *dynamics*[28], earlier introduced by Weger (1999, p. 46). This I do to emphasize that a process of change can be applied to any of the mentioned variables, not only on the change of position, to which the term *movement* could rather lead.

Thinking of dynamics occurring in data DVs, they can be planned by the designer as an animation. Furthermore, they can be triggered by an interaction like a mouse-over or a mouse click, or reversely be used as a trigger for an interaction like, for example, blinking. DiBiase et al. (1992) searched for an extension of Bertin's visual variables to be able to include the temporal dimension into dynamic maps. In this context, they (1992) suggest adding *interaction*, as "the empowerment of the viewer to modify a data display" (p. 205). I integrated this additional aspect into the visual description system used for this dissertation. But as the interaction itself has no visual appearance (but only either the previous or resulting state of the DV), this variable has a special status in the description

24 My translation. Original term: "la trajectoire" (2005, p. 190)
25 My translation. Original term: "la vitesse/fréquence" (2005, p. 190)
26 My translation. Original term: "la phase" (2005, p. 190)
27 My translation. Original term: "la durée"(2005, p. 190)
28 My translation. Original term: "la dynamique" (2005, p. 187)

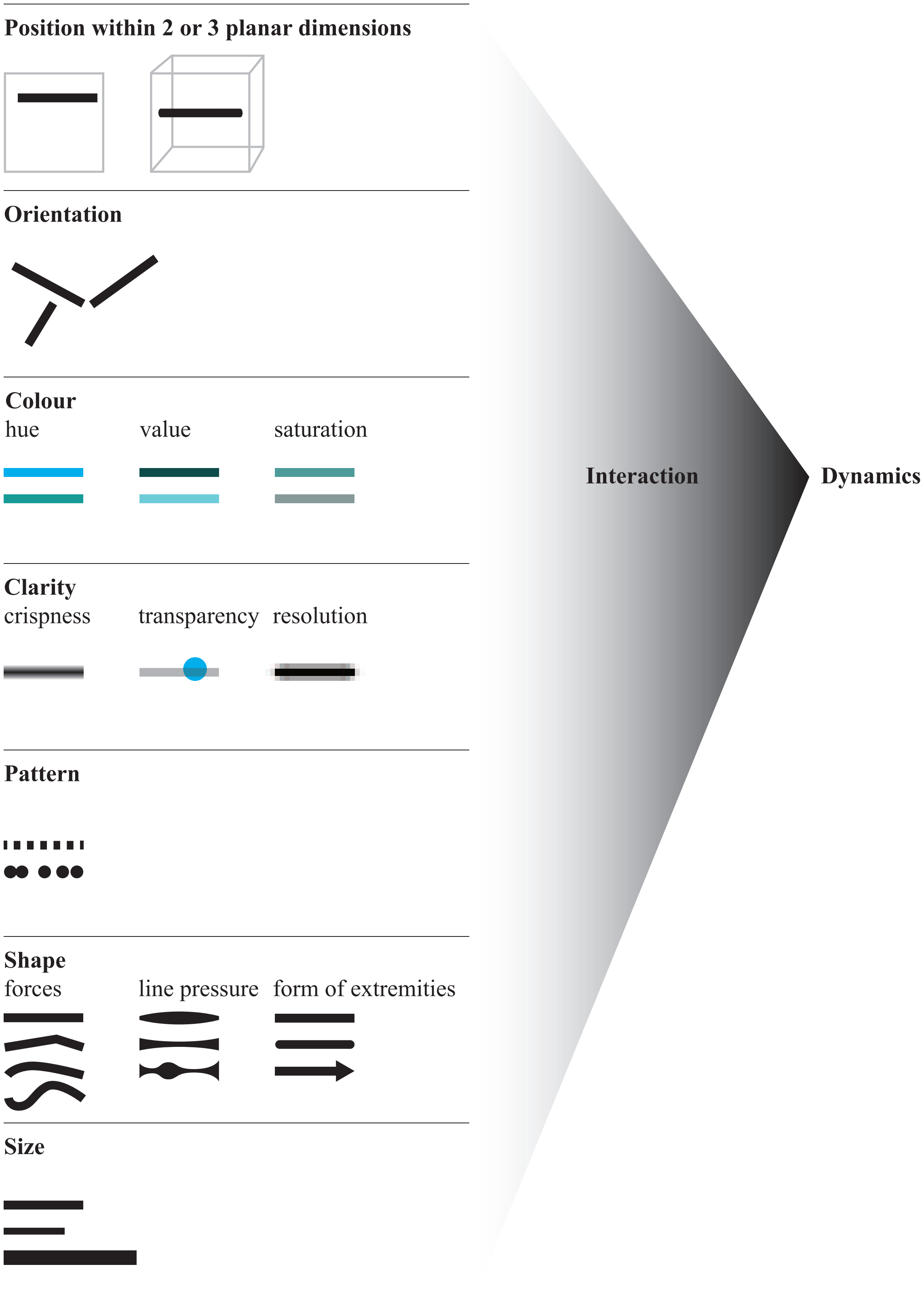
scheme. The same applies to the variable dynamics, which only affect the other visual variables used.

For the description of lines, which have quite a prominent length (W. Wong, 1993, p. 45), Bertin's visual variable *shape* is specifically interesting. For the analyses presented in Articles II and III it was necessary to describe the lines' thickness change along the length, their curvature or edges, and what their ends look like. I therefore defined three subdimensions of *shape*: *forces*, *line pressure* and *form of extremities*. For the first two subdimensions, Kandinsky (1926/1928) laid the foundation; the last one stems from Wong (W. Wong, 1993). Whether a line is straight, curved or bent depends, according to Kandinsky (1926/1928), on the forces that are applied when transforming a point to a line (p. 51). If only one force is applied, the result is a straight line. If more forces are applied, either at the same time or alternatingly, curvature or edges occur along the line. The pressure that is applied on the drawing tool defines the line's breadth, as Kandinsky further explains (p. 83). Thus, if the same pressure is used along the line, its breadth stays the same. If the pressure changes along the length of the line, its breadth becomes inconsistent. Now turning to the last subdimension of *shape*, Wong's (1993) description of the extremities as a characteristic of lines is relevant. He stated that the extremities "may be negligible when the line is very thin. But if the line is quite broad, the shapes of its extremities may become prominent. They may be square, round, pointed, or any simple shape" (p. 45). Thus, through the subdimension *form of extremities*, it is possible to describe arrows and tapering lines, for example (although that overlaps to some extent with the subdimension *line pressure*). Figure 12 shows an overview of the visual variables used for the description of the visual appearance of lines in Articles II and III.

In addition to defining the visual variables relevant to the investigations of this dissertation, it was necessary to define manifestation categories. This was needed because, for both investigations, I counted the number of occurrences of specific visual characteristics. The definition of the used manifestation categories for each article was done during an iterative process, starting with trying to describe lines of a few example visualizations that were not part of the corpus, and continuing until the completion of the inter-rater reliability test. This allowed for adjustment of the manifestation categories until they were fine-grained enough to describe the overall visual appearance of the lines and particularly the visual characteristics used for signalling modality and narrativity if that was the case.

Figure 12

Overview of the Included Basic Visual Variables and Examples of Their Effects on Lines



Note. The concrete lines shown here are only some examples. An extensive list is not possible because of the unlimited possibilities and nuances.

With other analysis foci, the manifestation categories in Table 2 of Article II and Table 2 in Article III would maybe have been defined differently. Also, for each visual variable, more fine-grained manifestation categories could have been defined (like, e.g., describing exact colour values). But since the resulting categories both stemmed from the earlier presented literature as well as from analytical observation of real use-cases, and proved useful in the inter-rater reliability studies, I determined them as described in the two articles.

7.4.2 Inter-Rater Reliability Studies

Before the corpus analyses presented in Article II and III could be performed, inter-rater reliability studies were conducted. This was done to estimate how consistent and reliable the coding (and therefore the stated questions and the offered answer categories) is. Double coding by two independent analysts was necessary for those questions that contain judgement variables (Brezina, 2018, p. 269). For practical reasons, this was done with 25 DVs, which is approximately 15 % of the corpus. The coders were the author of this dissertation (rater A) and the main doctoral supervisor (rater B).

The coding itself was done using Excel. In practice, two Excel documents (one for each corpus analysis) with three tables were built up, which were used for the inter-rater reliability studies. The first table of each document served for the semiotic analysis, the second one for the description of the visual appearance of the lines in focus, and the third one for defining which visual variable was rated as being used to signal modality or narrativity (if that was the case). Appendices D and H contain the coding instructions, which were used for the inter-rater reliability studies. Because of the word count limitations of Articles II and III, the procedure of the inter-rater reliability studies is described there only very briefly. Therefore, hereinafter I present the method details for both articles.

7.4.2.1 Inter-Rater Reliability Study for Article II

For Article II, the inter-rater reliability study was performed in two steps: first, a sample of 20 randomly selected DVs was analysed, assuring that visualizations stemming from websites awarded with both awards and several years were in the sample. The randomization was done by applying the "=RAND()" function in Excel to the list of 163 DVs in Appendix B, and then taking the first 20 of those randomly ordered ones. A brief look at the results showed that only a few (in the case of rater A) or none (in the case of the rater B) of the DVs were rated as signalling any kind of uncertainty or modality. The analysis method developed (as described in Appendix D), however, offers to do a very fine-grained analysis of exactly these issues. Therefore, neither rater assessed the sample of 20 DVs as representative enough to show the (dis-)agreement of the results of this pre-test. That's why in a second step, the sample size was increased by five more DVs, that were double-coded (using the same analysis method as before). This time, the author of this dissertation (rater A) chose five consecutive items of the same randomly ordered list used before, where at least one of them (according to rater A) signalled some kind of modality in her eyes. Rater B was given this information too but didn't know which or whether several DVs met this criterion, and how and what kind of modality rater A thought was signalled.

After having collected the data of this inter-rater reliability study for 25 DVs (sample size = 25), the results were evaluated in the software R (The R Foundation for Statistical Computing, 2018). The answers of both raters of the questions in column E, G, I, K, L, P, Q, R, S and U (as described in Appendix D) were checked for interrater reliability. This was done by calculating Gwet's AC_1 and Gwet's AC_2 coefficient (Gwet, 2014, pp. 40, 83) using the *gac* function of the *rel* package (Martire, n.d.) within R. For all questions with nominal answer categories (columns E, G, I, K, P, Q, S) Gwet's AC_1 coefficient (unweighted) was calculated. For questions which resulted in ordinal data (columns L, R and U), Gwet's AC_2 coefficient (weighted, applying linear weights) was calculated. Columns 3–4 in Table 9 show the resulting coefficients and standard errors for each question.

To interpret these results, the benchmarking method proposed by Gwet (Gwet, 2014, pp. 173–181), complemented by information given by Gwet in personal mail correspondence on 9-17 July 2019, was used. All details concerning this benchmarking process are presented in Appendices E and F. Column 5 of Table 9 shows the final benchmarking results. The two coders had either substantial agreement or higher. For most questions, almost perfect agreement was reached. As described in Article II, the coding method was generally approved because of these high agreement levels. However, the two raters had a follow-up discussion after the pre-test, which revealed improvement possibilities through a few small adjustments to the coding scheme. Therefore, these changes to the coding scheme were made, following an iterative method. The result of this process was the final coding scheme (see Appendix G).

Table 9

Results of the Calculations of Gwet's AC_1 and AC_2 Coefficient of the Inter-Rater Reliability Study of Article II, as Well as the Final Agreement Level of Each Question

Question column; question as described in Appendix D	**Coefficient calculated**	**Coefficient value (COEFF)**	**Standard error (SE)**	**Final agreement level (= interval name of first cumulative benchmark probability exceeding 0.95 – see Appendix E)**
E; Which coding orientation needs to be applied when viewing the connecting line?	AC_1	1	0	perfect
G; Is there any form of intended uncertainty in terms of lowered probability (that what the connecting line depicts didn't, doesn't/won't for sure happen or wasn't/isn't/won't be like that) or lowered reliability represented through the line in focus?	AC_1	0.864743	0.081975	substantial
I; Is it explicitly verbally stated on the detail or on the global level that data uncertainty is represented within this detail statement of the data visualization? (If no, then the next field is NULL)	AC_1	0.954914	0.045915	almost perfect
K; Based on the previous questions, does the line in focus signal modality in terms of intended lowered reliability?	AC_1	0.864743	0.081975	substantial
L; If yes, which modality value is signalled?	AC_2	0.912127	0.056049	almost perfect
P; Which coding orientation needs to be applied to the visual expression of the overall statement?	AC_1	0.913043	0.059514	almost perfect
Q; Does the choice of visual style signal any sense of modality in terms of intended lowered reliability?	AC_1	0.958368	0.042421	almost perfect
R; If yes, which modality value is signalled?	AC_2	0.98625	0.01389	almost perfect
S; Is it explicitly verbally stated on the detail or global level that data uncertainty is represented in the overall statement of the DV?	AC_1	0.954914	0.045915	almost perfect

U; Looking back at all the answers to the previous questions (both on the detail and the global level, maybe including intended modality (in terms of lowered reliability) or data uncertainty expressions), what kind of overall modality (in terms of lowered reliability) profile does the DV have?	AC_2	0.906503	0.056092	almost perfect

Note. Number of raters: 2; Sample size = 25 DVs; Confidence level: 95%.

7.4.2.2 Inter-Rater Reliability Study for Article III

The inter-rater reliability study performed previous to the corpus analysis focusing on narrativity (presented in Article III) followed almost the same procedure as the one presented in Section 7.4.2.1 for Article II. A sample of 25 DVs was selected randomly from the corpus of 163 DVs (see Appendix B), using the same Excel function as described in Section 7.4.2.1. Both raters got coding instructions (see Appendix H), the necessary files showing the DVs to be analysed and the empty Excel file to be filled out. After having collected the data from both raters, the agreement was calculated in R (The R Foundation for Statistical Computing, 2018) in the same way as for Article II. Columns 3-4 of Table 10 present the results for each question.

In the next step, the same benchmarking method was applied as in the inter-rater reliability study for Article II (see Section 7.4.2.1 as well as Appendix E). Column 5 of Table 10 shows the final agreement levels for each question. For further details concerning this calculation, see Appendix I.

As can be seen in column 5 of Table 10, the agreement level for the first question was almost perfect. In comparison, the agreement levels of the second and third questions were substantial and moderate respectively. Thus, the agreement declined the further detailed the question was (which is logical considering that the answers to the questions were partially dependent on each other). In addition, the follow-up discussion between the raters revealed theoretical ambiguities that made it very hard to answer the second and third questions. Therefore, but also because of other reasons presented in Article III, these two questions were not used for the single-coded analysis of the whole corpus. The coding instructions were adjusted according to this decision and can be found in Appendix J.

Table 10

Results of the Calculations of Gwet's AC_1 Coefficient of the Inter-RaterReliability Study of Article III

Question column; question as described in Appendix H	Coefficient calculated	Coefficient value (COEFF)	Standard error (SE)	Final agreement level
E; Does the visual appearance of the focused connecting line signal a narrative or a conceptional representational structure?	AC_1	0.94678	0.052157	almost perfect
F; If it signals a narrative, does it signal an action process or a conversion process?	AC_1	0.890049	0.075583	substantial
G; If it signals an action process, does it signal a transactional or a non-transactional one, or an event?	AC_1	0.65243	0.11944	moderate

Note. Number of raters: 2; Sample size = 25 data visualizations; Confidence level: 95%.

7.4.3 Single-Coded Corpus Analyses

The final coding schemes (see Appendices G and J) were used to perform two analyses of all 163 DVs in the corpus. One of them focuses on how modality is signalled by the connecting lines in the corpus (as presented in Article II). The other focuses on whether and how narrativity is signalled by the connecting lines in the corpus (as presented in Article III). This section shows how I conducted the corpus analyses from a practice point of view, including how I integrated the results into the database and performed the data analysis.

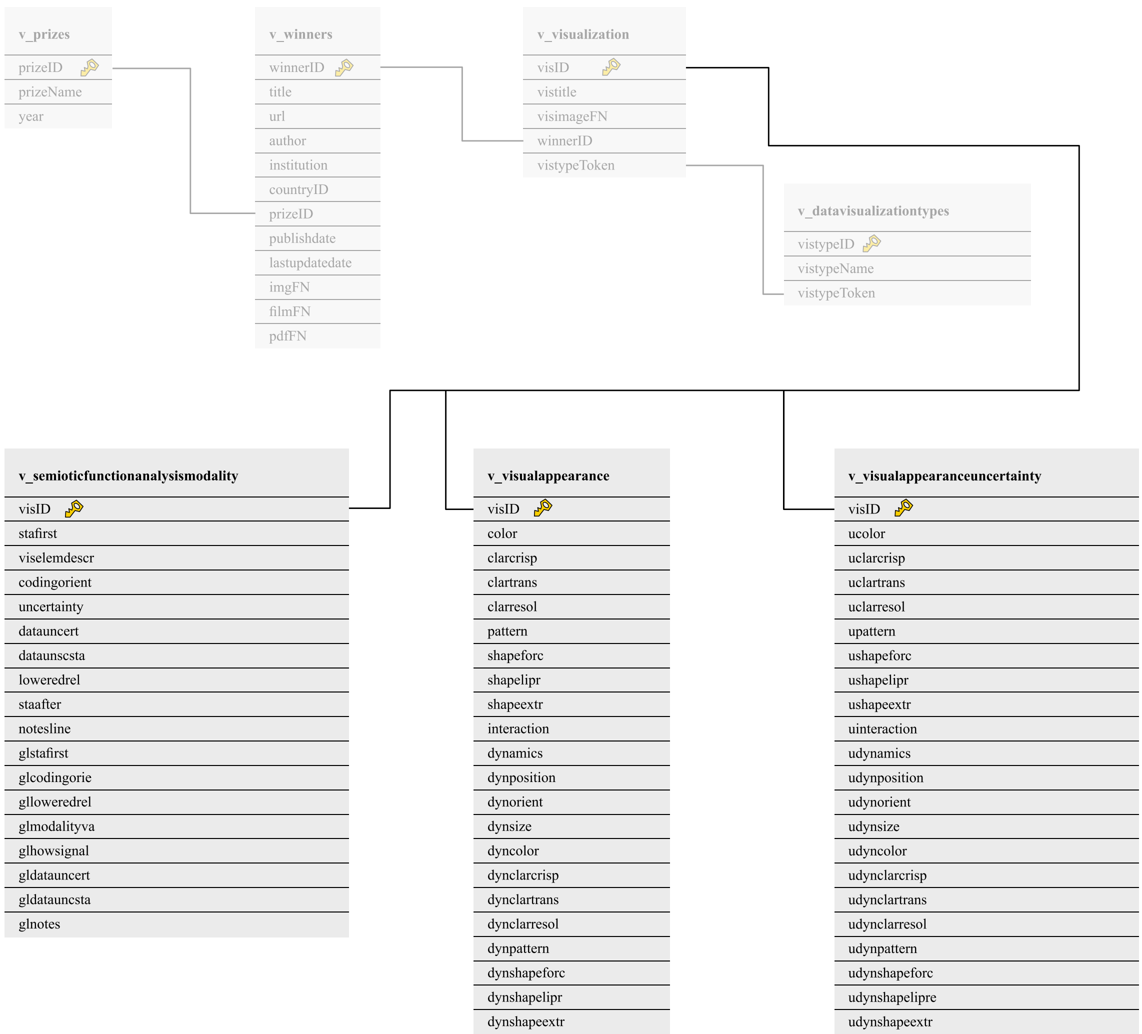
As already mentioned, the first step after both inter-rater reliability studies was to adjust the final coding schemes according to the decisions made. After this, the Excel documents for the single-coded analyses were developed. The Excel documents used for the inter-rater reliability studies (see 7.4.2) constituted the basis, only some questions and answer options were adjusted so that they matched with the final coding instructions.

Then each DV in the corpus was single-coded by the author of this dissertation, and the applicable answers recorded in the two Excel documents. This was done by looking at the locally stored screenshots, screencasts and PDFs (the filenames of which are organized in the database, see Figure 11). If instances of doubt occurred, the final codings were discussed with the second rater of the inter-rater reliability test.

The corpus analysis results from the resulting Excel files were then imported into the relational database. For this to be possible, the database structure, as shown in Figure 11, was enwidened with several empty tables prepared for importing the final codings. For the analysis focusing on modality, three tables were added, corresponding to the semiotic analysis, the visual analysis and the modality-marking visual variables of the lines in focus (if applicable). Figure 13 shows the resulting database structure. After the adjustment of the database structure, the coding results were imported.

Figure 13

Extension of the Relational Database of Figure 11



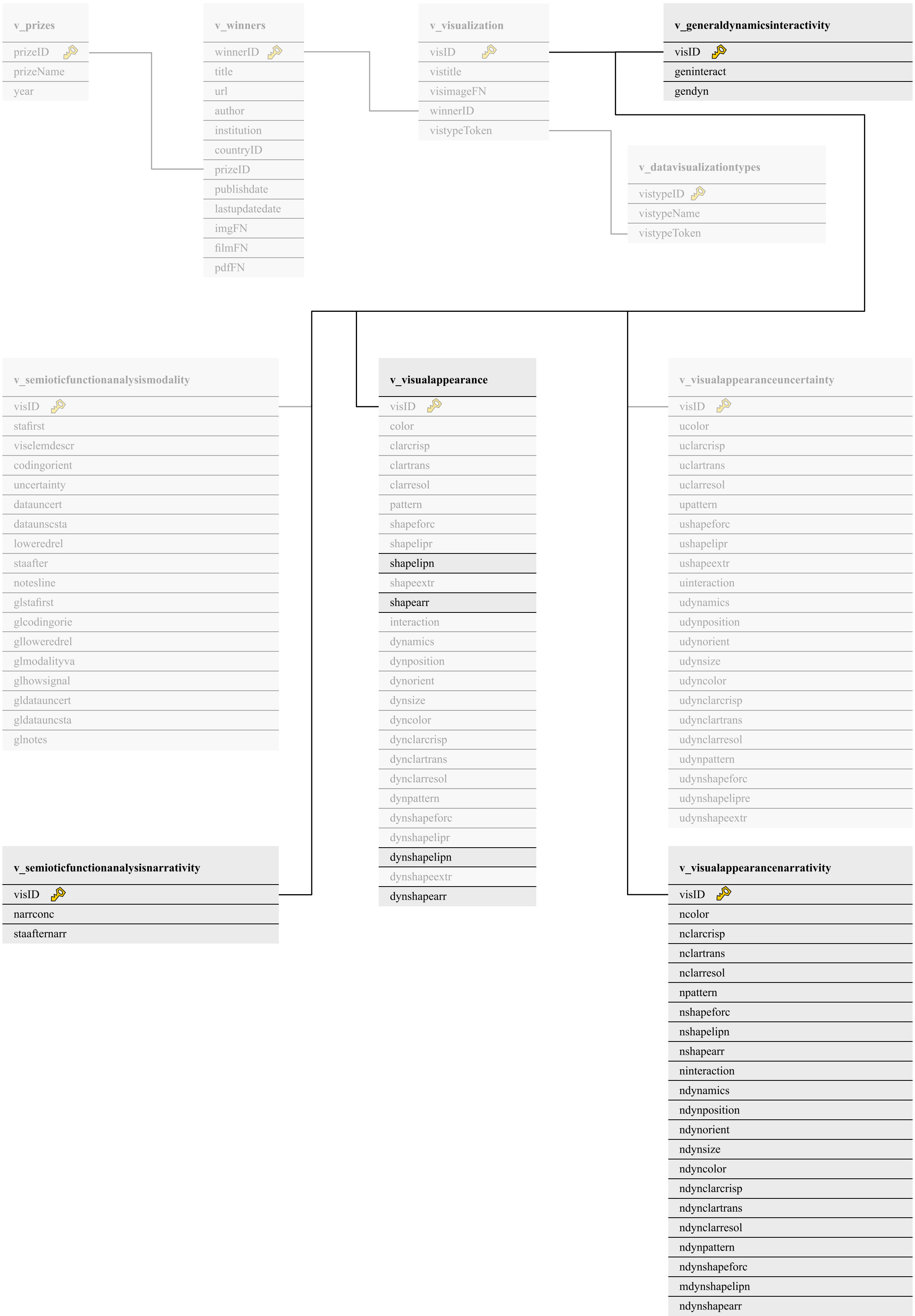
Note. The additional part is marked with full opacity and represents the tables filled out during the corpus analysis focusing on modality.

For the corpus analysis focusing on narrativity, the database was extended once again after these codings were finished. Figure 14 shows the changes in this last extension step. Three new empty tables were prepared, and one existing one was extended. The three new empty tables were prepared for the following: one for inserting the codings of the semiotic analysis (saying whether the line signals a narrative or conceptual representational structure), one for inserting the narrativity-marking visual variables of the lines in focus (if applicable), and one for inserting the general information about whether in the whole DV any dynamic or interactive features were used. The table containing the description of the visual appearance of the focused lines, that was already set up for the corpus analysis focusing on modality, was reused and extended. The extension was necessary because, in this corpus analysis, some other manifestation categories were used for the visual variables (see Table 2 in Articles II and III). Figure 15 presents the final structure of the database.

The coding results were then imported into the database, resulting in the final stage of the database. Having both the metadata of the DVs and the corpus analysis results in the relational database made it possible to perform the data analysis using SQL commands. This was done separately for each corpus analysis. To be sure that the codings were consistent, the codings were regularly checked by looking at the screenshots, screencasts and PDFs. If any irregularities were found, the codings saved in the database were adjusted. Subsequently, the SQL commands for the analysis were executed again. All SQL commands and the results were saved in two data analysis diary documents – one for each corpus analysis. The results (frequency counts of whether and how modality and narrativity are signalled by the focused connecting lines of the DVs in this corpus) and the conclusions are presented in Articles II and III.

Figure 14

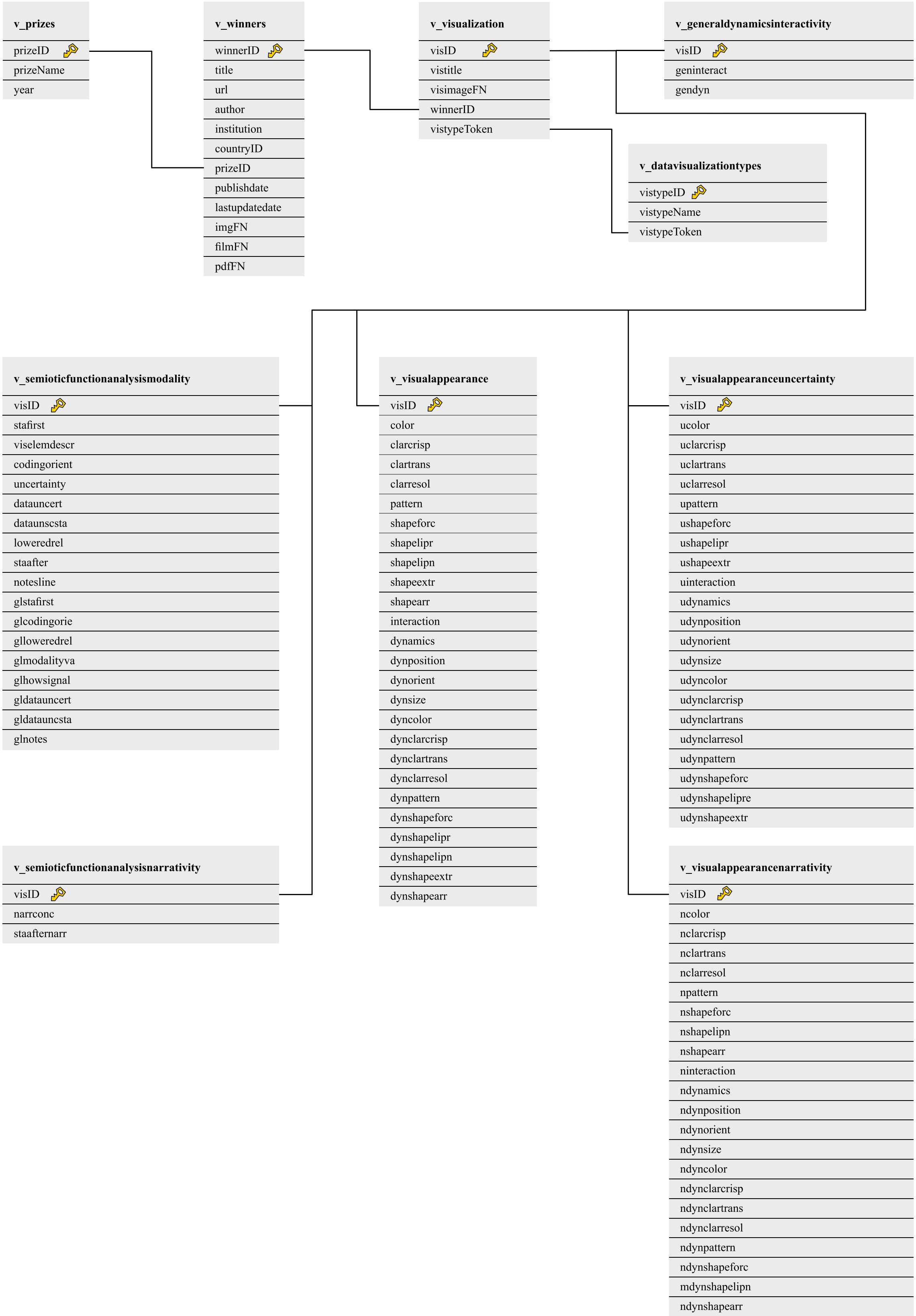
Last Extension Step of the Relational Database of Figure 13



Note. The additional part is marked with full opacity, and represents the tables or columns filled out during the corpus analysis focusing on narrativity.

Figure 15

Final Database Structure



8 Ethical Considerations

Within this dissertation, the following ethical aspects have required special attention:

(1) the distinction between public and private content,

(2) the citation of DVs accessible online and the copyright to screenshots taken from these websites,

(3) the data sharing possibilities of the corpus and the research results.

I have strictly followed the Guidelines for Research Ethics in the Social Sciences, Humanities, Law and Theology (NESH, 2016) and the Guide to Internet Research (NESH, 2019) by The National Committee for Research Ethics in the Social Sciences and the Humanities (NESH). When working with DVs stemming from websites as the central research material, it is sometimes helpful to use screenshots of them in the resulting publications. In this way, visual examples can be used to complement verbal explanations, that would otherwise stay abstract and difficult to understand. All DVs of the corpus created for this dissertation as well as all those that were cited or reprinted in the three articles are publicly available and do not contain any private information, which would be the case if the DVs had derived from personal blogs or the like. Apart from that, another reassuring factor concerning at least the Kantar Awards can be mentioned: in the rules of these awards it is stated that the submitters are responsible for ensuring that the content of their visualizations does not reveal any private information of third parties (The Information is Beautiful Awards Ltd, 2020). The producers of the DVs are willing to make their work more popular, as this is the logical consequence of the press work done around such awards. Therefore, the general question of whether the DVs can be used for research without any consent from the parties covered (NESH, 2019, p. 9), is superfluous.

However, the issue of authorship and copyright has to be considered. “Image producers have the right to benefit from their creations” (Pauwels, 2011a, p. 17). Therefore, if copyrighted images are reproduced in publications, written permission to reprint the content needs to be requested from the copyright holder

(American Psychological Association, 2010, p. 173). This was done for each reprinted screenshot of the copyrighted DVs, through contacting the copyright holders and obtaining written permission and licences if needed. The visual material is reproduced unchanged, precise references to the sources are provided, and the names of the creators are given when so is desired. The same "good citation practice" (NESH, 2016, p. 28) was applied to verbal-only mentions of specific DVs.

The corpus that I built up for this dissertation, as well as the results, could also be useful for other researchers working in the same field. Therefore, I investigated the possibilities for sharing these data. As "sharing of data is important for verification and re-use of research material" (NESH, 2019, p. 17), I ideally would have published the whole database (including the metadata of all DVs and the results of both corpus analyses) as well as the screenshots, screencasts and PDFs that show the DVs themselves. However, these visual materials could not be published due to the ownership and publication rights connected to the DVs. And having a database referring to images or videos that can't be reached is only of little use. Therefore, I have refrained from publishing the database as such. Instead, I decided to publish the award winner lists of the relevant years (or at least the relevant parts) and the final DV list of the corpus including the URLs to the original websites in the appendix to this dissertation. Moreover, I encourage the readers of Articles II and III to contact me through a little note in case they need any detailed research results. In this way, the legal framework can be examined for the specific upcoming case. Thus, the corpus database and all screenshots, screencasts and PDFs are only stored locally and not accessible to anybody else.

As I strictly followed the laws and regulations concerning access to the corpus and the raw research data, it was not possible to integrate the database as the backend of a website with an interface that makes it possible to search for images of DVs with a specific kind of lines. Such a queryable database showing research results was, for example, developed and published by Hill (n.d.). This design research tool, as he calls it, makes it possible to search through works of fine art that convey uncertainty by applying several filters. Making my corpus and the research results accessible in a similar way would have been a highly valuable step not only for other researchers but also for practitioners in the DV domain. But to do that, I would have had to obtain the consent of the copyright holders of each DV in the entire corpus to republish their content. As this was not feasible, this way of research dissemination was not possible.

9 Summaries of the Articles

This chapter includes summaries of the three articles presented in Chapter 10.

9.1 Summary of Article I
What a Line Can Say: Investigating the Semiotic Potential of the Connecting Line in Data Visualizations

The aim of this article is to investigate what semiotic functions a connecting line in a DV can have, in addition to the basic function of indicating a connection between two elements. The article presents the theoretical basis for this dissertation and shows how key concepts from social semiotic theory can be used to analyse the semiotic potential of connecting lines in DVs. The result of this article is the identification of four semiotic functions connecting lines can have, which are demonstrated with exemplary DVs. First, connecting lines can signal how certain a connection is (as a modality marker). The presented example is a map showing the future path of a hurricane with a dashed line, where the interruptions signal uncertainty. Second, connecting lines can indicate either a narrative or a conceptual claim. If the line indicates a direction (such as an arrow), it may serve as a vector and thus point to a narrative claim. Third, connecting lines can, often in combination with other semiotic elements, contribute to the construction of cohesion, creating a textual whole. Depending on the configuration of lines and their connected elements, for example, a sequence, a network, or a hierarchy may be indicated. Fourth, connecting lines can regulate the reader's position to the represented objects. The frame size and the perspective used for the DV, and thus also for the connecting line(s) both have an effect on the relation between the reader and what is represented. The article concludes with a call for further empirical, corpus-based research to investigate if and how the four identified semiotic functions are realised in current DVs and whether any graphical conventions can be observed.

9.2 Summary of Article II Modality and Uncertainty in Data Visualizations: A Corpus Approach to the Use of Connecting Lines

This article concentrates on the first semiotic function of connecting lines identified in Article I, namely the indication of modality and uncertainty. It (1) contains a theoretical discussion concerning the relation between modality and uncertainty, (2) introduces an analysis method for these concepts on different levels of DVs, and (3) presents a corpus analysis. As a result of the theoretical discussion, I found that the concepts of uncertainty (as it is discussed in the practice field of DV) and modality (as it is discussed by linguists and semioticians) are intertwined, the former being a wider concept than the latter. These findings were used to set up the analysis method for the corpus analysis. The corpus was specifically developed for this dissertation, and contains 163 current, publicly available digital DVs that were awarded between 2015 and 2018. Each DV in the corpus has one or more connecting lines in the leading role of communicating the DV's meaning. The corpus analysis was designed to investigate what graphical variations and conventions can be observed, concerning the visual expression of modality (in terms of lowered probability and reliability) through connecting lines. The findings reveal that several visual line characteristics are used to signal lowered probability and reliability. Lowered probability in this corpus is most often signalled by patterns that lead to interruptions in the continuous shape of the line. The same characteristics are rarely used for other purposes, pointing to a graphical convention applied to signal lowered probability through line interruptions. The study did not reveal any conventions concerning the signalling of lowered reliability through connecting lines. A more varied and unsystematic use of line characteristics was observed for that purpose.

9.3 Summary of Article III Arrows and Their Modern Versions: Narrativity Signalled by Lines in Data Visualizations

This article concentrates on the second semiotic function of connecting lines identified in Article I, that is, the indication of narrative and conceptual claims, respectively. It presents a corpus analysis based on the social semiotic concept of representational structures. The corpus analysed for Article II was reused in this study. This time, I explored what visual line characteristics are applied in the corpus to indicate direction, and thus to signal a narrative claim. The findings reveal that several visual line characteristics are used for this purpose, sometimes even in combination (e.g., a line that is growing and also has an arrowhead). Viewed individually, dynamic features (above all size changes, i.e., lines growing in one direction) and arrowheads or similar graphic elements indicating direction (like the icon of a plane) are most prevalent. These results point to a graphic convention to signal narrativity through these two groups of line characteristics. In conclusion, the visualization practices observed in this corpus appear to be influenced both by the long tradition of analogue visualization and the advent of digital production techniques and output devices.

10 Articles

10.1 Article I: What a Line Can Say: Investigating the Semiotic Potential of the Connecting Line in Data Visualizations

Published here with double pagination and with the layout from Amsterdam University Press.

Reference

Lechner, V. E. (2020). What a line can say: Investigating the semiotic potential of the connecting line in data visualizations. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society* (pp. 329–346). Amsterdam University Press. http://oapen.org/search?identifier=1007903

20. What a line can say: Investigating the semiotic potential of the connecting line in data visualizations

Verena Elisabeth Lechner

Abstract
The line is a graphical element widely used in data visualizations, its purpose often being to signal a connection between other visual elements. Based on social semiotic theory, this article investigates what semiotic functions connecting lines can have and how these functions can be related to variations in form. The results show that, in addition to the basic function of connecting elements, such lines can also indicate the level of certainty, direct the viewer to read the information either as a narrative or a conceptual claim, indicate patterns of cohesion, and regulate the viewer's position. These findings allow for further empirical research on the formation of visual conventions.

Keywords: Visual variables; Relation; Link; Metafunction; Modality; Arrow

Introduction

New digital forms of data visualization, as they appear for example on online newspaper pages and the webpages of organizations, companies, and private persons, offer the possibility to make data accessible for specialists as well as the broad public. The particular ways in which such graphical forms make meaning to the readers contribute to their social power in society, as Krippendorff states: 'We do not react to the physical properties of things, but act on what they mean to us' (1998, pp. 01_8).

A central task of various types of data visualizations (such as network visualizations, route maps, and others) is to show how different visual

Engebretsen, M. and H. Kennedy (eds.), *Data Visualization in Society*. Amsterdam: Amsterdam University Press, 2020
DOI 10.5117/9789463722902_CH20

elements are connected. Such connections are often represented by lines, which are in the focus of this chapter.[1] This chapter deals with the meaning potential of this basic element of the language of graphics, and it is thus concerned with the detail level of data visualizations. Although several examples will be given, this chapter stays on the theoretical level, whilst also opening up for practical investigations.

The use of graphical lines to represent connections between elements is an old technique (Bertin, 2011; Brinton, 1939) but still ubiquitous in current data visualizations. Nevertheless, their potential for making meaning has changed in the course of time, as the options for visual representation have increased, especially through the advent of digital production techniques and output devices. Today, the fact that connections are often represented by lines in data visualizations can be observed not only in the number of published data visualizations that have this characteristic (see Figure 20.1 as an example). It can also be observed within several tools available for digital creation of data visualizations (e.g. D3.js, Tableau, R). However, what functions these connecting lines have, and what effects different visual appearances of the lines have on their meaning potential, are issues which have not been widely researched.[2]

This chapter asks: What semiotic functions can a connecting line in a data visualization have, in addition to the basic function of indicating a connection between two visual elements?

Raising this question is necessary, not only for the scientific community, in order to generate more knowledge about visual language, but also for practitioners on the production side, in order to raise their awareness on how to communicate as nuanced and clearly as possible with their readers. Thus, a central aim for the chapter is to offer a language for discussing the functions of smaller elements within data visualizations. This is fundamental, because the meaning potential of visual elements informs important decisions in the design process.[3]

1 Besides the connective function, lines may have many other functions, e.g. to define contours, to separate elements, to lead the eye, or to function as a base element for textures and patterns (Poulin, 2012, p. 29).

2 It is possible to find literature about meaning potentials for different visual appearances of lines in general (cf. Habermann, 2015, p. 649; Horn, 1998, pp. 147-148; Ware, 2013, p. 225). However, the suggested meaning potentials are not directly transferable to connecting lines within the context of digital data visualizations.

3 It has to be noted here that other reasons than the desired function (e.g. the circumstances in the data visualization developer tools or aesthetical reasons) can influence the designer's decisions.

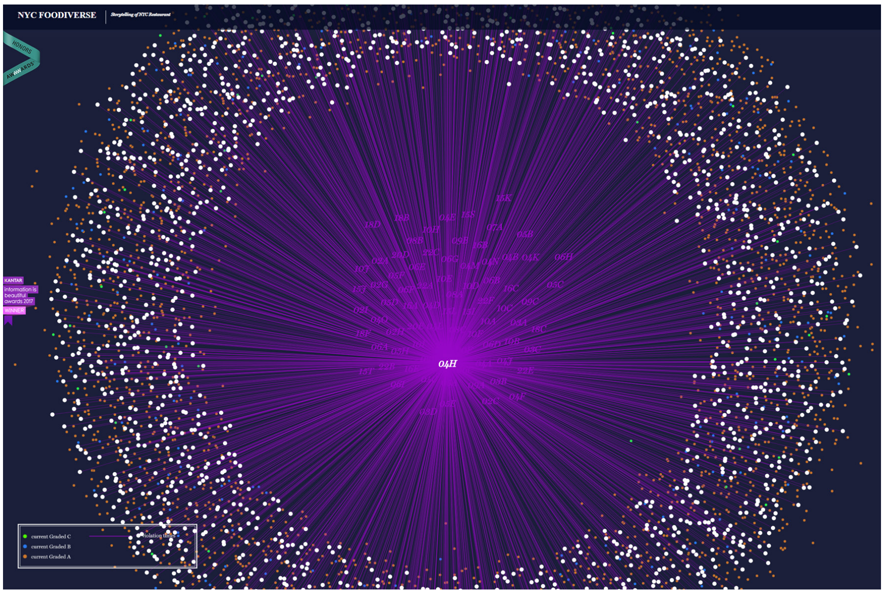


Figure 20.1. Example of a data visualization using lines to represent the connections between sanitary problems (central group of purple letter and number codes) and the restaurants in Manhattan, NYC they occurred in, represented as dots in the outer circle. From 'NYC FOODIVERSE' by W. Su, 2017 (http://nycfoodiverse.com). Copyright 2017 by W. Su. Reprinted with permission.

The approach chosen to answer the question will be presented in three steps. I will start with the central element, the connecting line itself, and describe especially *how* the line connects. In the second step, the elements that are being connected are discussed, in other words, *what* the line connects. In the last part, I describe how the connecting line is *integrated into the whole* data visualization and thus contributes to the creation of larger structures of information. However, before these steps can be taken, I need to make some terminological clarifications and outline the theoretical framework of the discussion.

Terminological considerations

The two terms that identify the object of study, namely *line* and *connection*, are used in many contexts and with a number of different meanings. Within the field of graphics, the French cartographer Jacques Bertin identified the line as one of three basic elements in the language of graphics, together with the point, and the area (2011, p. 271). Decades earlier, the Russian painter Wassily Kandinsky named the point and the line as the two elements that 'constitute the conclusive material for an independent kind of painting—graphic' (1947, p. 20). Concerning the formal characteristics of the line, he stated that the line is a product of the moving point (1947, p. 57). This movement is what provides the line with its main formal characteristics. Wucius Wong points out that the breadth of a line is 'extremely narrow' and 'its length is quite prominent' (1993, p. 45). However, as the gestalt laws state, single elements grouped in a certain way, or incomplete lines, can also be perceived as lines (Lauesen, 2005, pp. 68-69). Thus, in this chapter the term *line* refers to all kinds of visible lines, including incomplete lines and arrangements of visual elements that can be perceived as lines.

Returning to the central term *connecting line*, this refers to lines which have the basic function of establishing or indicating a connection. This means that it must be possible to identify the two parts of a connection: the single *components* and the *relationship*, as Bertin describes it (2011, p. 271), or the *nodes* and the *connector*, to mention the terms used by Engelhardt (2002, p. 40). The terminology used to describe the phenomenon of connections varies, since many synonyms exist, such as *relation*, *relationship*, *link,* and *tie* (Fergusson, 1992, p. 88). The terms are used slightly differently in various corners of the field (Brinton, 1939, pp. 43-72; Engelhardt, 2002, p. 40; Richards, 1984, p. 3/21; Ware, 2013, pp. 221-226). However, Kress and van Leeuwen, who look at graphics from a linguistic perspective similar to this chapter, use the compound term *connecting line* (2006, p. 59), which I have chosen to adopt.

Theoretical framework

Data visualizations, like other types of semiotic material, offer a specific way to communicate meaning. In order to make meaning out of a data visualization, the reader has to apply certain rules, which help him or her to decode what the producer of the data visualization wanted to communicate. On the other side, the producer of the data visualization most likely also had similar rules in mind when deciding on this specific form of visual representation. The meaning potential carried by the visual forms through the application of such shared rules defines the social function of the forms.

In the understanding of how these rules evolve, traditional semiotics and social semiotics differ on certain central aspects. In the former case, rules are seen as predefined and more or less consistent, and the communicating persons have to learn these rules before they are able to apply them, either in production or in interpretation (Hodge & Kress, 1988, p. 12). In contrast to that, in social semiotics, as van Leeuwen (2005, pp. 47-48) describes, people actively participating in social activities are seen as the ones who generate these rules—on the basis of certain culturally shared codes. He further explains that semiosis is an ongoing process, where the sign users themselves have the power to influence and change the rules. This again implies that those rules are seen to be rather unstable and to a high degree dependent on the social situation.

Returning to the case of data visualizations, which are often produced for a large and diverse target group, we can assume that some rules exist, connecting forms to meanings. But they might be somewhat different from how they were years ago and might also be dependent on the social context. M. A. K. Halliday laid the theoretical basis for seeing text as 'a sociological event, a semiotic encounter through which the meanings that constitute the social system are exchanged' (1978, p. 139, emphasis deleted). However, in the centre of his research stands verbal language. In their seminal work *Reading Images*: *The Grammar of Visual Design* (2006, p. 2), Theo van Leeuwen and Gunther Kress state that visual structures, just like linguistic structures, invite a particular interpretation, that is formed by experience and social interaction.[4]

From these theoretical abstractions, we can conclude that the process of meaning-making in contexts involving data visualization is a process where certain culturally formed, relatively stable codes and conventions interplay with a set of more unstable, situated rules concerning the exact meaning of the visual forms displayed. This interplay also defines the meaning of connecting lines, and calls for empirical research to investigate which semiotic functions are conventionalized and which are not.

Halliday defined three universal functions in verbal language, also known as 'metafunction[s]' (2004, p. 30), understood as different aspects of the meaning potentials of a clause. Any clause, any verbal utterance, carries all three metafunctions simultaneously: the 'ideational' (what is said about the world), the 'interpersonal' (how social relations between the participants are constructed), and the 'textual metafunction' (how the parts construct a coherent whole) (pp. 30-31). As Kress and van Leeuwen adapted Halliday's concept of social semiotics to other semiotic modes, they also applied the concept of these metafunctions to the analysis of visual expressions (2006, p. 13), and during the last decades, their work has been adapted by many other researchers. Yet, for every new social

4 It should be underlined that they also reflected critically on this comparison between visual and verbal structures and pointed out that this similarity easily can be overemphasized (see also Kress & van Leeuwen, 2006, p. 76).

semiotic study, the systems of meaning making have to be defined again in order to make a systematic analysis possible. This is the case because different types of visual material offer different semiotic choices and need to be interpreted in different ways. Thus, it is also necessary to define the system of choices activated in the kind of visual material investigated in this book. To develop the basis for that, focusing on connecting lines, is the contribution of this chapter.

Towards an analytical procedure

Functions related to the connecting line itself

Having given a brief insight into terminology and the theoretical framework for the study, I will now present a proposed method for analysing types and functions of connecting lines. Starting with the central element, the connecting line itself, its main function first has to be pointed out. This function is already implied in the word *connecting,* and therefore works as a basic selection criterion for the kind of lines that are to be investigated. As Clive James Richards noted, a line can have a verb-like function, and in the verbal translation of a figure showing two letters with a line in between, he states: 'A is connected to B' (1984, p. 3/21). In the words of Halliday, this corresponds to the ideational meaning, which says something about a process, or 'goings-on' (2004, p. 170). More precisely, the line represents the process itself. Secondly, the connecting line might also say something about the associated circumstances, e.g. whether the connection is strong or weak. The third component of a process—the participants—is determined at the ends of the line, showing what is connected.

Following the proposed analytical structure, we can summarize that a line *connects* certain objects in a *certain manner*. It represents a connection in a specific way, and it can, among other things, also point to the certainty of this connection.[5] In verbal language we have several alternatives to express the certainty of a piece of information, building up the modality system of the language in question (Halliday, 2004, p. 147). This system offers means to express the level of certainty that the speaker wants to give a certain claim—e.g. choosing between *This is probably true* and *This is certainly true*.

5 In his visual grammar of relationship representations Colin Ware named, for example, the strength of a connection as a characteristic that could possibly be expressed by different line weights (2013, p. 225).

Hodge and Kress assumed that modality markers can also appear in other kinds of media, although they considered them to be not so clearly articulated as the ones in verbal language (1988, pp. 121-122). Kress and van Leeuwen further investigated modality in terms of visual communication, including examples like illustrations, photographs, and pieces of art (2006, pp. 159-180). Because modality markers are closely related to the social participants in the communication process and are used to build shared truths, they consider modality as a phenomenon to be categorized as part of the interpersonal rather than the ideational metafunction (2006, pp. 159-160).

Van Leeuwen notes that different coding systems can have different kinds of *coding orientations*—like the 'naturalistic', 'technological', 'sensory', and 'abstract' orientations (van Leeuwen, 2005, pp. 168-170). As he understands it, these different orientations mean that the scales of modality may have different types of markers, or criteria for what is regarded true and realistic, and what is not. A line graph may show little details of the background, compared to a photograph, but that does not indicate that what is represented in the diagram is not true (p. 167). Within the abstract coding system, 'visual truth is abstract truth' (p. 168). 'The more an image [...] represents the general pattern underlying superficially different specific instances, the higher its modality from the point of view of the abstract truth. This is expressed by reduced articulation', he explains further (p. 168). This means that if a data visualization is seen as being part of this coding system, an abstract way of visualizing data conveys an impression of truth.

Visualizations of past and future paths of hurricanes can serve as an example of data visualizations that often contain a degree of uncertainty, such as the data visualization *Irma is following a well-worn path* (Dottle, King, & Koeze, 2017).[6] Here, the future, uncertain path of Hurricane Irma is shown as a dashed line, within a shape surrounded by another dashed line. The interruption of the lines therefore serves as the modality marker within this example. Here, the lack of sufficient data is signalled visually in the data visualization. In other cases, the data available may be precise and sufficient, but for certain reasons (like privacy protection) the visualizations are intentionally made to look imprecise, through the application of uncertainty markers (Dasgupta, Chen, & Kosara, 2012, p. 1022). Yet, although it might appear clear in the example of Hurricane Irma's path, there does not exist any general and recognized description of how certainty is expressed in graphical material through different forms of connecting lines. Sometimes, like in *Musicmap* (Crauwels, 2016, see https://musicmap.

6 See https://fivethirtyeight.com/features/what-lies-in-irmas-path/.

info/), dashed lines are used for purely compositional reasons, in order to separate them from other lines. The same can be said about other visual variables shaping the physical appearance of the line (like colour, shape, etc.). How these visual variables indicate specific types of connections, whether through convention or explicit explanation, is an issue that calls for extensive empirical investigation.

Functions related *to what* the line connects

As stated earlier, to recognize a connecting line as such, it must be possible for the reader to identify not only the line itself, but also the connected components. When looking at different types of data visualizations, it becomes apparent that sometimes lines are used to connect two different elements (as in network diagrams, see Figure 20.1), whereas in other types the lines connect two different states of the same element (as in route maps). In either case, the function type in question belongs to the ideational type, saying something about states in the world.

In order to trace the graphical lines to the natural, non-digital world, a relevant source is Tim Ingold (2007, pp. 41-43), who writes about lines from an ethnological viewpoint. He divides lines into five groups, two of which are called 'threads' and 'traces' (p. 41).[7] In his account, *threads* (such as a washing line, an electrical circuit, a tightrope, or a skein of yarn) seem to correspond to the former group of lines, connecting two different elements. *Traces* (such as a scratched line or the slime trail of a snail), on the other hand, relate to the connecting lines of the second type, connecting different states of the same element.

The way that connecting lines relate the connected components to each other forms their *representational structure*, a concept investigated by Kress and van Leeuwen (2006) within many different visual media. The two main categories into which they divide their investigated material are 'narrative structures' and 'conceptual structures' (p. 79). What is represented in narrative structures are 'unfolding actions and events, processes of change, transitory spatial arrangements' (p. 79). They contain 'vectors' (p. 59) which show a direction. In data visualizations, connecting lines can work as vectors when the direction is made explicit, e.g. by an arrowhead or a tapering body, as in Figure 20.3. Conceptual structures, on the other hand, have no vectors, and represent 'participants in terms of their more generalized

7 The three other groups he calls 'cuts, cracks and creases' (2007, p. 44), 'ghostly lines' (p. 47), and 'lines that don't fit' (p. 50).

and more or less stable and timeless essence, in terms of class, or structure or meaning' (p. 79). In other words, narrative structures always contain a form of action, whereas conceptual structures describe a phenomenon in a certain state. Although originating in different disciplines, it seems obvious that Ingold's traces to a certain degree correspond to the concept of narrative structures, while threads correspond more closely to the concept of conceptual structures.[8] Based on both sources, I suggest that connecting lines have the function of directing the viewer to read the information either as a narrative or as a conceptual claim, and that the way they do this through their visual appearance in current data visualization design is an issue that calls for both theoretical and empirical investigation.

Functions related to the line as part of larger text units

After having started this investigation on the micro level focusing on the connecting line itself, and then extending it to the connected units, it is now time to have a look at the surrounding context, that is, the wholeness of the data visualization. Some data visualization types, like network visualizations or tree diagrams, traditionally contain many connecting lines, often even lines interconnected with each other. Other data visualization types, like flow maps, might either show only one or a few lines, which are not necessarily interconnected (although they may cross each other). Such examples make it obvious that connecting lines contribute to, and are integrated in a bigger whole, a composition of semiotic elements. This observation is a starting point for analyses that focus on the textual, also called the compositional metafunction. At this level, *cohesion* is a core concept.

Linguists working on the discourse level have a long tradition of describing connectedness in verbal texts (Sanders & Pander Maat, 2006, p. 591). For the English language for example, Halliday and Hasan published their pioneering book *Cohesion in English* already in 1976, of which some main ideas shall be explained in the following (1994, pp. 1-4). According to them, what makes a text be regarded as such, is that it forms a recognizable, coherent unit of meaning. For that, it needs to have meaning relations that combine the single text units. These are called 'cohesive properties' (1994, p. 4). These properties come into action when 'the interpretation of some element in the discourse is dependent on that of another' (1994, p. 4, emphasis deleted). In other words,

8 Both Ingold and Kress and van Leeuwen emphasize that it is not always possible to distinguish their research material exhaustively with their categories. They rather see them as a tool for describing what is represented (Ingold, 2007, p. 50; Kress & van Leeuwen, 2006, p. 86).

a cohesive text contains connections between the single elements of the text, which help the reader to understand the meaning of the entire text.

However, the phenomenon of cohesion is manifest not only in verbal texts. Theo van Leeuwen has investigated forms of cohesion in the field of multimodal texts (2005, pp. 179-268), the principles of which will be shortly introduced here. He lists four ways of constructing cohesion, namely: 'composition', 'rhythm', 'information linking', and 'dialogue' (p. 179). *Composition,* he explains, works with the placement of elements in space. For van Leeuwen, whether an element is placed at the bottom or the top of a page, to the left or to the right, in the centre or in the margin, has an impact on its meaning potential. This impact is often based on metaphors from the physical world. Composition, as he continues, is the spatial equivalent to *rhythm.* This in turn is formed by a transition between two opposing states repeated in the dimension of time, such as soft and loud, fast and slow, big and small, and so on. *Information linking* has to do with the ways that one piece of information can be related to another piece. A *dialogic* structure, as the fourth form of cohesion he suggests, appears when more than one voice is perceived either simultaneously or sequentially, like in a spoken dialogue, or in a film track, where the flow of images and the music track may establish a dialogue (pp. 179-268).

When investigating the cohesive functions of connecting lines in digital data visualization, I propose to focus on *composition* and *information linking*, for a number of reasons.

Composition is built up by the spatial arrangement of the constituting elements in the visual object. In the case of data visualizations containing several connecting lines, the conscious placement of such connections, which include the connecting line as well as the connected elements, can be used to imply a certain meaning potential. On a macro level, the composition of these elements can build up specific types of data visualizations and help to define the roles of the connected elements. The configuration of connecting lines can indicate, for example, sequences, hierarchies, or networks, offering very different roles for each of the involved visual elements (see Figure 20.2).

In some types of interactive data visualization, the user is enabled to change the placement of the nodes and its connections manually. The visualizations in the report *Panama papers—the power players* (ICIJ, 2017), for instance, offer this affordance. (Figure 20.3 is a static screenshot of one of them.)[9]

Information linking, as another way to construct cohesion (van Leeuwen, 2005, pp. 219-247) shall here be discussed in further detail. Borrowing the

9 See https://panamapapers.icij.org/the_power_players/

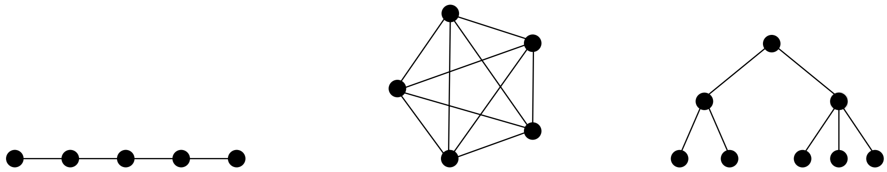


Figure 20.2. Three exemplary compositions of connecting lines and their connected elements. Illustration by V. E. Lechner.

linguistic concept of conjunctions from cohesion within verbal texts (as described in Halliday & Hasan, 1994, pp. 336-338) van Leeuwen states that links are 'temporal, logical or additive' (2005, p. 222). He further explains that if a temporal link occurs, this points to the fact that the two single pieces of information happen either at different points in time or in parallel. A logical link highlights that one of the information pieces 'gives a reason for, a condition of, or a comparison with the information in the first item' (p. 223). If it is not a temporal or logical link, yet the one item adds information to that given by the other, he concludes that additive linking takes place.[10]

Whereas in verbal language, the different linking types can be determined because of explicitly used words like conjunctions, in data visualizations, such 'cohesive tie[s]' (1994, p. 329), as Halliday and Hasan call them, might not always be so obvious. But as van Leeuwen shows with examples of multimodal, non-linear texts (2005, pp. 226-247), they do exist and have to be found by the reader to form the storyline. For this process, the surrounding context plays an important role, as it might influence which linking type might be the most relevant in a specific data visualization.

In Figure 20.3, we see a visualization where the connecting lines indicate a combination of additive and logical linking. The connections around Sigmundur Davíð Gunnlaugsson (former prime minister of Iceland), could be verbally translated to: *Gunnlaugsson is registered in address X and is a shareholder of company Y. Similarly, his wife is also registered at the same address* and *is also a shareholder of the same company* (*which was registered by a consulting firm*). The cohesion markers *and* and *similarly* point to additive respectively logical linking. The example shows that translating the data visualization into text may help to detect the ways in which lines are used as cohesion markers.

As the previous examples show, *composition* and *information linking* are relevant when studying cohesion formed by graphical lines in data

10 Examples for cohesion markers within verbal text: *then, next, meanwhile* (temporal linking); *because, for that reason, otherwise* (logical linking); *and, or* (additive linking). These examples can be found in van Leeuwen (2005, pp. 222-224) as well as in Halliday & Hasan (1994, pp. 336-338).

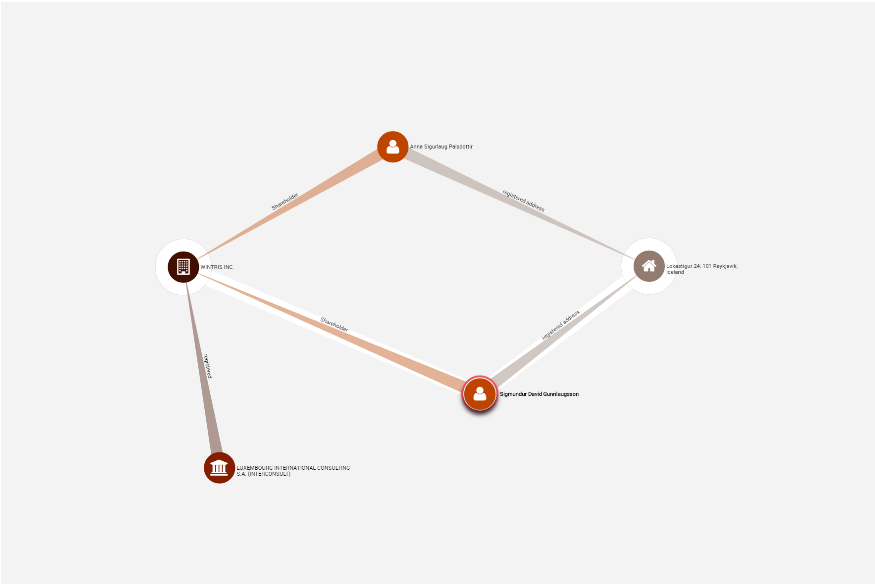


Figure 20.3. Visualization of the connections related to Sigmundur Davíð Gunnlaugsson. From 'Panama Papers—The Power Players' by The International Consortium of Investigative Journalists, 2017 (https://www.icij.org/investigations/panama-papers/the-power-players/). Copyright 2017 by ICIJ. Reprinted with permission.

visualizations. However, *rhythm* and *dialogue* can also form cohesion in such forms of textual expressions.

Rhythm can be established through visual representations of processes occurring over time, analogue to the rhythmic structures perceived in music (van Leeuwen, 2005, p. 182). In data visualizations, this form of cohesive structure can be perceived through observations of visual repetitions and certain patterns of such. It can either be shown by an animation or by presenting these repetitions in a linear sequence in a static presentation. Connecting lines can play a role in such a rhythmic organization of a data visualization. A visualization in the news site of *The Washington Post* shows flight patterns after the Brussels attacks on March 22, 2016 (Muyskens, 2016)[11]. In the early morning, all planes are flying directly to Brussels airport and build up a regular pattern of moving lines. The lines connect the planes to geographical points. But suddenly their flight routes change and the lines representing them develop an irregularity because the planes turn back before entering the airport. This example points to the fact that it is not always the coherent pattern itself that forms the most interesting feature

11 https://www.washingtonpost.com/news/wonk/wp/2016/03/23/watch-what-happened-to-flight-patterns-in-the-moments-after-brussels-attacks/

of a (visual) text, but rather the instances of violation of the pattern. The rhythm is disrupted, and attention is attracted.

In many publicly available data visualizations, a *dialogue* between different semiotic modes appears, between the visual forms, verbal elements, numbers, and sometimes dynamic modes like music or speech. How this dialogue is organized in time and space is interesting to analyse, but not so relevant for the investigation of connecting lines.

Exploring the functions of the line in relation to the total data visualization, we may also look at two aspects of interpersonal meaning, namely the ways in which the reader's position is regulated through *frame size* and *perspective*. According to Kress and van Leeuwen (2006, pp. 124-129) social distance between a human represented on an image (e.g. a photo) and the viewer of the image is managed by different frame sizes, 'close distance', 'middle distance', and 'long distance' (pp. 125-126). They further suggest that a similar set of relations is possible between the viewer and depicted non-human elements. In data visualizations, the connecting lines can be presented from a very far distant position (showing much of the surrounding context) or from very near, just as if the viewer could touch them. In interactive data visualizations, where the user is able to zoom, this could even be changed manually (as in *Musicmap* (Crauwels, 2016)).[12] Such interactive mechanisms offer the reader a position as an active participant in the communication, being able to choose the frame size and therefore also the position from where the data visualization is observed.

Besides frame size, the chosen perspective also influences the relation between the viewer and the represented objects. Data visualizations are often presented in a direct frontal or a top-down angle, which adds to their aura of objectivity, whereas other angles rather indicate subjectivity (Kress & van Leeuwen, 2006, pp. 135-151). Placing the connecting line in a two- or three-dimensional space makes it possible for the connecting line to indicate perspective. In data visualizations, a top-down angle is e.g. often used for route maps, where the movement of certain objects is shown on a map. One such example is the first visualization in the news article *Bussed out: How America moves its homeless* (Outside in America team, Bremer, & Wu, 2017), showing the route of a homeless person relocating in the US.[13] In another visualization from the same article, the same kind of geographical movements are presented as curved lines viewed from a frontal perspective

12 see http://www.musicmap.info

13 see http://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study

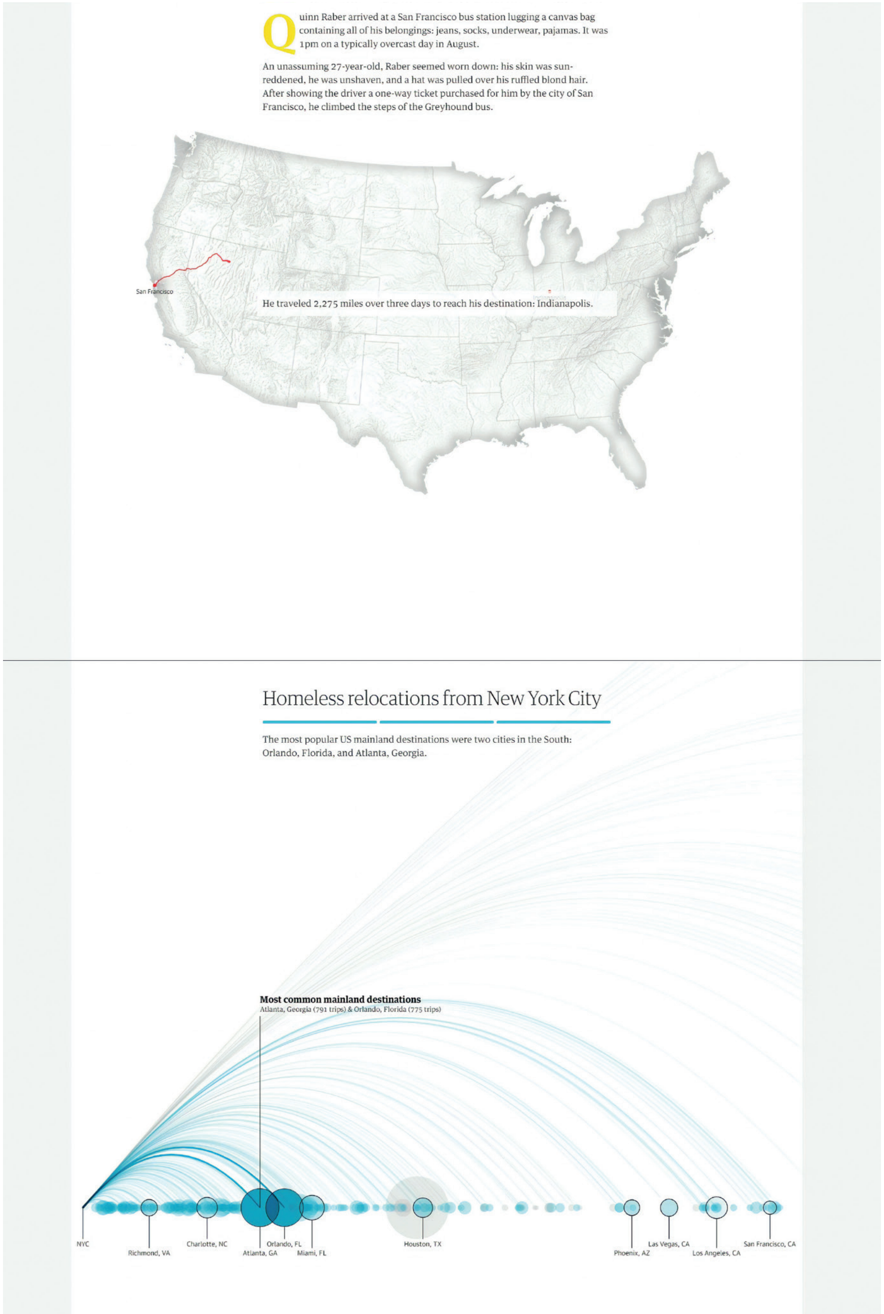


Figure 20.4. Two visualizations of spatial movement, using a top-down angle in a route map (upper picture) and a frontal perspective in an arc diagram (lower picture). From 'Bussed out: How America moves its homeless' by Outside in America team, N. Bremer and S. Wu, 2017 (http://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study). Copyright 2017 by *The Guardian*. Reprinted with permission.

(see Figure 20.4). If we apply Kress and van Leeuwen's principle, both perspectives indicate an objective representation of reality. Similar to frame size, some interactive data visualizations, such as Kim Albrecht's *Cosmic Web* (n.d.) also include possibilities to manually change the perspective.[14]

Conclusion

Beyond the basic function of connecting, four semiotic functions of connecting lines in the context of data visualizations have been identified and described in this chapter.[15] Connecting lines can potentially be used:

1) To indicate the level of certainty of a specific connection as a modality marker.
   Both the visual scales of the modality markers as well as what the exact values indicate in a certain context need to be investigated further on a corpus of data visualizations before they can be used for analysing single data visualization examples.
2) To direct the viewer to read the information either as a narrative or as a conceptual claim—how things *develop* or how things *are*.
   It should be possible to identify the two types of representational structures with the help of the surrounding context of the data visualization. What kinds of sub-categorization are possible and reasonable can only be discovered with the help of a corpus of data visualizations.
3) To indicate patterns of cohesion in a data visualization, and to indicate the role of particular objects in the context of the whole.
   Composition and information linking are especially relevant when investigating a data visualization as a cohesive textual unit.
4) To regulate the reader's position, by regulating the physical relation between the viewer and the connecting line(s).
   The concepts of frame size and perspective, such as proposed by Kress and van Leeuwen (2006), are directly applicable. However, the effects on the reception side call for further research.

In order to see if and how these potential functions are realized in current data visualization design, empirical research on larger corpora of data visualizations is demanded. Such studies would also offer insights in the

14 See http://cosmicweb.kimalbrecht.com/viz/#1

15 This list is not meant to be exhaustive, but it may serve as a starting point for further investigations.

evolving process of conventionalization—the forming of rules for making meaning through data visualizations, widely shared on both the production side and the viewer side. The stronger the conventions become, the stronger data visualizations' role will be in society, because they will afford a more nuanced communication. The relevance of these fields of knowledge can be demonstrated with the social impact of the following two examples. If a data visualization showing the potential future path of a hurricane (which might be indicated by connecting lines as in *Irma is following a well-worn path* (Dottle, King, & Koeze, 2017)) is interpreted incorrectly, this might have an impact on whether or not people decide to leave their homes. Similarly, if a data visualization about problems of a disadvantaged group in society (such as *Bussed out: How America moves its homeless* (Outside in America team et al., 2017), where connecting lines indicate the journey of homeless people taking part in a relocation programme) provokes a long social distance instead of compassion, social awareness might not be developed, and action might not be taken.

References

Albrecht, K. (n.d.). *Cosmic web* [Project page]. Retrieved March 31, 2018 from http://cosmicweb.kimalbrecht.com/viz/#1

Bertin, J. (2011). *Semiology of graphics: Diagrams, networks, maps* (W. J. Berg, Trans.). Redlands, CA: ESRI Press.

Brinton, W. C. (1939). *Graphic presentation*. New York: Brinton Associates.

Crauwels, K. (2016). *Musicmap: The genealogy and history of popular music genres*. Version 1.0.1. Retrieved April 21, 2018 from http://www.musicmap.info

Dasgupta, A., Chen, M., & Kosara, R. (2012). Conceptualizing visual uncertainty in parallel coordinates. *Computer Graphics Forum, 31*(3), 1015-1024). https://doi.org/10.1111/j.1467-8659.2012.03094.x

Dottle, R., King, R., & Koeze, E. (2017, September 8). What lies in Irma's path [Blog post]. Retrieved from https://fivethirtyeight.com/features/what-lies-in-irmas-path/

Engelhardt, Y. (2002). *The language of graphics. A framework for the analysis of syntax and meaning in maps, charts and diagrams* (Doctoral thesis, University of Amsterdam). Retrieved from http://dare.uva.nl/search?arno.record.id=105970

Fergusson, R. (Ed.). (1992). *The Penguin dictionary of English synonyms and antonyms* (revised). London: Penguin Books.

Habermann, H. (2015). *Grundlagen der Gestaltung: Industrie-Design / Kommunikations-Design*. Göttingen, Germany: Cuvillier Verlag.

Halliday, M. A. K. (2004). *An introduction to functional grammar* (3rd ed). London: Arnold.

Halliday, M. A. K. (1978). *Language as social semiotic: the social interpretation of language and meaning*. London: Arnold.

Halliday, M. A. K., & Hasan, R. (1994). *Cohesion in English* (13th impression). New York: Longman Group Limited.

Hodge, R. I. V., & Kress, G. (1988). *Social semiotics*. New York: Cornell University Press.

Horn, R. E. (1998). *Visual language: Global communication for the 21st century*. Bainbridge Island, WA: MacroVU.

ICIJ. (2017). Panama Papers—The Power Players [Investigation project page]. Retrieved July 30, 2018 from https://panamapapers.icij.org/the_power_players/

Ingold, T. (2007). *Lines: A Brief History*. Abingdon: Routledge.

Kandinsky, W. (1947 [1926]). *Point and line to plane: Contribution to the analysis of the pictorial elements*. (H. Dearstyne & H. Rebay, Trans.). New York: Solomon R. Guggenheim Foundation. Retrieved from http://archive.org/details/pointlinetoplaneookand

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). London: Routledge.

Krippendorff, K. (1998, December). *Design discourse: a way to redesign design*. Keynote speech presented at the Society for Science of Design Studies, Tokyo, Japan. Retrieved from https://repository.upenn.edu/asc_papers/227

Lauesen, S. (2005). *User interface design: A software engineering perspective*. Harlow: Pearson Education.

Muyskens, J. (2016, March 23). Watch what happened to flight patterns in the moments after Brussels attacks. *The Washington Post.* Retrieved from https://www.washingtonpost.com/news/wonk/wp/2016/03/23/watch-what-happened-to-flight-patterns-in-the-moments-after-brussels-attacks/

Outside in America team, Bremer, N., & Wu, S. (2017, December 20). Bussed out: how America moves its homeless. *The Guardian.* Retrieved from http://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study

Poulin, R. (2012). *The language of graphic design: An illustrated handbook for understanding fundamental design principles*. Beverly, MA: Rockport.

Richards, C. J. (1984). *Diagrammatics: An investigation aimed at providing a theoretical framework for studying diagrams and for establishing a taxonomy of their fundamental modes of graphic organization* (Doctoral thesis). Royal College of Art, London, United Kingdom.

Sanders, T., & Pander Maat, H. (2006). Cohesion and coherence: Linguistic approaches. In: *Encyclopedia of Language & Linguistics* (2nd ed.). (pp. 591-595). https://doi.org/10.1016/B0-08-044854-2/00497-1

Su, W. (2017a). NYC FOODIVERSE. Retrieved July 31, 2018 from https://jiahao01121.github.io/Restaurant_data_viz/second_draft_mar_25/index.html

Su, W. (2017b). NYC FOODIVERSE—Storytelling of NYC Restaurant. Retrieved April 25, 2018 from http://nycfoodiverse.com/

van Leeuwen, T. (2005). *Introducing social semiotics*. London & New York: Routledge.

Ware, C. (2013). *Information visualization: Perception for design* (3rd ed.). Waltham, MA: Morgan Kaufmann.

Wong, W. (1993). *Principles of form and design*. New York: John Wiley & Sons.

About the author

Verena Elisabeth Lechner is currently a PhD student at the University of Agder, Norway. She holds an MA in InterMedia from the University of Applied Sciences Vorarlberg, Austria and is especially interested in the visual appearances and meaning potential of graphic elements within data visualizations.

10.2 Article II: Modality and Uncertainty in Data Visualizations: A Corpus Approach to the Use of Connecting Lines

Published here with double pagination and with the layout from Springer.

Reference

Lechner, V. E. (2020). Modality and uncertainty in data visualizations: A corpus approach to the use of connecting lines. In A.-V. Pietarinen, P. Chapman, L. Bosveld-de Smet, V. Giardino, J. Corter, & S. Linker (Eds.), *Lecture notes in computer science: Vol. 12169. Diagrammatic representation and inference.* Diagrams 2020 (pp. 110–127). Springer. https://doi.org/10.1007/978-3-030-54249-8_9

Modality and Uncertainty in Data Visualizations: A Corpus Approach to the Use of Connecting Lines

Verena Elisabeth Lechner(✉)

University of Agder, 4630 Kristiansand, Norway
verena.lechner@uia.no

Abstract. In data visualizations, connecting lines may have various semiotic functions, including the semiotic potential of indicating modality and uncertainty. The goal of this article is to find out how this semiotic potential is realized in current best practices of data visualizations and what conventions exist for the visual manifestations of these functions. This issue is addressed by using a corpus-based approach and a two-level analysis method within a social semiotic framework. First, the article offers a theoretical discussion on how the concepts of modality and uncertainty interrelate. Second, a method for investigating how these concepts are visualized at different levels is presented. Third, a corpus analysis including 163 award-winning data visualizations is presented. The results indicate the existence of certain conventions for visual modality markers, and thus offer new insights relevant for both design theory and practice.

Keywords: Uncertainty · Modality · Social semiotics · Visual variables · Line · Connection · Reliability · Probability

1 Introduction

Data visualization (further abbreviated as *DV*) is a rapidly developing visual means of communication, strongly influenced by the advent of new digital technologies [1]. The amount of accessible data is greater than ever [2] and the forms of representation are constantly being developed further [3]. Consequently, the desire to express a variety of meanings through DV is increasing, and so are the graphical opportunities to do it. Such 'new modes of production bring with them new affordances' [4], which means that the conventions that connect visual expressions to culturally shared meanings are constantly under development. A specific graphical element, namely the line that is used to connect two entities (further referred to as *connecting line*), appears in many different

[1] The Peutinger Map, known as the first route map (366-225 BC) [3] and a multiple time-series graph showing planetary movements over time (10th century) [5] are historic visualizations that use connecting lines as their central graphical element.

This work was supported by the Research Council of Norway under Grant number 259536.

A.-V. Pietarinen et al. (Eds.): Diagrams 2020, LNAI 12169, pp. 110–127, 2020.
https://doi.org/10.1007/978-3-030-54249-8_9

visualization types (such as line charts, network diagrams, route maps etc.) and not just since the digital age.[1] However, the possibilities to signify specific meanings with such connecting lines have increased, because the application of transparency, interaction effects, animations etc., has become much easier in the digital era.

In the practice field of DV, *uncertainty* is a much-disputed topic[2]. One reason for that is that many big and complex datasets available today include elements of uncertainty related to confidence, variability, trends etc. [2]. The consequence for DV designers (here used as a collective term for all persons included in the DV production process) is that they have to find ways to visualize this uncertainty. Especially when designing visualizations for lay audiences, depicting uncertainty still remains a challenge, so big that it sometimes is not visualized at all [10].

Modality, as investigated by linguists and semioticians in verbal and visual text, is a concept that to some extent overlaps with uncertainty, as it is discussed in the practice field of DV. However, modality has not been brought into that field so far.

In the following, I will present the different perspectives on modality, both from a functional grammar point of view [following 11] and from a multimodality point of view [following 4 and 12] and relate the concept of modality to the concept of uncertainty. This is done because the linguistic concept of modality is very elaborated, and it is a hypothesis underlying this study, that it is relevant and useful also in the investigation of DVs. Following this theoretical trajectory, a two-level analysis method of modality and uncertainty in DVs will be presented.

This method is applied in the second, analytical part of the article. A corpus analysis of 163 award-winning DVs that include connecting lines, is presented. The focus of analysis is whether and how modality and uncertainty are expressed through connecting lines. Summing up, this article aims for three goals: (1) to clarify the relation between the concepts of modality and uncertainty in the field of DV, (2) to present a two-level method of analysis of modality and uncertainty in this text type, (3) to reveal graphical variations and conventions concerning the expression of modality and uncertainty through connecting lines within a corpus of award-winning DVs.

Corpus-based studies on current digital DVs in general and particularly those focusing on single graphical elements are still rare. Possible reasons for that may be the low availability of ready-to-use corpora and that the methods for accurately and time-effectively analyse such material are still at beginning stages [13–15].

2 Theoretical Perspectives on Modality and Uncertainty

2.1 Modality in Verbal Language

As a discussion capturing the breadth of works around the concept of modality is well outside the scope of this article I shall here only briefly introduce the work of the linguist Michael Halliday, who extended the system of modality with several aspects [16] relevant in this context.

[2] This can be stated because uncertainty was subject to many presentations at the IEEE Vis conference 2018 and 2019 [6, 7] as well as by the large number of publications published lately [additionally to the ones already cited in this article [e.g.: 8, 9].

Halliday sees modality as 'the speaker's judgement, or request of the judgement of the listener, on the status of what is being said' [11]. Modality 'construe[s] the region of uncertainty that lies between "yes" and "no"' [11] and is therefore an 'expression of indeterminacy' [11]. His system of modality – as applied on the clause level – includes four variables: the 'modality type', 'value', 'polarity' and 'orientation' [11]. I will further go deeper into the first two of these variables. The *modality type* 'modalization' [11] is most relevant in the context of DVs because it counts clauses that indicate some degree of a proposition's probability or usuality. Indications of probability, verbally expressed, for example, with adverbs (modal disjuncts) like *certainly, probably* or *possibly,* play an important role in some DVs, where an element of uncertainty is aimed to be communicated. They express a high, median and low *modality value*, respectively, which are the three modal judgement options suggested by Halliday [11]. Demonstrated with an example of Halliday [11]: *It certainly is* expresses a higher probability of this proposition than *It possibly is,* but both lie in between *It is* and *It isn't*.

2.2 Modality in Visual Material

As seen above, Halliday looked at the ways in which single words or word groups can express different degrees of probability. When it comes to modality in visual material, Gunther Kress and Theo van Leeuwen have borrowed the basic concepts from Halliday's functional grammar [12, 16]. The different levels of modality (modality value) are defined on scales of modality markers, such as colour saturation [12]. What constitutes a modality marker and where exactly on the scale the highest or lowest modality value is determined, is dependent on the 'coding orientation' [12]. Coding orientation, as Kress and Van Leeuwen further explain, refers to what counts as real in different social practices. Four types are named: the 'technological', 'sensory', 'abstract' and 'naturalistic coding orientation' [12].

In contexts where the semiotic content is a 'general pattern' or a 'deeper "essence" of what it depicts' [16] (as it often is in DVs), an abstract coding orientation will be applied. In such cases, semiotic reduction is crucial. This means a DV is valued as realistic if the most 'reduced articulation' [16] possible is used. A photo, on the other hand, is, according to Van Leeuwen [16], judged realistic if the colours, the articulation of depth, light and shadow, detail and background etc. are natural. Thus, a naturalistic coding orientation is applied.

By introducing the concept of coding orientation to different types of visual material, the issue of the 'construal and evaluation of the reliability of messages' [4] is focused. This constitutes a different aspect of a statement than probability. Thus, expressions of probability (*it will probably rain tomorrow*) and reliability (*you can believe me when I say that it will rain tomorrow)* have to be considered separately. However, especially in the context of statements realized by DVs, expressions of probability and reliability may be combined (*you can believe me when I say that it will probably rain tomorrow).* Moreover, it should be noted that the exemplary visual analyses carried out by Kress and Van Leeuwen regard the visual representation mainly as a whole [17] and therefore evaluate whether it, in its entirety, represents the 'given "proposition" (…) as true or not' [12]. In contrast to that, Halliday [11] looks at modality expressions on the clause level which means that the modality value of single sentences within a verbal text can vary.

2.3 Relating Modality Theory to the Analysis of Data Visualizations

As Halliday's statement that modality 'construe[s] the region of uncertainty that lies between "yes" and "no"' [11] implies, modality and uncertainty are intertwined. However, uncertainty is not only a research object for linguists and semioticians, but also widely disputed within the practice field of DV. Uncertainty is in that context related to different stages of the DV communication process. As a basis for the production of a DV, the designer has collected data about an aspect of the world that is either certain or uncertain. If the data is uncertain, this is what Dasgupta et al. [18] call data uncertainty. Data uncertainty may be caused by several reasons, like measurement imprecision, incompleteness of data (including missing values, sampling, aggregation), inference (including predictions, modeling and describing past events), disagreement and data incredibility [19].

During the design phase, the designer must decide what level of certainty that is most expedient to signal. The designer can decide to signal a high or low level of probability and reliability – or not to signal modality at all. After that decision, visual techniques for intendedly signalling a certain level of probability and reliability are chosen and applied by the designer. The results can be seen as visual expressions of modality. In most cases, what Dasgupta et al. [18] call *visual uncertainty* correlates with an intention to express a lowered level of probability (based on *data uncertainty*) or reliability. But it can also be a result of an unintended or unconscious application of visual forms that by convention or by earlier experience are associated with uncertainty by readers.

Summing up, *uncertainty* is a wider concept than *modality*, because it includes all factors causing uncertainty on the side of the reader, whether or not intended by the producer. In the present study, I am only interested in the visual expressions of modality that relate to lowered probability or reliability.

This comparative discussion of uncertainty (as discussed in the practice field of DV) and modality (as discussed by linguists and semioticians) allows for applying a more nuanced vocabulary when talking about uncertainty in DVs. It also allows for developing a detailed analysis method of modality in DVs, as presented below. The method is designed to answer the following research questions:

- How is lowered probability and reliability expressed by connecting lines in a corpus of award-winning, digital DVs?
- Does the corpus indicate any clear conventions concerning this issue?

3 A Two-Level Analysis Approach to Modality in Data Visualizations

3.1 Visual Segmentation

I will in the following propose a two-level approach to the investigation of modality in DVs. The two levels, further called *detail* and *global level*, refer to what parts of the DV that are in focus. How this visual segmentation is done, is inspired by Morten Boeriis' [17] dynamic functional rank scale. In *Tekstzoom* [17] he claims that a visual text can have several modality profiles on different text levels, and differentiates between four different text levels. For analysing modality in DVs, I propose that distinguishing between two zoom levels is sufficient.

At the detail level, only single graphical elements, like single lines or points, and the associated words, are considered (see right part of Fig. 1). This unit is comparable to a verbal sentence, as a part of a whole text. Here, we are interested in how these graphical elements – together with associated words – signal a certain level of probability and reliability, related to the detail statement they represent.

At the global level, the whole visualization (which may be integrated into a larger multimodal text including more verbal text or other visualizations) is focused (see left, the black part of Fig. 1). The pertinent question on this level is whether and how the choice of visual style signals that the visualization is a true reflection of an aspect of the world or not. The issue of coding orientation is here central, considering e.g. the effect that a hand drawing might have, compared to a digitally produced DV, regarding reliability. However, it may also be possible to find verbal hints of data uncertainty (expressing lowered probability) that concern not only the detail statement, but also the global statement of the whole DV. These verbal hints may be found within the global level, or in the surrounding co-text, as it might exist e.g. in a news article (see the grey area in Fig. 1).

Such a separation into two text levels allows for the investigation of whether and how single graphical elements, as well as the visual style of the whole visualization, signal modality.

Although this study focuses on the detail level, due to the connecting line constituting the study object, it is important to understand this model as a holistic concept. Boeriis claims that the overall modality of a text is a product of all modality profiles on all levels [17]. In other words, modality expressions on different text levels influence each other. However, how exactly this influence takes place and what effect it has on the overall modality profile is not a focused issue in this study.

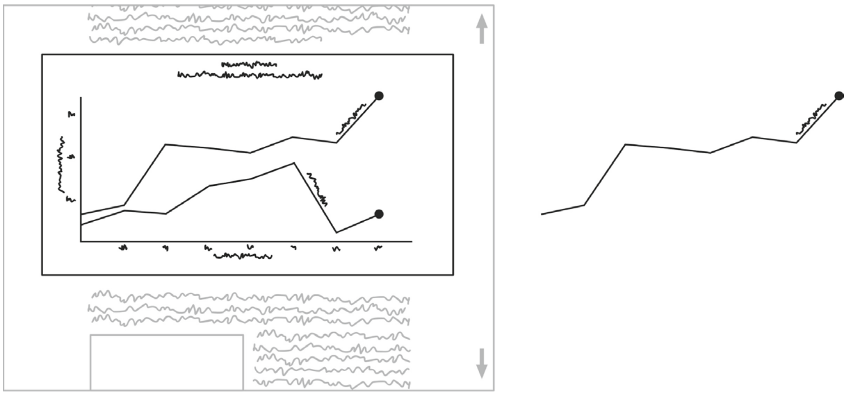


Fig. 1. Left: abstract representation of a line graph (black part = global level) and the co-text (grey) within a website; Right: only the detail level.

3.2 Operationalizing the Theory

Based on the two proposed levels for the DV analysis, Table 1 introduces concrete questions for an analysis of modality in DVs as well as the answer options. It should be understood as an extensible method offer, that may be adjusted to fit also analyses of other semiotic material or other research foci.

Table 1. Questions and answer options for an analysis of modality in DVs.

Nr.	Question	Answer options
Questions on the detail level:		
1	Only focus on the graphical elements that convey the main statement of the DV. If there are more of the same kind, decide for one exemplary unit, comparable to a clause. How could this clause be formulated?	Verbal statement
2	What kind of graphical element(s) represent(s) this statement?	Description of the visual element(s)
3	Which coding orientation needs to be applied when viewing this/these visual element(s)?	Abstract-, naturalistic-, technological-, sensory coding orientation [16] or none of these
4	What do(es) the graphical element(s) look like?	Description of the visual appearance of the visual element(s) in focus, see Table 2
5	Does the visual appearance of the graphical element(s) indicate any form of modality in terms of lowered probability or lowered reliability?	Yes or no
6	If yes, how is this lowered probability and/or lowered reliability signalled visually?	Description of the visual variables used
7	Is it explicitly verbally stated on detail or on global level of the DV, or in the co-text, that data uncertainty is represented within this detail statement of the DV?	Yes or no
8	If it is explicitly verbally stated that data uncertainty is represented, how?	Concrete formulation
9	Based on the former questions, do(es) the graphical element(s), signal modality in terms of lowered reliability?	Yes or no

(*continued*)

Table 1. (*continued*)

Nr.	Question	Answer options
10	Does the verbal sentence from question 1 need a reformulation, considering modality in terms of lowered probability and reliability, if any of them are expressed? If yes, which?	Verbal statement
Questions on the global level:		
11	What is the overall statement of this DV?	Verbal statement
12	Which coding orientation needs to be applied to the visual expression of the overall statement?	Abstract-, naturalistic-, technological-, sensory coding orientation [16] or none of these
13	Does the choice of visual style signal any sense of modality in terms of lowered reliability?	Yes or no
14	If yes, which modality value is signalled?	High, medium or low modality value
15	If yes, how does it signal a high, medium or low modality value?	Description of the visual clues that underly this decision [based on the descriptions of the coding orientations of 16]
16	Is it explicitly verbally stated on the detail or global level of the DV, or in the co-text, that data uncertainty is represented in the overall statement of the DV?	Yes or no
17	If it is explicitly verbally stated that data uncertainty is represented, how?	Verbal statement

Note. The verbal statements that are the answers to questions 1, 10, and 11 are only meant to serve as proposals. Visual language cannot be directly translated to verbal language or the other way around [12, 20]. However, the formulation of these sentences is considered helpful for investigating the semiotic potentials.

3.3 Description of the Visual Appearance of Connecting Lines

In Table 2, I suggest a set of visual variables and manifestation categories that can be used when focusing on connecting lines on a detail level. They are based on the system of 'visual variables' suggested by Jacques Bertin [21], as well as other scholars [22–27], who developed Bertin's visual variables further or contributed to a nuanced description of the visual appearance of lines. Figure 2 shows some visual examples to each visual variable of Table 2.

Table 2. The visual variables a line can have, and a suggestion of manifestation categories.

Nr.	Variable	Manifestation categories					
1	Position	Left out because dependent on DV specifications like scales, dimension of the DV, screen and window					
2	Orientation						
3	Size						
4	Colour	No variation (single - coloured)	Abrupt variation	Smooth transition			
5	Clarity: crispness	High	Lowered				
6	Clarity: transparen-cy	None	Low, medium or high				
7	Clarity: resolution	High	Lowered				
8	Pattern	Continu-ous line	Dashed /dotted line	Irregu-larly dashed/ dotted line	Change between continuous line and dashe / dotted line	Change between continuous line and large interrup-tions(s)	Change be-tween contin-uous line, dashed/dotted line and large interruption(s)
9	Shape: forces	One force (straight)	Two forces, curved	Two forces, bent	Three or more forces, curved	Three or more forces, bent	
10	Shape: line pressure	Consistent	Inconsistent				
11	Shape: form of extremities	No and direction-signifying extremity	Direction-signifying extremity (like e.g., an arrowhead)				
12	Interaction	Not possible	Possible				
13	Dynamics	Not available	Available, in one or more other visual variables (1-12)				

3.4 How to Identify Visual Indications of Lowered Probability and Reliability

Based on existing literature, we can assume there are three ways to identify visual indications of lowered probability and reliability in DVs. First, some visualization types are specifically developed to represent data uncertainty. Second, users may judge a visual element as an indication of uncertainty based on an analogy to the 'experiential world' [12]. Third, the user judgement may be based on criteria for what is real in the coding orientation applied.

Within the field of statistics, error bars and several newer visualization types, like gradient plots, violin plots or fan plots, are designed for indicating data uncertainty [28]. Also other visualization types can express data uncertainty, as is the case e.g. in various kinds of weather forecasts (see Hullman et al. [29] for other examples). However, in this study, focusing on the semiotic functions of connecting lines, it is most relevant to consider ways to identify signals of lowered probability and reliability on the detail level.

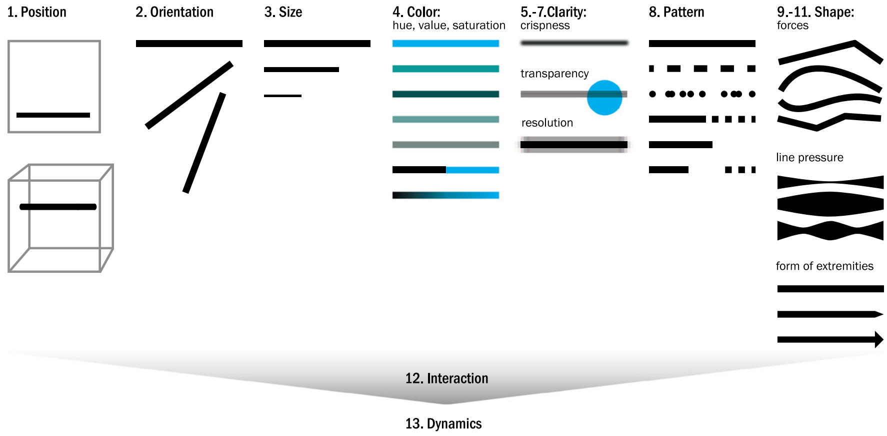


Fig. 2. Some examples to each of the visual variables from Table 2.

A first hint of potentially signalled lowered probability or reliability (referring to question 5 of Table 1) can be found when the line in focus resembles directly or metaphorically what the uncertainty indicates [30]. Analogies to our 'experiential world' [12] can be the reason why certain characteristics intuitively are interpreted as signs of uncertainty. The *sketchiness* of hand-drawn lines may metaphorically signal uncertainty [30], as well as the visual degradation of the line (through blur), since 'the harder it is to see …, the more uncertain it appears' [31]. Thus, blurry, sketchy, animated lines or lines with a pattern that leads to interruption (e.g., dashed lines) and lines with certain colour characteristics (e.g., low saturation) can indicate uncertainty [30, 31]. Also, if the visual appearance of the line changes along the length, this may be a hint of an indication of uncertainty. To an analyst, these aspects have to be considered, together with the coding orientation in use.

Given that one needs to apply an abstract coding orientation when analysing a line in a DV, the question to ask is: Is this the most 'reduced articulation' to represent the 'general pattern' or 'the deeper "essence" of what it depicts' [16] or not? Depending on the DV type and context, a line with the characteristics of the 3rd column of Table 2 (a straight, single-coloured, continuous, non-transparent etc. line) is counted as using the most reduced articulation. Whenever a more elaborated visual appearance is used, and other reasons behind this specific visual appearance can be ruled out, the line visually signals lowered probability or reliability. Such reasons can be: a) the intention to differentiate between different categories by different kinds of lines (as seen in Fig. 7); b) the intention to create a certain aesthetic effect, or c) the technical production tools favouring that kind of visual appearance.

In order to differentiate between signalled lowered probability and lowered reliability, it is often helpful to observe clues in the verbal text. If the visually depicted modality represents data uncertainty (and therefore lowered probability), the visual signal will normally be accompanied by explicit verbal clues (e.g., *forecast*, *scenario*, *95% confidence*). If that is not the case, and yet, the line visually signals some kind of modality, the analyst can conclude that the line signals lowered reliability. This conclusion can be based on the existence of 'intermodal tension' [32], i.e. that the verbal and the visual modes offer different, incompatible information. Engebretsen also states, that the conventions within 'genres focusing on informativity and fact-oriented learning … points [*sic*] toward a rhetoric of clarity and unambiguousness' [32]. Thus, unclarity and polysemy within visualizations have a negative impact on the reliability. In practical analysis, it can be difficult to judge whether incidents of such tension represent an intended use of modality or an unintended visual uncertainty expression.

4 Corpus Analysis

4.1 Data Selection and Database Setup

The method suggested in the previous section was applied to a corpus of 163 DVs. Due to the focus on the connecting line in this study, only the detail level was included in the analysis. The DVs were collected from the winner lists of the 2015, 2016 and 2017 Kantar Information is Beautiful Awards[3] [33] and the Malofiej Awards number 24, 25 and 26[4] [34]. All DVs but one were targeted to the general public and were published in online news media or other channels of public information. All winners with publicly available digital DVs (at the date of data collection) that contained one or more central DVs with one or more connecting line(s) in the leading role of communicating the DV's meaning were selected. The result of this filtering process was 163 single DVs stemming from 105 award-winning websites[5]. To establish a stable data basis for the analysis, over 400 screenshots, PDF documents and screencasts were created and organized in a relational database.

Due to the nature of the World Wide Web, it is impossible to claim that this corpus is a representative sample of the whole population of DVs with the characteristics mentioned above. Thus, the results of this analysis can by no means be used to generate valid statements about the whole population. However, this corpus contains a broad variety of DVs produced during the named timeframe in the western world, and the results of the research based on this material can be seen as a good approximation of how DVs in these countries have been developed in this specific time frame. Moreover, such awards raise publicity, and these DVs are judged by experts as 'best practices' and viewed by a broad audience, including practicioners. Therefore, they are expected to serve as models and to have strong convention forming abilities.

4.2 Method

Each DV was coded according to the method proposed in Section *A two-level analysis approach to modality in data visualizations*, using a detailed coding scheme. The detailed coding scheme contained the same questions and answer options as those in Table 1, with a description of criteria for choosing each option. Before that, an inter-rater reliability

[3] This includes projects created between May 2014 and September 2017.
[4] These projects were published between 2015 and 2017.
[5] A full list of all included DVs, the coding instructions and results of the inter-rater reliability study and the final coding scheme can be requested from the author.

study of a random sample of 25 DVs (approx. 15% of the corpus) was performed for the questions that contain judgement variables. This was necessary in order to 'estimate how reliable the categorisation (coding) is' [35], and therefore make sure that the stated questions and offered answer categories are precise and adequate. Two coders (a second coder and me) used the same coding instructions and worked independently. With the answers of both coders, Gwet's AC_1 and Gwet's AC_2 coefficient [36] were calculated. Results showed that for all questions, the coders had either substantial agreement or higher when analysed according to Gwet's benchmarking method [36][6]. This amount of agreement was deemed sufficient and the coding method was generally approved.

However, follow-up discussions between the two raters after the pre-test and also during the start of the single-coded analysis revealed that a few small adjustments of the coding scheme would still improve the rating process. Following an iterative method, these changes were made, resulting in the final coding scheme, that was then applied to the whole corpus. In instances of doubt, the second rater of the inter-rater reliability test was contacted to discuss the final codings. The (single-coded) analysis of the whole corpus then made it possible to generate frequency counts of whether and how modality is signalled in this corpus with connecting lines.

4.3 Analysis Findings

This section presents the results of the analysis on the detail level, using question 1 to 10 in Table 1. Due to the selection criteria for this corpus, the main statement in each DV is represented through graphical lines. For each DV, only one line is focused in the analysis. For all except two of the 163 lines in focus, an abstract coding orientation needs to be applied. For the final two, a naturalistic coding orientation is the most suitable.

As shown in Fig. 3, the connecting line in the focus of 26 (18 + 8) DVs out of 163 are found to indicate modality (lowered probability or lowered reliability). Within 33 (15 + 18) DVs, it is explicitly stated verbally that data uncertainty is represented within the detail statement represented through the focused connecting line. However, in only 18 DVs modality is signalled *both* visually through the connecting line in focus *and* through a corresponding verbal clue for data uncertainty. These 18 lines are therefore considered to signal lowered probability, while the reliability is not reduced. Under the earlier presented assumption that intermodal tension causes lowered reliability, this means that, on the detail level, the focused lines of 23 DVs (8 + 15) are found to be included in an instance of lowered reliability.

68% of the 41 DVs that verbally and/or visually signal modality on the detail level, are either route maps (41%) or line graphs (27%). The high occurrence of these two DV types also reflects the fact that these two types are the most common ones in this corpus (36% route maps, 23% line graphs).

I will now look at what visual variables of the connecting lines signal what kind of modality. When the lines in focus signal lowered probability (based on data uncertainty), different manifestation categories of the visual variable pattern were the most commonly used – especially those with pattern changes (see column three Table 3).

[6] The benchmarking method proposed in Gwet (2014: 173–181) was complemented by information given by Gwet in personal mail correspondence on 9–17 July 2019.

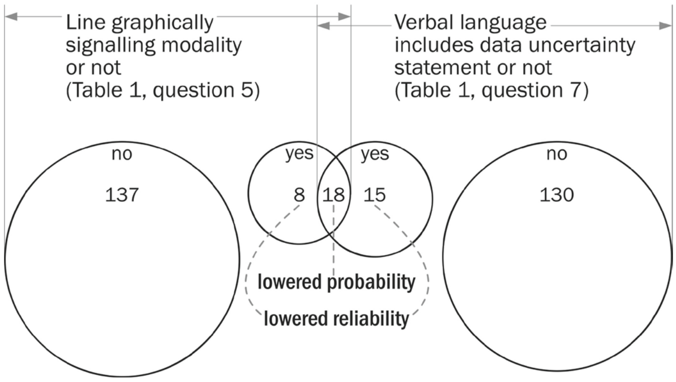


Fig. 3. Distribution of visually and/or verbally signalled modality.

Changes between a continuous line and large interruption(s) and between a continuous line and a dashed/dotted line are used 6 and 5 times respectively. Figure 4 presents an example of the latter. However, also other visual characteristics, namely transparency, lowered crispness, colour variations, inconsistent line pressure, three or more forces (curved) or dynamics in the size are used for that purpose.

Table 3. Distribution of visual characteristics of connecting lines used to signal modality (lowered probability and reliability) and the distribution of the same characteristics being used for other purposes. Note that in some DVs, several visual characteristics are used simultaneously to signal modality. N for each line in this table is 163 connecting lines from 163 DVs.

Visual variable used for signalling either lowered probability (data uncertainty) or reliability (answer to question 6 of Table 1)	Manifestation category used (answer to question 4 of Table 1)	Instances of visual characteristics signalling lowered probability (based on data uncertainty)	Instances of visual characteristics signalling lowered reliability	Instances of visual characteristics used for other purposes than signalling lowered probability or reliability
Pattern	Change between continuous line and dashed/dotted line	5	2	4
Pattern	Change between continuous line and large interruptions(s)	6	–	4

(*continued*)

Table 3. (*continued*)

Visual variable used for signalling either lowered probability (data uncertainty) or reliability (answer to question 6 of Table 1)	Manifestation category used (answer to question 4 of Table 1)	Instances of visual characteristics signalling lowered probability (based on data uncertainty)	Instances of visual characteristics signalling lowered reliability	Instances of visual characteristics used for other purposes than signalling lowered probability or reliability
Pattern	Change between continuous line, dashed/dotted line and large interruption(s)	2	–	–
Pattern	Dashed/dotted line	2	–	9
Pattern	Irregularly dashed/dotted line	–	1	4
Clarity: transparency	Low, medium or high	3	1	35
Clarity: crispness	Lowered	1	–	–
Colour	Abrupt variation	2	–	8
Colour	Smooth transition	1	–	15
Shape: forces	Two forces, curved	–	2	18
Shape: forces	Three or more forces, curved	1	1	47
Shape: line pressure	Inconsistent	3	–	13
Dynamics in size	Yes	1	1	50
Dynamics in orientation	Yes	–	1	16
Dynamics in position	Yes	–	1	15

When we look at how lowered reliability is signalled by the focused connecting lines (see column four Table 3), the results reveal that the visual variable *pattern* does not have such a prominent role. The pattern and the curviness of the lines signal lowered reliability three times. However, dynamics (in size, orientation and position) and transparency are also found once each.

An example of a visualization where curvature signals lowered reliability can be found in *An interactive visualization of every line in Hamilton* [38, see Fig. 5]. Here, the

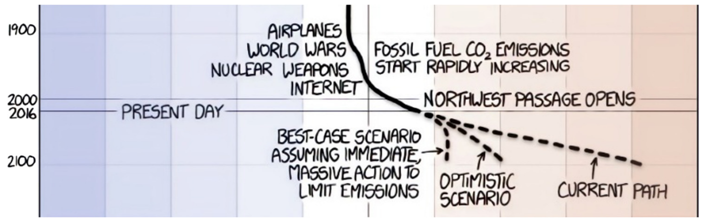


Fig. 4. Screenshot from *A timeline of earth's average temperature*, indicating data uncertainty by pattern change to a dashed line. © Randall Munroe [37]. Distributed under CC BY-NC 2.5.

semiotic motivation behind some connecting lines being curved, while others are straight, is not clear. Because the DV does not use the most reduced articulation possible (while applying an abstract coding orientation), it is rated as expressing lowered reliability.

As shown in Table 3, most of the visual characteristics of lines used to express modality, are not used exclusively for that purpose. Column five shows how many times the visual characteristics highlighted in column three and four are used for other purposes. For instance, in the visualization *The Stories Behind a Line* [39], different categories of transport means are visualized through different dashed/dotted lines (see Fig. 7). Another example of dashed lines not signalling modality is found in *Syrian war explained in 5 min* [40: 5:00, see Fig. 6]. Here, the animated dashes iconically represent moving bombs.

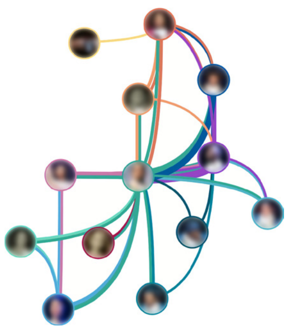


Fig. 5. Screenshot from *An interactive visualization of every line in Hamilton*, where curvature indicates lowered reliability. © Shirley Wu [38]. Reproduced with permission. Photos are blurred for copyright reasons.

Fig. 6. Abstract representation of a film frame of *Syrian war explained in 5 min* [40: 5:00]. The dashes move towards the square field named 'rebels', iconically representing moving bombs.

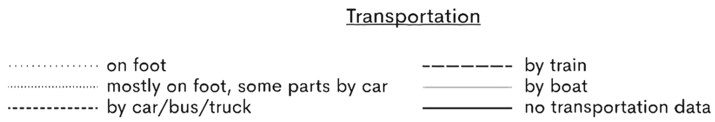


Fig. 7. Screenshot of the legend of *The Stories Behind a Line*, using interrupted lines for different categories of transport means. © Federica Fragapane, designed in collaboration with Alex Piacentini [39]. Reproduced with permission.

4.4 Limitations of the Results

Because most of the DVs were only single-coded and some questions contain judgement variables, it has to be kept in mind that my cultural background and previous knowledge might have influenced the interpretation. To counter this, the coding instructions were developed as detailed as possible, strictly followed and the inter-rater reliability study was performed.

Moreover, since only one connecting line was focused on the detail level of each DV, even if sometimes one DV contained more connecting lines, it is possible that the results could have changed if I had chosen to focus on other lines. Therefore, I have been careful when reporting these numbers, to refer only to the connecting lines 'in focus', not to all connecting lines in the material.

4.5 Implications and Conclusion

Within this corpus of 163 DVs, out of the 41 visualizations indicating some kind of modality on the detail level, 23 exhibit cases of intermodal tension. This number indicates that intermodal tension, meaning that the verbal and the visual resources offer conflicting signals, is fairly common in this field of DV-based communication. One implication of this finding is that the potential for DV designers to avoid unintended ambiguity by giving more attention to multimodal coherence is high.

The results further indicate a convention saying that *pattern change* is well suited for visually signalling data uncertainty, corresponding to the modality category *lowered probability.* Why pattern change – in the shaping of connecting lines in DVs – is emerging as a conventionalized signal of modality, may have several reasons. First, it must be assumed that pattern change potentially signals modality based on an analogy to the 'experiential world' [12]. Furthermore, the use of patterns, or larger interruptions, is not expected as a typical line form in any DV type (unlike e.g. curvature, which is common in spline graphs for instance), thus such characteristics are free to use as modality markers. Moreover, it is technically easy with most design tools to apply different patterns to a graphical line (unlike e.g. dynamics). Last, patterns are possible to use also in two-coloured DVs, and they are printable and drawable analogously, which points to a long application history. For signalling *lowered reliability*, however, no such convention was traced, as the results show a more varied and unsystematic use of characteristics indicating this kind of modality.

Summing up, the study reveals that various visual characteristics of connecting lines are used to signal modality in this corpus of award-winning DVs. However, *pattern change* is used more often than any of the other variables found in the corpus. Due to the relatively low number of observations in this corpus, it is impossible to provide practitioners with a simple recipe for what visual clues are most effectively applied to signal modality in DVs. Nonetheless, the results provide an overview of the current practices in using lines for indicating modality, which is helpful for practitioners to make informed design decisions.

5 Further Research

In this article, a method for analysing modality in DVs is presented, based on a body of pre-existing theory and terminology around modality and uncertainty. A newly collected corpus of digital DVs is analysed with the suggested method, offering detailed knowledge about how certain visual characteristics of the graphical line are used for signalling modality. The findings indicate certain conventions regarding the semiotic potential of the graphical line in relation to modality. Such insights are valuable both for designers and scholars in relevant fields, as they contribute to the colouring of some of the white spots on the map over a graphical language still in its making. However, more empirical research is needed in order to draw a more detailed and reliable map over the field of multimodal modality. The findings, as well as the methodology presented in this study, will hopefully be a contribution to this future work.

References

1. Kirk, A.: Data Visualization: A Successful Design Process. Packt Publishing, Birmingham (2012)
2. Bonneau, G.-P., et al.: Overview and state-of-the-art of uncertainty visualization. In: Hansen, C.D., Chen, M., Johnson, C.R., Kaufman, A.E., Hagen, H. (eds.) Scientific Visualization. MV, pp. 3–27. Springer, London (2014). https://doi.org/10.1007/978-1-4471-6497-5_1
3. Friendly, M., Denis, D.F.: Timeline. http://www.datavis.ca/milestones. Accessed 12 Feb 2019
4. Ravelli, L.J., Van Leeuwen, T.: Modality in the digital age. Vis. Commun. **17** (2018). https://doi.org/10.1177/1470357218764436
5. Friendly, M.: A brief history of data visualization. In: Chen, C., Härdle, W., Unwin, A. (eds.) Handbook of Data Visualization. SHCS, pp. 15–56. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-33037-0_2
6. IEEE: VIS 2018 program. http://ieeevis.org/attachments/vis18-program.pdf. Accessed 15 Jan 2019
7. IEEE: VIS 2019 program. http://ieeevis.org/year/2019/info/papers-sessions. Accessed 16 Oct 2019
8. Brodlie, K., Allendes Osorio, R., Lopes, A.: A review of uncertainty in data visualization. In: Dill, J., Earnshaw, R., Kasik, D., Vince, J., Wong, P. (eds.) Expanding the Frontiers of Visual Analytics and Visualization, pp. 81–109. Springer, London (2012). https://doi.org/10.1007/978-1-4471-2804-5_6
9. Kinkeldey, C., MacEachren, A.M., Schiewe, J.: How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. Cartogr. J. **51**, 372–386 (2014). https://doi.org/10.1179/1743277414Y.0000000099
10. Hullman, J.: Why authors don't visualize uncertainty. IEEE TVCG **26**, 130–139 (2020). https://doi.org/10.1109/TVCG.2019.2934287

11. Halliday, M.A.K.: An Introduction to Functional Grammar. 3rd edn. Revised by Matthiessen, C.M.I.M. Arnold, London (2004)
12. Kress, G., Van Leeuwen, T.: Reading Images: The Grammar of Visual Design. 2nd edn. Routledge, Abingdon (2006)
13. Bateman, J., McDonald, D., Hiippala, T., Couto-Vale, D., Costetchi, E.: Systemic functional linguistics and computation. New directions, new challenges. In: The Cambridge Handbook of Systemic Functional Linguistics, pp. 561–586. Cambridge University Press (2019)
14. Kembhavi, A., Salvato, M., Kolve, E., Seo, M., Hajishirzi, H., Farhadi, A.: A diagram is worth a dozen images. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 235–251. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_15
15. Hiippala, T., et al.: AI2D-RST: a multimodal corpus of 1000 primary school science diagrams (Forthcoming)
16. Van Leeuwen, T.: Introducing Social Semiotics. Routledge, Abingdon (2005)
17. Boeriis, M.: Tekstzoom - om en dynamisk funktionel rangstruktur i visuelle tekster. In: Hestbæk Andersen, T., Boeriis, M. (eds.) Nordisk socialsemiotik - multimodale, pædagogiske og sprogvidenskabelige landvindinger, pp. 131–153. University Press of Southern Denmark, Odense (2012)
18. Dasgupta, A., Chen, M., Kosara, R.: Conceptualizing visual uncertainty in parallel coordinates. In: Computer Graphics Forum, pp. 1015–1024. Blackwell, Vienna (2012). https://doi.org/10.1111/j.1467-8659.2012.03094.x
19. Skeels, M., Lee, B., Smith, G., Robertson, G.: Revealing uncertainty for information visualization. In: AVI 2008 Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 376–379. ACM, New York (2008). https://doi.org/10.1145/1385569.1385637
20. Saint-Martin, F.: Semiotics of Visual Language. Indiana University Press, Bloomington (1990)
21. Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps. ESRI Press, Redlands (2011)
22. DiBiase, D., MacEachren, A.M., Krygier, J.B., Reeves, C.: Animation and the role of map design in scientific visualization. Cartogr. Geogr. Inf. Syst. **19**, 201–214 (1992). https://doi.org/10.1559/152304092783721295
23. Kandinsky, W.: Punkt und Linie zu Fläche: Beitrag zur Analyse der malerischen Elemente. Albert Langen, München (1926)
24. MacEachren, A.M.: How Maps Work: Representation, Visualization, and Design. Guilford Press, New York (1995)
25. Saulnier, A.: La perception du mouvement dans les systèmes de visualisation d'informations. In: Proceedings of the 17th Conference on l'Interaction Homme-Machine, pp. 185–192. ACM, Toulouse (2005). https://doi.org/10.1145/1148550.1148574
26. Weger, G.: Cartographie Volume 1: Sémiologie graphique et conception cartographique. École nationale des science géographiques, Marne la Vallée (1999)
27. Wong, W.: Principles of Form and Design. Wiley, New York (1993)
28. Cairo, A.: The Truthful Art: Data, Charts, and Maps for Communication. New Riders, Thousand Oaks (2016)
29. Hullman, J., Qiao, X., Correll, M., Kale, A., Kay, M.: In pursuit of error: a survey of uncertainty visualization evaluation. IEEE TVCG **25**, 903–913 (2018). https://doi.org/10.1109/TVCG.2018.2864889
30. Boukhelifa, N., Bezerianos, A., Isenberg, T., Fekete, J.-D.: Evaluating sketchy lines for the visualization of qualitative uncertainty. Research Centre Saclay - Île-de-France, Le Chesnay Cedex (2012)
31. Tak, S., Toet, A., van Erp, J.: The perception of visual uncertainty representation by non-experts. IEEE TVCG **20**, 935–943 (2014). https://doi.org/10.1109/TVCG.2013.247

32. Engebretsen, M.: Balancing cohesion and tension in multimodal rhetoric. An interdisciplinary approach to the study of semiotic complexity. Learn. Media Technol. **37**, 145–162 (2012). https://doi.org/10.1080/17439884.2012.655745
33. The Information is Beautiful Awards Ltd.: KANTAR Information is Beautiful Awards. https://www.informationisbeautifulawards.com/. Accessed 31 Mar 2018
34. Malofiej Infographic World Summit: Malofiej Infographic World Summit. https://www.malofiejgraphics.com/awards/. Accessed 05 Nov 2018
35. Brezina, V.: Statistical choices in corpus-based discourse analysis. In: Taylor, C., Marchi, A. (eds.) Corpus Approaches to Discourse: A Critical Review, pp. 259–280. Routledge, Abingdon (2018)
36. Gwet, K.L.: Handbook of Inter-rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. 4th edn. Advanced Analytics. LLC, Gaithersburg (2014)
37. Munroe, R.: A timeline of earth's average temperature. https://xkcd.com/1732/. Accessed 24 Apr 2018
38. Wu, S.: An interactive visualization of every line in Hamilton. https://pudding.cool/2017/03/hamilton/index.html. Accessed 29 Oct 2019
39. Fragapane, F.: The stories behind a line. A project in collaboration with Alex Piacentini. http://www.storiesbehindaline.com. Accessed 23 July 2019
40. Fisher, M., Harris, J.: Syrian war explained in 5 minutes. https://www.youtube.com/watch?v=qxzMa7j6LN0. Accessed 20 June 2018

10.3 Article III: Arrows and Their Modern Versions: Narrativity Signalled by Lines in Data Visualizations

Published here with double pagination in manuscript layout, slightly adjusted for better readability. This manuscript is the revised version (minor revisions) of an article that was resubmitted to the Journal of Visual Literacy (https://www.tandfonline.com/loi/rjvl20).

Reference

Lechner, V. E. (2020). *Arrows and their modern versions: Narrativity signalled by lines in data visualizations*. [Manuscript submitted for publication]

Arrows and their modern versions: Narrativity signalled by lines in data visualizations

Verena Elisabeth Lechner

For creating and reading data visualizations, visual literacy is crucial. This article advances the knowledge about graphical variations and conventions related to the basic graphical element of the graphical line used as a connector in data visualizations. Some visual characteristics of connecting lines can be used to show directionality and thus signal a narrative claim. Arrowheads may be one way to do so. However, particularly in a digital environment, other techniques may be used as well. This article presents a corpus study investigating the ways in which narrativity is signalled by connecting lines in current, publicly available digital data visualizations. The central connecting lines of 163 award-winning data visualizations are analysed with focus on their visual forms and how they represent actions and situations. The repeated occurrence of some visual techniques points to conventions formed both by a long tradition of analogue visualization and the advent of digital production techniques and output devices. The presented results are relevant to researchers, educators and practitioners in the data visualization field, as they provide novel empirical data on the use of an omnipresent graphical element.

Keywords:
line; arrow; narrativity; social semiotics; visual variables; data visualization

Introduction

Data visualizations (further abbreviated as *DVs*) are used in many different contexts, such as journalism and public information, as well as in workplaces and education (Kennedy & Engebretsen, 2020). Being able to read and create DVs requires visual literacy skills (Kędra, 2018), which includes knowledge about visual conventions – the meanings that, by social agreement, are associated with specific configurations of the components of visual language. These conventions are constantly under development, since they are influenced by technological developments (Kennedy et al., 2016).

This article aims to contribute to the available knowledge about emerging graphical conventions in the DV domain. It focuses on one graphical element used in many different DV types (such as line graphs, network diagrams or route maps): the graphical line, used as a connector – further referred to as *connecting line*. Some connecting lines have visual characteristics that indicate directionality, and thus direct the viewer to read the information as a narrative claim: some kind of action is taking place (Lechner, 2020). This study offers an empirical investigation of what graphical variations of connecting lines are used to indicate direction, and whether any established conventions for such use can be observed.

The line constitutes one of the three most basic graphic elements, besides the point and the area (Bertin, 1967/2011). Thus, research on the meaning potentials of the line provides knowledge about a frequently used graphic element. Apart from that, two aspects of the current media situation call for such research. First, digital DV production tools make it possible to use more visual techniques for signalling directionality than those available through analogue tools. Arrows, for instance, have been used for centuries in printed books for abstractly indicating direction. An early example can be found in a book about hydraulics from 1737, where the arrow signals a push movement (see Fig. 1, Belidor, 1737, plate 4). Today, DV designers have more opportunities. How and to what extent these are used is, however, unclear. Second, digital DVs can easily be distributed worldwide and gain a broad readership. An assumption underlying this article is that, through the advent of digital display and distribution tools, socially agreed meaning systems for visual signs have developed within big and heterogeneous cultural groups. This facilitates and demands corpus studies that include current DVs produced for readers in many countries.

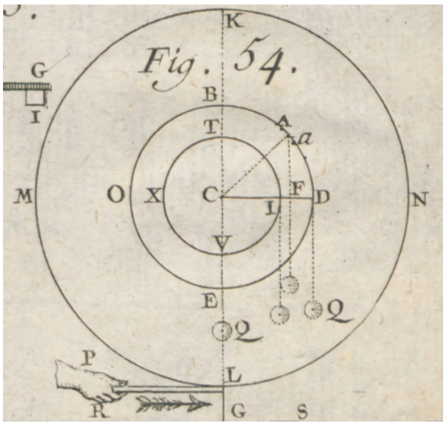


Figure 1. Picture detail of an illustration from 1737, where the arrow signals a pushing movement (Belidor, 1737, plate 4; in the public domain).

In this study, the corpus used contains 163 digital, award-winning DVs with one or more connecting lines in the leading role of communicating the DV's meaning. This article asks: Which graphical variations of connecting lines, expressing narrativity, can be observed in this corpus and with what frequency and combinations are they used? Based on the corpus analysis, the article discusses whether the findings indicate evolving conventions regarding this aspect of visual literacy. The generated knowledge is relevant for visual literacy training in general, for amateur and professional DV designers as well as for DV production tool developers and for researchers interested in visual communication.

Literature review

Typically, narrativity in digital DVs is mentioned in the research literature alongside storytelling, looking not only at single graphical elements within DVs, but rather taking a broad view on all multimodal elements on websites and how they build a story (see, e.g., Bach et al., 2018). Some researchers have looked at narrative elements in DVs using corpora similar to the one used in this article, albeit smaller – containing 45 to

80 data-driven stories (McKenna et al., 2017; Segel & Heer, 2010; Stolper et al., 2016, 2018; Weber, 2020).

Regarding the topic of lines indicating narrativity, the arrow may be the visual feature examined most frequently and within several scientific fields. Descriptions of the semiotic meaning potential that arrows can have can be found in Horn (1998). Van der Waarde and Westendorp (2000) specifically look at the functions of arrows applied in user instructions. More recently, arrows are discussed by Wong (2011). He mentions graphical conventions in scientific diagrams in the field of molecular biology, where arrows with different line lengths and line forms are used as symbols for what he refers to as 'promoter[s]' or 'primer[s]' (p. 701).

Arrows in diagrams have also been investigated in cognitive science. Heiser and Tversky (2006) have done experiments on the function of arrows in the production and interpretation of mechanical diagrams, testing the effects of the absence and presence of arrows in the interpretation of such diagrams. The results show that diagrams with arrows were described with significantly more functional units (using verbs like *enter*, *open*, *close*, *push*) than those without arrows (where verbs like *is* and *are* were more often used instead). Also, when testing how participants sketch diagrams according to either structural or functional descriptions, the participants produced diagrams with arrows more often when they had functional descriptions than when they had structural ones. Moreover, Tversky et al. (2008) have compared animation to other static ways of telling stories within diagrams, like using arrows. The overall conclusion of their research review is that using animation has no convincing benefits.

In the field of human-computer interaction, too, researchers have investigated connecting lines used as direction-signifying elements in visualizations. A relevant example is Koylu and Guo (2017), who tested five kinds of asymmetric lines in flow maps. The results revealed the potential usefulness of all five kinds, thus they recommend that designers make task-dependent design decisions.

In the field of social semiotics, which is the underlying theoretical basis of this article, lines in general are examined only to a very limited degree. The graphic line used as a boundary line was discussed by Cameron (2011). Regarding direction-indicating lines, Fuller (2002) examined the arrow as a wayfinding tool at airports, pointing out the arrow's function as a 'tool and trope of the control society' (p. 242).

Summing up, direction-signifying lines have already been researched within several fields. However, a systematic investigation of lines used for this purpose in current digital DVs is still missing. New digital production tools and distribution platforms, offering DV designers opportunities that go far beyond using static arrows, call for new approaches to the study of graphic lines and their semiotic functions. This article offers an original contribution to the fields of visual literacy, social semiotics and DV design, by:

(1) presenting a method for developing and analysing a corpus of DVs containing connecting lines in the leading role of communicating the DV's meaning;
(2) presenting the results of a corpus analysis based on social semiotic theory, focusing on whether and how the lines in focus indicate narrative claims by their specific visual appearance.

Theoretical background

Social semiotics

Within social semiotics, the rules of semiotic systems are regarded as flexible and modifiable. 'They are made by people, they come in different kinds, and they change over time' (Van Leeuwen, 2005, p. 53). The concepts presented in the ground-breaking work, *Reading Images* (Kress & Van Leeuwen, 1996/2006), offer a basis for investigating visual conventions in Western culture. Kress and Van Leeuwen's classification system of representational structures (1996/2006, pp. 42–113) is central to the study presented in this article.

Representational structures

The term *representational structure* refers to the choices that sign-makers have for how to represent 'objects and their relations in a world outside the representational system' (Kress & Van Leeuwen, 1996/2006, p. 42). Relevant to this investigation, Kress and Van Leeuwen have used as examples, among other text types such as art pieces or advertisements, abstract diagrams. They draw on Arnheim (1982), who made the broadest underlying distinction: 'We shall distinguish between volumes and vectors, between being and acting' (p. 154). Kress and Van Leeuwen also use the term *vector* to explain the difference between the two main representational structures: narrative and conceptual structures. They state that if a vector (an indicator of directionality, e.g., an arrow) is present, the visual representation has a narrative structure. It shows the participants as '*doing* something to or for each other' (1996/2006, p. 59). Accordingly, if no vector is present, the visual representation has a conceptual structure, showing no action, but merely the existence or state of something.

At this stage, the relationship between the following key terms needs to be clarified: *narrative*, *action* and *directionality*. A narrative presents a story, which is based on at least one action, unfolding in time and space (Weber, 2020). A vector is a sign that signals such action (a process transforming the situation) through an indication of directionality, like an arrow representing a movement in a specific direction. Consequently, in this study, lines that indicate direction are conceived as graphical units signalling action, and, thus, narrativity.

The two (narrative and conceptual) representational structures introduced by Kress and Van Leeuwen can be further distinguished into various process types. Since this article focuses on narrativity, only the process types of the narrative representational structure are considered. These are *action processes*, *reactional processes*, *mental* and *verbal processes* and *conversion processes* (1996/2006). For the present study, reactional processes, as well as mental and verbal processes, are omitted because their kinds of vectors are different from graphical lines (see Kress & Van Leeuwen, 1996/2006, pp. 67–68 for further details).

Kress and Van Leeuwen distinguish between three types of action processes: *non-transactional* and *transactional action processes* and *events* (1996/2006). Conversion processes, on the other hand, have no subcategories. Figure 2 shows the categories and relationships between the representational structures that are potentially relevant to the present study. Whether this system, as presented by Kress and Van Leeuwen, has satisfactory explanatory power in such a study of concrete DVs is among the issues that are discussed in the method section.

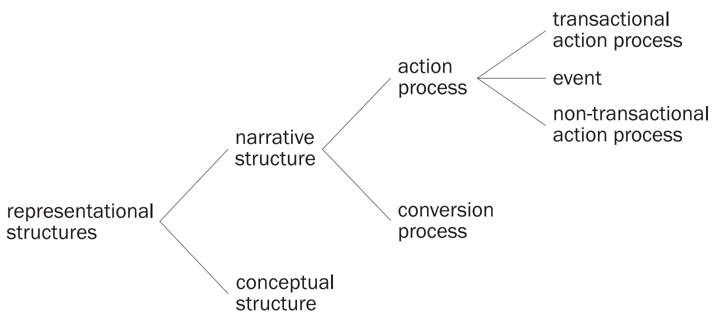


Figure 2. Dendrogram showing the categories of representational structures applied in this investigation (inspired by Kress & Van Leeuwen, 1996/2006, pp. 45–74).

As Kress and Van Leeuwen (1996/2006) state, ‘different kinds of narrative process can be distinguished on the basis of the kinds of vector and the number and kind

7

of participants involved' (p. 63). They identify two participants, namely the Actor and the Goal. A vector (indicating the direction) connects them. The Actor they define as 'the participant from which the vector emanates' (1996/2006, p. 63) and the Goal as 'the participant at whom or which the vector is directed, hence it is also the participant *to whom* or which the action is done, or *at whom* or which the action is aimed' (1996/2006, p. 64). The given visual examples of such Goals are the destination to which a pointing hand is directed, and the destination of a movement on a map, visualized as an arrow and an endpoint.

These participants are further used to differentiate between non-transactional processes, transactional processes, events and conversion processes. Figure 3 illustrates the differences between these categories of narrative representational structures. Transactional processes have an Actor and a Goal, non-transactional processes have only an Actor (and no Goal), the action is 'not "done to" or "aimed at" anyone or anything' (Kress & Van Leeuwen, 1996/2006, p. 63). Events have only a Goal (and no Actor). As they explain, 'who or what makes it happen' (1996/2006, p. 64) is invisible. Following these explanations, an illustration showing a man throwing a ball at a woman, with both the woman and the man visible, represents a transactional process. If the woman is not visible (e.g., because the image was cropped), one can refer to it as a non-transactional process. If the man is not visible, but it's clear that the ball's path is aimed at the woman, it is an event. In a conversion process several transactional processes are combined, and the state of what is passed over to the next process changes. This is the case, for example, in food chain diagrams.

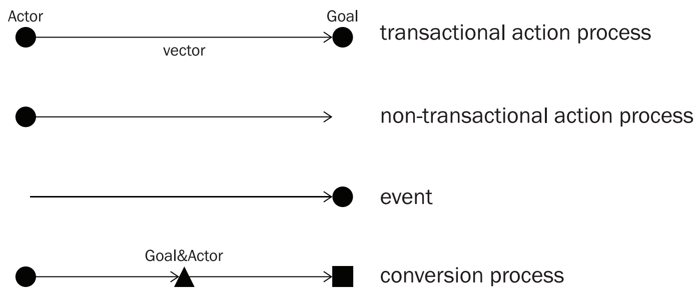


Figure 3. Abstract illustration of the narrative process types presented in Figure 2. (inspired by Kress & Van Leeuwen, 1996/2006, pp. 63–69).

Following a deductive approach, the different process types shown in Figure 3 were taken as a starting point for the categorization procedure and applied to the analysis of graphic lines.

Method

Corpus setup

The investigated corpus consists of 163 DVs collected from the winners' lists of the Kantar Information is Beautiful Awards 2015, 2016 and 2017 (The Information is Beautiful Awards Ltd, 2018) and the Malofiej Awards 24, 25 and 26 (Malofiej Infographic World Summit, n.d.). Consequently, the DVs were either created between May 2014 and September 2017 (for those stemming from the winners' lists of the Kantar Information is Beautiful Awards) or published between 2015 and 2017 (for those stemming from the winners' lists of the Malofiej Awards), but all were awarded between 2015 and 2018. Following the study's focus on digital DVs (meaning DVs that are developed for a presentation on digital systems), the winners' lists were searched for digital projects that were publicly available at the date of data collection. All projects that fulfilled these two requirements and contained one or more central DVs with one or more connecting lines in the leading role of communicating the DV's intended meaning (further called *DVs with connecting lines*) were selected. As a result of this filtering process, 105 award-winning projects were identified, containing 163 single DVs with connecting lines.

A relational database containing over 400 screenshots, PDF documents and screencasts of these 163 DVs as well as metadata (such as information about the price, award year, project name, URL and author) was set up. Moreover, the 163 DVs covering a wide range of themes were categorized according to DV type. The most frequent DV types occurring in the corpus are route maps (59 instances), line graphs (37 instances) and connection maps (18 instances). Of all other occurring DV types, six or fewer instances were identified. The target group of all DVs but one was the general public, following the fact that they were published in online news media or similar public information channels. All award-winning projects had verbal surrounding text in either English, Spanish, Portuguese, German, French or Norwegian. Thus, it can be assumed that the corpus fits the interests of this study to the investigation of emerging graphical conventions in Western culture.

Considering the limitations of this corpus, it can by no means be called representative of the whole population of publicly accessible DVs on the World Wide Web. However, the corpus does contain a broad variety of systematically collected DVs targeted towards a general, Western audience. Therefore, the results of the corpus analysis offer a good approximation of the ways in which such DVs were designed in the regarded timeframe. Moreover, the selected DVs presumably play a key role in the shaping of graphical conventions, being considered best practices by experts and hence used as models by practitioners.

Analysis procedure and coding scheme setup

The analysis of whether and how narrativity is signalled by connecting lines in this corpus was conducted in four steps:

(1) Identification of the line in focus and its endpoints (if visible) for each DV.

(2) Description of the visual appearance of the focused line.

(3) Identification of the representational structure signalled by the line.

(4) Description of the way that narrativity is shown visually (if that was the case).

Table 1 shows the questionnaire applied to each DV and each focused connecting line, respectively.

As can be seen from Table 1, only the visually signalled representational structure of the connecting line in focus was analysed in this study. No account was taken of whether the focused connecting line was part of any other major structures or processes. This is relevant to mention, because the representational structure of a single connecting line in a DV may differ from the representational structure of the DV as a whole.

A set of visual variables and manifestation categories was used for the visual description of the focused connecting lines (Table 1, Question 2), as presented in Table 2. They are based on the system of ‘visual variables’ suggested by Bertin (1967/2011, p. 42) and other scholars (DiBiase et al., 1992; Kandinsky, 1926/1928; MacEachren, 1995; Saulnier, 2005; Weger, 1999; W. Wong, 1993). These other scholars built on Bertin’s system or elaborated the vocabulary used to describe the visual appearance of lines. Particularly relevant to this study is the consideration of dynamic and interactive characteristics, which Bertin excluded due to his focus on graphics displayed on analogue, static surfaces.

Table 1. Questions and answer options for the analysis of representational structure signalled by connecting lines.

No.	Question	Answer options
1	Only focus on the connecting line (and its endpoints if applicable) that conveys the main statement of the DV. If there are more of the same kind, decide for one exemplary unit, comparable to a clause. How could this clause be formulated?	verbal statement
2	What does the focused connecting line look like?	description of the connecting line, see Table 2
3	Does the visual appearance of the focused connecting line signal a narrative or a conceptual representational structure?	narrative or conceptual representational structure, depending on whether the line is a vector or not; or ambiguous
4*	If it signals a narrative, does it signal an action process or a conversion process?	action process or conversion process
5*	If it signals an action process, does it signal a transactional or a non-transactional one, or an event?	transactional action process, non-transactional action process or event
6	If the visual appearance of the connecting line signals a narrative representational structure, how does it do so?	description of the visual variables (and manifestation categories) used
7	Try to reformulate the verbal sentence from Question 1, so that it best represents the type of representational structure you decided on in Questions 3 to 5.	verbal statement

Note. The verbal statement answering Question 1 is only used to determine which connecting line is focused on. The verbal statement answering Question 7 is only used to explain the answers given to Questions 3 to 5, though keeping in mind that a direct translation from visual language to verbal language is illusory, due to the different affordances of the two semiotic modes (Kress & Van Leeuwen, 1996/2006; Saint-Martin, 1987/1990). *Questions 4 and 5 were only applied in the pre-test, for reasons explained later.

The manifestation categories of some visual variables in Table 2 are customized to describe directionality (e.g., *shape*: *arrowheads or similar*), and are based on an iterative process. During that process, the list of visual variables and the corresponding manifestation categories described in the literature were iteratively modified based on empirical observations of the DVs in the corpus. Other manifestation categories could be more adequate when focusing on other semiotic potentials than narrativity.

Table 2. The visual variables a line can have, and a suggestion of manifestation categories.

No.	Variable	Manifestation categories					
1	position	left out in step two (description of visual appearance), because these variables are dependent on DV specifications like scales, dimension of the DV, screen and window					
2	orientation						
3	size						
4	colour	no variation (single-coloured)		abrupt variation		smooth transition	
5	clarity: crispness	high			lowered		
6	clarity: transparency	none			low, medium or high		
7	clarity: resolution	high			lowered		
8	pattern	continuous line	dashed /dotted line	irreg-ularly dashed/ dotted line	change between continuous line and dashed/ dotted line	change between continuous line and large inter-ruption(s)	change between continuous line, dashed/ dotted line and large in-terruption(s)
9	shape: forces	one force (straight)	two forces, curved		two forces, bent	three or more forces, curved	three or more forces, bent
10	shape: line pressure	consistent		inconsistent, several variations		inconsistent, tapering to-wards one end	
11	shape: arrowheads or similar	no arrow-head(s) or similar	arrowhead or similar at one end		arrow-heads or similar at both ends	arrowhead(s) or similar along the line	arrow(head) (s) or similar besides the line
12	interaction	not possible			possible		
13	dynamics	not available			available, in one or more other visual variables (1-12)		

Note. The extension *or similar* within the visual variable number 11 corresponds to other signs that can be used instead of arrowheads for signalling direction, for example, these may be icons of planes or other objects generally known to have a clear movement direction.

In the last stage of this analytic procedure, the visual variables and manifestation categories of Table 2 used to indicate direction (and thus a narrative structure), were registered. Moreover, the DVs were marked for any interactive or dynamic elements in general (apart from the single line). This is relevant to interpret the frequency of dynamic and interactive lines in the corpus.

A coding scheme was set up for the corpus analysis, explaining in detail the answer options for each of the four steps. Before applying this coding scheme to the whole corpus, an inter-rater reliability test was performed to estimate the coding reliability (Brezina, 2018).

Inter-rater reliability test

A random sample of 25 DVs (approx. 15% of the corpus) was analysed by two independent coders (the author and a second coder). Both used the same coding scheme, based on the procedure described in the previous section. Codings were compared and Gwet's AC_1 coefficient (Gwet, 2014) was calculated for the questions that contain judgement variables (Questions 3-5 of Table 1), using the software R (Martire, n.d.; The R Foundation for Statistical Computing, 2018). Finally, Gwet's benchmarking method was applied to determine the agreement level (Gwet, 2014; personal mail correspondence with Gwet, July 9-17, 2019). Results showed that the agreement for Question 3 was almost perfect, for Questions 4 and 5 it was substantial or moderate, respectively (see Table 3).

Table 3. Summary of the inter-rater reliability test results.

Question no. of Table 1	AC_1 coefficient	Standard error	Final agreement level
3	0.946780	0.052157	almost perfect
4	0.890049	0.075583	substantial
5	0.652430	0.119440	moderate

Note. Number of raters: 2; Sample size = 25 DVs; Confidence level: 95%.

In a follow-up discussion between the raters, the possible reasons for the lower agreement levels of Questions 4 and 5 were investigated. This revealed some ambiguities regarding how to determine the subcategories of narrative representational structures. These ambiguities are induced by the comparison of the 'Goal' (Kress & Van Leeuwen, 1996/2006, pp. 63–65) in visual presentations with the grammatical object in verbal language in the underlying literature. Kress and Van Leeuwen present the action displayed visually in transactional action processes (which have a Goal) as analogue to transitive verbs (verbs that take an object). Likewise, non-transactional action processes are presented as analogue to intransitive verbs. This analogy is, however, imprecise, because what is defined as an object in verbal language is different from Kress and Van Leeuwen's broader explanation of the Goal in a visual expression. In the examples they show, it comes clear that *Goal* also includes what, in verbal language, is called 'circumstantial adverbials', phrases that 'answer questions such as *when*, ***where*** [emphasis added], *how*, *why*' (Dypedahl & Hasselgård, 2018, p. 87) (for a concrete example see 'Fig 2.17 Gulf War Diagram' in Kress & Van Leeuwen, 1996/2006, p. 65).

Due to these categorical ambiguities, and to the accompanying results of the inter-rater reliability test, only the broadest distinction level that Kress and Van Leeuwen suggest was used in the analysis of the entire corpus, namely the one that distinguishes between *narrative* and *conceptual structures*. This decision did not substantially weaken the validity of the analysis, as the main aim of the study was to investigate how connecting lines signal narrativity, not to reveal all subcategories of narrative structures that a connecting line can signify. Thus, the coding scheme used for the inter-rater reliability test was adjusted according to this decision and then used for the single-coded analysis of the whole corpus[1]. The codings were integrated into the relational database and the data analysis was then performed using Structured Query

1 Both coding schemes, more detailed results of the inter-rater reliability test and the list of all DVs included in the corpus, can be requested from the author.

Language (SQL). The data analysis resulted in frequency counts, reflecting whether and how the focused lines signal a narrative representational structure. No inferential statistical analysis was performed due to the limited representativeness of the DVs in the corpus for the whole population of publicly accessible DVs on the World Wide Web (see section *Corpus setup*).

Results

The results of the corpus analysis are reported in two steps. First, the results from the whole corpus of 163 DVs are presented. This includes an overview of how many of the focused lines signal which representational structure, of the DV types in the corpus, and of any instances of dynamic features and interactivity. In the second step, only the DVs containing a line that signals a narrative representational structure (further abbreviated as *DVs with vectors*) are focused on. The main aim of this step is to reveal what visual characteristics are used for signalling narrativity through the connecting lines in focus.

Of all 163 DVs in the corpus, the focused lines in 71 DVs signal a narrative representational structure, whereas 92 signal a conceptual structure. This means that the focused lines in 71 DVs were identified as vectors, indicating direction and thus representing a narrative structure. As presented in Table 4, most of these 71 DVs are route maps, followed by connection maps, line graphs and flow maps. All other DV types occur only once in this selection of DVs. Looking at this distribution from a broader perspective, 57 of the 71 DVs (36+16+5; that is ~80%) contain a map (in this article used as a broad term, referring to an iconic, visual representation of a fragment of the earth). When looking at the whole corpus, these DV types (route maps, connection maps, flow maps) account for only ~51% of the 163 DVs. Recapitulating, these results indicate that connecting lines which function as vectors are more prevalent in DV types based on maps than in other, more abstract DV types – although they occasionally also appear in other DV types.

Table 4. Distribution of DV types in DVs with vectors.

DV type	**Number (total = 71)**
route map	36
connection map	16
line graph	9
flow map	5
bump-chart	1
dumb-bell chart	1
hybrid of timeline and line graph (spline graph)	1
hybrid of linear process diagram and dendrogram	1
hybrid of arc diagram and bubble chart	1

Fifty of the 71 DVs with vectors have either dynamic or interactive features, or both, in parts of the DV other than the focused line itself. Such features include, amongst others, animated graphic objects, labels appearing through mouse-over and buttons that adjust or filter the data shown in the DV.

The following part deals with the visual characteristics of the line that are used to signal narrativity. As presented in Table 5, there are manifestation categories of five visual variables used to signal narrativity in this corpus. Dynamic features, that is, changes in size, position or pattern, represent the most prevalent group of visual characteristics, with 46 instances. Using broad descriptions, there are 36 times that the focused lines grow in length, six times they move and four times the repeated elements that constitute a pattern (in these cases, dashes or dots) move and thus indicate narrativity. The second most prevalent group of characteristics are the arrowheads or similar graphic elements indicating direction, with 39 instances. They occur 25 times at one end of the line, 10 times along the line, three times beside the line and once at both ends. Smooth colour transitions and transparency (seven instances each) and interaction (three instances) are also applied to signal narrativity through connecting lines in this corpus.

Table 5. Distribution of visual characteristics used to signal narrativity through connecting lines.

Count	Variable	Manifestation category	Broad description of the detected phenomenon
25	shape: arrowheads or similar	arrowhead or similar at one end	arrowhead(s) or similar direction-signifying graphic elements involved
10	shape: arrowheads or similar	arrowhead(s) or similar along the line	
3	shape: arrowheads or similar	arrow(head)(s) or similar besides the line	
1	shape: arrowheads or similar	arrowheads or similar at both ends	
7	colour	smooth transition	both combined: fading tail
7	clarity: transparency	low, medium or high	
3	interaction	possible	
36	dynamics in size	available	growing line
6	dynamics in position	available	moving line
4	dynamics in pattern	available	moving dashes or dots

Note. In some DVs, several visual characteristics are used simultaneously to signal narrativity. Therefore, the numbers add up to a total of more than 71.

If only one visual characteristic is used to indicate direction in a focused line, this is most often done by applying arrowheads or similar (25 instances, all variants included) or a line growing in length (22 instances) (see Table 6). Figure 4 presents a connection map where the icon of the plane is the only visual characteristic used as a direction-indicating element (comparable to an arrowhead). This technique was applied in a DV from the corpus, published by the Washington Post (Granados et al., 2016). In the DV presented in Figure 5, only the increase in line length is used to signal narrativity in a route map showing a person's movement (Outside in America team et al., 2017).

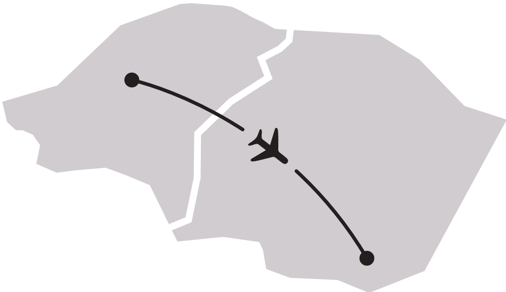


Figure 4. Schematic representation of a static connection map using an icon of a plane to signal narrativity (as similarly applied in Granados et al., 2016).

Figure 5. Screenshot of a dynamic route map showing a person's movement by a line growing in length (Outside in America team et al., 2017; © Guardian News & Media Ltd 2020, reprinted with permission).

Twenty times, a combination of several visual characteristics (as presented in Table 5) is used simultaneously within one focused line to signal narrativity, for example, some lines have a fading tail, are growing in length and moving at the same time. Table 6 shows which combinations of visual characteristics were found in the corpus, and how often this is the case.

Table 6. Combinations of the visual characteristics of the focused lines used to signal narrativity.

No. of character-istics used	Combined visual variables and manifestation categories	Broad description of the resulting phenomenon	No. of DVs (total = 71)
4	colour - smooth transition clarity: transparency - low, medium or high dynamics in position - available dynamics in size - available	growing and moving line with a fading tail	1
3	colour - smooth transition clarity: transparency - low, medium or high dynamics in position - available	moving line with a fading tail	3
3	colour - smooth transition clarity: transparency - low, medium or high shape - arrowhead or similar at one end	arrow with a fading tail	3
3	shape - arrowhead or similar at one end interaction - possible dynamics in size - available	arrow growing through interaction	3
2	dynamics in position - available dynamics in size - available	growing and moving line	1
2	dynamics in size - available dynamics in pattern - available	growing line with moving dashes	1
2	shape - arrowhead or similar at one end dynamics in size - available	growing line with arrowhead or similar	6
2	shape - arrowhead(s) or similar along the line dynamics in size - available	growing line with direction-indicating elements along the line	2
1	dynamics in size - available	growing line	22
1	shape - arrowhead or similar at one end	self-explanatory	13
1	shape - arrowhead(s) or similar along the line	self-explanatory	8
1	dynamics in pattern - available	line with moving dashes or dots	3
1	shape - arrow(head)(s) or similar besides the line	self-explanatory	3
1	shape - arrowheads or similar at both ends	self-explanatory	1
1	dynamics in position - available	moving line	1

A fading tail, achieved through a smooth colour transition to transparency, occurs several times together with other visual characteristics of the lines. If such a fading tail is used together with dynamic features (this appears four times in the corpus), the lines don’t have arrowheads or similar. Figure 6 shows such an example (Outside in America team et al., 2017). Relocations of homeless people are visualized here through lines with a fading tail that change their position. In static DVs, all three lines with fading tails are combined with an arrowhead.

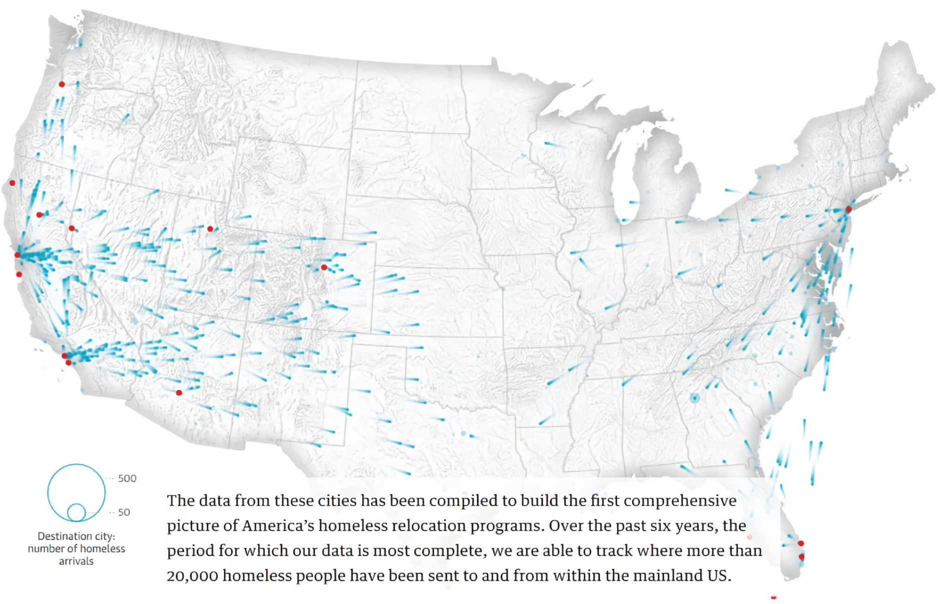


Figure 6. Screenshot of a dynamic connection map showing relocations of homeless people by moving lines with fading tails (Outside in America team et al., 2017; © Guardian News & Media Ltd 2020, reprinted with permission).

Arrowheads or similar at one end or along the line are used 11 times in combination with lines that are growing in length. Three of them also apply interactive techniques to signal narrativity. In these cases, the length of the line increases based on the reader’s scrolling. Thus, the reader can *feel* and control the action taking place. Two such examples can be found in an article about the transport of a rocket from Samara to the Vostochny Space Launch Centre (Rossiya Segodnya, 2016). A red line with an arrowhead, growing in length through scrolling, indicates narrativity.

Moving pattern parts (either dashes or dots) are used iconically four times to show a flow of items, like moving bombs, ships or people. Figure 7 shows a screenshot of one of them, where moving dashes represent people migrating to England and Wales (CLEVER°FRANKE, 2015). In one of the four cases, a dynamic change in line size is used in addition to the constant flow of items.

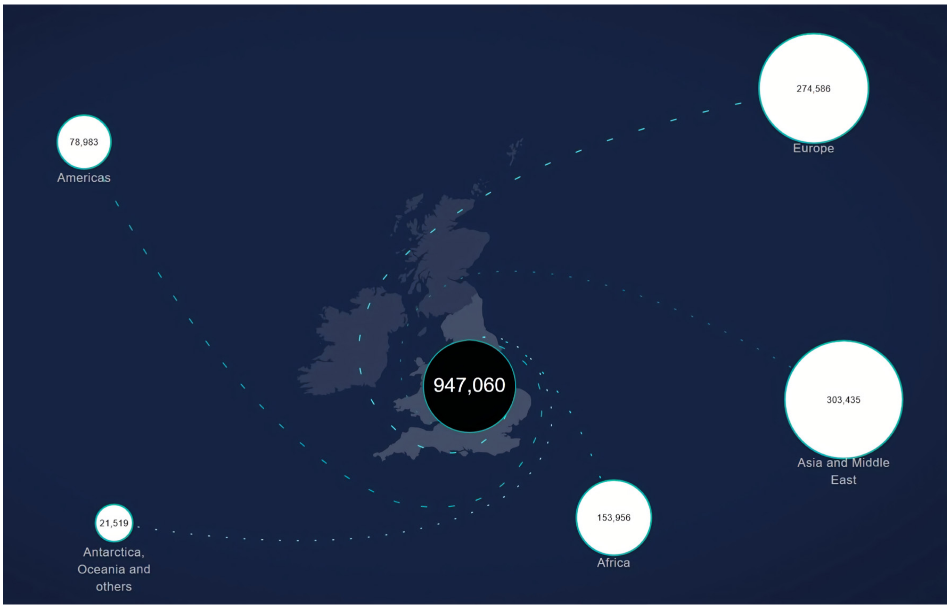


Figure 7. Screenshot of a dynamic connection map using moving dashes to represent migration to England and Wales (CLEVER°FRANKE, 2015; © 2015 by CLEVER°FRANKE, reprinted with permission).

Discussion

The results of the analysis reveal that the designers of the DVs in this corpus used various visual techniques to signal narrative claims with connecting lines. Sometimes they even applied a combination of several visual features for that purpose. However, two groups of visual characteristics stand out from the rest, due to the high frequency of use:

(1) arrow(head)s or similar direction-indicating graphic elements

(2) dynamic elements, above all, dynamics in line size (i.e., growing lines).

22

This result suggests that arrow(head)s or similar direction-indicating graphic elements, which are not dependent on digital production and display tools, are still commonly used in digital DV production. However, dynamic features that were not available in an analogue environment, particularly growing lines, seem to constitute an emerging alternative for indicating the same semiotic meaning potential. That reflects the social semiotic view of semiotic systems constantly being in flux, influenced by developments in technology and society (Van Leeuwen, 2005).

Another interesting finding concerns the use of fading tails to signal narrativity. It was used both in static (three times) and dynamic (four times) DVs, always accompanied by other techniques to signal narrativity. An additional arrowhead (or similar) at one end was used in the three static DVs, increasing line length (once) and line movement (four times) was used in addition to the fading tail in the four dynamic ones. Why the designers of all these cases assume that the technique of a fading tail needs additional visual hints in order to signify direction is an interesting question. One can assume that the effect of such fading tails is rooted in the readers' experiences from the physical world, for example, from watching an aircraft's condensation trail. Sometimes the aircraft is still visible at one end of the trail, sometimes not. Fading tails can also occur in photographs of movement using long exposure times. These examples suggest that a fading tail alone should be enough to signify direction. However, the analysis indicates that the designers of this corpus' award winning DVs don't regard this convention as strong enough to be reliable when communicating to a general audience. Thus, they add other visual features to strengthen their message.

Limitations and future research

Due to the setup of the corpus, the study and its results refer only to design practices found in the Western world within a specific time frame. Coding inconsistencies cannot be completely ruled out because of the coding method used (single-coding of most of the DVs), and the author's cultural background needs to be considered as an influencing factor. To counteract this issue, the detailed coding instructions were strictly followed, the codings were iteratively checked and an inter-rater reliability test was conducted. Furthermore, the fact that only one connecting line per DV was analysed (although more connecting lines might have been available) implies that other results could have been found if other lines were analysed. Thus, the results only represent the focused lines, which were consistently chosen because they convey the DV's main statement (see Table 1). The last limitation to mention is that the presented findings are only as fine-grained as the applied categorization system has allowed. Further investigations of each of the manifestation categories of Table 2 could have been made, or other categorization systems could have been developed. One example of a more fine-grained system to describe the shape of contour lines can be found in Johannessen (2017).

An issue calling for further research is the different kinds of arrowheads and similar direction-indicating elements applied in DVs. In this study, no difference was made between the graphical form of an arrowhead and an icon of a plane (as in Figure 4), or between different kinds of arrowheads. Such investigations are promising, especially when seen in relation to the historical context. Arrows such as the one displayed in Figure 1, for instance, may be hard to find in modern digital DVs. However, for such investigations, a larger corpus needs to be established to achieve a larger data basis for the single visual characteristics. Future research could also explore graphical conventions further by investigating more, older or newer DVs, or by looking into semiotic potentials other than directionality and narrativity, by using similar methodology.

Conclusion

This article presented a corpus study investigating the graphical variations of connecting lines that are used to signal narrativity in 163 publicly available digital DVs awarded between 2015 and 2018. It investigated which graphical variations of connecting lines, expressing narrativity, can be observed in that corpus and with what frequency and combinations they are used. The results demonstrate that the manifestation categories of five visual variables were used to signal narrativity in 71 DVs. Twenty times, two to four graphical characteristics were combined and applied in one line. Arrow(head)s (or similar direction-indicating graphic elements) and dynamics in line length are the two most prevalent techniques for signalling direction and indicating a narrative structure. Other visual techniques to signal narrativity are also used, but much less frequently. This result points to the existence of a graphical convention for signalling direction and narrativity through the two mentioned techniques. Due to that convention, these techniques may be easily understandable for a general audience. Other visual characteristics are obviously also applicable, however, from a visual literacy point of view, these may not be as well-known to readers and thus not so intuitively interpreted. Still, following conventions is not the only way to design effective DVs. The aesthetic aspect, for example, also plays an important role in that respect, motivating readers to take the time to read the DV (Brinch, 2020). The results of this analysis shall therefore not be turned into a recipe for how to design connecting lines to signal narrativity but constitute only a descriptive overview of what may be common to readers. In this sense, the findings contribute to the field of visual literacy, relevant in the contexts of education and DV design practice as well as DV tool development. Additionally, the findings may be used as a starting point for further investigations on graphical conventions.

Funding details

This work was supported by the Research Council of Norway under Grant number 259536.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Arnheim, R. (1982). *The power of the center: A study of composition in the visual arts*. University of California Press.

Bach, B., Stefaner, M., Boy, J., Drucker, S., Bartram, L., Wood, J., Ciuccarelli, P., & Engelhardt, Y. (2018). Narrative design patterns for data-driven storytelling. In N. H. Riche, C. Hurter, N. Diakopoulos, & S. Carpendale (Eds.), *Data-driven storytelling* (pp. 107–133). CRC Press/Taylor & Francis Group.

Belidor, B. F. de. (1737). *Architecture hydraulique, ou l'art de conduire, d'elever et de menager les eaux pour les différens besoins de la vie* (Vol. 1). Charles-Antoine Jombert; ETH-Bibliothek Zürich, Rar 6729. https://doi.org/10.3931/e-rara-27740

Bertin, J. (2011). *Semiology of graphics: Diagrams, networks, maps* (W. J. Berg, Trans.; Original title: "Semiologie graphique: Les diagrammes, les réseaux, les cartes"). ESRI Press. (Original work published 1967)

Brezina, V. (2018). Statistical choices in corpus-based discourse analysis. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 259–280). Routledge.

Brinch, S. (2020). What we talk about when we talk about beautiful data visualizations. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society* (pp. 259–275). Amsterdam University Press. http://oapen.org/search?identifier=1007903

Cameron, A. (2011). Ground zero – The semiotics of the boundary line. *Social Semiotics*, *21*(3), 417–434. https://doi.org/10.1080/10350330.2011.564391

CLEVER°FRANKE. (2015). *Migration in the census. A project for the Migration Observatory/Oxford University as part of Seeing Data (http://seeingdata.org/)*. http://seeingdata.cleverfranke.com/census/

DiBiase, D., MacEachren, A. M., Krygier, J. B., & Reeves, C. (1992). Animation and the role of map design in scientific visualization. *Cartography and Geographic Information Systems*, *19*(4), 201–214. https://doi.org/10.1559/152304092783721295

Dypedahl, M., & Hasselgård, H. (2018). *Introducing English grammar* (3rd ed.). Fagbokforlaget.

Fuller, G. (2002). The arrow--Directional semiotics: Wayfinding in transit. *Social Semiotics*, *12*(3), 231–244. https://doi.org/10.1080/10350330216376

Granados, S., Murphy, Z., Schaul, K., & Faiola, A. (2016, October 14). *Fenced out. A new age of walls—Episode 2* [Online newspaper]. The Washington Post. https://www.washingtonpost.com/graphics/world/border-barriers/europe-refugee-crisis-border-control/

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. (4th ed.). Advanced Analytics, LLC.

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science*, *30*(3), 581–592. https://doi.org/10.1207/s15516709cog0000_70

Horn, R. E. (1998). *Visual language. Global communication for the 21st century*. MacroVU.

Johannessen, C. M. (2017). Experiential meaning potential in the Topaz Energy logo: A framework for graphemic and graphetic analysis of graphic logo design. *Social Semiotics*, *27*(1), 1–20. https://doi.org/10.1080/10350330.2016.1187880

Kandinsky, W. (1928). *Punkt und Linie zu Fläche: Beitrag zur Analyse der malerischen Elemente* (2nd unchanged ed.). Albert Langen. http://archive.org/details/punktun00kand (Original work published 1926)

Kędra, J. (2018). What does it mean to be visually literate? Examination of visual literacy definitions in a context of higher education. *Journal of Visual Literacy*, *37*(2), 67–84. https://doi.org/10.1080/1051144X.2018.1492234

Kennedy, H., & Engebretsen, M. (2020). Introduction: The relationships between graphs, charts, maps, meanings, feelings, engagements. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society* (pp. 19–32). Amsterdam University Press. http://oapen.org/search?identifier=1007903

Kennedy, H., Hill, R. L., Aiello, G., & Allen, W. (2016). The work that visualisation conventions do. *Information, Communication & Society*, *19*(6), 715–735. https://doi.org/10.1080/1369118X.2016.1153126

Koylu, C., & Guo, D. (2017). Design and evaluation of line symbolizations for origin–destination flow maps. *Information Visualization*, *16*(4), 309–331. https://doi.org/10.1177/1473871616681375

Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge. (Original work published 1996)

Lechner, V. E. (2020). What a line can say: Investigating the semiotic potential of the connecting line in data visualizations. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society* (pp. 329–346). Amsterdam University Press. http://oapen.org/search?identifier=1007903

MacEachren, A. M. (1995). *How maps work: Representation, visualization, and design*. Guilford Press.

Malofiej Infographic World Summit. (n.d.). *Malofiej Infographic World Summit*. Retrieved 5 November 2018, from https://www.malofiejgraphics.com/awards/

Martire, R. (n.d.). *gac: Gwet's AC1 and AC2*. R Documentation. Retrieved 11 November 2019, from https://www.rdocumentation.org/packages/rel/versions/1.3.1/topics/gac

McKenna, S., Henry Riche, N., Lee, B., Boy, J., & Meyer, M. (2017). Visual narrative flow: Exploring factors shaping data visualization story reading experiences. *Computer Graphics Forum*, *36*(3), 377–387. https://doi.org/10.1111/cgf.13195

Outside in America team, Bremer, N., & Wu, S. (2017, December 20). *Bussed out: How America moves its homeless* [Digital newspaper page]. The Guardian. http://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study

Rossiya Segodnya. (2016, April 25). *Vostochny space launch center* (V. Grodsky, Trans.). РИА Новости. https://ria.ru/20160425/1054997772.html

Saint-Martin, F. (1990). *Semiotics of Visual Language* (Original title: 'Sémiologie du langage visuel'). Indiana University Press. (Original work published 1987)

Saulnier, A. (2005). La perception du mouvement dans les systèmes de visualisation d'informations. *Proceedings of the 17th Conference on l'Interaction Homme-Machine*, 185–192. https://doi.org/10.1145/1148550.1148574

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Stolper, C. D., Lee, B., Riche, N. H., & Stasko, J. (2016). *Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories* (Technical Report MSR-TR-2016-14). Microsoft Research. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/MSR-TR-2016-14-Storytelling-Techniques.pdf

Stolper, C. D., Lee, B., Riche, N. H., & Stasko, J. (2018). Data-driven storytelling techniques. In N. H. Riche, C. Hurter, N. Diakopoulos, & S. Carpendale (Eds.), *Data-driven storytelling* (pp. 85–105). CRC Press/Taylor & Francis Group.

The Information is Beautiful Awards Ltd. (2018). *KANTAR Information is Beautiful Awards*. https://www.informationisbeautifulawards.com/

The R Foundation for Statistical Computing. (2018). *R* (Version 3.5.1) [Computer software]. https://www.r-project.org/

Tversky, B., Heiser, J., Mackenzie, R., Lozano, S., & Morrison, J. (2008). Enriching animations. In R. Lowe & W. Schnotz (Eds.), *Learning with animation: Research implications for design* (pp. 263–285). Cambridge University Press.

Van der Waarde, K., & Westendorp, P. (2000, November). *The functions of arrows in user instructions*. IIID Expert forum on manual design, Eskilstuna, Sweden.

Van Leeuwen, T. (2005). *Introducing social semiotics*. Routledge.

Weber, W. (2020). Exploring narrativity in data visualization in journalism. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society* (pp. 295–311). Amsterdam University Press. http://oapen.org/search?identifier=1007903

Weger, G. (1999). *Cartographie Volume 1: Sémiologie graphique et conception cartographique*. École nationale des sciences géographiques. http://cours-fad-public.ensg.eu/pluginfile.php/1313/mod_resource/content/1/carto_vol1.pdf

Wong, B. (2011). Points of view: Arrows. *Nature Methods*, *8*(9), 701–701. https://doi.org/10.1038/nmeth.1676

Wong, W. (1993). *Principles of Form and Design*. John Wiley & Sons.

11 Reflections on Methodological Decisions

Every research project requires decisions on how to conduct the research, which in turn influence the outcome and research contributions. In the following, I reflect on the decisions I took along the different research steps by addressing three topics: (1) the general decision to use a social semiotic framework as the basis for my analyses, a framework originally developed for research on verbal language[29], (2) the decision to take a corpus approach in general, and (3) the manual coding process and its intersubjective validity.

Social semiotic theory, and several of its key concepts, is used as a basis for the method developed to conduct two corpus analyses. This decision was made because I had experienced social semiotics as a useful, systematic and very detailed tool to analyse DVs, and the exemplary visual text analyses presented by Kress and Van Leeuwen (1996/2006) were a good starting point. However, to be able to apply these concepts to the connecting lines in DVs, I also considered findings from other, related fields of research. This was necessary in particular to evaluate whether the connecting lines signal modality/uncertainty and to be able to verbally describe the visual appearance of the lines (see Chapter 5 and Section 7.4.1). In general, I experienced writing about visual texts on such a level of graphical detail as challenging. In this situation, social semiotic theory was helpful in my investigation of what semiotic functions the connecting lines might have, and I enjoyed the level of descriptive detail this theory offers. In Articles II and III I have demonstrated the usefulness of social semiotic concepts as a basis for the corpus analyses, at least on certain textual levels. In the practical process of coding the connecting lines present in the corpus, I sometimes encountered the descriptive detail as being too high, making it hard to decide which subcategory a line belonged to. Therefore, I refrained from using all detailed categorization levels offered by the theory, for example, of the representational structures used

29 This part relates to the points of critique presented in Section 4.5.

in Article III. On a very general level, I could probably have identified and investigated other semiotic potentials if I had based the study on frameworks other than social semiotics. The four semiotic functions identified in Article I are therefore not an exhaustive list, as commented on in Article I. Subsequently, one may ask what advantage was gained by using a social semiotic approach in this dissertation or, in other words, what "we see … now that we did not so do beforehand using the tools from other scholarly fields" (Machin, 2013, p. 349). In retrospect, perhaps the most striking new perspective was gained through the social semiotic view on uncertainty, building on the concept of modality as discussed by linguists and semioticians. The theoretical discussion on that issue (as presented in Article II) contributed significantly to a new, more inclusive understanding of uncertainty in DVs.

Reflecting on the decision to use a corpus approach, this was obviously not the only option to address the overall research questions of this dissertation. At least for assessing the semiotic functions of the connecting lines and the conventions regulating the relation between form and function, empirical testing in a lab environment or interviewing practitioners, for example, would also have been possible ways to go. However, it was necessary to first look at a systematically collected sample of DVs, given the interest in what graphical forms of connecting lines are currently used for specific semiotic functions in this kind of visual text material. Other approaches, such as the ones mentioned above, could have been combined with the corpus approach. Due to the time limitations and the lack of a pre-existing corpus and detailed analysis method, I decided to carry out only the corpus analyses. These may constitute the basis for further research, as described in Section 12.3.

The limitations of the specific corpus design have already been commented on in Articles II, III and Section 7.3.2, and I will therefore not elaborate on those in this section. Instead, I conclude this section with a reflection on the manual coding process and the intersubjective validity of the coding. Due to the lack of automated (machine-trained) analysis tools for single graphical elements in images, the DVs were coded manually. This was time-consuming, and additionally, there is always a danger of inconsistencies when human readers analyse large amounts of material. But I found that manual coding also bears advantages. Human raters have a specific background knowledge, based on previous experience, that may prove useful when coding. Such background knowledge had a significant impact, particularly on the corpus analysis presented in Article III.

My general knowledge helped me to identify narrativity indicators that automated tools may not have found. Icons or realistic depictions of planes (see Article III, Figure 4), for example, look quite different from arrowheads. In our "experiential world" (Kress & Van Leeuwen, 1996/2006, p. 46), however, it is unlikely that planes remain in the same location in the sky or that they fly backwards. Thus, they could be identified as direction indicators comparable to arrowheads.

A challenge related to ratings done by humans rather than computers is to ensure rating consistency and reliability. As described in Section 7.4.2, Articles II and III, I approached this challenge with performing inter-rater reliability tests. Approximately 15% of the DVs in the corpus were double-coded; the remaining ones were only single-coded in both analyses. Without a doubt, a higher number of raters would have been ideal, both for the inter-rater reliability tests and for the final complete ratings. However, performing the coding required theoretical background knowledge, and keeping apart graphical form and semiotic function of the focused lines on such level of detail was challenging. Letting laypeople do the coding with these coding schemes would have been possible only with extensive training. Or otherwise, random answers, inconsistencies and errors would have been expected. This is also one reason why I refrained from letting raters from Amazon Mechanical Turk (https://www.mturk.com/) do the rating, as has been done by other studies needing ratings of visual text (such as Boukhelifa et al., 2012; Hiippala et al., 2019; Kembhavi et al., 2016b). However, nothing speaks against taking a more experimental perspective, including more raters, on some detail issues in future research (see Section 12.3).

12 Conclusion

The aim of this dissertation was to investigate from a social semiotic perspective what different semiotic functions connecting lines in digital DVs may have, to find empirical evidence for how these semiotic functions are graphically realized and whether conventions can be observed. Four possible semiotic functions of connecting lines are identified in the theoretical part of the study (as presented in Article I), two of which (namely to signal modality and narrativity) are investigated in more detail, theoretically and empirically (as presented in Articles II and III).

The results of the two corpus analyses show that a broad range of visual characteristics of the studied connecting lines are used to signal modality and narrativity. The results also support the hypothesis underlying this dissertation, that certain graphical conventions exist to regulate how certain meaning potentials can be signalled in digital DVs and how these signs should be interpreted. Concerning modality, the findings of the corpus analysis presented in Article II indicate a convention for signalling lowered probability (based on data uncertainty) through pattern change (in other words, interruptions of the line). Lowered reliability, as the second aspect of modality, was signalled by a wider range of visual characteristics of the lines. Therefore, no clear graphical conventions were observed concerning this kind of meaning potential. Concerning narrativity, the findings of the corpus analysis presented in Article III point to a convention for signalling narrativity through lines by applying arrow(head)s or similar direction-indicating graphic elements and dynamic features, above all, dynamics in line size (i.e., growing lines).

In the following, I present the overall research contributions of this dissertation, expected implications for DV practice, and suggestions for further studies

12.1 Research Contributions

The key contributions of this dissertation are empirical, connected to the two corpus analyses presented in Articles II and III. The results of these analyses provide empirical evidence for how the two semiotic potentials investigated (modality and narrativity) are realized by connecting lines in the DVs of the systematically collected corpus. Based on these results, I have revealed indications of current conventions for how modality and narrativity are expressed by connecting lines in current, publicly available digital DVs. This dissertation provides the first corpus-based investigation of this kind of visual text, focusing exclusively on the semiotic potentials of a single graphical element. The results from the study widen our understanding of current semiotic practices, which are relevant to the fields of design and media studies as well as the visual literacy field and to practitioners (see Section 12.2).

Besides these empirical contributions, the dissertation also offers theoretical and methodological contributions specifically relevant to the fields of social semiotics and multimodality studies. As for the theoretical contributions, this dissertation provides the first social semiotic view on lines in DVs, specifically on what semiotic functions connecting lines in DVs can have. Furthermore, the dissertation introduces the concept of modality to the discussion concerning uncertainty in DVs. Bringing fields of theory and practice together, the study offers a comprehensive understanding of the relation between modality and uncertainty, and reliability and probability, respectively. From a more general point of view, the dissertation contributes to a more fine-grained vocabulary for describing the semiotic functions and the visual appearance of lines in digital environments.

From a methodological perspective, the dissertation offers an analysis method for investigating modality and narrativity in DV corpora, that may be further used and adjusted for other research foci (see Section 12.3). In relation to this method, it further offers a DV corpus with 163 systematically selected DVs that are organized in a relational database, including metadata and analysis results as a stable data basis. This corpus may be reused for other investigations or serve as an example for future corpus developments. Its setup as a relational database made it necessary to work in a strict, systematic way, which makes future extensions and further data analyses easily possible.

12.2 Implications for DV Practice

The empirical findings of the dissertation have several possible implications, particularly for the following four groups of practitioners: DV designers, DV tool developers, practitioners setting up DV design guidelines and educators teaching visual literacy. The way these four groups may be influenced by the research presented in this dissertation is explained in greater detail in the following. Evidently, a prerequisite for this to happen is that a targeted dissemination of the research results takes place.

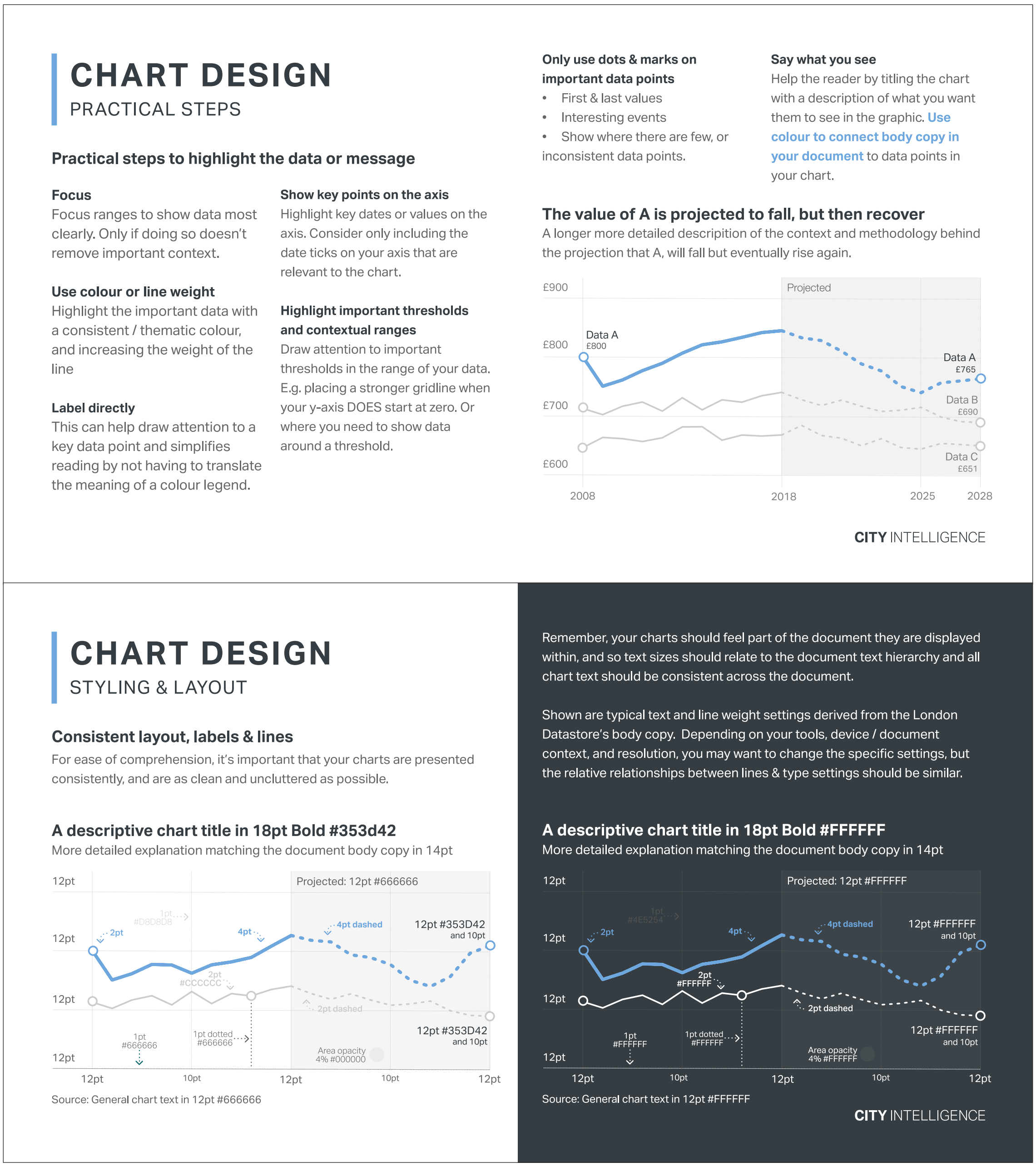
Both amateur and professional DV designers make design decisions, consciously or unconsciously, when developing a new visualization. The knowledge generated in this dissertation creates awareness of the consequences that detailed design decisions have on the meaning potential of the DVs. Likewise, it may contribute to a more fine-tuned graphical articulation being used in future DVs.

The design decisions that practitioners can take depend heavily on the tools they have at hand, and on what kinds of graphical choices these tools offer. Thus, DV tool developers strongly influence the diversity of graphical expressions used in DVs, and the development of graphical conventions. Little innovation and change in graphical conventions may happen if the DV design tools make easily available only the graphical expressions that had already been used before the advent of digital production and display tools. The same effect occurs if such expressions are pre-selected in the tools. Conversely, if DV design tools offer too many choices, there is a risk of chaotic appliance of the visual characteristics of the elements of DVs. The research presented in this dissertation illustrates the space of possible visual characteristics of connecting lines, as well as current conventions for what meaning potentials they may have – guiding tool developers to balance between too few and too many options.

DV design style guides, which Cesal defines as “standards for formatting and designing representations of information, like charts, graphs, tables, and diagrams” (2019b), have recently been developed by various companies and organizations. They are developed as an extension to their often already existing design standards (Cesal, 2019b). The list of such guides published by Cesal (2019a) includes guidelines of companies and organizations like IBM (2016), Adobe (2020), Google (n.d.), BBC (2019) and the London City Intelligence unit (Brondbjerg & GLA City Intelligence, 2019). The latter, for example, even contains recommendations and examples of how to use the lines’ characteristics for signalling certain kinds of meanings. Figure 16 shows how, when following these guidelines, an uncertain projection may be visualized in a line graph. On the one hand, the results of this dissertation may serve as a source of information for people developing such style guides. On the other hand, the dissertation also points to the influence of this group of persons on the diversity of graphical expression, similar to the influence of DV tool designers.

Figure 16

Page 8 and 10 From the London City Intelligence Data Design Guidelines



Note. From “City Intelligence Data Design Guidelines” by M. Brondbjerg, GLA City Intelligence, 2019 (https://data.london.gov.uk/blog/city-intelligence-data-design-guidelines/). Retrieved April 30, 2020. Copyright 2019 by Greater London Authority. Reprinted under the Open Government Licence

The skills needed for reading DVs are gained not only “through informal processes by observation of norms and behaviour” (Kędra, 2018, p. 73), but instead, further education is necessary. For educators teaching visual literacy, it is key to have access to sound knowledge on the graphical possibilities and graphical conventions regulating the use of certain visual characteristics of graphical elements (such as lines) to indicate a certain meaning potential. Therefore, this dissertation will hopefully affect the field of visual literacy and influence what educators may teach.

12.3 Suggestions for Further Research

The research presented in this dissertation could be followed up by investigations of various related issues. In the following, a few options that seem promising are outlined, bearing in mind that this list does not aim to be exhaustive.

Based on the corpus that was set up for this dissertation, more fine-grained categories of the visual variables could be introduced, describing what such lines look like. One might, for instance, investigate which different kinds of direction-indicating endings of lines (like arrows or icons of planes) are used to signal narrativity. Maybe, after having used abstract arrow endings for a long time, one could observe a move to more naturalistic depictions of direction-indicating elements? Or, to mention an example connected to Article II, one could have differentiated between only a few and many forces (i.e., curves or jags) when describing the form of the lines. This would have come closer to the investigations on logos, done by Johannessen (2018). One could ask whether and with what effect sketchy lines (such ones that have many forces) were used. Also, not only the detail level, but also the global level of DVs (as described in Article II) could be included in new corpus analyses of modality. In this case, one could study how the visual expressions on the different levels influence each other. Semiotic potentials other than modality and narrativity could, of course, be investigated as well.

When thinking of extending the given corpus with additional DVs, one might investigate the historic development of line use by adding both older and newer examples. The setup of the corpus as a relational database easily allows for extensions and performing various other queries over several tables. By adding DVs that represent visual cultures other than the Western one, one could also study cultural differences. Adding DVs where graphic elements other than the connecting line are the most central, these other graphic elements and their abilities to signal certain semiotic potentials might be studied. Considering the challenges represented by the corpus approach taken in this dissertation, one could also think of getting more, independent raters to further verify the results of the corpus analyses presented in Articles II and III.

Finally, it would be interesting to include the practitioners' and users' perspectives more by using approaches other than the corpus-based one. Issues calling for research in such cases include the users' reactions to different ways of indicating modality or narrativity; how practitioners prefer to signal modality and narrativity by lines or other graphic elements and why; where in the area of tension between strictly following conventions and trying out unconventional ways of depiction the practitioners see themselves; and what their experiences are in that respect. Possible methods for such investigations include experimental eye-tracking studies, interviews or focus groups, for instance.

Reflecting on the abovementioned opportunities for further research, it becomes clear that there is a broad range of work left to be done, and that the theoretical, methodological, and empirical work presented in this dissertation represents a possible starting point for many studies along this line of research. Regardless of which threads of further research are picked up first, it will be interesting to see how future DVs are designed and how the involved parties use their opportunities to shape the graphical conventions that influence how we – and those coming after us – understand the world.

References

Adobe. (2020). *Data visualization fundamentals*. Spectrum. https://spectrum.adobe.com/page/data-visualization-fundamentals/

Aiello, G. (2020). Inventorizing, situating, transforming: Social semiotics and data visualization. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society* (pp. 49–62). Amsterdam University Press. http://oapen.org/search?identifier=1007903

American Psychological Association (Ed.). (2010). *Publication manual of the American Psychological Association* (6th ed). American Psychological Association.

Andersen, T. H., Boeriis, M., Maagerø, E., & Tønnessen, E. S. (2015). *Social semiotics: Key figures, new directions*. Routledge.

Arnheim, R. (1982). *The power of the center: A study of composition in the visual arts*. University of California Press.

Avgerinou, M. D. (2009). Re-viewing visual literacy in the "Bain d' Images" era. *TechTrends*, *53*(2), 28–34. https://doi.org/10.1007/s11528-009-0264-z

Axel-Tober, K., & Gergel, R. (2015). Modality and mood in Formal Syntactic approaches. In J. Nuyts & J. Van der Auwera (Eds.), *The Oxford handbook of modality and mood*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199591435.013.21

Bach, B., Stefaner, M., Boy, J., Drucker, S., Bartram, L., Wood, J., Ciuccarelli, P., & Engelhardt, Y. (2018). Narrative design patterns for data-driven storytelling. In N. H. Riche, C. Hurter, N. Diakopoulos, & S. Carpendale (Eds.), *Data-driven storytelling* (pp. 107–133). CRC Press/Taylor & Francis Group.

Bateman, J. A., Delin, J., & Henschel, R. (2004). Multimodality and empiricism. Preparing for a corpus-based approach to the study of multimodal meaning-making. In E. Ventola, C. Charles, & M. Kaltenbacher (Eds.), *Perspectives on Multimodality* (pp. 65–87). John Benjamins. https://doi.org/10.1075/ddcs.6.06bat

Bateman, J. A., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundations, research and analysis: A problem-oriented introduction*. De Gruyter Mouton.

BBC. (2019). *BBC visual and data journalism cookbook for R graphics*. https://bbc.github.io/rcookbook/

Bell, P. (2001). Content analysis of visual images. In T. Van Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 10–34). SAGE.

Bernstein, B. (1981). Codes, modalities, and the process of cultural reproduction: A model. *Language in Society*, *10*(3), 327–363.

Bertin, J. (1967). *Semiologie graphique: Les diagrammes, les réseaux, les cartes*. Mouton/Gauthier-Villars.

Bertin, J. (2011). *Semiology of graphics: Diagrams, networks, maps* (W. J. Berg, Trans.; Original title: "Semiologie graphique: Les diagrammes, les réseaux, les cartes"). ESRI Press. (Original work published 1967)

Björkvall, A. (2009). *Den visuella texten: Multimodal analys i praktiken*. Hallgren & Fallgren.

Bonneau, G.-P., Hege, H.-C., Johnson, C. R., Oliveira, M. M., Potter, K., Rheingans, P., & Schulz, T. (2014). Overview and state-of-the-art of uncertainty visualization. In C. D. Hansen, M. Chen, C. R. Johnson, A. E. Kaufman, & H. Hagen (Eds.), *Scientific Visualization* (pp. 3–27). Springer. https://doi.org/10.1007/978-1-4471-6497-5_1

Boogaart, R., & Fortuin, E. (2015). Modality and mood in Cognitive Linguistics and Construction Grammars. In J. Nuyts & J. Van der Auwera (Eds.), *The Oxford handbook of modality and mood*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199591435.013.23

Boukhelifa, N., Bezerianos, A., Isenberg, T., & Fekete, J.-D. (2012). *Evaluating sketchy lines for the visualization of qualitative uncertainty* (Inria Research Report No. 7910). Research Centre Saclay - Île-de-France. https://tobias.isenberg.cc/personal/papers/Boukhelifa_2012_ESL.pdf

Brezina, V. (2018). Statistical choices in corpus-based discourse analysis. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 259–280). Routledge.

Brinton, W. C. (1914). *Graphic methods for presenting facts*. The Engineering Magazine Company.

Brinton, W. C. (1939). *Graphic presentation*. Brinton Associates.

Brondbjerg, M., & GLA City Intelligence. (2019). *City intelligence data design guidelines. (Intelligence-Data-Styleguide_v2_0.pdf)*. https://data.london.gov.uk/blog/city-intelligence-data-design-guidelines/

Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. New Riders.

Cameron, A. (2011). Ground zero – the semiotics of the boundary line. *Social Semiotics*, *21*(3), 417–434. https://doi.org/10.1080/10350330.2011.564391

Canadian Association of Journalists. (2018). *CAJ Awards*. http://caj.ca/Awards

Cesal, A. (2019a). *Data visualization style guidelines*. Medium - Nightingale. https://docs.google.com/spreadsheets/d/1F1gm5QLXh3USC8ZFx_M9TXYxmD-X5JLDD0oJATRTuIE/edit#gid=1679646668

Cesal, A. (2019b, July 10). *What are data visualization style guidelines?* Medium – Nightingale. https://medium.com/nightingale/style-guidelines-92ebe166addc

Chalmers, M., & Pierquin, M. (Eds.). (2003). *Oxford Business French Dictionary: French-English*. Oxford University Press. https://doi.org/10.1093/acref/9780191739491.001.0001

Collin, F. (2013). Social Constructivism. In B. Kaldis (Ed.), *Encyclopedia of philosophy and the social sciences* (Vol. 2). SAGE. http://web.a.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzU5MjU1N19fQU41?sid=207c5e38-55e5-4380-9a79-2e022baf9810@sessionmgr4010&vid=0&format=EK&lpid=n13&rid=0

Connect. (2003). In T. F. Hoad (Ed.), *Concise Oxford dictionary of English etymology* (online version). Oxford University Press. https://www.oxfordreference.com/view/10.1093/acref/9780192830982.001.0001/acref-9780192830982-e-3266?rskey=yuPdIF&result=1 (Original work published 1996)

Dasgupta, A., Chen, M., & Kosara, R. (2012). Conceptualizing visual uncertainty in parallel coordinates. *Computer Graphics Forum*, *31*/3, 1015–1024. https://doi.org/10.1111/j.1467-8659.2012.03094.x

Delanty, G., & Strydom, P. (2003). *Philosophies of social science. The classic and contemporary readings*. Open University Press.

DiBiase, D., MacEachren, A. M., Krygier, J. B., & Reeves, C. (1992). Animation and the role of map design in scientific visualization. *Cartography and Geographic Information Systems*, *19*(4), 201–214. https://doi.org/10.1559/152304092783721295

Engebretsen, M., & Weber, W. (2018). Graphic modes: The visual representation of data. In C. Cotter & D. Perrin (Eds.), *The Routledge handbook of language and media* (pp. 277–295). Routledge. https://doi.org/10.4324/9781315673134

Engelhardt, Y. (2002). *The language of graphics. A framework for the analysis of syntax and meaning in maps, charts and diagrams* [Doctoral dissertation, University of Amsterdam]. UvA-DARE. http://dare.uva.nl/search?arno.record.id=105970

Euclid, & Heath, T. L. (1956). *The thirteen books of Euclid's elements: Introduction and books I, II* (2nd ed., Vol. 1). Dover. (Original work published n.d.)

Fast Company & Inc. (2018). *Innovation By Design*. Fast Company. https://www.fastcompany.com/innovation-by-design/2018

Fergusson, R. (Ed.). (1992). *The Penguin dictionary of English synonyms and antonyms* (revised). Penguin Books.

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten* (2nd ed.). Analytics Press.

Forceville, C. J. (1999). Educating the eye? Kress and Van Leeuwen's Reading Images: The Grammar of Visual Design (1996). *Language and Literature: International Journal of Stylistics*, *8*(2), 163–178. https://doi.org/10.1177/096394709900800204

Forceville, C. J. (2007). Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook. *Journal of Pragmatics*, *39*(6), 1235–1238. https://doi.org/10.1016/j.pragma.2007.02.007

Forceville, C. J. (2010). Review of The Routledge Handbook of Multimodal Analysis. *Journal of Pragmatics*, *42*(9), 2604–2608. https://doi.org/10.1016/j.pragma.2010.03.003

Friendly, M., & Denis, D. F. (2001). *Timeline*. Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization. http://www.datavis.ca/milestones

Friman, S. (2009). M.A.K. Halliday. In S. Kolstrup, G. Agger, P. Jauert, & K. Schrøder (Eds.), *Medie- og kommunikationsleksikon* (pp. 192–193). Samfundslitteratur.

Fuller, G. (2002). The arrow--Directional semiotics: Wayfinding in transit. *Social Semiotics*, *12*(3), 231–244. https://doi.org/10.1080/10350330216376

Global Editors Network. (2018). *Data Journalism Awards*. Data Journalism Awards. https://www.datajournalismawards.org/

Google. (n.d.). *Data visualization*. Material Design. https://material.io/design/communication/data-visualization.html#style

Green, M. (1991, October 22). *Toward a perceptual science of multidimensional data visualization: Bertin and beyond. Presented as "The visual psychophysics of data visualization."* 2nd IEEE Conference on Visualization, Workshop on multivariate, multiparameter, multidimensional systems, San Diego, CA. http://graphics.stanford.edu/courses/cs448b-06-winter/papers/Green_Towards.pdf

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. (4th ed.). Advanced Analytics, LLC.

Habermann, H. (2015). *Grundlagen der Gestaltung: Industrie-Design / Kommunikations-Design*. Cuvillier Verlag.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.

Halliday, M. A. K. (2004). *An introduction to functional grammar* (3rd ed. revised by Matthiessen, C.M.I.M.). Arnold (Hodder Education). (Original work published 1985)

Halliday, M. A. K., & Hasan, R. (1994). *Cohesion in English* (13th impression). Longman Group. (Original work published 1976)

Hammer, N. (2008). *Mediendesign für Studium und Beruf: Grundlagenwissen und Entwurfssystematik in Layout, Typografie und Farbgestaltung*. Springer.

Hayward, S. (1984). *Computers for animation*. Focal Press. http://archive.org/details/computersforanim00hayw

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science*, *30*(3), 581–592. https://doi.org/10.1207/s15516709cog0000_70

Hiippala, T. (2020, May 5). *AI2D-RST dataset*. https://github.com/thiippal/AI2D-RST (Original work published 2019)

Hiippala, T., Alikhani, M., Haverinen, J., Kalliokoski, T., Logacheva, E., Orekhova, S., Tuomainen, A., Stone, M., & Bateman, J. A. (2019). *AI2D-RST: A multimodal corpus of 1000 primary school science diagrams*. https://arxiv.org/abs/1912.03879

Hill, A. (n.d.). *Design Research: Visualizing Uncertainty*. Uncertainty. https://uncertainty.io/art/

Hodge, R., & Kress, G. (1988). *Social semiotics*. Cornell University Press.

Holsanova, J. (in press). Uncovering scientific and multimodal literacy through audio description. *Journal of Visual Literacy*, *39*(3).

Horn, R. E. (1998). *Visual language: Global communication for the 21st century*. MacroVU.

Hullman, J. (2019). Confronting unknowns. *Scientific American*, *321*(3), 80–83.

Hullman, J. (2020). Why authors don't visualize uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 130–139. https://doi.org/10.1109/TVCG.2019.2934287

IBM. (2016, January 15). *Art meets science: Data visualization guidelines*. IBM Design Language (V1). http://www.ibm.com/design/v1/language/experience/data-visualization

Iliinsky, N., & Steele, J. (2011). *Designing data visualizations*. O'Reilly Media.

Ingold, T. (2007). *Lines: A Brief History*. Routledge.

Jensen, K. B. (1995). *The social semiotics of mass communication*. SAGE.

Jewitt, C. (2009). An introduction to multimodality. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 14–27). Routledge.

Jewitt, C., & Oyama, R. (2001). Visual meaning: A social semiotic approach. In T. Van Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 134–156). SAGE.

Johannessen, C. M. (2017). Experiential meaning potential in the Topaz Energy logo: A framework for graphemic and graphetic analysis of graphic logo design. *Social Semiotics*, *27*(1), 1–20. https://doi.org/10.1080/10350330.2016.1187880

Johannessen, C. M. (2018). The challenge of simple graphics for multimodal studies: Articulation and time scales in fuel retail logos. *Visual Communication*, *17*(2), 163–185. https://doi.org/10.1177/1470357217746811

Johannessen, C. M., & Van Leeuwen, T. (2018). (Ir)Regularity. In C. M. Johannessen & T. Van Leeuwen (Eds.), *The materiality of writing: A trace-making perspective* (pp. 175–192). Routledge.

Kandinsky, W. (1928). *Punkt und Linie zu Fläche: Beitrag zur Analyse der malerischen Elemente* (2nd unchanged ed.). Albert Langen. http://archive.org/details/punktun00kand (Original work published 1926)

Kandinsky, W. (1947). *Point and line to plane: Contribution to the analysis of the pictorial elements* (H. Dearstyne & H. Rebay, Trans.; Original-title: "Punkt und Linie zu Fläche"). Solomon R. Guggenheim Foundation. http://archive.org/details/pointlinetoplane00kand (Original work published 1926)

Kaufmann, M., & Kaufmann, S. (2015). Modality and mood in Formal Semantics. In J. Nuyts & J. Van der Auwera (Eds.), *The Oxford handbook of modality and mood*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199591435.013.24

Kędra, J. (2018). What does it mean to be visually literate? Examination of visual literacy definitions in a context of higher education. *Journal of Visual Literacy*, *37*(2), 67–84. https://doi.org/10.1080/1051144X.2018.1492234

Kembhavi, A., Salvato, M., Kolve, E., Seo, M., Hajishirzi, H., & Farhadi, A. (2016a). *AI2 Diagram Dataset*. Allen Institute for AI. https://allenai.org/data/[id]?id=diagrams

Kembhavi, A., Salvato, M., Kolve, E., Seo, M., Hajishirzi, H., & Farhadi, A. (2016b). A diagram is worth a dozen images. *Proceedings of the 14th ECCV. Lecture Notes in Computer Science*, *9908*, 235–251. https://doi.org/10.1007/978-3-319-46493-0_15

Kennedy, H., & Engebretsen, M. (2020). Introduction: The relationships between graphs, charts, maps, meanings, feelings, engagements. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society* (pp. 19–32). Amsterdam University Press. http://oapen.org/search?identifier=1007903

Kennedy, H., Hill, R. L., Aiello, G., & Allen, W. (2016). The work that visualisation conventions do. *Information, Communication & Society*, *19*(6), 715–735. https://doi.org/10.1080/1369118X.2016.1153126

Kinkeldey, C., MacEachren, A. M., & Schiewe, J. (2014). How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, *51*(4), 372–386. https://doi.org/10.1179/1743277414Y.0000000099

Kinkeldey, C., & Senaratne, H. (2018). Representing Uncertainty. *The Geographic Information Science & Technology Body of Knowledge*, 2018(Q2). https://doi.org/10.22224/gistbok/2018.2.3

Kirk, A. (2016). *Data visualisation: A handbook for data driven design*. SAGE.

Klee, P. (1991). Beiträge zur bildnerischen Formlehre. In *Kunst, Lehre: Aufsätze, Vorträge, Rezensionen und Beiträge zur bildnerischen Formlehre* (2nd ed., pp. 91–314). Reclam. (Original work published 1921–1922)

Kløvstad Langberg, Ø. (2019, March 5). Kinas ledighet har knapt rikket seg på 15 år. At noe ikke stemmer, er åpenbart. *Aftenposten*. https://www.aftenposten.no/article/ap-gPqk6a.html

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, *3*(3), 185–226. https://doi.org/10.1002/acp.2350030302

Koylu, C., & Guo, D. (2017). Design and evaluation of line symbolizations for origin–destination flow maps. *Information Visualization*, *16*(4), 309–331. https://doi.org/10.1177/1473871616681375

Krämer, S. (2016). *Figuration, Anschauung, Erkenntnis: Grundlinien einer Diagrammatologie*. Suhrkamp.

Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge. (Original work published 1996)

Lauesen, S. (2005). *User interface design: A software engineering perspective*. Pearson Education/Addison Wesley.

Ledin, P., & Machin, D. (2018). *Doing visual analysis: From theory to practice*. SAGE.

Line. (2003a). In T. F. Hoad (Ed.), *Concise Oxford dictionary of English etymology* (online version). Oxford University Press. https://www.oxfordreference.com/view/10.1093/acref/9780192830982.001.0001/acref-9780192830982-e-8761?rskey=TlRF88&result=3 (Original work published 1996)

Line. (2003b). In T. F. Hoad (Ed.), *Concise Oxford dictionary of English etymology* (online version). Oxford University Press. https://www.oxfordreference.com/view/10.1093/acref/9780192830982.001.0001/acref-9780192830982-e-8762?rskey=TlRF88&result=1 (Original work published 1996)

Link. (2003). In T. F. Hoad (Ed.), *Concise Oxford dictionary of English etymology* (online version). Oxford University Press. https://www.oxfordreference.com/view/10.1093/acref/9780192830982.001.0001/acref-9780192830982-e-8779?rskey=iItx8D&result=2 (Original work published 1996)

Liu, L., Padilla, L., Creem-Regehr, S. H., & House, D. H. (2019). Visualizing uncertain tropical cyclone predictions using representative samples from ensembles of forecast tracks. *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 882–891. https://doi.org/10.1109/TVCG.2018.2865193

Lotman, J. (1977). *The structure of the artistic text* (R. Vroon, Trans.). University of Michigan. https://monoskop.org/images/3/3e/Lotman_Jurij_The_Structure_of_the_Artistic_Text_1977.pdf (Original work published 1971)

Lu, M., Wang, S., Lanir, J., Fish, N., Yue, Y., Cohen-Or, D., & Huang, H. (2020). Winglets: Visualizing association with uncertainty in multi-class scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 770–779. https://doi.org/10.1109/TVCG.2019.2934811

MacEachren, A. M. (1992). Visualizing uncertain information. *Cartographic Perspectives*, *13*, 10–19. https://doi.org/10.14714/CP13.1000

MacEachren, A. M. (1995). *How maps work: Representation, visualization, and design*. Guilford Press.

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2496–2505. https://doi.org/10.1109/TVCG.2012.279

Machin, D. (2009). Multimodality and theories of the visual. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 181–190). Routledge.

Machin, D. (2013). What is multimodal critical discourse studies? *Critical Discourse Studies*, *10*(4), 347–355. https://doi.org/10.1080/17405904.2013.813770

Mainberger, S., & Ramharter, E. (Eds.). (2017). *Linienwissen und Liniendenken*. De Gruyter.

Malofiej Infographic World Summit. (n.d.-a). *Awards – Rules Malofiej 27*. Retrieved November 5, 2018, from https://www.malofiejgraphics.com/awards-rules/

Malofiej Infographic World Summit. (n.d.-b). *Malofiej 24: Awards list*. Retrieved November 26, 2018, from https://www.malofiejgraphics.com/wp-content/uploads/2016/03/M24_-AWARDS-LIST_OK.pdf

Malofiej Infographic World Summit. (n.d.-c). *Malofiej 25: Awards list*. Retrieved November 26, 2018, from https://www.malofiejgraphics.com/wp-content/uploads/2017/03/M25_AwardsList_OK3.pdf

Malofiej Infographic World Summit. (n.d.-d). *Malofiej 26: Awards list*. Retrieved November 26, 2018, from https://www.malofiejgraphics.com/wp-content/uploads/2018/03/M26-Awards-List-DEF.pdf

Malofiej Infographic World Summit. (n.d.-e). *Malofiej Infographic World Summit*. Retrieved November 5, 2018, from https://www.malofiejgraphics.com/awards/

Martire, R. (n.d.). *gac: Gwet's AC1 and AC2*. R Documentation. Retrieved November 11, 2019, from https://www.rdocumentation.org/packages/rel/versions/1.3.1/topics/gac

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McGranaghan, M. (1993). A cartographic view of spatial data quality. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *30*(2–3), 8–19. https://doi.org/10.3138/310V-0067-7570-6566

McKenna, S., Henry Riche, N., Lee, B., Boy, J., & Meyer, M. (2017). Visual narrative flow: Exploring factors shaping data visualization story reading experiences. *Computer Graphics Forum*, *36*(3), 377–387. https://doi.org/10.1111/cgf.13195

Meeks, E. (n.d.). *Semiotic*. Retrieved May 7, 2020, from https://semiotic.nteract.io/guides/sketchy-painty-patterns

Meeks, E. (2017, September 11). Sketchy Data Visualization in Semiotic. *Medium*. https://medium.com/@Elijah_Meeks/sketchy-data-visualization-in-semiotic-5811a52f59bc

Morrison, J. L. (1974). A theoretical framework for cartographic generalization with emphasis on the process of symbolization. In *International Yearbook of Cartography* (Vol. 14, pp. 115–127). Kirschbaum.

Munzner, T. (2014). *Visualization analysis and design*. CRC Press/Taylor & Francis Group.

NESH. (2016). *Guidelines for research ethics in the social sciences, humanities, law and theology* (4th ed.). https://www.etikkom.no/globalassets/documents/english-publications/60127_fek_guidelines_nesh_digital_corr.pdf

NESH. (2019). *A guide to internet research ethics* (2nd ed.). https://www.etikkom.no/globalassets/documents/publikasjoner-som-pdf/forskningsetisk-veileder-for-internettforskning/a-guide-to-internet-research-ethics.pdf

Neuman, W. L. (2006). *Social research methods: Qualitative and quantitative approaches* (6th ed). Pearson Education.

Nielsen, J. (1995). *Multimedia and hypertext: The internet and beyond* (2nd ed.). Morgan Kaufmann/Academic Press.

NxtMedia. (n.d.). *NODA Awards*. Retrieved November 5, 2018, from https://noda.media/noda-awards/

Online News Association. (n.d.). *The Online Journalism Awards*. Retrieved November 5, 2018, from https://awards.journalists.org/

Oxford University Press. (n.d.). *The Oxford English Corpus*. Oxford Dictionaries | English. Retrieved November 2, 2018, from https://en.oxforddictionaries.com/explore/oxford-english-corpus

Pauwels, L. (2011a). An integrated conceptual framework for visual social research. In E. Margolis & L. Pauwels (Eds.), *The SAGE handbook of visual research methods* (pp. 3–23). SAGE.

Pauwels, L. (2011b). Researching websites as social and cultural expressions: Methodological predicaments and a multimodal model for analysis. In E. Margolis & L. Pauwels (Eds.), *The SAGE handbook of visual research methods* (pp. 570–589). SAGE.

Peirce, C. S. (1935). *Collected papers of Charles Sanders Peirce: Vol. 5 (Pragmatism and pragmaticism)* (C. Hartshorne & P. Weiss, Eds.). Belknap Press of Harvard University Press.

Peirce, C. S. (1960). *Collected papers of Charles Sanders Peirce: Vol. 2 (Elements of logic)* (C. Hartshorne & P. Weiss, Eds.). Belknap Press of Harvard University Press.

Poulin, R. (2018). *The language of graphic design: An illustrated handbook for understanding fundamental design principles* (revised and updated). Rockport.

Ravelli, L. J., & Van Leeuwen, T. (2018). Modality in the digital age. *Visual Communication*, *17*(3), 277–297. https://doi.org/10.1177/1470357218764436

Reinsel, D., Gantz, J., & Rydning, J. (2018). *The digitization of the world from edge to core. An IDC white paper*. IDC. https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

Relate. (2003). In T. F. Hoad (Ed.), *Concise Oxford dictionary of English etymology* (online version). Oxford University Press. https://www.oxfordreference.com/view/10.1093/acref/9780192830982.001.0001/acref-9780192830982-e-12658?rskey=n2DuWR&result=1 (Original work published 1996)

Richards, C. J. (1984). *Diagrammatics: An investigation aimed at providing a theoretical framework for studying diagrams and for establishing a taxonomy of their fundamental modes of graphic organization* [Doctoral dissertation, Royal College of Art London]. EThOS. https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.335273

Riche, N. H., Hurter, C., Diakopoulos, N., & Carpendale, S. (2018). Introduction. In N. H. Riche, C. Hurter, N. Diakopoulos, & S. Carpendale (Eds.), *Data-driven storytelling* (pp. 1–15). CRC Press/Taylor & Francis Group.

Saint-Martin, F. (1990). *Semiotics of Visual Language* (Original title: "Sémiologie du langage visuel"). Indiana University Press. (Original work published 1987)

Sanders, T., & Pander Maat, H. (2006). Cohesion and coherence: Linguistic approaches. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed., pp. 591–595). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00497-1

Saulnier, A. (2005). La perception du mouvement dans les systèmes de visualisation d'informations. *Proceedings of the 17th Conference on l'Interaction Homme-Machine*, 185–192. https://doi.org/10.1145/1148550.1148574

Saulnier, A., Thievre, J., & Viaud, M.-L. (2006). La perception du mouvement dans la visualisation: Le cas des graphes. Motion perception in visualization: The case of graphs. *Revue d'Interaction Homme-Machine*, *7*(2), 55–78.

Saussure, F. de. (1983). *Course in general linguistics* (C. Bally, A. Sechehaye, & A. Riedlinger, Eds.; R. Harris, Trans.; Original title: "Cours de linguistique générale."). Duckworth. (Original work published 1916)

Schrøder, K. (2009). Semiologi/Semiotik. In S. Kolstrup, G. Agger, P. Jauert, & K. Schrøder (Eds.), *Medie- og kommunikationsleksikon* (pp. 480–483). Samfundslitteratur.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Sinclair, J. M. (2004). *Trust the text: Language, corpus and discourse* (R. Carter, Ed.). Routledge.

Skaggs, S. (2017). *FireSigns. A semiotic theory for graphic design*. MIT Press.

Skeels, M., Lee, B., Smith, G., & Robertson, G. (2008). Revealing uncertainty for information visualization. *AVI '08 Proceedings of the Working Conference on Advanced Visual Interfaces*, 376–379. https://doi.org/10.1145/1385569.1385637

SND. (n.d.-a). *The Best of Digital Design. A competition of SND, the world's foremost advocate of excellence in news design*. Retrieved November 5, 2018, from https://www.snd.org/bodd/

SND. (n.d.-b). *The Best of News Design. Showcasing the medal winners of the SND Creative Competition*. Retrieved November 5, 2018, from https://www.snd.org/bond/

Statista. (2018, December). *Volume of data/information created worldwide from 2010-2025*. Statista. https://www.statista.com/statistics/871513/worldwide-data-created/

Stokes, J. (2003). *How to do media and cultural studies*. SAGE.

Stolper, C. D., Lee, B., Riche, N. H., & Stasko, J. (2016). *Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories* (Technical Report MSR-TR-2016-14). Microsoft Research. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/MSR-TR-2016-14-Storytelling-Techniques.pdf

Stolper, C. D., Lee, B., Riche, N. H., & Stasko, J. (2018). Data-driven storytelling techniques. In N. H. Riche, C. Hurter, N. Diakopoulos, & S. Carpendale (Eds.), *Data-driven storytelling* (pp. 85–105). CRC Press/Taylor & Francis Group.

Su, W. (2017). *NYC FOODIVERSE - Storytelling of NYC Restaurant*. http://nycfoodiverse.com/

Taboada, M. (2019). Cohesion and conjunction. In G. Thompson, W. L. Bowcher, L. Fontaine, & D. Schönthal (Eds.), *The Cambridge handbook of systemic functional linguistics* (pp. 311–332). Cambridge University Press.

Tak, S., Toet, A., & Van Erp, J. (2014). The perception of visual uncertainty representation by non-experts. *IEEE Transactions on Visualization and Computer Graphics*, *20*(6), 935–943. https://doi.org/10.1109/TVCG.2013.247

The Information is Beautiful Awards Ltd. (2015). *2015 – The Winners*. Kantar Information Is Beautiful Awards. https://www.informationisbeautifulawards.com/news/116-2015-the-winners

The Information is Beautiful Awards Ltd. (2016a). *2016 – The Winners*. Kantar Information Is Beautiful Awards. https://www.informationisbeautifulawards.com/news/188-2016-the-winners

The Information is Beautiful Awards Ltd. (2016b). *Now open: The Kantar Information is Beautiful Awards 2016*. Information Is Beautiful. https://informationisbeautiful.net/2016/now-open-kantar-information-is-beautiful-awards-2016/

The Information is Beautiful Awards Ltd. (2017). *Winners 2017*. Kantar Information Is Beautiful Awards. https://www.informationisbeautifulawards.com/news/259-winners-2017

The Information is Beautiful Awards Ltd. (2018a). *KANTAR Information is Beautiful Awards*. https://www.informationisbeautifulawards.com/

The Information is Beautiful Awards Ltd. (2018b). *Showcase*. Kantar Information Is Beautiful Awards. https://www.informationisbeautifulawards.com/showcase?action=index-&award=2016&controller=showcase&page=1&pcategory=short-list&type=awards

The Information is Beautiful Awards Ltd. (2020). *Awards Terms & Conditions*. https://www.informationisbeautifulawards.com/awards-terms-and-conditions

The R Foundation for Statistical Computing. (2018). *R* (Version 3.5.1) [Computer software]. https://www.r-project.org/

Thomas, M. (2014). Evidence and circularity in multimodal discourse analysis. *Visual Communication*, *13*(2), 163–189. https://doi.org/10.1177/1470357213516725

Threadgold, T. (1986). The semiotics of Vološinov, Halliday and Eco. *American Journal of Semiotics*, *4*(3/4), 107–142. https://doi.org/10.5840/ajs198643/423

Toet, A., Van Erp, J. B. F., Boertjes, E. M., & Van Buuren, S. (2019). Graphical uncertainty representations for ensemble predictions. *Information Visualization*, *18*(4), 373–383. https://doi.org/10.1177/1473871618807121

Tversky, B., Heiser, J., Mackenzie, R., Lozano, S., & Morrison, J. (2008). Enriching animations. In R. Lowe & W. Schnotz (Eds.), *Learning with animation: Research implications for design* (pp. 263–285). Cambridge University Press.

Unknown, Previous owners: Celtis, K., Peutinger, K., & Savoyen-Carignan, E. von. (ca. 1200). *Tabula Peutingeriana* [Map]. Österreichische Nationalbibliothek. http://data.onb.ac.at/rep/10002029

Unsworth, L. (1997). Scaffolding reading of science explanations: Accessing the grammatical and visual forms of specialized knowledge. *Reading*, *31*(3), 30–42. https://doi.org/10.1111/1467-9345.00061

Van der Auwera, J., & Aguilar, A. Z. (2015). The history of modality and mood. In J. Nuyts & J. Van der Auwera (Eds.), *The Oxford handbook of modality and mood*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199591435.013.4

Van der Waarde, K., & Westendorp, P. (2000, November). *The functions of arrows in user instructions*. IIID Expert forum on manual design, Eskilstuna, Sweden.

Van Leeuwen, T. (1999). *Speech, music, sound*. Macmillan.

Van Leeuwen, T. (2005). *Introducing social semiotics*. Routledge.

WAN-IFRA. (2018). *World Digital Media Awards*. WAN-IFRA Events. https://events.wan-ifra.org/awards

Ware, C. (2013). *Information visualization. Perception for design* (3rd ed.). Morgan Kaufmann.

Weber, W. (2020). Exploring narrativity in data visualization in journalism. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society* (pp. 295–311). Amsterdam University Press. http://oapen.org/search?identifier=1007903

Webster, J. J. (2019). Key terms in the SFL model. In G. Thompson, W. L. Bowcher, L. Fontaine, & D. Schönthal (Eds.), *The Cambridge handbook of systemic functional linguistics* (pp. 35–54). Cambridge University Press.

Weger, G. (1999). *Cartographie Volume 1: Sémiologie graphique et conception cartographique*. École nationale des sciences géographiques. http://cours-fad-public.ensg.eu/pluginfile.php/1313/mod_resource/content/1/carto_vol1.pdf

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt: II. *Psychologische Forschung. Zeitschrift für Psychologie und ihre Grenzwissenschaften. Festschrift für Carl Stumpf.*, *4*, 301–350. https://doi.org/10.1007/BF00410640

White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library Trends*, *55*(1), 22–45. https://doi.org/10.1353/lib.2006.0053

Wilke, C. O. (2019). *Fundamentals of data visualization: A primer on making informative and compelling figures*. O'Reilly Media.

Wilson, A. A., & Landon-Hays, M. (2016). A social semiotic analysis of instructional images across academic disciplines. *Visual Communication*, *15*(1), 3–31. https://doi.org/10.1177/1470357215609213

Wong, B. (2011). Points of view: Arrows. *Nature Methods*, *8*(9), 701–701. https://doi.org/10.1038/nmeth.1676

Wong, W. (1993). *Principles of form and design*. John Wiley & Sons.

Wood, J., Isenberg, P., Isenberg, T., Dykes, J., Boukhelifa, N., & Slingsby, A. (2012). Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2749–2758. https://doi.org/10.1109/TVCG.2012.262

Appendices

Appendix A: Lists of Award-Winners Relevant for the Corpus

Table 1

Winner List of the Malofiej Awards 24, Including Only Those Winners That Were Categorized as "Online" on the Original Winner List (Malofiej Infographic World Summit, n.d.-b)

No.	Title of the winner	DV (1) / info-graphic (0)	Connecting line in main role (1) / subordinate role (2) / not present (0)	Award-winner no. in Appendix B
SPECIAL AWARDS				
'Miguel Urabayen' Award for the Best Map				
1	Unaffordable country: Where can you afford to buy a house?	1	0	
Climate Change and Environmental Commitment				
2*	What's Really Warming the World?	1	1	6
Human Rights				
3	The Next to Die: National, State, Case link	1	1	37
Equality and Woman's Promotion				
4	This Is How Fast America Changes Its Mind	1	1	38
ONLINE GRAPHICS				
Breaking News				
5	Germanwings: The last minutes of Germanwings flight 4U9525	1	1	39
6	How the Hajj Stampede Unfolded	1	1	40
7	Investigating the Philadelphia Amtrak Train Derailment	1	1	41
8	How Missing Jet's Debris Could Have Floated to Réunion	1	1	42
9	What Happened on the Germanwings Flight	1	1	43
10	Extent of the Damage From the Nepal Earthquake	not accessible on July 25, 2018		
11	Amtrak derailment	1	1	44
Features				
12	Homan Square: a portrait of Chicago's detainees	1	0	
13/3	The Next to Die: National, State, Case link	see above		
14	Flag Stories	0	0	
15	How Syrians Are Dying	1	0	

16	Courts Are Shaking Up House Elections in 2016	1	0	
17/4	This Is How Fast America Changes Its Mind	see above		
18	Illegal Tusk Trade	1	1	45
19/1	Unaffordable country: Where can you afford to buy a house?	see above		
20	What China Has Been Building in the South China Sea	0	0	
21	A visual guide to 75 years of major refugee crises…	1	1	46
22	Europe Deporting	1	0	
23*	Failure Factories: Why Pinellas County is the worst place in Florida	1	1	5
24	Rethinking Detroit	1	0	
25	The New New York Skyline	0	0	
26	Over 1,100 Health Data Breaches, but Few Fines	1	0	
27	A 3-D View of a Chart That Predicts The Economic Future: The Yield Curve	1	1	47
28	How to Catch a Spoofer	1	2	
29	How China's economic slowdown could weigh on the rest	1	1	48
30	Moscow Marathon results	URL not found on July 24, 2018		
31	The Dawn Wall: El Capitan's Most Unwelcoming Route	1	1	49
32	The Complete History Of the NFL	1	1	50
33	Ceni's Free Kicks in 7 Steps	1	1	51
34	Tour de France: the climb of Alpe d'Huez	1	1	52
35	Wayne Rooney breaks Bobby Charlton's England goalscoring record	1	1	53
36	Workers' Comp Benefits: How Much is a Limb Worth?	1	0	
37/2	What's Really Warming the World?	see above		
38	How measles spreads when kids get vaccinated - and when they don't	1	0	
39	Which Flight Will Get You There Fastest?	1	1	105
40	Greenland Is Melting Away	1	0	
41	Star Wars: The Force Accounted	1	0	
42	A tortured heap of towers': the London skyline of tomorrow	0	0	
43	Bieber, Diplo & Skrillex Make A Hit	0	0	
44*	Every death in Game of Thrones	0	0	
Portfolios				
45	The New York Times Breaking News Portfolio: 1, 2, 3, 4, 5, 6	These aren't single DV websites that were awarded, but portfolios with several subpages and many DVs.		
46	The Guardian Portfolio Breaking News 1, 2, 3, 4, 5, 6, 7, 8, 9, 10			
47	Berliner Morgenpost Interactive Team			
48	FiveThirtyEight Interactive team portfolio: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10			
49	The New York Times Features Portfolio: 1, 2, 3, 4, 5, 6, 7, 8, 9			
50	The New York Times Portfolio: Derek Watkins: 1, 2, 3, 4, 5, 6,			
51	Bloomberg News Why Climate Change Is True			

Criteria				
52/23	Failure Factories: Why Pinellas County is the worst place in Florida to be black and go to public school	see above		
53	The Climate Change Pledges Are In. Will They Fix Anything?	1	1	54
54/33	Ceni's Free Kicks in 7 Steps	see above		
55	How ISIS Expands	0	0	
56/14	Flag Stories	see above		
57	Scaling Everest	0	0	
58	Blue whale: The Largest animal on Earth (app)	0	0	
59	You Draw It: How Family Income Predicts Children's College Chances	1	1	55
Social Media Graphics				
60	The shape-shifting US income distribution	1	0	
61	Darren Sharper Facebook Video	0	0	
62	How Missing Jet's Debris Could Have Floated to Réunion	1	1	56
No. of award winners included into the corpus: 23; Of these 23: No. of award winners that were also awarded by the Kantar Information is Beautiful Award: 2.				

Note. Those winners, the DV(s) of which have been selected for the corpus, are marked with grey shading and the field in the last column contains the award-winner no. as in Appendix B.
Winners that were awarded both by the Kantar Information is Beautiful Awards and by the Malofiej Awards are marked with "*" in the first column. Winners that were awarded twice in the same award, are marked with "/no. of the first award in this list" in the first column.
Due to space limits, further details about authors, institution, country and prize type are omitted in this table.

Table 2

Winner List of the Malofiej Awards 25, Including Only Those Winners That Were Categorized as "Online" on the Original Winner List (Malofiej Infographic World Summit, n.d.-c)

No.	Title of the winner	DV (1) / infographic (0)	Connecting line in main role (1) / subordinate role (2) / not present (0)	Award-winner no. in Appendix B
BEST OF SHOW AWARDS				
1	Olympic Races Social Series	1	0	
SPECIAL AWARDS				
Miguel Urabayen Award for the Best Map				
2	The Two Americas 2016	1	0	
Climate Change and Environmental Commitment Best Graphics Award				
3	A Sharp Increase in 'Sunny Day' Flooding	1	0	
Human Rights Best Graphics Award				
4	A new age of walls	1	1	57
Equality and Woman's Promotion Best Graphics Award				
5	Vagina Dispatches	0	0	

ONLINE GRAPHICS				
Breaking News				
6	A Trail of Terror in Nice, Block by Block	0	1	
7	1,000 golds	1	0	
8	Missed the final presidential debate? We've mapped it out for you	1	1	58
9	What Happened Inside the Orlando Nightclub	0	1	
10	What lawmakers said after orlando	1	0	
11	What two body language experts saw at the second presidential debate	0	1	
12	Interactivo: todas las mentiras de Trump y Clinton frente a frente en un vistazo	1	0	
13	Choice and Chance	0	0	
14	The Keith Scott Shooting: A 3-D Reconstruction	1	1	59
15	The Chapecoense aircraft crashed because it ran out of fuel while waiting to land	1	1	60
16	The Turøy accident	1	1	61
Features				
17	Lost Cause: Seeing America Through the Losing Candidates' Map	1	0	
18	Five Years of Fukushima	1	0	
19	Palmyra: Liberated Pearl	0	1	
20	Bolts from the blue	1	2	
21	The UK in Europe: a visual guide to Brexit	1	2	
22	Gun	1	0	
23/4	A New Age of Walls	see above		
24	Who's winning? Find out how your income compares with every other generation	1	1	62
25/3	A Sharp Increase In 'Sunny Day' Flooding	see above		
26	Money, Race and Success: How Your School District Compares	1	1	63
27/2	The Two Americas of 2016	see above		
28	Swing countries	1	1	64
29	Sequence of wrong choices	1	1	65
30	Election maps lie	1	0	
31	Terror's daily reality	not accessible on 25 July 2018		
32	MH370	1	0	
33	Fortress of Fences	1	1	66
34	EU referendum: full results and analysis	1	0	
35	London mayoral election: how to make the most of your second vote	1	0	
36	Who supports Donald Trump? The new Republican center of gravity	1	2	
37	Since Sandra	1	0	
38	Penn Station Reborn	0	0	
39	A Bullet Could Hit Me and My Kids Anytime	0	0	
40	Popularity	1	1	67
41	Wikipedia editors	1	0	
42	When was America great	1	1	68
43	A Closer look at Calais	1	0	

44	These candidates live farthest away from their voters	1	1	69
45	Amazon Doesn’t Consider the Race of Its Customers. Should It?	1	1	70
46	40 Percent of the Buildings in Manhattan Could Not Be Built Today	0	0	
47	Nine years of waiting for St. Petersburg football arena	1	1	71
48	Manufacturing jobs are returning to some places. But these jobs are different	1	1	72
49*	Marathon 2016 in time-lapse	1	1	35
50	The Fine Line: Olympian Simone Biles	1	1	73
51	Sailing at wind	1	1	74
52	Decisive Moments at the Rio Olympics, Frame By Frame	0	1	
53	How Usain Bolt Came From Behind Again to Win Gold	1	0	
54	Anniversary Soccer World Cup 1986	1	1	75
55	Pitch by pitch: How Clayton Kershaw dominates hitters	1	1	76
56	How evictions have laid bare Rio’s real Olympic legacy	1	2	
57	Olympic Races, in Your Neighborhood	1	1	77
58	Sizing up Olympics	1	0	
59	A Bear’s Eye View of Yellowstone	1	1	78
60	Rewind the Red Planet	0	0	
61	Vostochny Space Launch Center	1	1	79
62	The dark side of Guardian comments	1	1	80
63	Seeking Pluto’s Frigid Heart	0	0	
64	Why 27 Million Are Still Uninsured Under Obamacare	1	0	
65	A deadly crisis: mapping the spread of America’s drug overdose epidemic	1	2	
66	Here’s $100. Can you win $1.5 billion at Powerball?	1	0	
67	Every Lata Mangeshkar song in one graphic	1	0	
68	The artists Prince ushered into the spotlight	0	0	
Portfolios				
69	The New York Times Breaking News Portfolio	These aren’t single DV websites that were awarded, but portfolios with several subpages and many DVs.		
70	The Washington Post Breaking News Portfolio			
71	Berliner Morgenpost Features Portfolio			
72	FiveThirtyEight Features Portfolio			
73	Hindustan Times Features Portfolio			
74	The Washington Post Features Portfolio			
75	The New York Times Features Portfolio			
76	Reuters News Graphics Service Features Portfolio			
77	Los Angeles Times Features Portfolio			
78	The Guardian Features Portfolio			
79	Quartz Features Portfolio			
Criteria				
80	How One of the Deadliest Hajj Accidents Unfolded	1	1	81

81	Electing the next US president: The numbers game	1	0	
82	Forcibly Displaced	1	0	
83*	This Chart Shows Who Marries CEOs, Doctors, Chefs and Janitors	1	1	17
Social Media Graphics				
84	Breaking the Black Box	0	0	
85	Major bombings in Turkey	1	0	
86	Explaining Tor on Twitter	0	0	
87/1	Olympic Races Social Series	see above		
88	Post Graphics Donald Trump's hand gestures	0	0	
Specials				
89	Segredos Olímpicos. Olympic Secrets - A series of 28 videos shown an often invisible	0	0	
90	At Death's Door. An investigation into capital punishment in India	1	0	
No. of award winners included into the corpus: 27; Of these 27: No. of award winners that were also awarded by the Kantar Information is Beautiful Award: 2.				

Note. Those winners, the DV(s) of which have been selected for the corpus, are marked with grey shading and the field in the last column contains the award-winner no. as in Appendix B.
Winners that were awarded both by the Kantar Information is Beautiful Awards and by the Malofiej Awards are marked with "*" in the first column. Winners that were awarded twice in the same award, are marked with "/no. of the first award in this list" in the first column.
Due to space limits, further details about authors, institution, country and prize type are omitted in this table.

Table 3

Winner List of the Malofiej Awards 26, Including Only Those Winners That Were Categorized as "Online" on the Original Winner List (Malofiej Infographic World Summit, n.d.-d)

No.	Title of the winner (or award type respectively)	DV (1) / infographic (0)	Connecting line in main role (1) / subordinate role (2) / not present (0)	Award-winner no. in Appendix B
BEST OF SHOW AWARDS				
1	The Science of Hummingbirds	0	0	
MAIN AWARDS				
Miguel Urabayen Award for the Best Map				
2	Tracking Harvey's Destructive Path Through Texas and Louisiana	1	1	82
Climate Change and Environmental Commitment Best Graphic Award				
3	Antarctic Dispatches	1	1	83
Human Rights Best Graphic Awards				
4	Life in the camps	1	0	
Equality and Women's Promotion Best Graphic Awards				
5	Can we talk about the gender pay?	1	1	84
ONLINE GRAPHICS				
Breaking News				
6	Thousands Cried for Help as Houston Flooded	1	0	

7	Nine Rounds a Second: How the Las Vegas Gunman Outfitted a Rifle to Fire Faster	1	0	
8/2	Tracking Harvey's Destructive Path Through Texas and Louisiana	see above		
9	Emmanuel Macron's official portrait is a symbolic celebration of centrism	0	0	
10	What Lies in Irma's Path	1	1	85
11	Hundimiento Ara San Juan	not accessible on October 2, 2018		
12	How the Las Vegas shooting unfoilded	0	0	
13	How Northam gained in a more polarized Virginia	1	1	86
14	How cable networks reacted to Comey's hearing	1	0	
15	Inside the Las Vegas Gunman's Mandalay Bay Hotel Suite	0	0	
16	Mexico City Was Built on an Ancient Lake Bed. That Makes Earthquakes…	1	0	
Features				
17	North Korea: A rising threat	1	0	
18/4	Life in the Camps	see above		
19	Did the Turkish President's Security Detail Attack Protesters in Washington?	1	0	
20	Mass Exodus	1	0	
21	A deadly crossing	1	1	87
22	Bussed out - How America moves its homeless	1	1	88
23	Want to fix gun violence in America? Go local.	1	1	89
24	How key senators changed their positions to bring down Obamacare repeat	1	1	90
25	10 Minutes. 12 Gunfire Bursts. 30 Vides. Mapping the Las Vegas Massacre	0	0	
26	How California's Most Destructive Wildfire Spread, Hour by Hour	1	0	
27	Astronauts	1	0	
28	How death has changed over 100 years in Britain	1	0	
29	Incarceration by Executive Order	1	2	
30	What is North Korea trying to hit	1	2	
31	After Hurricane Maria, Puerto Rico is still in the dark	1	1	91
32	Trump's Nuclear Arsenal	1	0	
33	Outdated and Unreliable: FEMA's Faulty Flood Maps Put Homeowners at Risk	1	0	
34	Diving into Urban-Rural Prejudice	1	1	92
35	That's how Berlin is ticking along your line	1	1	93
36	If you're black	1	0	
37/5	Can we talk about gender pay?	see above		
38	Our Broken Economy, in One Simple Chart	1	1	94
39	110 N.F.L. Brains	1	0	
40	The United States of oil and gas	1	1	95
41/1	The Science of Hummingburds	see above		
42	Cracking the mystery of egg shape	1	1	96
43	Cassini's Grand Tour	1	1	97
44	Resurrecting a Dragon	0	0	
45	What Makes Mako Sharks the Fastest Sharks in the Ocean?	1	0	

46	Here’s every total solar eclipse happening in your lifetime.	1	1	98
47	Escribe tu “zip code” y descubre cómo se verá el eclipse solar e tu ciudad	1	1	99
48	The panda diaspora: where China shares ist bears	not accessible on October 2, 2018		
49	The tallest statues in the world	1	0	
50	The three-degree world: the cities that will be drowned by global warming	1	1	100
51	How Ed Sheeran Made ‘Shape of You’ the Year’s Biggest Track	0	0	
52	Which tech world did you grow up in?	1	1	101
53	Don’t waste your time at Disneyland. Here’s how to avoid the lines	1	1	102
54*	How do you draw a circle?	1	1	25
55	The Year In Language	1	1	103
56*	How to Fix a Toilet And Other Things We Couldn’t Do Without Search	1	0	
58	How To Spot A Front-Runner On The ‘Bachelor’ or ‘Bachelorette’	1	2	
Portfolios				
58/6	Thousands Cried for Help as Houston Flooded	see above		
59/12	Las Vegas Shooting	see above		
60	National Geographic Magazine Online Features Portfolio	These aren’t single DV websites that were awarded, but portfolios with several subpages and many DVs.		
61	Nexo Jornal Features Portfolio			
62/30	North Korea targets	see above		
63	The Next Bechdel Test	1	0	
64/20	Mass exodus	see above		
Criteria				
65	Robots Are Coming for These Wall Street Jobs	1	0	
66	Wonsan: Where tourism meets tanks	0	0	
67	Beautiful In English	1	1	104
68/43	Cassini’s Grand Tour	see above		
69	The Uber game	0	0	
70	Watch five hurricanes churn through the Atlantic in one month	1	0	
No. of award winners included into the corpus: 24; Of these 24: No. of award winners that were also awarded by the Kantar Information is Beautiful Award: 1.				

Note. Those winners, the DV(s) of which have been selected for the corpus, are marked with grey shading and the field in the last column contains the award-winner no. as in Appendix B.
Winners that were awarded both by the Kantar Information is Beautiful Awards and by the Malofiej Awards are marked with “*” in the first column. Winners that were awarded twice in the same award, are marked with “/no. of the first award in this list” in the first column.
Due to space limits, further details about authors, institution, country and prize type are omitted in this table.

Table 4

Winner List of the Kantar Information Is Beautiful Awards 2015 (The Information is Beautiful Awards Ltd, 2015)

No.	Title of the winner	DV (1) / info-graphic (0)	Digital (1) / non digital (0)	Connecting line in main role (1) / subordinate role (2) / not present (0)	Award-winner no. in Appendix B
Arts, Entertainment & Pop Culture					
1	Vaccines and Infectious Diseases	1	1	0	
2	A World of Languages	0	1	0	
3	Rise of Partisanship	1	1	1	1
4	Close The Gap	1	1	1	2
Infographic					
5	Rare Earth Elements	1	0	1	
6	UFO Sightings	1	0	1	
7	Valar Morghulis	1	1	0	
8	Rent, Salary, Price of Houses	not accessible on September 20, 2018			
Interactive					
9	How Ebola Spreads	not accessible on September 20, 2018			
10	The Counted: People Killed by Police in the United States in 2015	1	1	0	
11	Job Market Tracker	1	1	2	
12	Histography	1	1	0	
Motion Infographic					
13	The Fallen of WW2	1	1	0	
14	How Many Trees Are There in World	1	1	0	
15	Syrian War in 5 Minutes	1	1	1	3
16	Charting Culture	1	1	1	4
Data Journalism					
17	German Unification	1	1	2	
18*	Why Pinellas County Is The Worst Place in Florida to Be Black and Go to Public School	1	1	1	5
19*	What's Really Warming the World?	1	1	1	6
20	California's Getting Fracked	1	1	2	
Mini and Mobile Visualization					
21	How To Build A Human	0	1	0	
22	What's across the Ocean from You When You're at the Beach	0	1	0	
23	The Dude Map: How Americans Refer to Their Bros	1	1	0	
24	Here's How Much of Your Life the United States Has Been at War	1	1	0	
Free Dataviz Tool					
25	D3.js	These aren't single DV websites that were awarded, but tools.			
26	R				
27	Tableau Public				
28	Quadrigram				

Dataviz Website					
29	Visualising Data	These are websites with several subpages and many DV projects and single DVs. It is thus the collection that was awarded.			
30	Flowing Data				
31	Visualoop				
32	Visual Complexity				
Data Visualization Project					
33	Dear Data	This project includes many different kinds of DVs, comparable to several DV websites. It is thus not clear which one was awarded.			
34	On Broadway	1	1	0	
35	2014 Annual Report	1	0	2	
36	Surgeon Scorecard	1	1	0	
Best Team					
37	Berliner Morgen Post for: Here's How Loud It Is at Your Doorstep; M29 - The Bus Route of Contrasts; Where the Population of Europe is Growing - and Where's it's Declining	This is a team being awarded.			
Individual Achievement					
38	Visual Intro to Machine Learning, free congress	1	1	2	
Student					
39	for Freedom in Countries	This is a student being awarded.			
Community Vote - based on entire shortlist 2015					
40	The World in 2015	1	1	2	
41	Visualisation of Global Weather Conditions	1	1	1	7
42	Drones	1	1	0	
Commercial Client Project					
43	Migration in the Census and in the News	1	1	1	8
44/8	Rent, Salary, Price of Houses	see above			
45	LA Dodgers Digital Trading Room	not publicly available			
Internal Business Project					
46	London Squared Map: Making the City Easier to Read	not publicly available			
47	RPP Intranet	not publicly available			
48	DHL GoGreen Solutions Workbook	not publicly available			
Commercial Studio					
49	Accurat	These are studios being awarded.			
50	Domestic Streamers				
51	Periscopic				
52	Clever Franke				
Most beautiful					
53/33	Dear Data	see above			
No. of award winners included into the corpus: 8; Of these 8: No. of award winners that were also awarded by the Malofiej Awards (but are counted here): 2.					

Note. Those winners, the DV(s) of which have been selected for the corpus, are marked with grey shading and the field in the last column contains the award-winner no. as in Appendix B.
Winners that were awarded both by the Kantar Information is Beautiful Awards and by the Malofiej Awards are marked with "*" in the first column. Winners that were awarded twice in the same award, are marked with "/no. of the first award in this list" in the first column.
Due to space limits, further details about authors, institution, country and prize type are omitted in this table.

Table 5

Winner List of the Kantar Information Is Beautiful Awards 2016 (The Information is Beautiful Awards Ltd, 2016a)

No.	Title of the winner	DV (1) / info-graphic (0)	Digital (1) / non digital (0)	Connecting line in main role (1) / subordinate role (2) / not present (0)	Award-winner no. in Appendix B
Data visualization					
1	Earth Temperature Timeline	1	1	1	9
2	Projekt Ukko - Seasonal Wind Predictions for the Energy Sector	1	1	0	
3	Gun Deaths in America	1	1	0	
4	Income Inequality in LA & Chicago	1	0	0	
Infographic					
5	The Missing Migrants Map	1	0	2	
6	Swanh.net	0	1	0	
7	The Chart of Cosmic Exploration	0	0	1	
Dataviz Project					
8	Data Cuisine	This project includes many different kinds of DVs, comparable to several DV websites. It is thus not clear which one was awarded.			
9	Roads to Rome	1	1	1	10
10	PhotoViz	0	0	0	
11	Makeover Monday	This is a learning platform.			
12	Objektivno	This project includes many different kinds of DVs, comparable to several DV websites. It is thus not clear which one was awarded.			
Data Journalism					
13	Spies in the Skies	1	1	1	11
14	Science Isn't Broken	1	1	0	
15	Film Dialogue	1	1	2	
16	Crime in Context	1	1	1	12
17	The Math of Mass Shootings	1	1	1	13
Interactive					
18	ShipMap.org	1	1	1	14
19	What's Your Pay Gap?	1	1	1	15
20	Evolution of Music Taste	1	1	1	16
21*	Who Marries Whom?	1	1	1	17
22	The Network Behind the Cosmic Web	1	1	1	18
23	Neural Network Playground	1	1	1	19
Dataviz Website					
24	Flowing Data	These are websites with several subpages and many DV projects and single DVs. It is thus the collection that was awarded.			
25	Visualising Data				
26	Visualoop				
Commercial/biz Project					
27	Data USA	This project includes many different kinds of DVs, comparable to several DV websites. It is thus not clear which one was awarded.			

28	News Explorer	1	1	1	20
29	Animated Infographic Microsite on E-commerce	0	1	0	
30	Educational Pathways	not accessible on Mai 31, 2018			
Studio of the Year					
31	Poligraph	This is a studio being awarded.			
Outstanding Team					
32	FiveThirtyEight	These are teams being awarded.			
33	Bloomberg				
34	Guardian				
35	Quartz				
Community Awards					
36	The sum of the Parts	1	1	1	21
37	WTFViz	This project includes many different kinds of DVs, comparable to several DV websites. It is thus not clear which one was awarded.			
38	Toronto Symphony Orchestra Listening Guide	0	1	0	
Outstanding Individual					
39	Moritz Stefaner	This is a person being awarded.			
Rising Star					
40	Will Stahl-Timmins	These are persons being awarded.			
41	Nadieh Bremer				
Student Awards					
42	Herwig Scherabon	These are persons being awarded.			
43	Pei Ye Le				
44	Janet Chan				
45	Cordelia Morales Trevino, Gabriela Gonzalez Rubio Gutierrez				
Best Non-English Language					
46	Terre Urbaine	1	1	0	
Most Beautiful					
47/13	Spies in the Skies	see above			
No. of award winners included into the corpus: 13; Of these 13: No. of award winners that were also awarded by the Malofiej Awards (but are counted here): 1.					

Note. Those winners, the DV(s) of which have been selected for the corpus, are marked with grey shading and the field in the last column contains the award-winner no. as in Appendix B.
Winners that were awarded both by the Kantar Information is Beautiful Awards and by the Malofiej Awards are marked with "*" in the first column. Winners that were awarded twice in the same award, are marked with "/no. of the first award in this list" in the first column.
Due to space limits further details about authors, institution, country and prize type are omitted in this table.

Table 6

Winner List of the Kantar Information Is Beautiful Awards 2017 (The Information is Beautiful Awards Ltd, 2017)

No.	Title of the winner	DV (1) / info-graphic (0)	Digital (1) / non digital (0)	Connecting line in main role (1) / subordinate role (2) / not present (0)	Award-winner no. in Appendix B
Arts, Entertainment & Pop Culture					
1	The Unlikely Odds of Making it Big	1	1	1	22
2	OddityViz - a visual deconstruction of Space Oddity	1	0	1	
3	An Interactive Visualization of Every Line in Hamilton	1	1	1	23
4	Film Money	1	1	1	24
Current Affairs & Politics					
5	On their Way: the Journey of Foreign Fighters	0	0	1	
6	One Angry Bird	1	1	0	
7	Want to fix gun violence in America? Go local	1	1	2	
8	The Electoral College misrepresents every state, but not as much as you may think	1	1	0	
People, Language & Identity					
9*	How do you draw a circle?	1	1	1	25
10	What's it like to get trolled all day long?	1	1	0	
11	Data Futures	1	1	0	
12	Live in Clay	0	0	0	
Environments & Maps					
13	1812: When Napoleon Ventured East	1	1	1	26
14	Lights on & Lights out	1	1	0	
15	Peak Spotting	1	1	1	27
16	Travel Visa Inequality	1	1	1	28
Humanitarian / Global					
17	The Shadow Peace - The Nuclear Threat	1	1	0	
18	The Point of no return - How the world is adapting to climate change	1	1	1	29
19	How a Melting Arctic Changes Everything	1	1	1	30
20	Those Who Did not Cross	0	1	0	
21	The Stories Behind a Line	1	1	1	31
Science & Technology					
22	Science Paths	1	1	1	32
23	Why Are so Many Babies Born around 8:00 A.M.?	1	1	0	
24	Seeing Theory	This is a learning platform.			
25	Nutrition Label	1	0	0	
26	Apollo	0	0	1	

27	This is the tangled future of tech and transportation	1	1	1	33
Sports & Games					
28	Rhythm of Food	1	1	2	
29	Swimming World Records throughout History	1	1	0	
30	NYC Foodiverse	1	1	1	34
31	What city is the microbrew capital of the US?	1	1	2	
32*	Berlin Marathon 2016: how fast your city runs	1	1	1	35
Unusual					
33	Data Sketches in Twelve Installments	This project includes many different kinds of DVs, comparable to several DV websites. It is thus not clear which one was awarded.			
34	How to Fix a Toilet (And Other Things We Couldn't Do Without Search)	1	1	2	
35	Data Viz Project	This is a website with several subpages and many DV projects and single DVs. It is thus the collection that was awarded.			
36	Forma Fluens	1	1	1	36
Studio of the year					
37	Density Design	This is a studio being awarded.			
Visitor's Vote					
38	The Annual Report of the ERGO Hestia Group,"Network"	1	0	1	
39	Viz for Social Good	This is a website with several subpages and many DV projects and single DVs. It is thus the collection that was awarded.			
40	Han Chinese Opera	0	0	1	
41	Surfing in Italy	0	0	0	
Outstanding Individual					
42	Nadieh Bremer	This is a person being awarded.			
Rising Star					
43	Jory Fleming	These are persons being awarded.			
44	Amy Cesal				
45	Nam Wook Kim				
46	Giacomo Flaim				
Best Non-English Language Viz					
47/3	Berlin Marathon 2016: how fast your city runs	see above			
48	Women Farmers in the Loneliness of Onion Fields	0	1	0	
49	Creation of a Thangka	0	0	0	
50	All the government barbecues	1	1	0	
No. of award winners included into the corpus: 15; Of these 15: No. of award winners that were also awarded by the Malofiej Awards (but are counted here): 2.					

Note. Those winners, the DV(s) of which have been selected for the corpus, are marked with grey shading and the field in the last column contains the award-winner no. as in Appendix B.
Winners that were awarded both by the Kantar Information is Beautiful Awards and by the Malofiej Awards are marked with "*" in the first column. Winners that were awarded twice in the same award, are marked with "/no. of the first award in this list" in the first column.
Due to space limits further details about authors, institution, country and prize type are omitted in this table.

References

Malofiej Infographic World Summit. (n.d.-b). *Malofiej 24: Awards list.* Retrieved November 26, 2018, from https://www.malofiejgraphics.com/wp-content/uploads/2016/03/M24_-AWARDS-LIST_OK.pdf

Malofiej Infographic World Summit. (n.d.-c). *Malofiej 25: Awards list.* Retrieved November 26, 2018, from https://www.malofiejgraphics.com/wp-content/uploads/2017/03/M25_AwardsList_OK3.pdf

Malofiej Infographic World Summit. (n.d.-d). *Malofiej 26: Awards list.* Retrieved November 26, 2018, from https://www.malofiejgraphics.com/wp-content/uploads/2018/03/M26-Awards-List-DEF.pdf

The Information is Beautiful Awards Ltd. (2015). *2015 – The Winners.* Kantar Information Is Beautiful Awards. https://www.informationisbeautifulawards.com/news/116-2015-the-winners

The Information is Beautiful Awards Ltd. (2016a). *2016 – The Winners.* Kantar Information Is Beautiful Awards. https://www.informationisbeautifulawards.com/news/188-2016-the-winners

The Information is Beautiful Awards Ltd. (2017). *Winners 2017.* Kantar Information Is Beautiful Awards. https://www.informationisbeautifulawards.com/news/259-winners-2017

Appendix B: Corpus Overview of the Single DVs

Table 1

List of Single DVs, Stemming From the Award-Winning Projects Presented in Appendix A

DV no.	DV title in the database (DB), self-generated name, based on the information on the website	Award-winner no.	Award-winner title in the DB, based on the website title	URL of the award-winning website where the visualization was found	Award title, award year
1	Year: 2011	1	Rise of Partisanship	http://www.mamartino.com/projects/rise_of_partisanship	K, 2015
2	Labour force participation, largest gap, 2013	2	Close The Gap	http://closethegap.studiometric.co	K, 2015
3	Gulf steps up support 2013	3	Syrian War in 5 Minutes	https://www.youtube.com/watch?v=qxzMa7j6LN0	K, 2015
4	Russia strikes 2015	3	Syrian War in 5 Minutes	https://www.youtube.com/watch?v=qxzMa7j6LN0	K, 2015
5	Charting culture172 CE	4	Charting Culture	https://www.nature.com/news/humanity-s-cultural-history-captured-in-5-minute-film-1.15650#/ref-link-1	K, 2015
6	Most segregated schools in 2014	5	Why Pinellas County Is The Worst Place in Florida to Be Black and Go to Public School\Failure Factories: Why Pinellas County is the worst place in Florida	http://www.tampabay.com/projects/2015/investigations/pinellas-failure-factories/chart-failing-black-students	K, 2015
7	No, it really is greenhouse gases	6	What's Really Warming the World?	https://www.bloomberg.com/graphics/2015-whats-warming-the-world	K, 2015
8	Air, W3	7	Visualisation of Global Weather Conditions	https://earth.nullschool.net	K, 2015
9	Change of amount of coverage mentioning each migrant group	8	Migration in the Census and in the News	http://seeingdata.cleverfranke.com	K, 2015
10	Appearance of different people or organisations within news	8	Migration in the Census and in the News	http://seeingdata.cleverfranke.com	K, 2015
11	Migration to England and Wales	8	Migration in the Census and in the News	http://seeingdata.cleverfranke.com	K, 2015
12	Where the non-UK were born	8	Migration in the Census and in the News	http://seeingdata.cleverfranke.com	K, 2015
13	A timeline of earth's average temperature	9	Earth Temperature Timeline	https://xkcd.com/1732	K, 2016
14	Continents & countries	10	Roads to Rome	http://roadstorome.moovellab.com/countries	K, 2016

15	Street DNA graph	10	Roads to Rome	http://roadstorome.moovellab.com/countries	K, 2016
16	Flight tracks	11	Spies in the Skies	https://www.buzzfeed.com/peteraldhous/spies-in-the-skies?utm_term=.rlwl2JVE7#.hhzwqNE1O	K, 2016
17	Violent crimes reported to 68 police departments	12	Crime in Context	https://www.themarshallproject.org/2016/08/18/crime-in-context#.b18C7Pe0q	K, 2016
18	Crime trends since 1975	12	Crime in Context	https://www.themarshallproject.org/2016/08/18/crime-in-context#.b18C7Pe0q	K, 2016
19	Milwaukee: up 11 percentbetween 2010 and 2015	12	Crime in Context	https://www.themarshallproject.org/2016/08/18/crime-in-context#.b18C7Pe0q	K, 2016
20	Details about each shooting	13	The Math of Mass Shootings	https://www.washingtonpost.com/graphics/2018/national/mass-shootings-in-america/?utm_term=.477565b332b9	K, 2016
21	17. May 2012 09:00	14	ShipMap.org	https://www.shipmap.org	K, 2016
22	Women working as hand laborers and freight, stock and material movers	15	What's Your Pay Gap?	http://graphics.wsj.com/gender-pay-gap	K, 2016
23	Number one on february 04, '90	16	Evolution of Music Taste	https://pudding.cool/2017/03/music-history/index.html	K, 2016
24	Who is married to designers	17	Who Marries Whom?\This Chart Shows Who Marries CEOs, Doctors, Chefs and Janitors	https://www.bloomberg.com/graphics/2016-who-marries-whom	K, 2016
25	Cosmic Web, varying length	18	The Network Behind the Cosmic Web	http://cosmicweb.kimalbrecht.com/viz/#1	K, 2016
26	Neural Network	19	Neural Network Playground	http://playground.tensorflow.org/#activation=tanh&batchSize=10&dataset=circle&regDataset=reg-plane&learningRate=0.03&regularizationRate=0&noise=0&networkShape=4,2&seed=0.26939&showTestData=false&discretize=false&percTrainData=50&x=true&y=true&xTimesY=true&xSquared=false&ySquared=false&cosX=false&sinX=false&cosY=false&sinY=false&collectStats=false&problem=classification&initZero=false&hideText=false	K, 2016
27	Recent news about Vladimir Putin related to Donald Trump	20	News Explorer	http://news-explorer.mybluemix.net	K, 2016
28	The sum of the parts	21	The sum of the Parts	http://www.scmp.com/infographics/article/1913814/infographic-sum-parts	K, 2016
29	The different paths each band took to make it	22	The Unlikely Odds of Making it Big	https://pudding.cool/2017/01/making-it-big	K, 2017
30	Explore their stories	23	An Interactive Visualization of Every Line in Hamilton	https://pudding.cool/2017/03/hamilton/index.html	K, 2017

31	Top 20 biggest budgests - US Revenue in US$	24	Film Money	http://www.datamake.io/project/film-money	K, 2017
32	Random samples of circles from major language groups	25	How do you draw a circle?	https://qz.com/994486/the-way-you-draw-circles-says-a-lot-about-you	K, 2017
33	Tarutino maneuver	26	1812: When Napoleon Ventured East	https://1812.tass.ru/en	K, 2017
34	Scythian tactics	26	1812: When Napoleon Ventured East	https://1812.tass.ru/en	K, 2017
35	Day view animated map of train movements	27	Peak Spotting	http://www.nand.io/projects/clients/peak-spotting	K, 2017
36	Let's explore	28	Travel Visa Inequality	http://projects.christianlaesser.com/travel-visa-inequality	K, 2017
37	One step after the other: what happened	29	The Point of no return - How the world is adapting to climate change	https://www.behance.net/gallery/52898387/The-Point-of-No-Return	K, 2017
38	Talking - An evolving conversation: main themes and COPs	29	The Point of no return - How the world is adapting to climate change	https://www.behance.net/gallery/52898387/The-Point-of-No-Return	K, 2017
39	Adaptation programs founded during and after COPs	29	The Point of no return - How the world is adapting to climate change	https://www.behance.net/gallery/52898387/The-Point-of-No-Return	K, 2017
40	Top ten donors and top twenty recipients	29	The Point of no return - How the world is adapting to climate change	https://www.behance.net/gallery/52898387/The-Point-of-No-Return	K, 2017
41	Change in average annual temperatures	30	How a Melting Arctic Changes Everything	https://www.bloomberg.com/graphics/2017-arctic	K, 2017
42	Total degrees below zero	30	How a Melting Arctic Changes Everything	https://www.bloomberg.com/graphics/2017-arctic	K, 2017
43	Path of M.B., 18 years old	31	The Stories Behind a Line	http://www.storiesbehindaline.com	K, 2017
44	Distance of M.B., 18 years old	31	The Stories Behind a Line	http://www.storiesbehindaline.com	K, 2017
45	Quantifying the evolution of individual scientific impact	32	Science Paths	http://sciencepaths.kimalbrecht.com	K, 2017
46	The complicated web of the future of transportation	33	This is the tangled future of tech and transportation	http://nordic.businessinsider.com/web-of-ride-hailing-investments-2016-12?r=US&IR=T	K, 2017
47	Ashtray present in smoke-free area	34	NYC Foodiverse	http://nycfoodiverse.com	K, 2017
48	43. Berlin-Marathon im Zeitraffer	35	Berlin Marathon 2016: how fast your city runs	https://interaktiv.morgenpost.de/berlin-marathon-2016	K, 2017
49	Apple	36	Forma Fluens	http://www.formafluens.io	K, 2017
50	Texas far outpaces the other states	37	The Next to Die: National, State, Case link	https://www.themarshallproject.org/next-to-die	M, 24

51	Tracking the pace of social change	38	This Is How Fast America Changes Its Mind	https://www.bloomberg.com/graphics/2015-pace-of-social-change	M, 24
52	Interracial marriage	38	This Is How Fast America Changes Its Mind	https://www.bloomberg.com/graphics/2015-pace-of-social-change	M, 24
53	Final impact	39	Germanwings: The last minutes of Germanwings flight 4U9525	https://www.theguardian.com/world/ng-interactive/2015/mar/24/path-of-germanwings-flight-4u9525-interactive	M, 24
54	Location of stampede	40	How the Hajj Stampede Unfolded	https://www.nytimes.com/interactive/2015/09/24/world/middleeast/mecca-mina-stampede-hajj-maps.html	M, 24
55	The journey to Mecca	40	How the Hajj Stampede Unfolded	https://www.nytimes.com/interactive/2015/09/24/world/middleeast/mecca-mina-stampede-hajj-maps.html	M, 24
56	Engineer lost track of his location on the route	41	Investigating the Philadelphia Amtrak Train Derailment	http://www.nytimes.com/interactive/2015/05/13/us/investigating-the-philadelphia-amtrak-train-crash.html	M, 24
57	Ships have been scouring the sea bottom for months	42	How Missing Jet's Debris Could Have Floated to Réunion	http://www.nytimes.com/interactive/2014/03/17/world/asia/search-for-flight-370.html?_r=0	M, 24
58	Possible flight paths	42	How Missing Jet's Debris Could Have Floated to Réunion	http://www.nytimes.com/interactive/2014/03/17/world/asia/search-for-flight-370.html?_r=0	M, 24
59	The search moves north	42	How Missing Jet's Debris Could Have Floated to Réunion	http://www.nytimes.com/interactive/2014/03/17/world/asia/search-for-flight-370.html?_r=0	M, 24
60	Sorting out the clues	42	How Missing Jet's Debris Could Have Floated to Réunion	http://www.nytimes.com/interactive/2014/03/17/world/asia/search-for-flight-370.html?_r=0	M, 24
61	The descent	43	What Happened on the Germanwings Flight	http://www.nytimes.com/interactive/2015/03/24/world/europe/germanwings-plane-crash-map.html	M, 24
62	How high and fast the plane was flying when they lost contact	43	What Happened on the Germanwings Flight	http://www.nytimes.com/interactive/2015/03/24/world/europe/germanwings-plane-crash-map.html	M, 24
63	Where the debris are suggest the plane hit the ground	43	What Happened on the Germanwings Flight	http://www.nytimes.com/interactive/2015/03/24/world/europe/germanwings-plane-crash-map.html	M, 24
64	Where the aircraft took off and the controllers lost contact	43	What Happened on the Germanwings Flight	http://www.nytimes.com/interactive/2015/03/24/world/europe/germanwings-plane-crash-map.html	M, 24
65	Path of train	44	Amtrak derailment	https://www.washingtonpost.com/graphics/local/amtrak-derail	M, 24
66	Photo view angle	44	Amtrak derailment	https://www.washingtonpost.com/graphics/local/amtrak-derail	M, 24
67	Train 188 speed compared to average train speeds	44	Amtrak derailment	https://www.washingtonpost.com/graphics/local/amtrak-derail	M, 24
68	Days 1 to 16	45	Illegal Tusk Trade	https://www.nationalgeographic.com/tracking-ivory/map.html	M, 24

69	Trade route to consolidation hubs	45	Illegal Tusk Trade	https://www.nationalgeographic.com/tracking-ivory/map.html	M, 24
70	Trade route to Asian markets	45	Illegal Tusk Trade	https://www.nationalgeographic.com/tracking-ivory/map.html	M, 24
71	1940 to 1960 Post-World War II	46	A visual guide to 75 years of major refugee crises…	https://www.washingtonpost.com/graphics/world/historical-migrant-crisis/	M, 24
72	The yield curve	47	A 3-D View of a Chart That Predicts The Economic Future: The Yield Curve	http://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html	M, 24
73	A drop in China's demand could drag these countries down…	48	How China's economic slowdown could weigh on the rest	https://www.theguardian.com/world/ng-interactive/2015/aug/26/china-economic-slowdown-world-imports	M, 24
74	Around the world - asia	48	How China's economic slowdown could weigh on the rest	https://www.theguardian.com/world/ng-interactive/2015/aug/26/china-economic-slowdown-world-imports	M, 24
75	The Dawn Wall	49	The Dawn Wall: El Capitan's Most Unwelcoming Route	http://www.nytimes.com/interactive/2015/01/09/sports/the-dawn-wall-el-capitan.html	M, 24
76	Buffalo Bills	50	The Complete History Of the NFL	https://projects.fivethirtyeight.com/complete-history-of-the-nfl	M, 24
77	Avaliar a distância do gol	51	Ceni's Free Kicks in 7 Steps	http://app.globoesporte.globo.com/futebol/times/sao-paulo/as-faltas-de-ceni-em-7-passos	M, 24
78	Medir a força de acordo com o vento	51	Ceni's Free Kicks in 7 Steps	http://app.globoesporte.globo.com/futebol/times/sao-paulo/as-faltas-de-ceni-em-7-passos	M, 24
79	Bater por cima ao lado da barreira	51	Ceni's Free Kicks in 7 Steps	http://app.globoesporte.globo.com/futebol/times/sao-paulo/as-faltas-de-ceni-em-7-passos	M, 24
80	Bater na bola com a força ideal	51	Ceni's Free Kicks in 7 Steps	http://app.globoesporte.globo.com/futebol/times/sao-paulo/as-faltas-de-ceni-em-7-passos	M, 24
81	The climb	52	Tour de France: the climb of Alpe d'Huez	https://www.theguardian.com/sport/ng-interactive/2015/jul/23/tour-de-france-the-climb-of-alpe-dhuez-interactive	M, 24
82	Start	52	Tour de France: the climb of Alpe d'Huez	https://www.theguardian.com/sport/ng-interactive/2015/jul/23/tour-de-france-the-climb-of-alpe-dhuez-interactive	M, 24
83	Ronney's first goal	53	Wayne Rooney breaks Bobby Charlton's England goalscoring record	https://www.theguardian.com/football/ng-interactive/2015/sep/05/wayne-rooney-england-equals-record-goalscoring-charlton	M, 24
84	When they delivered	53	Wayne Rooney breaks Bobby Charlton's England goalscoring record	https://www.theguardian.com/football/ng-interactive/2015/sep/05/wayne-rooney-england-equals-record-goalscoring-charlton	M, 24
85	Business as usual vs. pledges vs. goal	54	The Climate Change Pledges Are In. Will They Fix Anything?	http://www.nytimes.com/interactive/2015/11/23/world/carbon-pledges.html	M, 24
86	Chart showing the aggregate choices of 78022 readers	55	You Draw It: How Family Income Predicts Children's College Chances	https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html	M, 24

87	How wing flap from missing MH370 plane could have gotten to Reunion	56	How Missing Jet's Debris Could Have Floated to Réunion	https://twitter.com/nytgraphics/status/628992155495858176	M, 24
88	Total number of borders with barriers, by year	57	A new age of walls	https://www.washingtonpost.com/graphics/world/border-barriers/global-illegal-immigration-prevention	M, 25
89	Migrant flow after the fence has gone up	57	A new age of walls	https://www.washingtonpost.com/graphics/world/border-barriers/global-illegal-immigration-prevention	M, 25
90	The fences did not stop migration	57	A new age of walls	https://www.washingtonpost.com/graphics/world/border-barriers/global-illegal-immigration-prevention	M, 25
91	Kurz's stops	57	A new age of walls	https://www.washingtonpost.com/graphics/world/border-barriers/global-illegal-immigration-prevention	M, 25
92	The final presidential depate mapped	58	Missed the final presidential debate? We've mapped it out for you	https://www.theguardian.com/us-news/ng-interactive/2016/oct/20/final-presidential-debate-clinton-trump-interactive	M, 25
93	Frame-by-frame analysis of the videos	59	The Keith Scott Shooting: A 3-D Reconstruction	https://www.nytimes.com/interactive/2016/09/25/us/charlotte-scott-shooting-video.html	M, 25
94	Possibilities to refuel the aircraft	60	The Chapecoense aircraft crashed because it ran out of fuel while waiting to land	https://www.univision.com/noticias/accidente-aereo/interactivo-reconstruccion-de-los-ultimos-minutos-del-vuelo-del-chapecoense-antes-de-estrellarse-por-falta-de-combustible	M, 25
95	The way of the plane from Santa Cruz	60	The Chapecoense aircraft crashed because it ran out of fuel while waiting to land	https://www.univision.com/noticias/accidente-aereo/interactivo-reconstruccion-de-los-ultimos-minutos-del-vuelo-del-chapecoense-antes-de-estrellarse-por-falta-de-combustible	M, 25
96	Animation of the last minutes	60	The Chapecoense aircraft crashed because it ran out of fuel while waiting to land	https://www.univision.com/noticias/accidente-aereo/interactivo-reconstruccion-de-los-ultimos-minutos-del-vuelo-del-chapecoense-antes-de-estrellarse-por-falta-de-combustible	M, 25
97	Path of the helicopter over the sea	61	The Turøy accident	https://www.vg.no/spesial/2016/hks241	M, 25
98	When the main rotor suddenly loosened	61	The Turøy accident	https://www.vg.no/spesial/2016/hks241	M, 25
99	Last position, accident site and where the rotor was found	61	The Turøy accident	https://www.vg.no/spesial/2016/hks241	M, 25
100	They are doing better than you!	62	Who's winning? Find out how your income compares with every other generation	https://www.theguardian.com/world/ng-interactive/2016/mar/07/whos-winning-find-out-how-your-income-compares-with-every-other-generation	M, 25
101	Difference between white, hispanic and black students	63	Money, Race and Success: How Your School District Compares	https://www.nytimes.com/interactive/2016/04/29/upshot/money-race-and-success-how-your-school-district-compares.html	M, 25

102	What the U.S. countries voted for	64	Swing countries	https://www.washingtonpost.com/graphics/politics/2016-election/swing-counties	M, 25
103	Path of the airplane	65	Sequence of wrong choices	http://app.globoesporte.globo.com/sc/futebol/times/chapecoense/sequencia-de-escolhas-erradas	M, 25
104	Actual and desired path	65	Sequence of wrong choices	http://app.globoesporte.globo.com/sc/futebol/times/chapecoense/sequencia-de-escolhas-erradas	M, 25
105	Primary migrant route	66	Fortress of Fences	https://www.reuters.com/investigates/special-report/migration/#story/38	M, 25
106	Comparison of popularity on different dates	67	Popularity	https://www.washingtonpost.com/graphics/politics/2016-election/unpopular	M, 25
107	Comparison of Hillary Clinton and Donald Trump	67	Popularity	https://www.washingtonpost.com/graphics/politics/2016-election/unpopular	M, 25
108	Income change	68	When was America great	https://www.washingtonpost.com/graphics/politics/2016-election/when-was-america-great	M, 25
109	Where the candidates live and where their electoral district is	69	These candidates live farthest away from their voters	https://interaktiv.morgenpost.de/waehlernaehe-berlin	M, 25
110	Percentage of residents eligible for same-day delivery	70	Amazon Doesn't Consider the Race of Its Customers. Should It?	https://www.bloomberg.com/graphics/2016-amazon-same-day	M, 25
111	Construction cost changes	71	Nine years of waiting for St. Petersburg football arena	http://zenit-arena.tass.com	M, 25
112	History of stadium's construction	71	Nine years of waiting for St. Petersburg football arena	http://zenit-arena.tass.com	M, 25
113	Manufacturing jobs are declining in…	72	Manufacturing jobs are returning to some places. But these jobs are different	https://www.washingtonpost.com/graphics/national/manufacturing-in-america	M, 25
114	At the peak of this pass, she'll clear nearly twice her own height	73	The Fine Line: Olympian Simone Biles	https://www.nytimes.com/interactive/2016/08/05/sports/olympics-gymnast-simone-biles.html	M, 25
115	Biles' powerful run	73	The Fine Line: Olympian Simone Biles	https://www.nytimes.com/interactive/2016/08/05/sports/olympics-gymnast-simone-biles.html	M, 25
116	O percurso olímpico	74	Sailing at wind	http://app.globoesporte.globo.com/olimpiadas/vela	M, 25
117	Gol del siglo	75	Anyversary Soccer World Cup 1986	https://www.lanacion.com.ar/1911398-mexico-86-el-gol-del-siglo-paso-por-paso-en-un-grafico-interactivo	M, 25
118	Pitch by pitch	76	Pitch by pitch: How Clayton Kershaw dominates hitters	http://www.latimes.com/projects/la-sp-kershaw-2016-pitch-by-pitch	M, 25

119	Explore races that begin at this adress	77	Olympic Races, in Your Neighborhood	https://www.nytimes.com/interactive/2016/08/18/sports/olympics/olympic-races-near-you.html	M, 25
120	The journey of the bear	78	A Bear's Eye View of Yellowstone	https://www.nationalgeographic.com/magazine/2016/05/yellowstone-national-parks-bears-video/?source=NGcomm	M, 25
121	How a rocket reaches Russia's first civilian launch site	79	Vostochny Space Launch Center	https://ria.ru/infografika/20160425/1054997772.html?lang=en	M, 25
122	Rocket at the space center	79	Vostochny Space Launch Center	https://ria.ru/infografika/20160425/1054997772.html?lang=en	M, 25
123	Articles written by female and male writers	80	The dark side of Guardian comments	https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments	M, 25
124	Which subjects attracted more abusive or disruptive comments	80	The dark side of Guardian comments	https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments	M, 25
125	Travel overview of Mr. Siddiqui	81	How One of the Deadliest Hajj Accidents Unfolded	https://www.nytimes.com/interactive/2016/09/06/world/middleeast/2015-hajj-stampede.html	M, 25
126	Which street Mr. Siddiqui's group took	81	How One of the Deadliest Hajj Accidents Unfolded	https://www.nytimes.com/interactive/2016/09/06/world/middleeast/2015-hajj-stampede.html	M, 25
127	Category 4 or higher hurricanes that have reached land	82	Tracking Harvey's Destructive Path Through Texas and Louisiana	https://www.nytimes.com/interactive/2017/08/24/us/hurricane-harvey-texas.html	M, 26
128	How the storm moved through the region	82	Tracking Harvey's Destructive Path Through Texas and Louisiana	https://www.nytimes.com/interactive/2017/08/24/us/hurricane-harvey-texas.html	M, 26
129	How the ice is moving	83	Antarctic Dispatches	https://www.nytimes.com/interactive/2017/05/18/climate/antarctica-ice-melt-climate-change.html	M, 26
130	Pay gap for every occupation	84	Can we talk about the gender pay?	https://www.washingtonpost.com/?utm_term=.10772e35c1ff	M, 26
131	How much median earnings for each job changed	84	Can we talk about the gender pay?	https://www.washingtonpost.com/?utm_term=.10772e35c1ff	M, 26
132	Irma is following a well-worn path	85	What Lies in Irma's Path	https://fivethirtyeight.com/features/what-lies-in-irmas-path	M, 26
133	Percent of vote 2013 and 2017	86	How Northam gained in a more polarized Virginia	https://www.washingtonpost.com/graphics/2017/politics/va-gov-election-results-precincts	M, 26
134	The deadliest day	87	A deadly crossing	http://fingfx.thomsonreuters.com/gfx/rngs/MYANMAR-ROHINGYA/010051JR3GY/index.html	M, 26
135	Where Raber travelled to reach his destination	88	Bussed out - How America moves its homeless	https://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study	M, 26
136	Where homeless people have been sent to and from	88	Bussed out - How America moves its homeless	https://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study	M, 26

137	Homeless relocations from New York City	88	Bussed out - How America moves its homeless	https://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study	M, 26
138	Most tickets recipients are relocated to places with a lower median income	88	Bussed out - How America moves its homeless	https://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study	M, 26
139	Dramatic racial disparities in homicide have persisted for decades	89	Want to fix gun violence in America? Go local.	https://www.theguardian.com/us-news/ng-interactive/2017/jan/09/special-report-fixing-gun-violence-in-america	M, 26
140	Positions of 11 key senators	90	How key senators changed their positions to bring down Obamacare repeat	https://www.washingtonpost.com/graphics/2017/politics/health-care-swing-votes	M, 26
141	Puerto Rico's longest power outages since 2000 compared to hurricane Maria	91	After Hurricane Maria, Puerto Rico is still in the dark	https://www.washingtonpost.com/graphics/2017/national/puerto-rico-hurricane-recovery/?utm_term=.68a43c43c08c	M, 26
142	Monthly electrical price per state since 2014	91	After Hurricane Maria, Puerto Rico is still in the dark	https://www.washingtonpost.com/graphics/2017/national/puerto-rico-hurricane-recovery/?utm_term=.68a43c43c08c	M, 26
143	Puerto Ric's labor force	91	After Hurricane Maria, Puerto Rico is still in the dark	https://www.washingtonpost.com/graphics/2017/national/puerto-rico-hurricane-recovery/?utm_term=.68a43c43c08c	M, 26
144	A muslim mayor in my community?	92	Diving into Urban-Rural Prejudice	https://www.zeit.de/feature/deutsche-bevoelkerung-stadt-land-unterschiede-vorurteile	M, 26
145	So tickt Berlin an deiner Linie	93	That's how Berlin is ticking along your line	https://interaktiv.morgenpost.de/berlin-an-deiner-linie	M, 26
146	S9 Jung und Alt	93	That's how Berlin is ticking along your line	https://interaktiv.morgenpost.de/berlin-an-deiner-linie	M, 26
147	Income growth	94	Our Broken Economy, in One Simple Chart	https://www.nytimes.com/interactive/2017/08/07/opinion/leonhardt-income-inequality.html	M, 26
148	Production on federal and Indian Islands	95	The United States of oil and gas	https://www.washingtonpost.com/graphics/national/united-states-of-oil	M, 26
149	Degree of asymmetry	96	Cracking the mystery of egg shape	http://vis.sciencemag.org/eggs	M, 26
150	Cassini's route	97	Cassini's Grand Tour	https://www.nationalgeographic.com/science/2017/09/cassini-saturn-nasa-3d-grand-tour/#storm	M, 26
151	Future total solar eclipses worldwide in your lifetime	98	Here's every total solar eclipse happening in your lifetime.	https://www.washingtonpost.com/graphics/national/eclipse/?utm_term=.eeea166bece0	M, 26
152	What the total solar eclipse in August will look like throughout the U.S.	98	Here's every total solar eclipse happening in your lifetime.	https://www.washingtonpost.com/graphics/national/eclipse/?utm_term=.eeea166bece0	M, 26
153	Total solar eclipse paths over the contiguous U.S.	98	Here's every total solar eclipse happening in your lifetime.	https://www.washingtonpost.com/graphics/national/eclipse/?utm_term=.eeea166bece0	M, 26

154	Trayectoria del eclipse total	99	Escribe tu “zip code” y descubre cómo se verá el eclipse solar e tu ciudad	https://www.univision.com/noticias/eclipse-solar/escribe-tu-zip-code-y-descubre-como-se-vera-el-eclipse-solar-en-tu-ciudad	M, 26
155	Dónde y cómo se verá el eclipse	99	Escribe tu “zip code” y descubre cómo se verá el eclipse solar e tu ciudad	https://www.univision.com/noticias/eclipse-solar/escribe-tu-zip-code-y-descubre-como-se-vera-el-eclipse-solar-en-tu-ciudad	M, 26
156	Global temperature change	100	The three-degree world: the cities that will be drowned by global warming	https://www.theguardian.com/cities/ng-interactive/2017/nov/03/three-degree-world-cities-drowned-global-warming	M, 26
157	How did people watch videos?	101	Which tech world did you grow up in?	https://www.washingtonpost.com/graphics/2017/entertainment/tech-generations/?utm_term=.80c545ba4036	M, 26
158	Average wait times for the busiest rides in both parks	102	Don’t waste your time at Disneyland. Here’s how to avoid the lines	http://www.latimes.com/projects/la-fi-disneyland-ride-wait-time/#ride-times	M, 26
159	Search interest 2015 - 2016	103	The Year In Language	https://googletrends.github.io/year-in-language	M, 26
160	Maximum search interest for historically popular words	103	The Year In Language	https://googletrends.github.io/year-in-language	M, 26
161	The similarities between the top 10 words per language	104	Beautiful In English	http://www.beautifulinenglish.com	M, 26
162	Route performance and flight volume of the airports	105	Which Flight Will Get You There Fastest?	https://projects.fivethirtyeight.com/flights	M, 24
163	How much faster or slower the airlines are	105	Which Flight Will Get You There Fastest?	https://projects.fivethirtyeight.com/flights	M, 24

Note. K = Kantar Information is Beautiful Awards; M = Malofiej Awards.
Those award-winners that were awarded both by the Malofiej Awards as well as by the Kantar Information is Beautiful Awards are only counted in the latter category.

Appendix C: Setup of the DV Type Repertoire

In the relevant literature as well as in production tools, several different collections and taxonomies can be found (e.g., Kirk, 2016; Munzner, 2014; RStudio, 2015; Tableau Software, Inc., n.d.) that could have been used for categorizing the DVs into various types. Thus, a decision had to be made about what to use in this specific categorization process. The decision fell on combining two different sources. The first one is Andy Kirk's *Data Visualization. A Handbook of Data Driven Design* (2016) as a printed source. It offers a repertoire of 49 different types including detailed descriptions of the representation, variations, alternatives and examples (Kirk, 2016). He states, that because the "vast permutations of different marks and attributes prevents any finite limit to how one might portray data visually" (p. 158), this is not meant to be an exhaustive list, but a "curated collection of some of the common and useful chart types being used across the field today" (p. 157).

Because not all DVs in the corpus could be assigned to one of these 49 types, Kirk's (2016) repertoire had to be widened. For that purpose, I used the website *datavizproject.com* (Ferdio ApS (www.ferdio.com), n.d. (status as of 14 August 2018)). At the date of access, it presented 154 different visualization types (see Figure 1). I considered this source a reasonable supplement to Kirk's repertoire, as it was a current, non-commercial project, developed by the Danish infographic and DV agency, Ferdio ApS. Another reason was that for each visualization type (each presented on a separate subpage), the authors offer current examples, including links to the original sources, which is a clear benefit of web-based sources in general. This opportunity to compare several examples was especially helpful in the process of assigning the categories to the examples.

Figure 1

Screenshot of the Overview of All the 154 DV Types Identified and Labelled in the DataVizProject

Note. From "The Data Viz Project" by Ferdio ApS, n.d. (http://datavizproject.com/). Copyright by Ferdio ApS [n.d.] Reprinted with permission

When comparing these two sources, it became apparent that for some visualization types, the naming as well as the granularity of subcategories varies. Therefore, a consistent naming system had to be set up and a decision had to be made about which granularity of subcategories to include. Focusing on the DV types found in my corpus, Table 1 shows a comprehensive comparison of the two different systems, with the terms and granularity I decided on as well as the reasons for the naming decision. The table is divided into two parts. The first part shows the DV types that could be directly assigned to the examples of my collection. The second part shows those DV types that were needed to describe hybrids (containing not only DV types that use connecting lines).

The result of this process of comparison and labelling was a repertoire of 21 DV types, that were necessary for the categorization of the award-winning DV examples of my collection. At this point, I would like to emphasise that this repertoire is not meant to be a comprehensive list of types that contain connecting lines as a central semiotic element. Rather, it resulted from the parallel process of viewing the examples of which the corpus is comprised. Therefore, there may be other relevant DV types that have not been taken into consideration here. That is the case either because examples of this DV type have not been submitted to one of the DV awards in the included years (or maybe have not even been produced), or because such visualizations have not been awarded by the juries.

Table 1

Overview of the Repertoire of DV Types Used for the Categorization

Visualization types that are directly assignable to the examples (pure ones)				
Term according to Kirk (2016, pp. 161–209)	**Term according to DataVizProject (Ferdio ApS (www. ferdio.com), n.d.)**	**Term used during the categorization process**	**Reason for the decision**	**Type no.**
flow map (“also known as connection map, route map, stream map, particle flow map” (Kirk, 2016, p. 206))	route map	route map	These three types have different purposes, which can also influence the graphical appearance of the connecting lines. Therefore, and also because several of all three categories are included in the collection, I consider the division into these three categories as useful.	1
	flow map	flow map		2
	connection map	connection map		3
line chart	line graph	line graph	The difference between a spline graph and line graph is only whether the connecting line is curved or not. There is no functional difference. Therefore, the term *line graph* is used here for both types. The term *graph* is preferred because it is the most constrained form of the descriptions of the four “types of ‘symbolic’ displays” (Kosslyn, 1989, p. 185) and shows a symmetrical “‘paired with’ relation” (Kosslyn, 1989, p. 186).	4
	spline graph			
node-link diagram (“also known as network diagram, graph, hairballs” (Kirk, 2016, p. 188))	network visualization	network diagram	As both sources speak of the same type, I chose the one that does not contain the terms *node* or *link*, in order to limit the use of the dissertation’s key terms (see Chapter 2 in the main part). The term *diagram* is preferred because it is more specific (see also Kosslyn, 1989, p. 186).	5
stream graph	stream graph	stream graph	The only difference between the two categories suggested by Ferdio ApS is that they are either sorted or stacked. In order to keep the number of types small, I decided to use the term stream graph for both types.	6
	sorted stream graph			
- (not mentioned)	lollipop chart	lollipop chart	No selection was needed.	7
- (not mentioned)	timeline	timeline	Whether the timeline is scaled or not has no effect on the connecting line. Therefore, the term *timeline* is used for both types.	8
	scaled timeline			
Sankey diagram (“also known as alluvial diagram” (Kirk, 2016, p. 190))	alluvial diagram	alluvial diagram	The term was suggested by both sources.	9
- (not mentioned)	sociogram	sociogram	There was no selection needed.	10
connected dot plot (“also known as Barbell chart, dumb-bell chart” (Kirk, 2016, p. 164))	dumbbell plot	dumb-bell chart	I decided on this version because the Oxford English Dictionary states dumb-bell to be the right spelling (“Dumb-Bell,” 2018). The term *chart* is preferred because it is one of the four “types of ‘symbolic’ displays” (Kosslyn, 1989, p. 185).	11

bump chart	bump chart	bump chart	The terms are equivalent.	12
parallel coordinates	parallel coordinates	parallel coordi-nates	The terms are equivalent.	13
slope graph chart (“also known as slope chart” (Kirk, 2016, p. 193))	slope chart	slope chart	The decision fell on the term that is suggested by both.	14
radar chart	radar diagram	radar chart	Bearing in mind the definitions of Stephen M. Kosslyn (1989, p. 186), the term chart fits better in this case.	15
-	-	direct trace visu-alization	The visualizations categorized with this made up term lie within a border area of DVs. Just like any DV, a direct trace visualization “visualizes non-visible data … or non-numeric conceptual schemes, and … provides a cognitive tool for discovery and exploration” (Engebretsen & Weber, 2018, p. 279) (as they, for instance, artificially overlay the traces of several participants’ strokes). Whether it is “algorithmically drawn” and therefore “easy to regenerate with different data” (Iliinsky & Steele, 2011, p. 7) is a matter of discussion. The reason is that the sin-gle lines are directly hand-drawn on the digital surface by the test participants, but nevertheless, the final visualization that overlaps all of them may be algo-rithmically created.	16
Visualization types needed to describe the basis types of the hybrids				
- (not mentioned)	linear process diagram	linear process diagram	No selection was needed.	17
dendrogram	dendrogram	dendrogram	The terms are equivalent.	18
bubble plot (“also known as bubble chart” (Kirk, 2016, p. 184))	bubble chart	bubble chart	This was a term that both suggested.	19
- (not mentioned)	arc diagram	arc diagram	No selection was needed.	20
chord diagram (“also known as radial network diagram, arc dia-gram (wrongly)” (Kirk, 2016, p. 189))	radial convergences (“also called: radial network” (Ferdio ApS (www.ferdio.com), n.d.)) / non-ribbon chord diagram: The difference between these two groups is unclear.	radial network diagram	The term *radial network* was used by both sources. The term *diagram* is preferred because it is more specific (see also Kosslyn, 1989, p. 186)	21

Note. Type number 16 is not based on either of the two DV type collections but was invented by me in order to be able to include visualization examples of that kind.

As described in Table 1, the DV types 17 to 21 were used to describe hybrids. Thus, combinations of several DV types were used for categorizing some of the DVs in the corpus. The final list of DV types (including these hybrids) is present-ed in Table 7 of Section 7.3.6 in the main part.

References

Dumb-bell. (2018). In *OED Oxford English dictionary*. Oxford University Press http://www.oed.com/view/Entry/58381

Engebretsen, M., & Weber, W. (2018). Graphic modes: The visual representation of data. In C. Cotter & D. Perrin (Eds.), *The Routledge handbook of language and media* (pp. 277–295). Routledge. https://doi.org/10.4324/9781315673134

Ferdio ApS (www.ferdio.com). (n.d.). DVP. Data Viz Project. http://datavizproject.com/

Iliinsky, N., & Steele, J. (2011). *Designing data visualizations*. O'Reilly Media.

Kirk, A. (2016). *Data visualisation: A handbook for data driven design*. SAGE.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3), 185–226. https://doi.org/10.1002/acp.2350030302

Munzner, T. (2014). *Visualization analysis and design*. CRC Press/Taylor & Francis Group.

RStudio. (2015). *Data Visualization with ggplot2: Cheat sheet* https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf

Tableau Software, Inc. (n.d.). *Build charts and analyze data*. Tableau Software Online Help. Retrieved August 17, 2018, from https://onlinehelp.tableau.com/current/pro/desktop/en-us/design_and_analyze.html

Appendix D: Article II: Coding Instructions of the Pre-Test

Coding instructions

for the inter-rater reliability pre-test of the corpus analysis
focusing on how modality and uncertainty are signalled by connecting lines

1. Introduction

Thank you for your participation in this pre-test. With your help, I will be able to find out to what extent our answers to the analysis questions are similar. The more consistent the answers between the raters of this pre-test are, the more reliable the method suggested to analyse the data visualization corpus can be considered (Bell, 2001, p. 21). If they are not consistent enough, the method (asked questions and answer categories) need to be rethought and adjusted. The focus of the whole study is to find out, how different kinds of uncertainty and modality are signified in data visualizations, further called DVs. More precisely, special attention is paid to connecting lines, which in all the DVs selected are the graphical elements conveying the main statement.

2. Definition of the units of analysis and the used documents

Within this pre-test, a randomly selected sample (25 items) of the entire corpus (containing 163 data visualizations) is analysed independently by two raters. In addition to this coding instruction document, you have been given the following documents:

(1) the Excel file “analysis_pretest.xlsx”
(2) the Excel file “visualizationlistforanalysis.xlsx
(3) a folder with JPG, PNG, PDF and mp4 documents

Let me explain what these documents are used for:

(1) This is the document where the coding takes place. It has three tabs, each of which contains a table that needs to be filled out. The first two columns of all three tables have the same content: the number and title of the DVs that are analysed in this pre-test. This means that each row stands for one DV. The columns from C onwards stand for one question each.

(2) The Excel file “visualizationlistforanalysis.xlsx” helps you to find the right documents in the document folder (3.). Because the DVs that are being analysed stem from websites, and their content or the access possibilities can be changed, I have made screenshots, and screencasts of each DV and website. All the filenames of these screenshots and screencasts, as well as the original link to the website where the DV stem from, are listed in this Excel file. Since more than just one DV can stem from one website, the column “visimageFN” contains the filenames of JPGs that define which DV is being investigated and coded within this row. All the following columns (“winnerID”, “imgFN”, “filmFN”, “pdfFN” and “url”) can therefore have the same content in several rows.

(3) This folder contains the research material, the filenames of which are documented in the Excel file “visualizationlistforanalysis.xlsx”.

3. First steps

(1) Open the Excel file “analysis_pretest.xlsx” (1.), go to the first tab.

(2) Look at which visualization is the first to analyse. (The columns “visID” and “vistitle” tell you this.) Let’s say this would be visID “16”, vistitle “Flight tracks”.

(3) Open the Excel file “visualizationlistforanalysis.xlsx” and go to the row with the same visID and vistitle. Mark it, for example, with using a filling colour for this particular row.

(4) Open Windows Explorer/Finder and navigate to the folder with all the research material (3.).

(5) Now search for the filenames written in the marked row (step 3.) and open these documents. In our example this would be “Spies_in_the_skies.JPG”, “spies_in_the_skies.png”, “Spies_in_the_skies.mp4” and “Spies_in_the_skies.pdf”. Maybe you would also like to open the

website these screenshots and screencasts have been taken from (in this case https://www.buzzfeed.com/peteraldhous/spies-in-the-skies?utm_term=.rlwl2JVE7#.hhzwqNE1O). But note that changes may have been made since the screenshots and screencasts were taken. That's why these documents, and not the online website, remain the main point of reference.

(6) Now go back to the Excel file "analysis_pretest.xlsx" (1.) and answer the first question to the DV investigated, by looking at the files opened in step 5.

(7) Go through all questions in all three tabs of the "analysis_pretest.xlsx" file (for this DV) and answer them. (The table in the second tab is already filled out, you may only want to check the answers randomly. In case there is any kind of uncertainty or modality expressed through the line, you will need to make use of the answers in the second tab.)

(8) Demark the row marked in step 3.

(9) Then repeat these steps for each DV by going back to step 2, finding the next DV and so on.

4. Descriptions of the questions and answer categories

In this section, the questions and answer categories available within the Excel file "analysis_pretest.xlsx" are explained. This is to make sure that the questions are understood by all raters, and that the answer categories are made clear. In the "analysis_pretest.xlsx" file, even though the titles of each column contain only an abbreviation, the full questions asked are noted in a comment attached to each column title field.

4.1 The "semioticfunctionanalysis" tab

As you can see in the first line of the table in this tab, the questions in this tab are organized into two sections: the detail level and the global level. This refers to the "zooming levels".

On the detail level, what is in the focus for answering the question is one particular connecting line, as well as the surrounding elements, related to this line (like labels e.g.). (In the words of Boeriis' dynamic functional rank scale (2012, p. 139) the detail level summarizes the component, unit and group rank.)

On the global level, the whole data visualization, including any surrounding co-text that's related to this visualization is evaluated. (In the words of Boeriis' dynamic functional rank scale (2012, p. 139) the global level summarizes the whole rank and the co-text.)

The following tables show the column character, column title, the complete question, answer options and descriptions of the answer options for each column. This can be used as a reference when filling out the tables in the Excel file "analysis_pretest.xlsx". This is the way the following table is built up:

Column, column title	**Question**	
Answer option 1	Description of the answer option 1	General explanation of the answer options, if necessary.
Answer option 2	Description of the answer option 2	

Questions and answer categories of the detail level:

C, stafirst	**Only focus on the graphical elements that convey the main statement of the DV. If there are more of the same kind, decide on one exemplary unit, comparable to a clause. Try to formulate a first draft of how this clause could be formulated.**	
This answer is already given.	Since this question is necessary to define which graphical element is in focus, and this needs to be the same for all raters, this answer is already given. It will be taken up again in column M.	
D, viselemdescr	**What kind of graphical element(s) represent(s) this statement?**	
free text	In this analysis the answer should always be "line and the connected point(s)" or "line". This means that either one or both connected points are made especially visible (e.g. through dots or other graphical elements), or not. In the centre of the main statement is a connection signalled by a line connecting something visible or invisible (further referred to as "connecting line" or "line in focus".	
E, codingorient	**Which coding orientation needs to be applied when viewing the connecting line?**	
abstract	If you would say, the more the line "represents the deeper 'essence'" of the statement or "the more it represents the general pattern underlying superficially different specific instances, the higher its modality" (Van Leeuwen, 2005, p. 168), then the answer option "abstract" has to be taken. (Or in other words: if a more reduced articulation expresses the above-named statement, then take this answer option.)	When is the connection (signified by the line) considered as true? These answer categories are derived from the coding orientations described by Van Leeuwen (2005, pp. 168–171). In case there is not clearly one or other category that fits, please choose the option that fits best in your opinion.
naturalistic	If you would say, the more the statement is represented as an "accurate external representation" (Van Leeuwen, 2005, p. 168), the more it is considered as true, then the answer option "naturalistic" has to be taken.	
technological	If you would say, the more the connecting line "can be used as a blueprint or aid for action" (Van Leeuwen, 2005, p. 168), the more it is considered as true, then this is the answer option to be taken. Here as well, a "strongly decreased articulation" (Van Leeuwen, 2005, p. 168) may be used.	
sensory	If you would say, the more "the effect of pleasure or displeasure created" (Van Leeuwen, 2005, p. 170) by the connecting line, the more the statement is considered as true, then take this answer option.	
none of these	If you think none of these other orientations can be applied, choose this answer option.	

F, lineappgener	What does the connecting line look like?
This needs to be filled out in the separate table in tab “visualappear-ance”.	For this pre-test, this table is already filled out.
G, uncertainty	**Is there any form of intended uncertainty in terms of lowered probability (that what the connecting line depicts didn’t, doesn’t/won’t for sure happen or wasn’t/isn’t/won’t be like that) or lowered reliability represented through the line in focus?**
yes	Select “yes”, if the line represents the statement in column C as uncertain and/or unrelia-ble to some extent.
no	Select “no”, if the statement in column C is represented as certain and reliable. If “no” is selected, the answer to the next question and respectively all questions in the tab “visual-appearanceuncertainty” is “NULL”.
H, yeshowvisvar	**If yes, how is this lowered probability and/or lowered reliability signalled visually? (Which of the visual variables are used to signal this uncertainty?)**
This needs to be filled out in the separate table in tab “visualappearan-ceuncertainty”.	
I, datauncert	**Is it explicitly verbally stated on the detail or on the global level that data uncertain-ty is represented within this detail statement of the data visualization? (If no, then the next field is NULL)**
yes	Select “yes” if you can find a verbal clue stating that the underlying data for this state-ment were uncertain. Words like “scenario”, “confidence interval”, “forecast” and are examples of such verbal clues.
no	Select “no” if there’s either no uncertainty signalled at all (see column G), or if you can’t find any verbal explanation of why the connecting line is signalling data uncertainty.
J, datauncsta	**If it is explicitly verbally stated on the detail or on the global level that data uncer-tainty is represented, how?**
NULL	Write “NULL” if the answer to the previous question was “no”.
free text	Write in the concrete formulation, for example, “95% confidence” or “Best-case scenario, optimistic scenario, current scenario”.
K, loweredrel	**Based on the previous questions, does the line in focus signal modality in terms of intended lowered reliability?**
yes	Select “yes”, if the answer to the question in column G was “yes”, and it’s not verbally explained why the connection is visualized as uncertain (no clue to data uncertainty).
no	Select “no”, if the connecting line does not signify any uncertainty at all (see column G), or if it’s explained verbally that data uncertainty is represented and the reason for the data uncertainty is made clear (see columns I and J).

<table>
<tr><td>L, modalityva</td><td colspan="2">If yes, which modality value is signalled?</td></tr>
<tr><td>NULL</td><td colspan="2">Select “NULL” if the answer to the question in column K was “no”. This means that there’s no modality (in terms of lowered reliability) signalled at all through that connecting line.</td></tr>
<tr><td>high modality value</td><td>Select this answer category if you would use the adverb “certainly” to describe the reliability of this detail statement in verbal language. This is the case when the following propositions about the visual appearance of the line in focus are fairly correct:
• abstract orientation: reduced articulation is used general pattern or deeper “essence” of what it depicts is represented
• naturalistic orientation: the graphical element in focus resembles the way we would see it if we saw it in reality
• technological orientation: the graphical element(s) in focus is practically useful, as blueprint or aid for action
• sensory orientation: the line(s) in focus has/have an effect of giving pleasure or displeasure
• none of these orientations: no theoretical basis is given here, please formulate your argumentation in column Y.</td><td rowspan="3">These three answer categories as well as the adverbs used in verbal language are based on Halliday (1985/2004, p. 620).

The descriptions of the answer categories are derived from the coding orientations described by Van Leeuwen (2005, pp. 168–171).</td></tr>
<tr><td>median modality value</td><td>Select this answer category if you would use the adverb “probably” to describe the reliability of this detail statement in verbal language. This is the case when the orientation explanations of the above field are rather not correct.</td></tr>
<tr><td>low modality value</td><td>Select this answer category, if you would use the adverb “possibly” to describe the reliability of this detail statement in verbal language. This is the case when the orientation explanations of the above field are not correct.</td></tr>
<tr><td>M, staafter</td><td colspan="2">Does the verbal sentence from column C need a reformulation, considering modality value and data uncertainty, if any of them are expressed? If yes, which?</td></tr>
<tr><td>NULL</td><td>Type “NULL” if the answers in columns G and K were “no”.</td><td rowspan="2">This question is asked only so that the former answers are thought through again. There’s no need to have exactly the same answers by all raters for this question. However, if data uncertainty or modality is expressed by this particular line, this shall also be tried to be expressed in the revised verbal statement.</td></tr>
<tr><td>free text</td><td>Type a revised verbal statement, if the connecting line signifies any kind of uncertainty or modality (in terms of lowered reliability). Use, for example, the adverbs referring to the single modality values, or the data uncertainty source explanations.</td></tr>
<tr><td>N, notesline</td><td colspan="2">If you have any other comments, that may be useful for a more qualitative investigation of this particular DV or the data analysis in the next step, or if anything was unclear, please write them down here.</td></tr>
<tr><td>free text</td><td colspan="2">Write down your comments in keywords.</td></tr>
<tr><td></td><td colspan="2">If you have no comments, leave this field empty.</td></tr>
</table>

Questions and answer categories of the global level:

O, glstafirst	**What is the overall statement of this data visualization?**	
This answer is already given.		
P, glcodingorie	**Which coding orientation needs to be applied to the visual expression of the overall statement?**	
abstract	If you would say the more the whole visualization "represents the deeper 'essence'" of the overall statement or "the more it represents the general pattern underlying superficially different specific instances, the higher its modality" (Van Leeuwen, 2005, p. 168), then the answer option "abstract" has to be taken. (Or in other words: if a more reduced articulation expresses the above-named meaning, then take this answer option.)	In contrast to the question in column E, here the whole data visualization is the object of attention. This means that the same or different coding orientations can be applied on the detail level (for the way the line in focus signals the detail statement as true) and on the global level (for the way the whole data visualization signals the overarching statement as true) These answer categories are derived from the coding orientations described by Van Leeuwen (2005, pp. 168–171) In case there is not clearly one or other category that fits, please choose the option that fits best in your opinion.
naturalistic	If you would say the more the overall statement is represented as an "accurate external representation" (Van Leeuwen, 2005, p. 168), the more it is considered as true, then the answer option "naturalistic" has to be taken.	
technological	If you would say the more this DV "can be used as a blueprint or aid for action" (Van Leeuwen, 2005, p. 168), the more it is considered as true, then this is the answer option to be taken. Here as well, a "strongly decreased articulation" (Van Leeuwen, 2005, p. 168) may be used.	
sensory	If you would say the more "the effect of pleasure or displeasure created" (Van Leeuwen, 2005, p. 170) by the DV, the more the overall statement is considered as true, then take this answer option.	
none of these	If you think none of these other orientations can be applied, choose this answer option.	
Q, glloweredrel	**Does the choice of visual style signal any sense of modality in terms of intended lowered reliability?**	
yes	Select "yes" if the overall statement of the DV is represented as not 100% reliable. This is the case if, depending on the coding orientation (column P), a less abstract/naturalistic/etc. way of visualizing the overall statement has been used.	
no	Select "no" if the overall statement of the DV is presented as totally reliable. This is the case if, depending on the coding orientation (column P), a maximally abstract/naturalistic/etc. way of visualizing the overall statement has been used.	

R, glmodalityva	**If yes, which modality value is signalled?**	
NULL	Select "NULL" if the answer to the question in column P was "no". This means that there's no modality (in terms of lowered reliability) signalled at all.	
high modality value	Select this answer category, if you would use the adverb "certainly" to describe the modality value in verbal language. This is the case when it's fairly correct that in: • abstract orientation: reduced articulation is used, general pattern or deeper "essence" of what it depicts is represented • naturalistic orientation: image resembles the way we would see it if we saw it in reality • technological orientation: DV is practically useful, as blueprint or aid for action • sensory orientation: DV has an effect of pleasure or displeasure • none of these orientations: no theoretical basis is given here, please formulate your argumentation in the column Y.	These three answer categories as well as the adverbs used in verbal language are based on Halliday (1985/2004, p. 620). The descriptions of the answer categories are derived from the coding orientations described by Van Leeuwen (2005, pp. 168–171).
median modality value	Select this answer category, if you would use the adverb "probably" to describe the modality value in verbal language. This is the case when the orientation explanations of the above field are not really correct.	
low modality value	Select this answer category, if you would use the adverb "possibly" to describe the modality value in verbal language. This is the case when the orientation explanations of the above field are not correct.	
S, gldatauncert	**Is it explicitly verbally stated on the detail or global level that data uncertainty is represented in the overall statement of the DV?**	
yes	Select "yes" if you can find a verbal clue stating on the global level that the underlying data for the overall statement were uncertain. Words like "scenario", "confidence interval" and "forecast" are examples of such verbal clues.	Since it's possible that verbal statements about data uncertainty are made on the global as well as on the detail level, both the question in column I as well as this one may be answered with "yes".
no	Select "no" if you can't find any verbal statement on the global level that the underlying data for the overall statement of the DV were uncertain.	
T, gldatauncsta	**If it is explicitly verbally stated on the detail or the global level that data uncertainty is represented, how?**	
NULL	Write "NULL" if the previous question was answered with "no".	
free text	Write down the concrete formulation, for example, "95% confidence" or "Best-case scenario, optimistic scenario, current scenario".	

U, gloverallmod	**Looking back at all the answers to the previous questions (both on the detail and the global level, maybe including intended modality (in terms of lowered reliability) or data uncertainty expressions), what kind of overall modality (in terms of lowered reliability) profile does the DV have?**	
no modality (in terms of lowered reliability)	Choose this answer if neither on the detail nor on the global level any kind of modality was signalled (if in columns K and Q “no” was the answer) and you therefore think that the (aspect of) the “experiential” (Kress & Van Leeuwen, 1996/2006, pp. 42, 46) world is presented as 100% reliable.	This question is based on Boeriis’ claim that the overall modality of a text is a product of all modality profiles on all levels (2012, p. 147). If on the detail or on the global level there is data uncertainty represented, this should be considered as an influencing factor too. The three last answer categories as well as the adverbs used in verbal language are based on Halliday (1985/2004, p. 620).
high modality value	Select this answer if you think the DV presents the (aspect of) the “experiential” (Kress & Van Leeuwen, 1996/2006, pp. 42, 46) world as not 100% reliable. The adverb used to describe the reliability in verbal language would be “certainly”.	
median modality value	Select this answer if you think the DV presents the (aspect of) the “experiential” (Kress & Van Leeuwen, 1996/2006, pp. 42, 46) world as not 100% reliable. The adverb used to describe the reliability in verbal language would be “probably”.	
low modality value	Select this answer if you think the DV presents the (aspect of) the “experiential” (Kress & Van Leeuwen, 1996/2006, pp. 42, 46) world as not 100% reliable. The adverb used to describe the reliability in verbal language would be “possibly”.	
V, glstaafter	**Does the verbal statement from column O need a reformulation, considering all the former questions (on the global as well as on the detail level)?**	
NULL	Type “NULL” if the verbal statement from column O doesn’t need any reformulation.	This question is only asked so that the previous answers are thought through again. There’s no need to have exactly the same answers by all raters for this question.
free text	Type a revised verbal statement, if the sentence from column O does not yet express all notions of lowered probability, data uncertainty and lowered reliability (if those are signified). Use, for example, the adverbs referring to the single modality values or the data uncertainty source explanations.	
W, glnotes	**If you have any other comments that may be useful for a more qualitative investigation of this particular DV or the data analysis in the next step, or if anything was unclear, please write this down here.**	
free text	Write down your comments in keywords.	
	If you have no comments, leave this field empty.	

4.2 The “visualappearance” tab

This table has been filled out in advance, because it is only about what the lines look like, not what they mean. That’s why the answers to this question are not considered to be variables that rely “on the analyst’s judgement and interpretation” (Brezina, 2018, p. 269). However, a list of what categories are available, and some additional explanations of what they mean are given in the following table. The question for the whole table is: What does the line (described in column C in the “semioticfunctionanalysis” table) look like?

column, column title	**long version of the column title**	**answer option**	**additional explanations if necessary**
C, colour	colour	no variation (single-coloured)	
		abrupt variation	
		smooth transition	
D, clarcrisp	clarity: crispness	high	These categories derive from MacEachren (1995, pp. 275–277) and are further described there.
		lowered	
E, clartrans	clarity: transparency	none	
		medium or high High transparency mean that it is still slightly visible.	
F, clarresol	clarity: resolution	high	
		lowered	
G, pattern	pattern	continuous line	
		interrupted line	No matter what form the different elements have. (E.g., the line may be dashed, dotted, or consist of at least two of any other elements.)
H, shapeforc	shape: forces	one force (straight)	The description of the line shape with the term *forces* derives from Kandinsky (1926/1928, p. 51).
		two forces, curved	
		two forces, bent	
		three or more forces, curved	
		three or more force, bent	
I, shapelipr	shape: line pressure	consistent	
		inconsistent	
J, shapeextr	shape: form of extremities	directly connected to the end point	This option has to be chosen, if there’s no change in shape of the extremity towards the end point.
		direction-signifying extremity	This may, for example, be an arrowhead, or a tapering body (decreasing line pressure).
K, interaction	interaction	not possible	Here the question is whether the visual appearance of the line changes through an action of the viewer (e.g., mouse-over or mouse-click). The interaction can take place with the line directly, or with other elements of the DV. If interaction is possible, but the visual appearance of the line does not change, “not possible” is the answer to take. (This means that choosing the answer “possible” here, has as a consequence that dynamics are available (next question).
		possible	
L, dynamics	dynamics	not available	If this answer is chosen, the following columns have to be answered with “NULL”.
		available	If this answer is chosen, the following columns have to be answered with either “yes” or “no”. The term *dynamics*, means that what the line looks like changes. (If only other elements of the DV change, then this answer is not chosen.)

M, dynposition	dynamics in the visual variable position	NULL	
		yes	If the whole line (meaning the starting and end point) moves, this option has to be chosen.
		no	
N, dynorient	dynamics in the visual variable orientation	NULL	
		yes	
		no	
O, dynsize	dynamics in the visual variable size	NULL	
		yes	If the line length changes, this option has to be chosen.
		no	
P, dyncolour	dynamics in the visual variable colour	NULL	
		yes	
		no	
Q, dynclarcrisp	dynamics in the visual variable clarity: crispness	NULL	
		yes	
		no	
R, dynclartrans	dynamics in the visual variable clarity: transparency	NULL	
		yes	If the line gets more or less transparent (with stages of transparency where the line is still visible but elements in the back are visible as well), this answer has to be chosen. If the line appears or disappears completely (at once without growing, without any stages of transparency where the line is still visible but elements in the back are visible as well) (e.g., because of the mouse position, clicking, scrolling or filtering), this answer is chosen as well.
		no	
S, cynclarresol	dynamics in the visual variable clarity: resolution	NULL	
		yes	
		no	
T, dynpattern	dynamics in the visual variable pattern	NULL	
		yes	If the visual appearance of single line elements (e.g., the length of the dashes of a dashed line) changes, then this answer has to be chosen.
		no	
U, dynshapeforc	dynamics in the visual variable shape: forces	NULL	
		yes	
		no	
V, dynshapelipr	dynamics in the visual variable shape: line pressure	NULL	
		yes	
		no	
W, dynshapeextr	dynamics in the visual variable shape: form of extremities	NULL	
		yes	
		no	

4.3 The “visualappearanceuncertainty” tab

This table has to be filled out by you. It is connected to column G in the “semioticfunctionanalysis” table. There you have answered whether the line represents any form of uncertainty. In this table you explain how, by choosing which visual variable (from the "visualappearance" table) was used to signify uncertainty. For all columns, the following three answer options are available:

NULL	Choose this answer if there is no uncertainty signified at all. (This is the case if the answer to column G in the "semioticfunctionanalysis" table was "No".)
used	Choose this answer if this visual variable is used to signify uncertainty.
not used	If this visual variable was not used to signify uncertainty, but one or more other visual variables were used for this purpose, then choose this answer.

5. General instructions for using the Excel files

- Only the Excel file "analysis_pretest.xlsx" will be used for calculating the inter-rater reliability. This means that no changes in the Excel file "visualizationlistforanalysis.xlsx" (like colour marking) will have any consequences.
- In the Excel file "analysis_pretest.xlsx", for most of the questions there are answer options offered. You can see them if you click first into the empty field, and then on the arrow appearing on the right side of the field. If there are such answer options available, only use these ones! Don't write any free text into such fields or add any letters or spaces after one answer option has been selected.
- Sometimes it may not be the case that there is clearly only one answer that could fit. In such a case, please choose the option that fits best in your opinion.

References

Bell, P. (2001). Content analysis of visual images. In T. Van Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 10–34). SAGE.

Boeriis, M. (2012). Tekstzoom: Om en dynamisk funktionel rangstruktur i visuelle tekster. In T. H. Andersen & M. Boeriis (Eds.), *Nordisk socialsemiotik: Multimodale, pædagogiske og sprogvidenskabelige landvindinger* (pp. 131–153). University Press of Southern Denmark.

Brezina, V. (2018). Statistical choices in corpus-based discourse analysis. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 259–280). Routledge.

Halliday, M. A. K. (2004). *An introduction to functional grammar* (3rd ed. revised by Matthiessen, C.M.I.M.). Arnold (Hodder Education). (Original work published 1985)

Kandinsky, W. (1928). *Punkt und Linie zu Fläche: Beitrag zur Analyse der malerischen Elemente* (2nd unchanged ed.). Albert Langen. http://archive.org/details/punktun00kand (Original work published 1926)

Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge. (Original work published 1996)

MacEachren, A. M. (1995). *How maps work: Representation, visualization, and design.* Guilford Press.

Van Leeuwen, T. (2005). *Introducing social semiotics.* Routledge.

Appendix E: Article II: Details Concerning the Benchmarking of the Inter-Rater-Reliability Study Results

The benchmarking method proposed by Gwet (Gwet, 2014, pp. 173–181), complemented by information given by Gwet in personal mail correspondence on 9-17 July 2019, was used to interpret the resulting Gwet's AC_1 and Gwet's AC_2 coefficients and the corresponding standard errors. As a basis for the benchmarking process, the Landis and Koch's Benchmark Scale (Landis & Koch, 1977) was chosen. This scale was slightly adjusted following Gwet's advice in personal mail correspondence on 17 July 2019, resulting in the intervals (a, b) presented in Table 1. "a" represents the lower value in the interval, "b" represents the higher one.

Table 1
Adjusted Landis and Koch Kappa's Benchmark Scale, Used as the Basis for the Interpretation of Gwet's AC_1 and AC_2 Coefficients

Kappa Statistic (interval from a to b)	**Strength of Agreement (benchmark level)**	**Comments to the fields to the left**
-1 to 0	Poor	lowest benchmark level and corresponding interval
0 to 0.2	Slight	
0.2 to 0.4	Fair	
0.4 to 0.6	Moderate	
0.6 to 0.8	Substantial	
0.8 to 1	Almost Perfect	highest benchmark level and corresponding interval
-1 to 1		total interval

After the determination of the benchmark scale, the Interval Membership Probability (IMP) was calculated in Microsoft Excel, as described by Gwet (2014, pp. 177–178):

IMP = NORM.S.DIST((COEFF-a)/SE,TRUE) - NORM.S.DIST((COEFF-b)/SE,TRUE)

This formula was applied to each question's (as listed in Table 9 of the main part, question columns E, G, I, K, L, P, Q, R, S and U) coefficient (COEFF) and standard error (SE) value (see Table 9 of the main part, column 3 and 4). For each question which didn't have perfect agreement (all but question E in Table 9 of the main part), the Interval Membership Probability (IMP) was calculated for all intervals listed in Table 1, starting from the highest to the lowest level. The IMP describes the probability, to which the coefficient lies within that interval. The tables of Appendix F present the IMP for each question and interval, as well as the Max IMP, revised IMP and the cumulative benchmark probability (as explained in the following) for each question.

The calculation of the maximum IMP value (Max IMP) was done following Gwet's further explanations of the method presented in Gwet 2014, given in the same mail correspondence mentioned before. It is calculated the same way as shown in the IMP equation, with the only difference being the use of the maximum interval from the lowest to the highest interval value (in the case of the Kappa Statistic from -1 to 1). Further, the revised IMP was calculated by dividing the IMP by the Max IMP (revised IMP = IMP / Max IMP).

Based on this, the cumulative benchmark probability was calculated (Gwet, 2014, p. 178). For this, the revised IMP values were summed up starting from the highest benchmark interval to the lowest. These cumulative benchmark probability values were finally used to determine the final agreement level of the adjusted Landis and Koch Kappa Benchmark Scale (see Table 1). The final agreement level is the "highest benchmark level associated with the smallest cumulative probability that exceeds 0.95" (Gwet, 2014, p. 178).

References

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. (4th ed.). Advanced Analytics, LLC.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. JSTOR. https://doi.org/10.2307/2529310

Appendix F: Article II: Calculation Results of Benchmarking the Inter-Rater-Reliability Study Results

Table 1

Calculation Results of Question Column G of Table 9 in the Main Part

Interval 1: almost perfect		**Interval 2: substantial**		**Interval 3: moderate**		**Interval 4: fair**		**Interval 5: slight**		**Interval 6: poor**	
COEFF	0.8647	COEFF	0.8647	COEFF	0.8647	COEFF	0.8647	COEFF	0.8647	COEFF	0.8647
a	0.8	a	0.6	a	0.4	a	0.2	a	0	a	-1
b	1	b	0.8	b	0.6	b	0.4	b	0.2	b	0
SE	0.0820	SE	0.0820	SE	0.0820	SE	0.0820	SE	0.0820	SE	0.0820
IMPa	0.7852	IMPa	0.9994	IMPa	0.1000	IMPa	1	IMPa	1	IMPa	1
IMPb	0.0495	IMPb	0.7852	IMPb	0.9994	IMPb	0.1000	IMPb	1	IMPb	1
IMP	0.7357	IMP	0.2142	IMP	0.0006	IMP	0.0000	IMP	0	IMP	0
Max IMP	0.9505	Max IMP	0.9505	Max IMP	0.9505	Max IMP	0.9505	Max IMP	0.9505	Max IMP	0.9505
IMP-Revised	0.7740	IMP-Revised	0.2254	IMP-Revised	0.0007	IMP-Revised	0.0000	IMP-Revised	0	IMP-Revised	0
cumulative benchmark probability:	0.7740	cumulative benchmark probability:	**0.9993**	cumulative benchmark probability:	0.1000	cumulative benchmark probability:	1	cumulative benchmark probability:	1	cumulative benchmark probability:	1

The value in the grey field defines the final agreement level because this is the "highest benchmark level associated with the smallest cumulative probability that exceeds 0.95" (Gwet, 2014, p. 178)

Note.
COEFF = result of the calculated coefficient (Gwet's AC_1 or AC_2);
a = lower bound of this interval;
b = upper bound of this interval;
SE = Standard Error;
IMP = Interval Membership Probability;
Max IMP = Maximum Interval Membership Probability.
These definitions also apply to the same abbreviations in all other tables of Appendix F.

References

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters.* (4th ed.). Advanced Analytics, LLC.

Table 2

Calculation Results of Question Column I of Table 9 in the Main Part

	Interval 1: almost perfect	Interval 2: substantial	Interval 3: moderate	Interval 4: fair	Interval 5: slight	Interval 6: poor
COEFF	0.9549	0.9549	0.9549	0.9549	0.9549	0.9549
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.0459	0.0459	0.0459	0.0459	0.0459	0.0459
IMPa	0.9996	1	1	1	1	1
IMPb	0.1631	0.9996	1	1	1	1
IMP	0.8366	0.0004	0.0	0	0	0
Max IMP	0.8369	0.8369	0.8369	0.8369	0.8369	0.8369
IMP-Revised	0.9996	0.0004	0.0	0	0	0
cumulative benchmark probability:	**0.9996**	1	1	1	1	1

Table 3

Calculation Results of Question Column K of Table 9 in the Main Part

	Interval 1: almost perfect	Interval 2: substantial	Interval 3: moderate	Interval 4: fair	Interval 5: slight	Interval 6: poor
COEFF	0.8647	0.8647	0.8647	0.8647	0.8647	0.8647
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.0820	0.0820	0.0820	0.0820	0.0820	0.0820
IMPa	0.7852	0.9994	0.1000	1	1	1
IMPb	0.0495	0.7852	0.9994	0.1000	1	1
IMP	0.7357	0.2142	0.0006	0.0000	0	0
Max IMP	0.9505	0.9505	0.9505	0.9505	0.9505	0.9505
IMP-Revised	0.7740	0.2254	0.0007	0.0000	0	0
cumulative benchmark probability:	0.7740	**0.9993**	0.1000	1	1	1

Table 4
Calculation Results of Question Column L of Table 9 in the Main Part

	Interval 1: almost perfect	Interval 2: substantial	Interval 3: moderate	Interval 4: fair	Interval 5: slight	Interval 6: poor
COEFF	0.9121	0.9121	0.9121	0.9121	0.9121	0.9121
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.0560	0.0560	0.0560	0.0560	0.0560	0.0560
IMPa	0.9773	0.1000	1	1	1	1
IMPb	0.0585	0.9773	0.1000	1	1	1
IMP	0.9188	0.0227	0.0000	0	0	0
Max IMP	0.9415	0.9415	0.9415	0.9415	0.9415	0.9415
IMP-Revised	0.9759	0.0241	0.0000	0	0	0
cumulative benchmark probability:	**0.9759**	0.1000	1	1	1	1

Table 5
Calculation Results of Question Column P of Table 9 in the Main Part

	Interval 1: almost perfect	Interval 2: substantial	Interval 3: moderate	Interval 4: fair	Interval 5: slight	Interval 6: poor
COEFF	0.9130	0.9130	0.9130	0.9130	0.9130	0.9130
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.0595	0.0595	0.0595	0.0595	0.0595	0.0595
IMPa	0.9712	0.1000	1	1	1	1
IMPb	0.0720	0.9712	0.1000	1	1	1
IMP	0.8993	0.0288	0.0000	0	0	0
Max IMP	0.9280	0.9280	0.9280	0.9280	0.9280	0.9280
IMP-Revised	0.9690	0.0310	0.0000	0	0	0
cumulative benchmark probability:	**0.9690**	0.1000	1	1	1	1

Table 6

Calculation Results of Question Column Q of Table 9 in the Main Part

	Interval 1: almost perfect	**Interval 2: substantial**	**Interval 3: moderate**	**Interval 4: fair**	**Interval 5: slight**	**Interval 6: poor**
COEFF	0.9584	0.9584	0.9584	0.9584	0.9584	0.9584
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.0424	0.0424	0.0424	0.0424	0.0424	0.0424
IMPa	0.9999	1	1	1	1	1
IMPb	0.1632	0.9999	1	1	1	1
IMP	0.8367	0.0001	0	0	0	0
Max IMP	0.8368	0.8368	0.8368	0.8368	0.8368	0.8368
IMP-Revised	0.9999	0.0001	0	0	0	0
cumulative benchmark probability:	**0.9999**	1	1	1	1	1

Table 7

Calculation Results of Question Column R of Table 9 in the Main Part

	Interval 1: almost perfect	**Interval 2: substantial**	**Interval 3: moderate**	**Interval 4: fair**	**Interval 5: slight**	**Interval 6: poor**
COEFF	0.9863	0.9863	0.9863	0.9863	0.9863	0.9863
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.0139	0.0139	0.0139	0.0139	0.0139	0.0139
IMPa	1	1	1	1	1	1
IMPb	0.1611	1	1	1	1	1
IMP	0.8389	0	0	0	0	0
Max IMP	0.8389	0.8389	0.8389	0.8389	0.8389	0.8389
IMP-Revised	1	0	0	0	0	0
cumulative benchmark probability:	**1**	1	1	1	1	1

Table 8

Calculation Results of Question Column S of Table 9 in the Main Part

	Interval 1: almost perfect	**Interval 2: substantial**	**Interval 3: moderate**	**Interval 4: fair**	**Interval 5: slight**	**Interval 6: poor**
COEFF	0.9549	0.9549	0.9549	0.9549	0.9549	0.9549
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.0459	0.0459	0.0459	0.0459	0.0459	0.0459
IMPa	0.9996	1	1	1	1	1
IMPb	0.1631	0.9996	1	1	1	1
IMP	0.8366	0.0004	0.0	0	0	0
Max IMP	0.8369	0.8369	0.8369	0.8369	0.8369	0.8369
IMP-Revised	0.9996	0.0004	0.0	0	0	0
cumulative benchmark probability:	**0.9996**	1	1	1	1	1

Table 9

Calculation Results of Question Column U of Table 9 in the Main Part

	Interval 1: almost perfect	**Interval 2: substantial**	**Interval 3: moderate**	**Interval 4: fair**	**Interval 5: slight**	**Interval 6: poor**
COEFF	0.9065	0.9065	0.9065	0.9065	0.9065	0.9065
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.0561	0.0561	0.0561	0.0561	0.0561	0.0561
IMPa	0.9712	0.1000	1	1	1	1
IMPb	0.0478	0.9712	0.1000	1	1	1
IMP	0.9234	0.0288	0.0000	0	0	0
Max IMP	0.9522	0.9522	0.9522	0.9522	0.9522	0.9522
IMP-Revised	0.9698	0.0302	0.0000	0	0	0
cumulative benchmark probability:	**0.9698**	0.1000	1	1	1	1

Appendix G: Article II: Final Coding Instructions

Coding scheme

for the single-coded corpus analysis
focusing on how modality and uncertainty are signalled by connecting lines

1. Definition of the units of analysis and the used documents

Within this analysis, the entire corpus is analysed by the main author of this dissertation. The following documents are used:

(1) the Excel file "analysis.xlsx"
(2) the Excel file "visualizationlistforanalysis.xlsx
(3) a folder with JPG, PNG, PDF and mp4 documents as described in the relational database.

This is what these documents are used for:

(1) This is the document where the coding takes place. It has three tabs, each of which contains a table that needs to be filled out. The first two columns of all three tables have the same content: the number and title of the DVs that are analysed. This means that each row stands for one DV. The columns from C onwards stand for one question each.
(2) The Excel file "visualizationlistforanalysis.xlsx" helps to find the right documents in the document folder (3.) and was directly exported from the relational database. It contains all the filenames of these screenshots and screencasts, as well as the original link to the website where the DV stems from, are listed here. Since more than just one DV can stem from one website, the column "visimageFN" contains the filenames of JPGs that define which DV is being investigated and

coded within this row. All the following columns ("winnerID", "imgFN", "filmFN", "pdfFN" and "url") can therefore have the same content in several rows.

(3) This folder contains the research material, the filenames of which are documented in the Excel file "visualizationlistforanalysis.xlsx".

2. First steps

(1) Open the Excel file "analysis.xlsx" (1.), going to the first tab.
(2) Look at which visualization is the first to analyse. (The columns "visID" and "vistitle" tell you this.) Let's say this would be visID "16", vistitle "Flight tracks".
(3) Open the Excel file "visualizationlistforanalysis.xlsx" and go to the row with the same visID and vistitle. Mark it, for example, by using a filling colour for this particular row.
(4) Open Windows Explorer/Finder and navigate to the folder with all the research material (3.).
(5) Search for the filenames written in the marked row (step 3.) and open these documents. In our example this would be "Spies_in_the_skies.JPG", "spies_in_the_skies.png", "Spies_in_the_skies.mp4" and "Spies_in_the_skies.pdf". Maybe also opening the website from which these screenshots and screencasts have been taken can be helpful (in this case https://www.buzzfeed.com/peteraldhous/spies-in-the-skies?utm_term=.rlwl2JVE7#.hhzwqNE1O). But it has be noted that changes may have been made since the screenshots and screencasts were taken. That's why these documents, and not the online website, stay the main point of reference.
(6) Go back to the Excel file "analysis.xlsx" (1.) and answer the first question for the DV investigated, by looking at the files opened in step 5.
(7) Go through all questions in all three tabs of the "analysis.xlsx" file (for this DV) and answer them.
(8) Demark the row marked in step 3.
(9) Then repeat these steps for each DV by going back to step 2, finding the next DV and so on.

3. Descriptions of the questions and answer categories

In this section, the questions and answer categories available within the Excel file "analysis.xlsx" are explained.

3.1 The "semioticfunctionanalysis" tab

The questions in this tab are organized into two sections: the detail level and the global level. This refers to the "zooming levels".

On the detail level, what is in the focus for answering the question is one particular connecting line, as well as the surrounding elements, related to this line (like labels, e.g.). (In the words of Boeriis' dynamic functional rank scale (2012, p. 139) the detail level summarizes the component, unit and group rank.)

On the global level, the whole data visualization, including any surrounding co-text that's related to this visualization, is evaluated. (In the words of Boeriis' dynamic functional rank scale (2012, p. 139) the global level summarizes the whole rank and the co-text.)

The following tables show the column character, column title, the complete question, answer options and descriptions of the answer options for each column. This can be used as a reference when filling out the tables in the Excel file "analysis.xlsx". This is the way the following table is built up:

Column, column title	**Question**	
Answer option 1	Description of the answer option 1	General explanation to the answer options, if necessary.
Answer option 2	Description of the answer option 2	

Questions and answer categories of the detail level:

C, stafirst	**Only focus on the graphical elements that convey the main statement of the DV. If there are more of the same kind, decide for one exemplary unit, comparable to a clause. How could this clause be formulated?**
free text	
D, viselemdescr	**What kind of graphical element(s) represent(s) this statement?**
free text	In this analysis the answer should always be "line and the connected point(s)" or "line". This means that either one or both connected points are made especially visible (e.g. through dots or other graphical elements), or not. In the centre of the main statement is a connection signalled by a line connecting something visible or invisible (further referred to as "connecting line" or "line in focus".

<table>
<tr><th>E, codingorient</th><th colspan="2">Which coding orientation needs to be applied when viewing the connecting line?</th></tr>
<tr><td>abstract</td><td>If you would say the more the line “represents the deeper ‘essence’” of the statement or “the more it represents the general pattern underlying superficially different specific instances, the higher its modality” (Van Leeuwen, 2005, p. 168), then the answer option “abstract” has to be taken. (Or in other words: if a more reduced articulation expresses the above-named statement, then take this answer option.)</td><td rowspan="5">When is the connection (signified by the line) considered as true?

These answer categories are derived from the coding orientations described by Van Leeuwen (2005, pp. 168–171).

In case there is not clearly one or other category that fits, please choose the option that fits best in your opinion.</td></tr>
<tr><td>naturalistic</td><td>If you would say the more the statement is represented as an “accurate external representation” (Van Leeuwen, 2005, p. 168), the more it is considered as true, then the answer option “naturalistic” has to be taken.</td></tr>
<tr><td>technological</td><td>If you would say the more the connecting line “can be used as a blueprint or aid for action” (Van Leeuwen, 2005, p. 168), the more it is considered as true, then this is the answer option to be taken. Here as well, a “strongly decreased articulation” (Van Leeuwen, 2005, p. 168) may be used.</td></tr>
<tr><td>sensory</td><td>If you would say the more “the effect of pleasure or displeasure created” (Van Leeuwen, 2005, p. 170) by the connecting line, the more the statement is considered as true, then take this answer option.</td></tr>
<tr><td>none of these</td><td>If you think that none of these other orientations can be applied, choose this answer option.</td></tr>
<tr><th>F, lineappgener</th><th colspan="2">What does the connecting line look like?</th></tr>
<tr><td>This needs to be filled out in the separate table in tab “visualappearance”.</td><td colspan="2"></td></tr>
<tr><th>G, uncertainty</th><th colspan="2">Does the visual appearance of the connecting line indicate any form of modality in terms of lowered probability (that what the connecting line depicts didn’t, doesn’t/won’t for sure happen or wasn’t/isn’t/won’t be like that) or lowered reliability?</th></tr>
<tr><td>yes</td><td colspan="2">Select “yes” if the line represents the statement in column C as uncertain and/or unreliable to some extent. This is the case when what the connecting line looks like has characteristics of one or more visual variables that signify some kind of uncertainty. (Which visual appearance is decided on is highly dependent on the coding orientation in use (see column E)). Some visual characteristics may be conventionalized to signal uncertainty, maybe stemming from experiences of the “experiential” (Kress & Van Leeuwen, 1996/2006, pp. 42, 46) world. Sometimes, a deviation from the standard visual appearance of line of the DV type in focus, or a deviation of a single line in comparison to the other lines (in a DV containing several connecting lines) can also point to an indication of uncertainty.

Moreover, the following reasons for the visual appearance have to be ruled out as the reason for this specific visual appearance: the depiction of dimensions other than probability or reliability (such as different categories by different kind of lines), the intention to create a certain visual effect, the technical production tools favouring that kind of visual appearance.</td></tr>
<tr><td>no</td><td colspan="2">Select “no” if the statement in column C is represented as certain and reliable. If “no” is selected, the answer to the next question and respectively all questions in the tab “visualappearanceuncertainty” is “NULL”.</td></tr>
</table>

H, yeshowvisvar	**If yes, how is this lowered probability and/or lowered reliability signalled visually? (Which of the visual variables are used to signal this uncertainty?)**	
This needs to be filled out in the separate table in tab "visualap-pearanceuncertainty"		
I, datauncert	**Is it explicitly verbally stated on detail or on global level of the data visualization, or in the co-text, that data uncertainty is represented within this detail statement of the data visualization? (If no, then the next field is NULL)**	
yes	Select "yes" if you can find a verbal clue stating that the underlying data for this statement was uncertain. Words like "scenario", "confidence interval", "forecast" are examples of such verbal clues.	
no	Select "no" if there's either no uncertainty signalled at all (see column G), or if you can't find any verbal explanation for why the connecting line is signalling data uncertainty.	
J, datauncsta	**If it is explicitly verbally stated that data uncertainty is represented, how?**	
NULL	Write "NULL" if the previous question was answered with "no".	
free text	Write in the concrete formulation, for example, "95% confidence" or "Best-case scenario, optimistic scenario, current scenario".	
K, loweredrel	**Based on the former questions, does the line in focus signal modality in terms of lowered reliability?**	
yes	Select "yes" if the answer to the question in column G was "yes", and it's not verbally explained why the connection is visualized as uncertain (no clue to data uncertainty). Or select "yes" if the verbal text states that there is data uncertainty represented (answer in column I = "yes") but the visual appearance of the connecting line doesn't signal any uncertainty (answer in column G = no)	If the verbal and visual text differ in whether they indicate uncertainty or not (and therefore trigger "intermodal tension" (Engebretsen, 2012), this causes lowered reliability.
no	Select "no", if the connecting line does not signify any uncertainty at all (see column G) and the verbal text doesn't signal that data uncertainty is represented (see column I), or if it's explained verbally that data uncertainty is represented, the reason for the data uncertainty is made clear (see columns I and J) and the visual appearance of the connecting line indicates uncertainty (answer in column G = "yes").	
L, staafter	**Does the verbal sentence from column C need a reformulation, considering modality in terms of lowered probability and reliability, if any of them are expressed? If yes, which?**	
NULL	Type "NULL" if this is not necessary.	This question is only asked so that the previous answers are thought through again.
free text	Type a revised verbal statement, if the sentence in column C doesn't yet express all notions of lowered probability, data uncertainty or lowered reliability (if any of those are signified by the connecting line in focus). Use e.g. the adverbs referring to the single modality values, or the data uncertainty source explanations.	
M, notesline	**If you have any other comments that may be useful for a more qualitative investigation of this particular DV or the data analysis in the next step, or if anything was unclear, please write them down here.**	
free text	Write down your comments in keywords.	
	If you have no comments, leave this field empty.	

Questions and answer categories of the global level:

<table>
<tr><th>N, glstafirst</th><th colspan="2">What is the overall statement of this data visualization?</th></tr>
<tr><td>free text</td><td colspan="2"></td></tr>
<tr><td>O, glcodingorie</td><td colspan="2">Which coding orientation needs to be applied to the visual expression of the overall statement?</td></tr>
<tr><td>abstract</td><td>If you would say the more the whole visualization “represents the deeper ‘essence’” of the overall statement or “the more it represents the general pattern underlying superficially different specific instances, the higher its modality” (Van Leeuwen, 2005, p. 168), then the answer option “abstract” has to be taken. (Or in other words: if a more reduced articulation expresses the above-named meaning, then take this answer option.)</td><td rowspan="5">In contrast to the question in column E, here the whole data visualization is the object of attention. This means that the same or different coding orientations can be applied on the detail level (for the way the line in focus signals the detail statement as true) and on the global level (for the way the whole data visualization signals the overarching statement as true).

These answer categories are derived from the coding orientations described by Van Leeuwen (2005, pp. 168–171).

In case there is not clearly one or other category that fits, please choose the option that fits best in your opinion.</td></tr>
<tr><td>naturalistic</td><td>If you would say the more the overall statement is represented as an “accurate external representation” (Van Leeuwen, 2005, p. 168), the more it is considered as true, then the answer option “naturalistic” has to be taken.</td></tr>
<tr><td>technological</td><td>If you would say the more this DV “can be used as a blueprint or aid for action” (Van Leeuwen, 2005, p. 168), the more it is considered as true, then this is the answer option to be taken. Here as well, a “strongly decreased articulation” (Van Leeuwen, 2005, p. 168) may be used.</td></tr>
<tr><td>sensory</td><td>If you would say the more “the effect of pleasure or displeasure created” (Van Leeuwen, 2005, p. 170) by the DV, the more the overall statement is considered as true, then take this answer option.</td></tr>
<tr><td>none of these</td><td>If you think that none of these other orientations can be applied, choose this answer option.</td></tr>
<tr><td>P, glloweredrel</td><td colspan="2">Does the choice of visual style signal any sense of modality in terms of lowered reliability?</td></tr>
<tr><td>yes</td><td colspan="2">Select “yes” if the visual style represents the overall statement of the DV as not 100% reliable. This is the case if, depending on the coding orientation (column P), a less abstract/naturalistic/etc. way of visualizing the overall statement has been used.</td></tr>
<tr><td>no</td><td colspan="2">Select “no” if the visual style represents the overall statement of the DV as totally reliable.This is the case if, depending on the coding orientation (column P), a maximally abstract/naturalistic/etc. way of visualizing the overall statement has been used.</td></tr>
</table>

Q, glmodalityva	**If yes, which modality value is signalled?**	
NULL	Select “NULL” if the answer to the question in column P was “no”. This means that there is no modality (in terms of lowered reliability) signalled at all.	
high modality value	Select this answer category, if you would use the adverb “certainly” to describe the modality value in verbal language. This is the case when it’s fairly correct that in: • abstract orientation: reduced articulation is used, general pattern or deeper “essence” of what it depicts is represented • naturalistic orientation: image resembles the way we would see it if we saw it in reality • technological orientation: DV is practically useful, as blueprint or aid for action • sensory orientation: DV has an effect of pleasure or displeasure • none of these orientations: no theoretical basis is given here, please formulate your argumentation in the column Y.	These three answer categories as well as the adverbs used in verbal language are based on Halliday (1985/2004, p. 620). The descriptions of the answer categories are derived from the coding orientations described by Van Leeuwen (2005, pp. 168–171).
median modality value	Select this answer category if you would use the adverb “probably” to describe the modality value in verbal language. This is the case when the orientation explanations of the above field are fairly incorrect	
low modality value	Select this answer category if you would use the adverb “possibly” to describe the modality value in verbal language. This is the case when the orientation explanations of the above field are not correct.	
R, glhowsignal	**If yes, how does it signal a high, medium or low modality value?**	
NULL	Select “NULL” if the answer to the question in column P was “no”.	
free text	Write down what visual clues indicate the modality value you decided on in column Q, based on the descriptions of the coding orientations of Van Leeuwen (2005, pp. 168–171).	
S, gldatauncert	**Is it explicitly verbally stated on the detail or global level of the data visualization, or in the co-text, that data uncertainty is represented in the overall statement of the DV?**	
yes	Select “yes” if you can find a verbal clue on the detail or the global level of the DV, or in the co-text, stating that the underlying data for the overall statement was uncertain. Words like “scenario”, “confidence interval”, “forecast” are examples of such verbal clues. If the statements on the detail and the global level are closely interconnected (e.g., because there is only one connecting line making the overall meaning of the DV), it may be the case that the same verbal statement prompts “yes” as an answer in columns I and S. Then columns J and T also contain the same verbal statement.	
no	Select “no” if you can’t find any verbal statement that the underlying data for the overall statement of the DV was uncertain.	
T, gldatauncsta	**If it is explicitly verbally stated that data uncertainty is represented, how?**	
NULL	Write “NULL” if the answer to the previous question was “no”.	
free text	Write down the concrete formulation, for example, “95% confidence” or “Best-case scenario, optimistic scenario, current scenario”.	
U, glnotes	**If you have any other comments that may be useful for a more qualitative investigation of this particular DV or the data analysis in the next step, or if anything was unclear, please write that down here.**	
free text	Write down your comments in keywords.	
	If you have no comments, leave this field empty.	

4.2 The “visualappearance” tab

The question for the whole table is: What does the line (described in column C in the “semioticfunctionanalysis” tab) look like?

column, column title	long version of the column title	answer option	additional explanations if necessary
C, colour	colour	no variation (single-coloured)	
		abrupt variation	
		smooth transition	If the line has a transition from a colour to no colour (transparent) then this category has to be chosen too, in combination with “clarity: transparency” “low, medium or high”.
D, clarcrisp	clarity: crispness	high	These categories derive from MacEachren (1995, pp. 275–277) and are further described there. High transparency means that it is still slightly visible.
		lowered	
E, clartrans	clarity: transparency	none	
		low, medium or high	
F, clarresol	clarity: resolution	high	
		lowered	
G, pattern	pattern	continuous line	
		dashed/dotted line	By this I mean all lines with small, regularly recurring interruptions. The elements forming the line may be either dashes or dots in the narrow sense, or other elements of any form that are regularly placed one after another forming a line.
		irregularly dashed/dotted line	By this I mean all lines with small, irregularly recurring interruptions. The elements forming the line may be either dashes or dots in the narrow sense, or other elements of any form that are irregularly placed one after another forming a line.
		change between continuous line and dashed/dotted line	
		change between continuous line and large interruption(s)	
		change between continuous line, dashed/dotted line and large interruption(s)	
H, shapeforc	shape: forces	one force (straight)	The description of the line shape with the term *forces* derives from Kandinsky (1926/1928, p. 51).
		two forces, curved	
		two forces, bent	
		three or more forces, curved	
		three or more force, bent	
I, shapelipr	shape: line pressure	consistent	
		inconsistent	

J, shapeextr	shape: form of extremities	no direction-signifying extremity	This option has to be chosen if there's no change in shape of the extremity towards the end point.
		direction-signifying extremity	This may, for example, be an arrowhead, or a tapering body (decreasing line pressure).
K, interaction	interaction	not possible	Here the question is whether the visual appearance of the line changes through an action of the viewer (e.g. mouse-over or mouse-click). The interaction can take place with the line directly, or with other elements of the DV. If interaction is possible, but the visual appearance of the line does not change, "not possible" is the answer to take. (This means that choosing the answer "possible" here, has as a consequence that dynamics are available (next question).
		possible	
L, dynamics	dynamics	not available	If this answer is chosen, the following columns have to be answered with "NULL".
		available	If this answer is chosen, the following columns have to be answered with either "yes" or "no". The term *dynamics* means that what the line looks like changes. (If only other elements of the DV change, then this answer is not chosen.)
M, dynposition	dynamics in the visual variable position	NULL	
		yes	If the whole line (meaning the starting and end point) moves, this option has to be chosen.
		no	
N, dynorient	dynamics in the visual variable orientation	NULL	
		yes	If the perspective on the line changes, then this option has to be taken.
		no	
O, dynsize	dynamics in the visual variable size	NULL	
		yes	If the line length changes, this option has to be chosen.
		no	
P, dyncolour	dynamics in the visual variable colour	NULL	
		yes	
		no	
Q, dynclarcrisp	dynamics in the visual variable clarity: crispness	NULL	
		yes	
		no	
R, dynclartrans	dynamics in the visual variable clarity: transparency	NULL	
		yes	If the line gets more or less transparent (with stages of transparency where the line is still visible but elements behind it are visible as well), this answer has to be chosen. If the line appears or disappears completely (at once without growing, without any stages of transparency where the line is still visible but elements behind it are visible as well, e.g., because of the mouse position, clicking, scrolling or filtering), this answer is chosen as well.
		no	
S, cynclarresol	dynamics in the visual variable clarity: resolution	NULL	
		yes	
		no	

T, dynpattern	dynamics in the visual variable pattern	NULL	
		yes	If the visual appearance of single line elements (e.g., the length of the dashes of a dashed line) changes, then this answer has to be chosen.
		no	
U, dynshapeforc	dynamics in the visual variable shape: forces	NULL	
		yes	
		no	
V, dynshapelipr	dynamics in the visual variable shape: line pressure	NULL	
		yes	
		no	
W, dynshapeextr	dynamics in the visual variable shape: form of extremities	NULL	
		yes	
		no	

5.3 The "visualappearanceuncertainty" tab

This table is connected to column G in the "semioticfunctionanalysis" table. Here the question of whether the line represents any form of uncertainty is answered. In this table it is explained how, by choosing which visual variable (from the "visualappearance" table) was used to signify uncertainty. For all columns, the following three answer options are available:

NULL	Choose this answer, if there is no uncertainty signified at all. (This is the case if the answer to column G in the "semioticfunctionanalysis" table was "no".)
used	Choose this answer, if this visual variable is used to signify uncertainty.
not used	If this visual variable was not used to signify uncertainty, but one or more other visual variables were used for this purpose, then choose this answer.

References

Boeriis, M. (2012). Tekstzoom: Om en dynamisk funktionel rangstruktur i visuelle tekster. In T. H. Andersen & M. Boeriis (Eds.), *Nordisk socialsemiotik: Multimodale, pædagogiske og sprogvidenskabelige landvindinger* (pp. 131–153). University Press of Southern Denmark.

Engebretsen, M. (2012). Balancing cohesion and tension in multimodal rhetoric. An interdisciplinary approach to the study of semiotic complexity. *Learning, Media and Technology*, 37(2), 145–162. https://doi.org/10.1080/17439884.2012.655745

Halliday, M. A. K. (2004). *An introduction to functional grammar* (3rd ed. revised by Matthiessen, C.M.I.M.). Arnold (Hodder Education). (Original work published 1985)

Kandinsky, W. (1928). *Punkt und Linie zu Fläche: Beitrag zur Analyse der malerischen Elemente* (2nd unchanged ed.). Albert Langen. http://archive.org/details/punktun00kand (Original work published 1926)

Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge. (Original work published 1996)

MacEachren, A. M. (1995). *How maps work: Representation, visualization, and design.* Guilford Press.

Van Leeuwen, T. (2005). *Introducing social semiotics.* Routledge.

Appendix H: Article III: Coding Instructions of the Pre-Test

Coding instructions

for the inter-rater reliability pre-test of the corpus analysis
focusing on how narrativity is signalled by connecting lines

1. Introduction

Thank you for your participation in this pre-test. With your help, I will be able to find out to what extent our answers to the analysis questions are similar. The more consistent the answers between the raters of this pre-test are, the more reliable the method suggested to analyse the data visualization corpus can be considered (Bell, 2001, p. 21). If they are not consistent enough, the method (asked questions and answer categories) need to be rethought and adjusted.

The focus of the whole study is to find out how narrativity is signified in data visualizations, further called DVs. More precisely, special attention is paid to connecting lines which, in all the DVs selected, are the graphical elements conveying the main statement.

2. Definition of the units of analysis and the used documents

Within this pre-test, a randomly selected sample (25 items) of the entire corpus (containing 163 data visualizations) is analysed by two raters independently. In addition to this coding instruction document, you have been given the following documents:

(1) the Excel file “analysis_narrativity_pretest.xlsx”
(2) the Excel file “visualizationlistforanalysis.xlsx
(3) a folder with JPG, PNG, PDF and mp4 documents

Let me explain what these documents are used for:

(1) This is the document where the coding takes place. It has three tabs, each of which contains a table that needs to be filled out. The first two columns of all three tables have the same content: the number and title of the DVs that are analysed in this pre-test. This means that each row stands for one DV. The columns from C onwards stand for one question each.

(2) The Excel file "visualizationlistforanalysis.xlsx" helps you to find the right documents in the document folder (3.). Because the DVs that are being analysed stem from websites, and the content or the access possibilities can be changed, I have made screenshots, and screencasts of each DV and website. All the filenames of these screenshots and screencasts, as well as the original link to the website where the DV stems from, are listed in this Excel file. Since more than just one DV can stem from one website, the column "visimageFN" contains the filenames of JPGs that define which DV is being investigated and coded within this row. All the following columns ("winnerID", "imgFN", "filmFN", "pdfFN" and "url") can therefore have the same content in several rows.

(3) This folder contains the research material, the filenames of which are documented in the Excel file "visualizationlistforanalysis.xlsx".

3. First steps

(1) Open the Excel file "analysis_narrativity_pretest.xlsx" (1.), go to the first tab.

(2) Look at which visualization is the first to analyse. (The columns "visID" and "vistitle" tell you this.) Let's say this would be visID "16", vistitle "Flight tracks".

(3) Open the Excel file "visualizationlistforanalysis.xlsx" and go to the row with the same visID and vistitle. Mark it, for example, by using a filling colour for this particular row.

(4) Open Windows Explorer/Finder and navigate to the folder with all the research material (3.).

(5) Now search for the filenames written in the marked row (step 3.) and open these documents. In our example this would be "Spies_in_the_skies.JPG", "spies_in_the_skies.png", "Spies_in_the_skies.mp4" and

"Spies_in_the_skies.pdf". Maybe you would also like to open the website these screenshots and screencasts have been taken from (in this case https://www.buzzfeed.com/peteraldhous/spies-in-the-skies?utm_term=.rlwl2JVE7#.hhzwqNE1O). But note that changes may have been made since the screenshots and screencasts were taken. That's why these documents, and not the online website, remain the main point of reference.

(6) Now go back to the Excel file "analysis_narrativity_pretest.xlsx" (1.) and answer the first question to the DV investigated, by looking at the files opened in step 5.

(7) Go through all questions in all three tabs of the "analysis_pretest.xlsx" file (for this DV) and answer them. (The table in the second tab is already filled out, you may only want to check the answers randomly. In case the line indicates narrativity, you will need to make use of the answers in the second tab.)

(8) Demark the row marked in step 3.

(9) Then repeat these steps for each DV by going back to step 2, finding the next DV and so on.

4. Descriptions of the questions and answer categories

In this section, the questions and answer categories available within the Excel file "analysis_narrativity_pretest.xlsx" are explained. This is to make sure that the questions are understood by all raters, and that the answer categories are made clear. In the "analysis_narrativity_pretest.xlsx" file, even though the titles of each column contain only an abbreviation, the full questions asked are noted in a comment attached to each column title field.

4.1 The "semioticfunctionanalysis" tab

The following tables show the column character, column title, the complete question, answer options and descriptions of the answer options for each column. This can be used as a reference when filling out the tables in the Excel file "analysis_narrativity_pretest.xlsx". This is the way the following table is built up:

Column, column title	**Question**	
Answer option 1	Description of the answer option 1	General explanation to the answer options, if necessary.
Answer option 2	Description of the answer option 2	

C, stafirst	**Only focus on the connecting line (and its endpoints if applicable) that conveys the main statement of the DV. If there are more of the same kind, decide for one exemplary unit, comparable to a clause. How could this clause be formulated?**	
This answer is already given.	Since this question is necessary to define which graphical element is in focus, and this needs to be the same for all raters, this answer is already given.	
D, lineappgener	**What does the focused connecting line look like?**	
This needs to be filled out in the separate table in tab "visualappearance".		
E, narrconc	**Does the visual appearance of the focused connecting line signal a narrative or a conceptual representational structure?**	
narrative representational structure	If the line is a vector (a line that visually indicates directionality), and therefore serves "to present unfolding actions and events, processes of change, transitory spatial arrangements" (Kress & Van Leeuwen, 1996/2006, p. 59), then take this answer option.	These answer categories are derived from Kress and Van Leeuwen (1996/2006, pp. 45–113).
conceptual representational structure	If the line is not a vector and therefore (together with its endpoints if visible) represents the "participants in terms of their class, structure or meaning, in other words, in terms of their generalized and more or less timeless essence" (Kress & Van Leeuwen, 1996/2006, p. 59), this answer option has to be taken.	
ambiguous	If it is unclear whether the line signals a narrative or a conceptual structure, this answer option has to be taken.	
F, actconvers	**If it signals a narrative, does it signal an action process or a conversion process?**	
action process	If there is an Actor (whether visible or not) that is doing something (to somebody or something), then this answer option has to be taken.	These answer categories are derived from Kress and Van Leeuwen (1996/2006, pp. 63–70)
conversion process	If what Kress and Van Leeuwen call a relay is present, then take this answer option. "Relays do not just pass on, in unchanged form, what they receive; they always also transform it." (1996/2006, p. 68)	
NULL	Select "NULL" if the answer to the question in column E was "conceptual representational structure" or "ambiguous".	

G, transnonev	If it signals an action process, does it signal a transactional or a non-transactional one, or an event?	
transactional process	If the action has two participants (an Actor and a Goal), it is a transactional process. In other words: if you would rather use a "transitive verb, a verb that takes an object (e.g. 'transport' or 'send' …)" (1996/2006, p. 65), to describe the action process in verbal language, then take this answer option.	These answer categories are derived from Kress and Van Leeuwen (1996/2006, pp. 63–66).
non-transactional process	If the action "has no 'Goal', is not 'done to' or 'aimed at' anyone or anything" (1996/2006, p. 63), this answer option has to be taken. In other words: if you would use an intransitive verb (a verb that doesn't need an object) to describe the action process in verbal language, take this answer option.	
event	If the visual representation of the action lacks an Actor (but only includes a Goal - "a visual analogue … of 'passive agent deletion'" (1996/2006, p. 64)), then take this answer option.	
NULL	Select "NULL" if the answer to the question in column E was "conceptual representational structure" or "ambiguous" or if the answer to the question in column F was "conversion process".	
H, hownarr	**If the visual appearance of the connecting line signals a narrative representational structure, how?**	
This needs to be filled out in the separate table in tab "visualappearancenarrativity".		
I, staafternarr	**Try to reformulate the verbal sentence from column C, so that it best represents the type of representational structure you decided for in columns E to G.**	
free text	The goal is not to find an exact verbal equivalent to the visual sign, but to try to reason the answers in column E to G.	

4.2 The "visualappearance" tab

This table has already been filled out in advance, because it is only about what the lines look like, not what they mean. That's why the answers to this question are not considered to be variables that rely "on the analyst's judgement and interpretation" (Brezina, 2018, p. 269). However, a list of what categories are available, and some additional explanations of what they mean are given in the following table. The question for the whole table is: What does the focused connecting line (described in column D in the "semioticfunctionanalysis" table) look like?

column, column title	**long version of the column title**	**answer option**	**additional explanations if necessary**
C, colour	colour	no variation (single-coloured)	
		abrupt variation	
		smooth transition	If the line has a transition from a colour to no colour (transparent) then this category has to be chosen too, in combination with "clarity: transparency" "low, medium or high".
D, clarcrisp	clarity: crispness	high	These categories derive from MacEachren (1995, pp. 275–277) and are further described there. High transparency means that it is still slightly visible.
		lowered	
E, clartrans	clarity: transparency	none	
		low, medium or high	
F, clarresol	clarity: resolution	high	
		lowered	
G, pattern	pattern	continuous line	
		dashed/dotted line	By this I mean all lines with small, regularly recurring interruptions. The elements forming the line may be either dashes or dots in the narrow sense, or other elements of any form that are regularly placed one after another forming a line.
		irregularly dashed/dotted line	By this I mean all lines with small, irregularly recurring interruptions. The elements forming the line may be either dashes or dots in the narrow sense, or other elements of any form that are irregularly placed one after another forming a line.
		change between continuous line and dashed/dotted line	
		change between continuous line and large interruption(s)	
		change between continuous line, dashed/dotted line and large interruption(s)	
H, shapeforc	shape: forces	one force (straight)	The description of the line shape with the term *forces* derives from Kandinsky (1926/1928, p. 51).
		two forces, curved	
		two forces, bent	
		three or more forces, curved	
		three or more force, bent	
I, shapelipn	shape: line pressure	consistent	
		inconsistent, several variations	
		inconsistent, tapering towards one end	Here the line pressure is decreasing towards one end, but apart from that the line has no other line pressure changes.

J, shapearr	shape: arrowheads or similar	no arrowhead(s) or similar	
		arrowhead or similar at one end	A similar element to an arrowhead could, for example, be an icon of a plane, where the moving direction is clear.
		arrowheads or similar at both ends	This option has to be chosen if an arrowhead or similar is placed on both ends of the line (and not on other places along the line).
		arrowhead(s) or similar along the line	This option has to be chosen if one or more arrowhead(s) or similar is placed along the line. If arrowheads or similar occur both at one end and along the line, this option has to be chosen too.
		arrow(head)(s) or similar besides the line	This option has to be chosen if one or more arrow(head)(s) or similar is placed parallel to the line.
K, interaction	interaction	not possible	Here the question is whether the visual appearance of the line changes through an action of the viewer (e.g., mouse-over or mouse-click). The interaction can take place with the line directly, or with other elements of the DV. If interaction is possible, but the visual appearance of the line does not change, “not possible” is the answer to take. (This means that choosing the answer “possible” here has the consequence that dynamics are available (next question).
		possible	
L, dynamics	dynamics	not available	If this answer is chosen, the following columns have to be answered with “NULL”.
		available	If this answer is chosen, the following columns have to be answered with either “yes” or “no”. The term *dynamics* means that what the line looks like changes. (If only other elements of the DV change, then this answer is not chosen.)
M, dynposition	dynamics in the visual variable position	NULL	
		yes	If the whole line (meaning the starting and end point) moves, this option has to be chosen.
		no	
N, dynorient	dynamics in the visual variable orientation	NULL	
		yes	If the perspective on the line changes, then this option has to be taken.
		no	
O, dynsize	dynamics in the visual variable size	NULL	
		yes	If the line length changes, this option has to be chosen.
		no	
P, dyncolour	dynamics in the visual variable colour	NULL	
		yes	
		no	
Q, dynclarcrisp	dynamics in the visual variable clarity: crispness	NULL	
		yes	
		no	

R, dynclartrans	dynamics in the visual variable clarity: transparency	NULL	
		yes	If the line gets more or less transparent (with stages of transparency where the line is still visible but elements behind it are visible as well), this answer has to be chosen. If the line appears or disappears completely (at once without growing, without any stages of transparency where the line is still visible but elements behind it are visible as well, e.g., because of the mouse position, clicking, scrolling or filtering), this answer is chosen as well.
		no	
S, dynclarresol	dynamics in the visual variable clarity: resolution	NULL	
		yes	
		no	
T, dynpattern	dynamics in the visual variable pattern	NULL	
		yes	If the visual appearance of single line elements (e.g., the length or position of the dashes of a dashed line) changes, then this answer has to be chosen.
		no	
U, dynshapeforc	dynamics in the visual variable shape: forces	NULL	
		yes	
		no	
V, dynshapelipn	dynamics in the visual variable shape: line pressure	NULL	
		yes	
		no	
W, dynshapearr	dynamics in the visual variable shape: arrowheads or similar	NULL	
		yes	
		no	

5.3 The "visualappearanceuncertainty" tab

This table has to be filled out by you. It is connected to the columns E to G in the "semioticfunctionanalysis" table. There you answered whether the line signals any kind of narrative structure. In this table you will explain how, by choosing which visual variable(s) (from the "visualappearance" table) was/were used to signal narrativity.

For all columns, the following three answer options are available:

NULL	Choose this answer if the focused line does not signal a narrative representational structure. (This is the case if the answer to column E in the "semioticfunctionanalysis" table was "conceptual representational structure".)
used	Choose this answer if this visual variable is used to signal narrativity.
not used	If this visual variable was not used to signal narrativity, but one or more other visual variables were used for this purpose, then choose this answer.

5. General instructions for using the Excel files

- Only the Excel file "analysis_narrativity_pretest.xlsx" will be used for calculating the inter-rater reliability. This means that no changes in the Excel file "visualizationlistforanalysis.xlsx" (like colour marking) will have any consequences.
- In the Excel file "analysis_narrativity_pretest.xlsx", for most of the questions there are answer options offered. You can see them if you click first into the empty field, and then on the arrow appearing on the right side of the field. If there are such answer options available, only use these ones! Don't write any free text into such fields or add any letters or spaces after one answer option has been selected.
- Sometimes it may not be the case that it's clearly only one answer that could fit. In such a case, please choose the option that fits best in your opinion.

References

Bell, P. (2001). Content analysis of visual images. In T. Van Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 10–34). SAGE.

Brezina, V. (2018). Statistical choices in corpus-based discourse analysis. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 259–280). Routledge.

Kandinsky, W. (1928). *Punkt und Linie zu Fläche: Beitrag zur Analyse der malerischen Elemente* (2nd unchanged ed.). Albert Langen. http://archive.org/details/punktun00kand (Original work published 1926)

Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge. (Original work published 1996)

MacEachren, A. M. (1995). *How maps work: Representation, visualization, and design.* Guilford Press.

Appendix I: Article III: Calculation Results of Benchmarking the Inter-Rater-Reliability Study Results

Table 1

Calculation Results of Question Column E of Table 10 in the Main Part

Interval 1: almost perfect		**Interval 2: substantial**		**Interval 3: moderate**		**Interval 4: fair**		**Interval 5: slight**		**Interval 6: poor**	
COEFF	0.9468	COEFF	0.9468	COEFF	0.9468	COEFF	0.9468	COEFF	0.9468	COEFF	0.9468
a	0.8	a	0.6	a	0.4	a	0.2	a	0	a	-1
b	1	b	0.8	b	0.6	b	0.4	b	0.2	b	0
SE	0.0522	SE	0.0522	SE	0.0522	SE	0.0522	SE	0.0522	SE	0.0522
IMPa	0.9976	IMPa	1	IMPa	1	IMPa	1	IMPa	1	IMPa	1
IMPb	0.1538	IMPb	0.9976	IMPb	1	IMPb	1	IMPb	1	IMPb	1
IMP	0.8438	IMP	0.0024	IMP	0.0	IMP	0	IMP	0	IMP	0
Max IMP	0.8462	Max IMP	0.8462	Max IMP	0.8462	Max IMP	0.8462	Max IMP	0.8462	Max IMP	0.8462
IMP-Revised	0.9971	IMP-Revised	0.0029	IMP-Revised	0.0	IMP-Revised	0	IMP-Revised	0	IMP-Revised	0
cumulative benchmark probability:	**0.9971**	cumulative benchmark probability:	1	cumulative benchmark probability:	1	cumulative benchmark probability:	1	cumulative benchmark probability:	1	cumulative benchmark probability:	1

The value in the grey field defines the final agreement level because this is the "highest benchmark level associated with the smallest cumulative probability that exceeds 0.95" (Gwet, 2014, p. 178)

Note.
COEFF = result of the calculated coefficient (Gwet's AC_1 or AC_2);
a = lower bound of this interval;
b = upper bound of this interval;
SE = Standard Error;
IMP = Interval Membership Probability;
Max IMP = Maximum Interval Membership Probability.
These definitions also apply to the same abbreviations in all other tables of Appendix I.

References

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters.* (4th ed.). Advanced Analytics, LLC.

Table 2

Calculation Results of Question Column F of Table 10 in the Main Part

	Interval 1: almost perfect	Interval 2: substantial	Interval 3: moderate	Interval 4: fair	Interval 5: slight	Interval 6: poor
COEFF	0.8900	0.8900	0.8900	0.8900	0.8900	0.8900
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.0756	0.0756	0.0756	0.0756	0.0756	0.0756
IMPa	0.8833	0.9999	1	1	1	1
IMPb	0.0729	0.8833	0.9999	1	1	1
IMP	0.8104	0.1167	0.0001	0.0	0	0
Max IMP	0.9271	0.9271	0.9271	0.9271	0.9271	0.9271
IMP-Revised	0.8741	0.1259	0.0001	0.0	0	0
cumulative benchmark probability:	0.8741	**0.9999**	1	1	1	1

Table 3

Calculation Results of Question Column G of Table 10 in the Main Part

	Interval 1: almost perfect	Interval 2: substantial	Interval 3: moderate	Interval 4: fair	Interval 5: slight	Interval 6: poor
COEFF	0.6524	0.6524	0.6524	0.6524	0.6524	0.6524
a	0.8	0.6	0.4	0.2	0	-1
b	1	0.8	0.6	0.4	0.2	0
SE	0.1194	0.1194	0.1194	0.1194	0.1194	0.1194
IMPa	0.1083	0.6697	0.9827	0.9999	0.1000	1
IMPb	0.0018	0.1083	0.6697	0.9827	0.9999	0.1000
IMP	0.1065	0.5613	0.3131	0.0172	0.0001	0.0000
Max IMP	0.9982	0.9982	0.9982	0.9982	0.9982	0.9982
IMP-Revised	0.1067	0.5624	0.3136	0.0172	0.0001	0.0000
cumulative benchmark probability:	0.1067	0.6691	**0.9827**	0.9999	0.1000	1

Appendix J: Article III: Final Coding Instructions

Coding scheme

for the single-coded corpus analysis
focusing on how narrativity is signalled by connecting lines

1. Definition of the units of analysis and the used documents

Within this analysis, the entire corpus is analysed by the author of this dissertation. The following documents are used:

(1) the Excel file “analysis_narrativity.xlsx”
(2) the Excel file “visualizationlistforanalysis.xlsx
(3) a folder with JPG, PNG, PDF and mp4 documents as described in the relational database.

Let me explain what these documents are used for:

(1) This is the document where the coding takes place. It has three tabs, each of which contains a table that needs to be filled out. The first two columns of all three tables have the same content: the number and title of the DVs that are analysed in this pre-test. This means that each row stands for one DV. The columns from C onwards stand for one question each.
(2) The Excel file „visualizationlistforanalysis.xlsx“ helps to find the right documents in the document folder (3.) and was directly exported from the relational database. It contains all the filenames of these screenshots and screencasts, as well as the original link to the website that the DV stems from. Since more than just one DV can stem from one

website, the column "visimageFN" contains the filenames of JPGs that define which DV is being investigated and coded within this row. All the following columns ("winnerID", "imgFN", "filmFN", "pd-fFN" and "url") can therefore have the same content in several rows.

(3) This folder contains the research material, the filenames of which are documented in the Excel file "visualizationlistforanalysis.xlsx".

3. First steps

(1) Open the Excel file "analysis_narrativity.xlsx" (1.), go to the first tab.

(2) Look at which visualization is the first to analyse. (The columns "visID" and "vistitle" tell you this.) Let's say this would be visID "16", vistitle "Flight tracks".

(3) Open the Excel file "visualizationlistforanalysis.xlsx" and go to the row with the same visID and vistitle. Mark it, for example, by using a filling colour for this particular row.

(4) Open Windows Explorer/Finder and navigate to the folder with all the research material (3.).

(5) Search for the filenames written in the marked row (step 3.) and open these documents. In our example this would be "Spies_in_the_skies.JPG", "spies_in_the_skies.png", "Spies_in_the_skies.mp4" and "Spies_in_the_skies.pdf". Maybe also opening the website these screenshots and screencasts have been taken from can be helpful (in this case https://www.buzzfeed.com/peteraldhous/spies-in-the-skies?utm_term=.rlwl2JVE7#.hhzwqNE1O). But it has to be noted that changes may have been made since the screenshots and screencasts were taken. That's why these documents, and not the online website, remain the main point of reference.

(6) Go back to the Excel file "analysis_narrativity_pretest.xlsx" (1.) and answer the first question to the DV investigated, by looking at the files opened in step 5.

(7) Go through all questions in all three tabs of the "analysis_pretest.xlsx" file (for this DV) and answer them. (The table in the second tab is already filled out, you may only want to check the answers randomly. In case the line indicates narrativity, you will need to make use of the answers in the second tab.)

(8) Demark the row marked in step 3.

(9) Then repeat these steps for each DV by going back to step 2, finding the next DV and so on.

4. Descriptions of the questions and answer categories

In this section, the questions and answer categories available within the Excel file "analysis_narrativity.xlsx" are explained.

4.1 The "semioticfunctionanalysis" tab

The following tables show the column character, column title, the complete question, answer options and descriptions of the answer options for each column. This can be used as a reference when filling out the tables in the Excel file "analysis_narrativity.xlsx". This is the way the following table is built up:

Column, column title	**Question**	
Answer option 1	Description of the answer option 1	General explanation to the answer options, if necessary.
Answer option 2	Description of the answer option 2	

C, stafirst	**Only focus on the connecting line (and its endpoints if applicable) that conveys the main statement of the DV. If there are more of the same kind, decide for one exemplary unit, comparable to a clause. How could this clause be formulated?**	
This answer is already given.	Since this question is necessary to define which graphical element is in focus, and this needs to be the same for all raters, this answer is already given.	
D, lineappgener	**What does the focused connecting line look like?**	
This needs to be filled out in the separate table in tab "visualappearance".		
E, narrconc	**Does the visual appearance of the focused connecting line signal a narrative or a conceptual representational structure?**	
narrative representational structure	If the line is a vector (a line that visually indicates directionality), and therefore serves "to present unfolding actions and events, processes of change, transitory spatial arrangements" (Kress & Van Leeuwen, 1996/2006, p. 59), then take this answer option.	These answer categories are derived from Kress and Van Leeuwen (1996/2006, pp. 45–113).
conceptual representational structure	If the line is not a vector and therefore (together with its endpoints if visible) represents the "participants in terms of their class, structure or meaning, in other words, in terms of their generalized and more or less timeless essence" (Kress & Van Leeuwen, 1996/2006, p. 59), this answer option has to be taken.	
ambiguous	If it is unclear whether the line signals a narrative or a conceptual structure, this answer option has to be taken.	

F, hownarr	**If the visual appearance of the connecting line signals a narrative representational structure, how does it do so?**
This needs to be filled out in the separate table in tab “visualappearancenarrativity”.	
G, staafternarr	**Try to reformulate the verbal sentence from column C, so that it best represents the type of representational structure you decided on in columns E to G.**
free text	The goal is not to find an exact verbal equivalent for the visual sign, but to try to reason the answers in column E.

4.2 The “visualappearance” tab

The question for the whole table is: What does the focused connecting line (described in column D in the “semioticfunctionanalysis” table) look like?

column, column title	long version of the column title	answer option	additional explanations if necessary
C, colour	colour	no variation (single-coloured)	
		abrupt variation	
		smooth transition	If the line has a transition from a colour to no colour (transparent) then this category has to be chosen too, in combination with “clarity: transparency” “low, medium or high”.
D, clarcrisp	clarity: crispness	high	These categories derive from MacEachren (1995, pp. 275–277) and are further described there. High transparency means that it is still slightly visible.
		lowered	
E, clartrans	clarity: transparency	none	
		low, medium or high	
F, clarresol	clarity: resolution	high	
		lowered	
G, pattern	pattern	continuous line	
		dashed/dotted line	By this I mean all lines with small, regularly recurring interruptions. The elements forming the line may be either dashes or dots in the narrow sense, or other elements of any form that are regularly placed one after another forming a line.
		irregularly dashed/dotted line	By this I mean all lines with small, irregularly recurring interruptions. The elements forming the line may be either dashes or dots in the narrow sense, or other elements of any form that are irregularly placed one after another forming a line.
		change between continuous line and dashed/dotted line	
		change between continuous line and large interruption(s)	
		change between continuous line, dashed/dotted line and large interruption(s)	

H, shapeforc	shape: forces	one force (straight)	The description of the line shape with the term *forces* derives from Kandinsky (1926/1928, p. 51).
		two forces, curved	
		two forces, bent	
		three or more forces, curved	
		three or more force, bent	
I, shapelipn	shape: line pressure	consistent	
		inconsistent, several variations	
		inconsistent, tapering towards one end	Here the line pressure is decreasing towards one end, but apart from that the line has no other line pressure changes.
J, shapearr	shape: arrowheads or similar	no arrowhead(s) or similar	
		arrowhead or similar at one end	A similar element to an arrowhead could, for example, be an icon of a plane, where the moving direction is clear. If from our experiential knowledge it is clear to which direction the element moves, then it is rated as an element similar to an arrowhead.
		arrowheads or similar at both ends	This option has to be chosen if an arrowhead or similar is placed on both ends of the line (and not on other places along the line).
		arrowhead(s) or similar along the line	This option has to be chosen if one or more arrowhead(s) or similar is placed along the line. If arrowheads or similar occur both at one end and along the line, this option has to be chosen too.
		arrow(head)(s) or similar besides the line	This option has to be chosen if one or more arrow(head)(s) or similar is placed parallel to the line.
K, interaction	interaction	not possible	Here the question is, whether the visual appearance of the line changes through an action of the viewer (e.g., mouse-over or mouse-click). The interaction can take place with the line directly, or with other elements of the DV. If interaction is possible, but the visual appearance of the line does not change, "not possible" is the answer to take. This means that choosing the answer "possible" here, has the consequence that dynamics are available (next question).
		possible	
L, dynamics	dynamics	not available	If this answer is chosen, the following columns have to be answered with "NULL".
		available	If this answer is chosen, the following columns have to be answered with either "yes" or "no". The term *dynamics* means that what the line looks like changes. (If only other elements of the DV change, then this answer is not chosen.)
M, dynposition	dynamics in the visual variable position	NULL	
		yes	If the whole line (meaning the starting and end point) moves, this option has to be chosen.
		no	
N, dynorient	dynamics in the visual variable orientation	NULL	
		yes	If the perspective on the line changes, then this option has to be taken.
		no	

O, dynsize	dynamics in the visual variable size	NULL	
		yes	If the line length changes, this option has to be chosen.
		no	
P, dyncolour	dynamics in the visual variable colour	NULL	
		yes	
		no	
Q, dynclarcrisp	dynamics in the visual variable clarity: crispness	NULL	
		yes	
		no	
R, dynclartrans	dynamics in the visual variable clarity: transparency	NULL	
		yes	If the line gets more or less transparent (with stages of transparency where the line is still visible but elements behind it are visible as well), this answer has to be chosen. If the line appears or disappears completely (at once without growing, without any stages of transparency where the line is still visible but elements behind it are visible as well) (e.g. because of the mouse position, clicking, scrolling or filtering), this answer is chosen as well.
		no	
S, dynclarresol	dynamics in the visual variable clarity: resolution	NULL	
		yes	
		no	
T, dynpattern	dynamics in the visual variable pattern	NULL	
		yes	If the visual appearance of single line elements (e.g., the length or position of the dashes of a dashed line) changes, then this answer has to be chosen.
		no	
U, dynshapeforc	dynamics in the visual variable shape: forces	NULL	
		yes	
		no	
V, dynshapelipn	dynamics in the visual variable shape: line pressure	NULL	
		yes	
		no	
W, dynshapearr	dynamics in the visual variable shape: arrowheads or similar	NULL	
		yes	
		no	
Y, geninteract	general interaction apart from with the focused line	yes	If any interaction is possible with any element of this DV apart from the focused line, then this answer has to be chosen. Only starting an animation through interaction (like scrolling or clicking a start button) doesn't count as an interaction here.
		no	
X, gendyn	general dynamics apart from the focused line	yes	If anything in this DV apart from the focused line is changing its state (either automatically through animation or through interaction), then this answer has to be chosen.
		no	

5.3 The “visualappearanceuncertainty” tab

This table is connected to column E in the “semioticfunctionanalysis” table. There you have answered the question of whether the line signals a narrative representational structure or not. In this table you will explain how, by choosing which visual variable(s) (from the “visualappearance” table) was/were used to signal narrativity.

For all columns, the following three answer options are available:

NULL	Choose this answer, if the focused line does not signal a narrative representational structure. (This is the case if the answer to column E in the “semioticfunctionanalysis” table was not “narrative representational structure”.)
used	Choose this answer, if this visual variable is used to signal narrativity.
not used	If this visual variable was not used to signal narrativity, but one or more other visual variables were used for this purpose, then choose this answer.

References

Kandinsky, W. (1928). *Punkt und Linie zu Fläche: Beitrag zur Analyse der malerischen Elemente* (2nd unchanged ed.). Albert Langen. http://archive.org/details/punktun00kand (Original work published 1926)

Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge. (Original work published 1996)

MacEachren, A. M. (1995). *How maps work: Representation, visualization, and design.* Guilford Press.