



Universiteit
Leiden
The Netherlands

Affordances and limitations of algorithmic criticism

Verhaar, P.A.F.

Citation

Verhaar, P. A. F. (2016, September 27). *Affordances and limitations of algorithmic criticism*. Retrieved from <https://hdl.handle.net/1887/43241>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/43241>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/43241> holds various files of this Leiden University dissertation

Author: Verhaar, Peter

Title: Affordances and limitations of algorithmic criticism

Issue Date: 2016-09-27

Chapter 9

Visualisation

9.1. Introduction

One of the most striking consequences of the advent of digital methodologies is the marked upsurge in the use of visualisations, and, simultaneously, an attenuation of the centrality of text. Conventional humanities research is dominated profoundly by the written word. Especially in fields such as literary studies and philology, text is frequently both the object and the outcome of academic research. Unlike research projects in the natural sciences, which tend to rely heavily on quantitative data, works of literary criticism have rarely employed visual displays such as graphs and charts for the organisation and the dissemination of knowledge. The growing use of visualisation is spurred both by the nature of the digital medium, and by the type of research that the digital medium makes possible. Firstly, while it is obviously possible to transfer images via paper-based books or journals, the ease with which images can be created, edited and disseminated in a digital environment is unequalled. In printed publications, texts and images are the only modalities which can be conveyed, but this limitation no longer exists in the digital space. The computer is in fact “a continuum in which many signs of representation can happily co-exist”.⁵⁴⁶ On the computer, it is often unclear “where the pictorial space ends and the verbal space begins”.⁵⁴⁷ Secondly, as was discussed in earlier chapters, the essential plasticity and computability of digital text has also inspired innovative forms of analysis, and the growing interest among literary scholars in quantification and in statistics has also urged scholars to explore whether or not the salient features of numerical data sets can be clarified and communicated effectively through images.

In most cases, the outcomes of quantitative analyses can be examined most effectively if these are presented visually. Statistical manipulations of quantitative data initially result in new numbers, but bare numbers displayed in a tabular form often fail to facilitate an efficient exploration of trends or of notable exceptions to these trends. An adroit use of visualisation techniques can help scholars to promptly discern relevant characteristics of voluminous data collections. The

⁵⁴⁶ Jay Bolter, *Writing Space: The Computer, Hypertext, and the History of Writing* (Hillsdale N.J.: L. Erlbaum Associates 1991), p. 60.

⁵⁴⁷ *Ibid.*, p. 74.

transfer of ideas via graphical displays derives much of its attraction from the fact that it is often perceived as a more immediate form of communication. Dehaene explains that viewing objects requires considerably less neural activity than reading.⁵⁴⁸ Unlike the capacity for vision, the human brain does not have an innate wiring for processing written texts, and the ability to read demands numerous new connections between distinct areas in the brain. Stan Ruecker stresses similarly that human beings have a great capacity for perceiving and processing visual information, as the human brain has “ecological advantages” that enable it to process “environmental features such as luminosity, motion and colour” with great speed.⁵⁴⁹ Merely from glancing at a diagram or a graph, human viewers can identify notable features almost immediately, and, in many cases, the focus is directed intuitively to outliers, colour contrasts, or other irregularities in the overall shape, if present. This visual orientation and the innate aptitudes for perceiving patterns and regularities can be leveraged expediently in the creation and the communication of knowledge.

Since studies in the field of literary criticism generally aim to illuminate the meaning of texts, Sinclair et al. argue that the value of scholarly tools for the visualisation of literary texts can be gauged “by determining how well it supports this interpretative activity”.⁵⁵⁰ This chapter concentrates more closely on the ways in which visualisation techniques can contribute to hermeneutic processes. This chapter initially offers a definition of data visualisation. Next, a classification is proposed of the various ways in which data about texts can be clarified visually. Finally, an assessment is made of the ways in which the patterns that emerge within visualisation can support or obstruct interpretation.

9.2. Definition

Providing a clear definition of the term ‘visualisation’ is complicated by the fact the word may refer to a wide range of visual phenomena. Following Manovich, a distinction can be made between visualisations which represent objects or phenomena directly and visualisations which primarily represent data about these objects or phenomena. In the first case, graphic displays are essentially renditions of the “a priori fixed spatial layout of the real physical objects such as a brain, a coastline, a galaxy”.⁵⁵¹ These types of visualisations, which Manovich refers to as ‘scientific

⁵⁴⁸ Stanislas Dehaene, *Reading in the Brain: The Science and Evolution of a Human Invention* (New York: Viking 2009).

⁵⁴⁹ Stan Ruecker, “Rich Prospect Browsing Interfaces”, in: *Affordances of Prospect for Academic Users of Interpretively Tagged Text Collections*, University of Alberta 2003, p. 5.

⁵⁵⁰ Stéfan Sinclair, Stan Ruecker & Milena Radzikowska, “Information Visualization for Humanities Scholars”, in: *Literary Studies in the Digital Age*, Modern Language Association of America 2013, n. pag.

⁵⁵¹ Lev Manovich, “What Is Visualization?”, in: *Visual Studies*, (2011), n.pag.

visualisations', are based on a graphical structure which has been observed in a particular physical environment. The existing shapes, colours and dimensions are reproduced in an abstracted form on a certain medium, often for the purpose of further analysis. Visualisations that mimic physical objects or phenomena can be distinguished from 'data visualisations' or 'information visualisations',⁵⁵² which generally involve "a mapping between discrete data and a visual representation".⁵⁵³ Maureen Stone views data visualisation similarly as "the field of study that uses interactive graphical tools to explore and present digitally represented data, that may be stimulated, measured, or archived".⁵⁵⁴ Information visualisations take a shapeless collection of data as input, and attempt to illuminate characteristics of this data set through a specific organisation or arrangement of graphic elements.

Using the adjective "visual" may be misleading, as written texts, like graphs and diagrams also consist of signs which are to be perceived through vision. The term 'data visualisation' can be defined more precisely as an operation through which specific properties of a data set are represented using the graphical modality. 'Modality', in short, refers to the manner in which the information that is conveyed through a medium is encoded. It consists of a distinct class of signs which senders and recipients can use to communicate. Modalities may be classified using the human senses that are needed to decode the information. Written text and images are both visual modalities, while music and braille are aural and tactile modalities, respectively. An important difference between texts and images is that the former uses arbitrary signs that form a linguistic code that must be learned, while images are available to anyone without formal instruction. A thorough discussion of the graphic modality is offered in Jacques Bertin's influential monograph *Semiology of Graphics*, which describes seven graphic variables, namely shape, scale, tonal value, texture, colour, orientation and location.⁵⁵⁵ The graphic modality communicates knowledge through dexterous combinations of these basic variables. Leland Wilkinson, in *The Grammar of Graphics*, argues in a similar fashion that complex graphics can be broken down into a limited number of basic graphic primitives. He explains that graphics can be created by mapping a data set to a perceivable physical representation, and that such representations consist of "aesthetic attributes". Building on the categorisation that was developed by Bertin, Wilkinson enumerates a large number of aesthetic attributes, including include position, size,

⁵⁵² The two terms will be treated synonymously.

⁵⁵³ Lev Manovich, "What Is Visualization?", p. 2.

⁵⁵⁴ Maureen Stone, "Information Visualization: Challenge for the Humanities", in: *Working Together or Apart : Promoting the next Generation of Digital Scholarship : Report of a Workshop Cosponsored by the Council on Library and Information Resources and the National Endowment for the Humanities*, Washington D.C.: 2009, p. 44.

⁵⁵⁵ Jacques Bertin, *Semiology of Graphics* (Madison: University of Wisconsin Press 1983).

shape, rotation, colour, saturation, orientation and blur. Wilkinson's attributes are grouped into five categories: "form, surface, motion, sound, and text".⁵⁵⁶

Data visualisation often involves a change in modality. The original objects which are studied generally employ a set of semiotic signs that differ from the signs that occur in the source that is ultimately represented by the visualisation. In the context of literary informatics, visualisation involves a process in which a message in the textual modality is represented and clarified through information encoded in the graphical modality. In previous chapters, it was explained that computer-based text analysis involves the application of a range of techniques which can convert a corpus of linear and discursive texts into discrete and structured data, and which can further analyse the resultant data set in a variety of ways. Visualisation is a specific type of processing in which the results of such analyses are subsequently represented via graphical primitives. Patterns created by variations in the use of the colours, shapes and locations are taken to represent differences and resemblances among the original texts.

In general, however, data visualisations do not consist exclusively of signs from the graphical modality. In his essay "The Rhetoric of the Image", Roland Barthes examines the various ways in which images can convey a meaning, and identifies three classes of messages.⁵⁵⁷ The literal message, first, consists of the concrete objects or events which are depicted. Next to the literal denotation, the image also has a coded or symbolic message, consisting of the connotations of the objects that are shown. Importantly, Barthes stresses that images often contain a linguistic message, which is made up of the words which occur in the image itself, or in the caption or in the heading that accompanies the image. Barthes' analysis was based principally on the messages produced by advertisement photographs, and it cannot be extended directly to explain the rhetoric of data visualisations. One important difference is that the non-textual elements in a data visualisation cannot establish denoted or symbolic messages independently. Unlike scientific visualisations, graphical renditions of data do not directly portray concrete objects or phenomena. Barthes maintains that, since the signifier and the signified of the objects that are visible in photographs are "quasi-tautological",⁵⁵⁸ the literal message can be referred to as "a message without a code". It may be argued that the graphical modality used in a data visualisation consists exclusively of codes which cannot produce a message in themselves. Specific permutations of graphical primitives principally convey patterns, exposing differences or similarities between various data values. The more precise meanings attributed to specific components of the graph are not codified and they are mostly unique to a given visualisation. They remain fickle and indeterminate until their particular signification is clarified via

⁵⁵⁶ Leland Wilkinson, *The Grammar of Graphics* (New York: Springer 2005), p. 118.

⁵⁵⁷ Roland Barthes, *Image, Music, Text* (New York: Hill and Wang 1977).

⁵⁵⁸ *Ibid.*, p. 154.

the textual modality. Since data visualisations are rarely self-explanatory, developers of visual displays generally need to add a legend or another form of textual support to explain what the various components of the graphic represent. The graphical modality thus depends on the textual modality to produce a meaning beyond the pattern itself. In the case of data visualisations, the text's ability to provide anchorage is vital, as, without it, graphical displays remain problematically devoid of meaning.

The nature and the function of visualisation can be clarified further by contrasting it with the function of typography. Whereas the application of typography and the use of visualisation both result in a certain visual presentation, there are also a number of important differences. The purpose of typography is to show the lexical codes of a text in a specific form, to clarify the logical structure of the text, and, more generally, to ensure that the text is legible and accessible. Bolter argues that typography ought to “make the letter unobtrusive”⁵⁵⁹ and to let readers focus on the contents of the message in the textual modality with as little distractions or obtrusions as possible. It is misleading, nonetheless, to suggest that typography is a neutral layer. Typographers present the text in a particular manner which, to a higher or lesser degree, influences the way in which texts are experienced. Typography, importantly, does not distort the linearity of the text. Linguistic signs derive part of their meaning from their placement within a specific context, and, if the linear order of the tokens is forfeited, this mostly means that readers can no longer decode the message. Whereas typography is concerned with the visual appearance of the text in its full linear form, data visualisation focuses on the production of a succinct and a non-linear rendition of data about texts. The transition to the graphical modality demands operations such as tokenisation and quantification, and, in this process, the linearity of the text is crucially discarded. In an information visualisation, the text can no longer be read in its entirety and the focus is singularly on data about the text. While typography is generally meant to guide the reader's attention effectively to the text and to its various logical components, data visualisation makes the text itself invisible and focuses exclusively on the forms that are produced out of data.

Stephen Few notes that graphical displays of data can be used for two purposes: “sense-making (also called data analysis) and communication”.⁵⁶⁰ Visualisations may initially help scholars to explore the characteristics of large data sets. Once conspicuous aspects have been identified scholars may also generate graphs and diagrams to communicate specific ideas to peers. Graphic displays can either be a research tool or a means of communication. The next section argues

⁵⁵⁹ Jay Bolter, *Writing Space: The Computer, Hypertext, and the History of Writing*, p. 63.

⁵⁶⁰ Stephen Few, “Data Visualization for Human Perception”, in: Mads Soegaard & Rikke Friis Dam (eds.), *The Encyclopedia of Human-Computer Interaction*, Aarhus: The Interaction Design Foundation 2014, n.pag.

that, while visual displays can be highly effective during the exploratory phase of research, they also have a number of characteristics that complicate the communication of scholarly knowledge.

9.3. Expressiveness and rhetoric of visualisations

While scholarly knowledge is conventionally disseminated via articles and monographs, the growing use of visual materials increasingly undermines the position of text as the privileged channel for the transfer of facts and of ideas. In a discussion of the distinction between the terms “illustration” and “visualisation”, Jessop asserts that “an illustration is intended merely to support a rhetorical device (usually textual)” and that a visualisation is critically “intended either to be the primary rhetorical device or serve as an alternative but parallel (rather than subordinate) rhetorical device”.⁵⁶¹ Jessop’s argument implies that data visualisations may partly or wholly supplant a textual publication. Visualisations are generally produced, like written texts, to transmit information, but an exclusive use of the graphical modality also complicates or obviates the transfer of particular types of messages. To understand the value of visualisation within the context of literary research, it is useful to concentrate initially on the means by which graphics can produce meaning, and on the type of messages they can encode.

An analysis of the components of visualisations may be based on conceptualisations of the visualisation process. Visualisations are created in various stages, and many of these stages can affect the manner in which graphics can convey information. Useful conceptualisations of the visualisation process are provided by Ben Fry,⁵⁶² by Leland Wilkinson and by Chen and Floridi.⁵⁶³ In a similar fashion, Hullmann and Diakopoulous distinguish four “editorial layers” that can impact the meaning of graphics.⁵⁶⁴ Using these various descriptions of the visualisation process as inspiration, it may be argued that the meaning of visualisations is determined most fundamentally by four aspects. All descriptions of the visualisation process concur that graphics are based on data. The data roughly

⁵⁶¹ M. Jessop, “Digital Visualization as a Scholarly Activity”, in: *Literary and Linguistic Computing*, 23:3 (2008), p. 283.

⁵⁶² Fry’s textbook *Visualizing Data* presents a sequence consisting of seven stages, in which creators of visualisations sequentially acquire, parse, filter, mine, represent and refine data. Eventually, viewers may also be enabled to interact with the visual display. See Ben Fry, *Visualizing Data* (Cambridge: O’Reilly Media Inc. 2008), p. 8.

⁵⁶³ Chen and Floridi divide the “visualisation pipeline” in steps such as “enriching & filtering”, “visual mapping”, “rendering” and “displaying”. See Min Chen & Luciano Floridi, “An Analysis of Information Visualisation”, in: *Synthese*, 190:16 (26 September 2012), p. 3422.

⁵⁶⁴ The nature of a visualisation is determined by the data that are shown, the way in which these data are “mapped to the visual domain”, the presence of annotations, and the degree of interactivity. See Jessica Hullman & Nicholas Diakopoulos, “Visualization Rhetoric: Framing Effects in Narrative Visualization”, in: *IEEE Transactions on Visualization and Computer Graphics*, (2011), p. 4.

indicate the domain which is described by the chart or the graph. Data visualisations can only represent those resources which were classified in Chapter 4 as structured annotations or as derived data. Captured data are essentially digital reproductions of texts, and these cannot be mapped directly to the visual domain.⁵⁶⁵ The data are visualised can represent data either about individual texts or about corpora in their entirety.

A second aspect which contributes to the meaning of the visualisation is the type of additional processing these data have undergone. The values may, for instance, be sorted, clustered or filtered. As was indicated, visualisations consist of a mapping between qualitative or quantitative data and a set of graphical primitives. Visualisations can be classified, thirdly, by considering how the values are mapped to the visual domain. The data may be represented using different aesthetic attributes,⁵⁶⁶ including shape, rotation, colour or position. Selecting a particular graphical primitive may imply a loss of information, as the granularity with which such primitives may be varied does not necessarily match the granularity of the original data. The meaning of the various visual components may be clarified using legends or labels.⁵⁶⁷ This third criterion is defined broadly, and also covers the usage of different types of scales and coordinate systems.⁵⁶⁸ A fourth aspect which

⁵⁶⁵ In his article "What is Information Visualisation?", Manovich argues that his original definition of information visualisation, as a mapping between discrete data and a graphical display, is complicated by the advent of a number of recent visualisation tools. Although Manovich does not use the term 'modality', he does stress that the reduction of information to an arrangement of graphical primitives is essential to the creation of a visualisation. On the basis of this argumentation, he concludes that a word cloud, for instance, is not a data visualisation. Manovich claims that, since the words that make up the text are also used directly in the visualisation, there is no actual reduction. A word cloud, consequently, is referred to as a 'direct visualisation'. It is a representation in which new forms are produced from the original media without any reduction and without a change in modality. This view, however, is abstruse, as it seems clear that word clouds and the captured data they are based on engage different modalities. While word clouds and natural language texts contain the same codes, these symbols evidently have different functions. The linear words order, which is necessary to understand the meaning of the original text, is, in most cases, discarded fully in the case of a word cloud. The symbols that are used in the latter type of display are primarily labels for metrics. In word clouds, differences in graphical variables such as scale and position are applied directly to the labels to clarify the values that have been calculated for these metrics. Manovich argues that "the two-stage process of first counting, or quantifying data, and then representing the results graphically" is no longer applicable in the case of direct visualisations. To create word clouds, however, it is still necessary to tokenise the text and to count frequencies of types. A word cloud is based on derived data and not on captured data.

⁵⁶⁶ Bertin's term "graphical primitives" and Wilkinson's term "aesthetic attributes" are considered synonymous.

⁵⁶⁷ The textual clarifications which are employed in a data visualisations can connected conceptually to what Barthes refers to as the "linguistic message" of the image.

⁵⁶⁸ Wilkinson explains that the scale of the graphic indicates the ratio between the actual distances in the data collection and the new distances in the graphic. Visualisations may also be based on different coordinate systems, which consist of "sets that locate points in space". The Cartesian coordinate system is used most commonly.

may result in further modifications of the meaning of the visualisation is the degree of interactivity. Sinclair et al. explain that static visualisations are “fundamentally tools for display”, and that interactive visualisations are also research tools, as they enable scholars to experiment with the different renditions of data sets. This exploratory process is often sequential and iterative, as researchers may “revisit previous steps at a later stage and make different choices, informed by the outcomes produced in the interim”.⁵⁶⁹ Whereas Jessop opines that interactivity ought to be viewed as a defining characteristic of visualisation,⁵⁷⁰ this thesis will use the term to refer both to fixed representations of data and to visualisations which can be manipulated flexibly.

In recent years, a large number of tools have become available for the analysis of texts and for the visualisation of the results of such analyses.⁵⁷¹ The current profusion of tools potentially complicates the task of matching visualisation methods to specific research questions. Using the four aspects which were discussed, namely, the type of data, the processing these data have undergone, the graphical primitives and the degree of interactivity, the main functionalities offered by available visualisation tools can be described effectively. Tools such as Voyant Type Frequency Chart and the Distribution tool in TaporWare, for instance, focus on word frequencies via bar chart or via line charts. In both tools, only one file may be uploaded simultaneously, and this makes the tool more suitable for document analysis than for corpus analysis. The tools divide the text into segments, and presents the frequencies of terms within these separate segments. In both applications, the sizes of the text segments may be specified interactively.

In relation to the third aspect that was discussed, the visualisation’s graphical primitives, a distinction can be made between diagrams which are based on conventional statistical models on the one hand and visual displays which explore new forms on the other. The first class of diagrams are based on visual models such as the bar chart, the line chart and the pie chart, which were first developed by William Playfair in the late eighteenth century.⁵⁷² These displays can normally be produced within statistical packages such as SPSS or R. A second class of visualisations clarifies data via innovative amalgamations of visual primitives. One example is Voyant Bubblelines, which is a tool for the creation of distribution graphs. It can clarify the frequencies of terms within distinct segments of individual texts. Voyant Bubblelines, interestingly, represents such values via the radiuses of circles.⁵⁷³ Stefanie Posavec’s project *Literary Organisms*, which visuali-

⁵⁶⁹ Stéfan Sinclair, Stan Ruecker & Milena Radzikowska, “Information Visualization for Humanities Scholars”, pp. 1–2.

⁵⁷⁰ M. Jessop, “Digital Visualization as a Scholarly Activity”, p. 283.

⁵⁷¹ The TAPOR directory of digital research tools lists more than 150 applications for the visualisation of data.

⁵⁷² Edward Tufte, *The Visual Display of Quantitative Information* (Cheshire: Graphics Press 1983).

⁵⁷³ <<http://voyant-tools.org/tool/Bubblelines/>> (11 March 2014)

ses Jack Kerouac's novel *On the Road*, also provides a striking example of an innovative type of visualisation.⁵⁷⁴ Each chapter of the book is represented as a flower, in which the size of the petals indicate the number of words. In addition, data were created manually about the various themes that are covered in each chapter, and these data about themes determine the colour of each flower.

While the informative value of standardised and innovative graphs may be roughly identical, their aesthetics often differ sharply. Hullmann and Diakopoulos note that specific graphical primitives have both a denotation and a connotation. The former term refers to the manner in which graphical primitives express data values. In a bar chart, the heights of the bars typically represent the values that were captured for a given variable. Connotation, conversely, refers to all the supplementary ways in which a graphical element may provoke meaning. Viewers of a bar chart, for instance, are likely to view the data as a "discrete rather than a temporal trend", and line graphs "tend to evoke temporal interpretations".⁵⁷⁵ It may be added that standardised graphics, such as bar charts and line charts connote quantification and objectivity, while the manually drawn visuals of Kerouac appear to belong more naturally within the realm of literary interpretation. For scholars who are sensitive to the aesthetic nature of graphics, Posavec's creative visualisations may bolster hermeneutic processes more cogently.

Visualisations, as was shown, can be dissected into their constituent meaningful components, and designers often have a wide range of options for each of these distinct aspects. Given a particular dataset, there is no predetermined way in which these values may be rendered graphically. Visualisations seldom represent data in a fully neutral manner. In response to Edward Tufte's axiom that graphics ought to "reveal data", Johanna Drucker stresses that data do not have an indisputable shape prior to their visualisation, and that "every graphic representation is a rhetorical device". Unlike scientific visualisations, which mimic an object or an environment whose visual characteristics are given a priori, diagrams that represent literary research data are largely designed during the data analysis phase. Creators of graphics often aim to convey particular ideas, and design their visualisation in such a way that they most effectively and most understandably express their predilections. In Ben Fry's description of the visualisation pipeline, the step that is labelled "refine" explicitly aims to further embellish or to improve the rhetorical effect of the graphics that are created initially.⁵⁷⁶ By altering aesthetic aspects such as shape, colour and width, or, by adjusting ranges on the horizontal or on the vertical axes, for instance, researchers can actively influence the dimensions and the overall appearance of the patterns that emerge. Via such manipulations of

⁵⁷⁴ <<http://www.stefanieposavec.co.uk/>> (10 March 2014)

⁵⁷⁵ Jessica Hullman & Nicholas Diakopoulos, "Visualization Rhetoric: Framing Effects in Narrative Visualization", p. 2237.

⁵⁷⁶ Ben Fry, *Visualizing Data*, p. 5.

visualisations, viewers may be guided into the direction of specific readings. Maureen Stone has noted that such manipulations may occasionally lead to misrepresentations of the data.⁵⁷⁷ Kathleen Kerr explains, in a similar fashion, that visualisations can be based either on a formalist and rational approach, or on a social constructionist and post-rationalist approach. The former approach assumes that the characteristics of data can be shown in a manner that is unproblematically truthful, and that visualisations can represent the inherent shape of the data set in an objective manner. The social constructionist approach emphasises that “the right format is negotiated in relation to the needs of both producer and user”.⁵⁷⁸ Graphics are designed by scholars working in a particular theoretical tradition, using tools that are based on particular epistemological assumptions, and for the purpose of highlighting particular aspects in the data set. What is included and what is excluded in a visualisation depends for a large part on the idiosyncratic interests and the methodological conventions of scholars.

Whereas scholars can usually influence the appearance of the forms that are produced out of data, making an explicit claim about the domain that is depicted, beyond the mere patterns that are produced, is generally onerous. Data visualisations primarily depict patterns, and they generally fail to express sustained arguments. Jay Bolter explains that images, in general, are “designed to identify objects ... and situations ... rather than convey a discursive message”.⁵⁷⁹ Defending an argument involves a temporal or a linear progression in which initial premises are followed by statements that can be concluded from earlier assertions. While different images can evidently be placed in a sequence in order to construct a narrative, the precise relations between the ideas and the concepts that are depicted is often difficult to convey explicitly, without taking recourse to the verbal modality. Persuasive argumentation also demands the possibility to express evaluative statements, or the capacity to categorically sanction or to dismiss an idea. Images can depict objects or phenomena, but, since there are no codified visual means of expressing approval or dismissal, it is often difficult to give a direct and unambiguous expression to the attitude of the creator of the graph towards the objects that are depicted. Visualisations, in short, are often unable to present complicated forms of inductive reasoning, or to invalidate counterarguments.

Data visualisations, crucially, convey patterns rather than sustained arguments. Precisely because of the dearth of explicit arguments, however, viewers are often empowered to develop their own readings of the visual information, within the constraints that are set by the creator of the graphic. Sarah Jones emphasises that textual descriptions of a scholar’s impression of a literary work habitually remain

⁵⁷⁷ Maureen Stone, “Information Visualization: Challenge for the Humanities”, pp. 46–47.

⁵⁷⁸ Kathleen Kerr & Waqas Javen, “Visualization and Rhetoric: Key Concerns for Utilizing Big Data in Humanities Research”, in *IEEE International Conference on Big Data*, (2013), p. 29.

⁵⁷⁹ Jay Bolter, *Writing Space: The Computer, Hypertext, and the History of Writing*, p. 61.

confined to those aspects which were considered worthy of attention by the author. Readers of a text follow a discursive trajectory which was fully designed beforehand on the basis of the interests of the author. Data visualisations, by contrast, can grant viewers the liberty to concentrate on the aspects of their choice. Compared to reading a text, inspecting a chart or a diagram is an experience that “belongs more to the reader”.⁵⁸⁰ The ability to provide direct access to visual renditions of data may be considered an exponent of the broader development which is referred to by Van der Weel as the “deferral of the interpretative burden”.⁵⁸¹ As digital research instruments continue to generate large quantities of raw data, scholars increasingly provide access to scholarly “semi-manufactures”. Visualisations do not explicitly state why particular shapes are relevant or significant, and they lack an explanation of the causes that underlie the patterns that are shown, relocating the the task of “sense-making” to consumers of these resources.⁵⁸² As discussed, it would be incorrect to suggest that data visualisation provides access to uninterpreted data. The data values themselves often result from subjective methodological decisions, and their graphical renditions are likewise the outcome of a biased selection of visualisation techniques and of subsequent subjective refinements.

The suspension of explication complicates Jessop’s suggestion that visualisations of data can serve as substitutes for textual publications. The question whether or not graphics can serve independently as legitimate products of scholarship can be connected to a broader discussion on the status of non-textual artefacts produced within digital humanities research. A scholarly engagement with computers generally results in a broad range of digital resources, such as databases, visualisations, software and scans. Many institutions that assess academic performance, including grant committees and tenure committees, frequently contend that the creation of such digital artefacts does not constitute scholarship in itself. In many cases, digital humanists can receive recognition for such digital products only when they conjointly publish a critical companion text. Mills Kelly has argued that genuine acts of scholarship ought to be “the result of original research”. It must propagate “an argument of some sort”, which needs to be “situated in a pre-existing conversation among scholars”.⁵⁸³ As data visualisations often lack such explicit arguments, it can be difficult to let them count as independent scholarly resources.

⁵⁸⁰ Sarah Jones, “When Computers Read: Literary Analysis and Digital Technology”, in: *Bulletin of the American Society of Information Science and Technology*, 38:4 (2012), p. 28.

⁵⁸¹ Adriaan van der Weel, “New Mediums: New Perspectives on Knowledge Production”, in: Wido van Peurzen, Ernst Thoutenhoofd, & Adriaan van der Weel (eds.), *Text Comparison and Digital Creativity*, Leiden: Brill 2010, p. 263.

⁵⁸² In this respect, data visualisations are similar to digital editions which provide access to all extant witnesses of a text. If the resource does not privilege a particular variant, readers are forced to find their own paths through the materials.

⁵⁸³ Mills Kelly, “Making Digital Scholarship Count”, in: Dan Cohen & Tom Scheinfeldt (eds.), *Hacking the Academy: A Book Crowdsourced in One Week*, Michigan: MPublishing 2011, p. 33.

The notion that visualisations are incapable of expressing interpretation has been contested, however, by Johanna Drucker. She urges scholars to “find graphical means of expressing interpretative complexity”, and also notes that standardised visualisation models, such as bar charts and pie charts, are inadequate in this respect, largely because of their roots in statistics and in the empirical sciences.⁵⁸⁴ The field of literary studies does not have a historical tradition of using visual displays, and digital humanists were consequently forced to borrow many tools and techniques from other disciplines, such as the social and the natural sciences. Drucker emphasises that such inherited visualisation techniques inappropriately follow the epistemological assumptions of disciplines that assume that properties of phenomena exist independently from the observer. Conventional graphical displays which represent values on a Cartesian coordinate system, or which represent values as distinct bars with equivalent sizes are considered inappropriate for the representation of the results of humanities research. The concepts and phenomena that scholars engage with critically are frequently highly complex, and they typically resist a reduction to a finite set of sharply separate categories. Specific phenomena, such as the experience of time in Woolf’s novel *Miss Dalloway*, for instance, cannot be represented simply though a linear timeline. Drucker has developed experimental diagrams herself which draw clear attention to the interpretative and subjective nature of humanistic enquiry, by using fluid borders in between categories or by replacing single broad categories with a multitude of smaller overlapping more flexible categories. Instead of Cartesian grids, Drucker proposes axes in which the temporal dimension is more transient. The bar charts she proposes do not display values in mutually exclusive categories, but values which can be in different categories simultaneously and values which can flow unstintingly outside of predefined categories.

Drucker’s diagrams may be viewed as a criticism, in a graphical form, of an oversimplified approach towards quantification in the humanities. Indeed, the reductions and the abstractions that underlie data creation are not always warranted. Properties cannot always be described via a single value and texts do not always fit neatly within predefined categories. Drucker proposes that the concept of data, approached in a reductive manner, must be replaced with the notion of interpretation, being the researcher’s subjective experience of the properties of aesthetic works. Evidently, such idiosyncratic impressions do not follow mathematical principles, and, consequently, these cannot be visualised using standardised statistical packages.

While Drucker’s experimental diagrams rightly attest to the problematic character of humanistic data, the suggestion that visualisations ought to express the full complexity of the phenomena that are studied can be challenged. Drucker argues that visualisations created for humanities research ought to represent

⁵⁸⁴ Johanna Drucker, “Humanities Approaches to Graphical Display”, s. 50.

subjective interpretations of phenomena rather than quantitative data which can be manipulated mathematically, but such renditions of interpretations are often difficult to interpret in themselves. Graphs are typically produced to enhance cognition, and to impart information about a specific scholarly domain. Drucker's humanistic visualisations, consisting of fluid categories and non-linear timelines, seem difficult to decode without an additional textual explanation. The graphs do organise specific components in a spatial arrangement, but it is unclear if these specific positions also represent a particular meaning. The diagrams, much like the works that are depicted, become resources that need to be interpreted. Since these visualisations are not based on a formal logic or a known code, and given the limited expressivity of the graphical modality, the viewer cannot know whether or not a particular interpretation is correct until a textual explanation is given. As the precise statements about the phenomena that are studied are often unclear, the diagrams primarily emphasise the notion that the creation of data requires nuance and flexibility. When graphical elements are used to communicate a critical scrutiny of the humanistic methodology, as in Drucker's proposal, this imposes an illocutionary weight on these graphs they can barely carry.

An additional difficulty inherent in graphs which represent the interpretation of phenomena is that they frustrate the emergence of new discoveries. Data visualisations are often created to marshal surprises. They may expose unexpected correlations or conspicuous disassociations, which can in turn galvanise new readings of the texts that are rendered graphically. The visualisations proposed by Drucker are steered wholly by existing interpretations, which, as can be assumed, were produced in a non-digital context. Evidently, the embossed presence of one scholar's interpretation discourages and restrains new exploratory hermeneutic analyses. Drucker's diagrams indeed usefully stress the convolution of particular humanistic phenomena, but, as the diagrams are ostensibly not based on logic, they remain difficult to interpret. Mathematics and statistics, by contrast, imply a standardised method for encoding and decoding graphical properties, and it seems judicious to continue to create diagrams which adhere to basic quantitative principles. Visual literacy, in many cases, demands graphical displays based on data which can be manipulated mathematically. Discussion about the complexity of the methodology, and, more specifically, about the difficulties that inhere in the definition and the application of descriptive categories, can arguably be expressed more effectively via the textual modality.

Visualisations, in short, communicate characteristics of quantitative data via particular constellations of graphical primitives. They may present data at different analytic levels and they may be static or interactive. While visualisations are less adept at communicating complex scholarly arguments, visualisations often stimulate their viewers to formulate interpretative claims on the basis of patterns. Gibbs and Owens stress that work with data in the humanities "can be exploratory

and deliberately without the mathematical rigour that social scientists must use to support their epistemological claims”.⁵⁸⁵ The capacity of data analyses to generate new ideas is generally more important than their ability to serve as conclusive evidence for particular statements, let alone than their ability to make such statements themselves. Visualisations used in the service of literary criticism are valuable mostly because of the fact that they can foster a hermeneutic engagement with the objects that are represented.

9.4. Interpretation

In *Simulacra and Simulation*, Jean Baudrillard has argued that postmodern society has become deeply reliant on models and on simulations. He uses the term “simulacrum” to refer to “models of a real without origin or reality”.⁵⁸⁶ While representation is based on the principle that the sign is a counterfeit of an original, simulacra are models of objects or phenomena which are fully contrived. They establish a “hyperreality”. Visualisations may likewise be viewed as simulacra. Helen Westgeest observes that, within the natural sciences, graphics are often created to give objects which are invisible in reality a perceivable presence. Researchers can visualise imperceptible phenomena by means of models, which can be described as “visual constructions”.⁵⁸⁷ Visualisations of data about literary texts likewise fabricate shapes that do not have an exact match in an objective reality. Data sets are evanescent and shapeless in themselves, and the process of visualisation fundamentally consists of inventing a physical layout for these data. Graphical displays introduce new forms, but they are produced to reveal characteristics of the originals which are represented. Their aim is mostly to produce an innovative perspective, and to allow for novel and compelling interpretations of these texts.

For interpretation to be possible, viewers initially ought to be able to decode the visual patterns that are observed into descriptive statements about the objects that are represented. They need to be able to formulate a set of statements which appear to be justified on the basis of what is shown in the visualisation. Stephen Few stresses that a graphic is “only successful to the degree that it encodes information in a manner that our eyes can discern and our brains can understand”.⁵⁸⁸ The value of visualisation tools may be assessed by evaluating how effectively they

⁵⁸⁵ Frederick W. Gibbs & Trevor J. Owens, “The Hermeneutics of Data and Historical Writing”, in: Kristen Nawrotzki & Jack Dougherty (eds.), *Writing History in the Digital Age*, MIT Press 2013, n.pag.

⁵⁸⁶ Jean Baudrillard, *Simulacra and Simulation* (Ann Arbor: University of Michigan Press 1994), p. 166.

⁵⁸⁷ Helen Westgeest, “Visualizing Research and Visual Communication of Research”, in: Helen Westgeest (ed.), *Making Research Visible to the World*, Amstelveen: Canon Foundation in Europe 2010, p. 12.

⁵⁸⁸ Stephen Few, “Data Visualization for Human Perception”, n.pag.

convey information. When data sets are voluminous and heterogeneous, it can evidently be taxing to represent the values in a manner that is both legible and meaningful. A comprehension of the information that is represented may additionally be impeded when viewers lack an understanding of the type of processing the data have undergone. One example of a tool which may clog cognition is Voyant Knots. The tool represents the corpus as a collection of twisted lines, and the “extent to which lines overlap indicates the level of correspondence or linkage between the terms”.⁵⁸⁹ Without a thorough understanding of the algorithm that underpins the visualisation, it is difficult to translate the visual patterns that can be generated in the tool to a univocal descriptive statement about the corpus that is depicted.

In the context of literary criticism, visualisations of data about texts can be valuable if these can spur a hermeneutic engagement. Interpretation, on a first level, entails an understanding of the central meaning or of the central theme of the text. If the term “theme” is understood narrowly as the direct denotation or as the literal “aboutness” of the text, explorations of themes may potentially be based on displays of significant or distinctive vocabulary. This approach is followed for instance, in Matthew Jockers’ book *Macroanalysis*.⁵⁹⁰ As was discussed in Chapter 5, poetic texts rarely express their abstract literary themes directly. Themes can, in many cases, be found exclusively via close reading. In writing that uses figurative language, the semantic context is generally too complex and too unpredictable for current semantic taggers or for topic modelling algorithms. It is cumbersome, for this reason, to devise visualisations which enable scholars to read the literary themes a collection of texts directly.

Whereas visualisations often fail to elucidate the immediate thematic concerns of a poem, graphical resources can help scholars to investigate the manner in which the language of a text contributes to the production of its central meaning. The aim of conventional close reading is often to describe the relationship between the text’s form and content. Data which are produced automatically typically focus on the language of the literary object, and operations such as filtering, clustering, sorting or distribution, performed at the level of the corpus as a whole, can often expose trends or regularities in an author’s use of literary devices or of linguistic constructions.

As was discussed in Chapter 4, the various functionalities which are provided by text analysis tools can be classified by making use of Unsworth’s concept of the scholarly primitives. Visualisation tools primarily provide support for “comparison” and for “discovery”. Graphics which display correlations, clusters or distributions centrally aim to expose differences and similarities between the texts within the collection and, as such, they centrally entail a comparison of texts.

⁵⁸⁹ <<http://docs.voyant-tools.org/tools/knots/>> (2 June 2015)

⁵⁹⁰ Matthew Jockers, *Macroanalysis : Digital Methods and Literary History*, p. 118.

Comparison initially discloses patterns within a collection of texts, but it may also lead to the discovery of a limited set of documents which are distinctive within the context of the collection. Visualisations of text corpora primarily support interpretation through their capacity to inspire more targeted forms of close reading. Visualisations of the formal features of texts often raise particular questions, and the answers to these questions can often be found solely by rereading the separate texts that shape the diagram. The formal properties of texts can be accounted for by relating these to other factors such as the genre, the historical context or the overall thematic concerns.

While the close reading that follows from discovery can be performed on the basis of the original linear texts, such analyses at the level of poems or at the level of verse lines can also be supported effectively via visualisations which represent data about individual poems. Visualisations at the micro-level have been developed within a number of projects. The Mandala browser which was developed by Dobson, Ruecker, Gabriele and Sinclair, primarily offers support for the clustering of the smaller units of a texts. The browser demands XML-encoded texts as input. On the basis of the encoding, the text is divided into its constituent structural components, such as paragraphs, verse lines, or, in the case of drama, speeches. Having created these smaller units, users can define search terms within the text's interface which can attract or repel specific fragments. In this way, specific groups can be created of text fragments which contain or which lack specific search terms. As the Mandala Browser transforms the single linear text into an assemblage of movable units, Brown et al. note that the application stimulates a traversal of "a major interface boundary". The verbal text is abandoned "in favour of working with configurations of coloured dots",⁵⁹¹ and scholars can focus predominantly on the formation and the exploration of patterns.

Visualisations of individual texts can similarly be produced within *Myopia* and the *Poem Visualizer*. In *Myopia*, data about literary devices initially need to be supplied manually via a customised set of TEI elements. The program can subsequently exploit these data to generate abstracted renditions of the encoded texts. *Myopia* displays the individual words of the poems as blocks, and it can be specified that the fill colours of these blocks must be determined by the occurrences of literary devices. The tool can illuminate the ways in which textual phenomena such as alliteration, rhyme or consonances are distributed over the various lines.⁵⁹² *Poem Viewer*, which was developed by A. Abdul-Rahman et al., likewise shows the full verse lines, combined with connectors at the end of each line which indicate the lines that rhyme. A small diagram is shown, furthermore, above each vowel in the text, to clarify the positions of phonetic articulations. Such

⁵⁹¹ Susan Brown et al., "Reading Orlando with the Mandala Browser: A Case Study in Algorithmic Criticism via Experimental Visualization", in: *Digital Studies / Le champ numérique*, 2:1, n.pag.

⁵⁹² Manish Chaturvedi et al., "Myopia: A Visualization Tool in Support of Close Reading", n.pag.

graphics effectively enable scholars to study developments in vowel positions and particular sonic recurrences.⁵⁹³

Unlike the Mandala browser, *Myopia* and the *Poem Viewer* display both the original linear texts and the data that are available about these texts. In a sense, these types of visualisations blend the textual modality with the graphical modality. They blur the distinction between reading texts via typography and viewing texts via visualisations. An important objective of visualisations is to establish a condensation, and to provide a succinct expression of the available data. In the visualisations that are created in *Myopia* and in the *Poem Viewer*, however, the graph is as extensive as the text itself. The graphics also display a degree of linearity, since the order in which the data elements are shown corresponds directly to the linear order in which these aspects occur in the original. Graphics which display the occurrences of literary devices within the same logical structure that is used in the original linear text can effectively illustrate the craftsmanship of a poet. Poetic effects frequently consist of specific arrangements of sounds or of words, and the exact nature of correspondences and recurrences can often be illuminated adequately by visualising occurrences of such devices at the level of individual verse lines.

Such visualisations at the micro-level, in which occurrences of literary devices are indicated within their original logical structure can similarly be created using the structured annotations that were created for the central case study of this thesis. Figure 9.1 clarifies the distribution of perfect rhyme and of alliteration in MacNeice's poem "Selva Oscura". Occurrences of perfect rhyme are shown in red and instances of alliteration are shown in blue. In some cases, the final phoneme sequences which produce the rhyme at the end of the verse lines also appear in the interior of verse lines. Such forms of internal rhyme have also been indicated using a red colour. The poem has a fairly regular structure, consisting of four stanzas of five lines each. In each stanza, only four lines rhyme. This formal structure, in which one line is left companionless, mirrors the loneliness of the poem and the central idea that a "house can be haunted by those who were never there / If there was where they were missed" (ll. 1-2). The poem synchronously makes a commanding use of alliteration. In many of the verse lines, and most frequently in the last two stanzas, the word which produces the end rhyme also alliterates with one or more other words on that same line. This is the case in the lines "One sudden shaft of light from the hidden sky" and "Perhaps suddenly too I strike a clearing and see". The alignment of rhyme and alliteration, and the harmonious ring that it gives to these lines, effectively braces the experience, captured in the final stanza, that

⁵⁹³ A. Abdul-Rahman et al., "Rule-based Visual Mappings – with a Case Study on Poetry Visualization", in: B. Preim, P. Rheingans, & H. Theisel (eds.), *Eurographics Conference on Visualization (EuroVis)*, (2013), p. 8.

the solitude can also be ended. In the final lines, the speaker sees that the door of the isolated house “swings open and a hand / Beckons to all the life my days allow”.

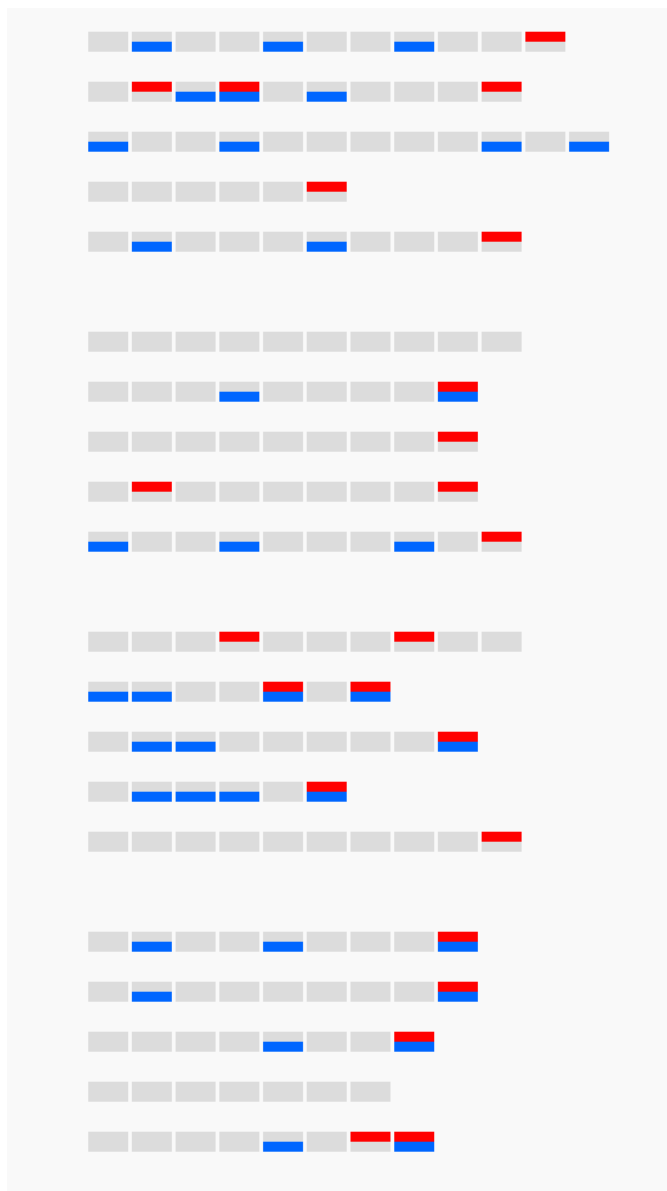


Figure 9.1. Visualisation of perfect rhyme and of alliteration in "Selva Oscura"

Figure 9.2 clarifies the distribution of rhyming sounds in the “The Sunlight on the Garden”. To create this diagram, the rhyming sounds that occur in the poem have been assigned a unique code. All words have been analysed, so that occurrences can be identified of both final rhyme and internal rhyme. The diagram effectively displays both the elegant rhyming scheme and the adept use of internal rhyme. A degree of editing was necessary, nevertheless, to ensure that the regularities in the rhyming patterns could be shown effectively. It was found, for example, that the regularity in the end rhymes could be shown most clearly if the verse lines were aligned to the right.

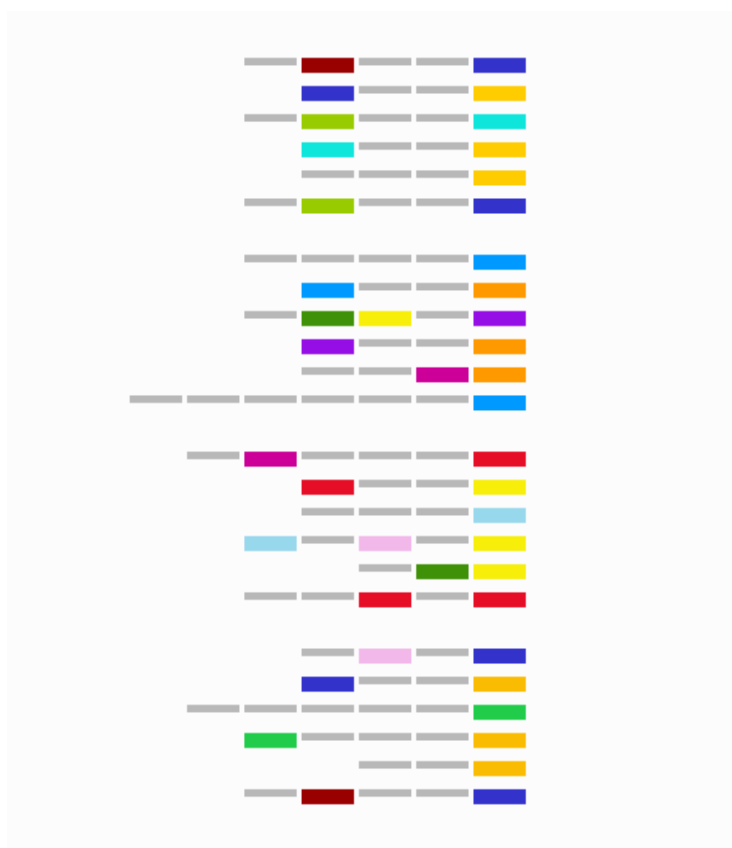


Figure 9.2. Rhyming sounds in “The Sunlight on the Garden”

Figure 9.3 is a similar representation of “The Glacier”. As is the case in figure 9.2, all the final phoneme sequences that occur more than once, both at the end and in the interior of a verse line, have been associated with a unique colour code. The colours have been assigned randomly. As can be seen in figure 9.3, the first section contains a large number of repeated sounds. Phoneme sequences are

repeated, for instance, in the words “who”, “through”, “two” and “you”, in “climb” and “fine” and in “they” and “day”. In this poem, the use of rhyming sounds is clearly supportive of the text’s central idea. “The Glacier” consists of two separate sections. The opening section focuses on the hectic nature of modern life, and it conveys the central idea that the traffic on the city streets moves so quickly that it paradoxically seems static and petrified. The closing section expresses a longing for a quieter alternative. The cascade of repeated rhyming sounds in the first half of the poem, in a sense, mimics the glacier that is depicted. The second section of the poem contains a smaller number of rhyming sounds and the lower half of the diagram consequently looks calmer. The calmness and the stasis is also underscored by the rime riche in “where” and “ware”. While patterns such as these are difficult to see in the regular verbal structure, such unequal distributions can be shown efficiently via visualisation.

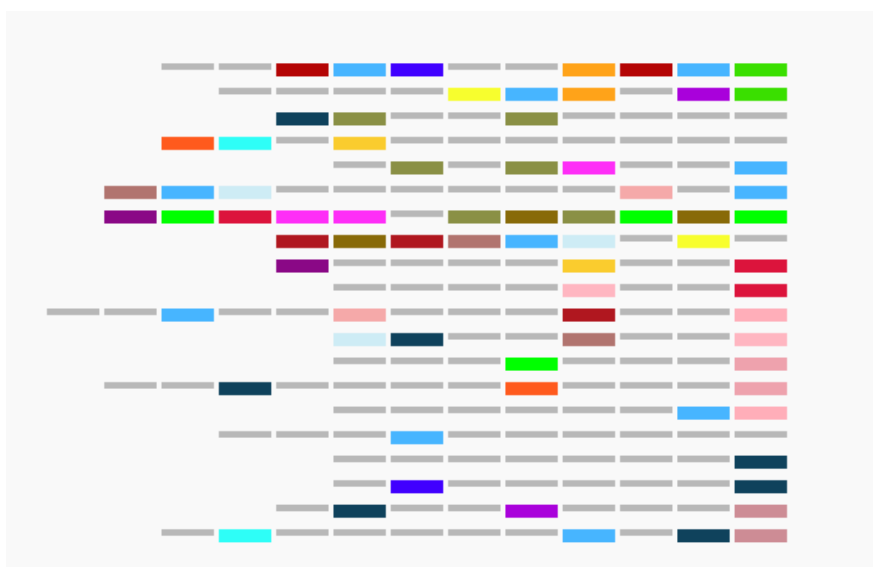


Figure 9.3. Rhyming sounds in “The Glacier”

Visualisations at a micro-level can also clarify the texture of the poem. The term *texture* generally refers to the manner in which the sounds that occur in a poem are combined into patterns. The sounds in vowels and in fricative or aspirated consonants are mostly soft, while plosives are often experienced as harsh. Figure 9.4 is a representation of the texture of “The Glacier”. To produce this graphic, each line has been divided into its separate syllables. Syllables containing plosives have been given a dark red colour, while the syllables which contain softer sounds have been assigned a light yellow colour. On the basis of this visualisation, the softness or the harshness of the sounds in the different poetic lines can be

studied directly. It can be seen that the first section of the poem uses many harsh sounds. This is especially the cases for lines 6 and 12 of the poem (“That you cannot catch the fraction of a chink between the two” and “Cannot bear to watch that catafalque creep down”). The second part, which expresses a longing for quietude, also uses much softer sounds. Line 18, for instance (“Eyes appraise the glazen life of Majolica ware”) mostly contains nasal and lateral consonants.

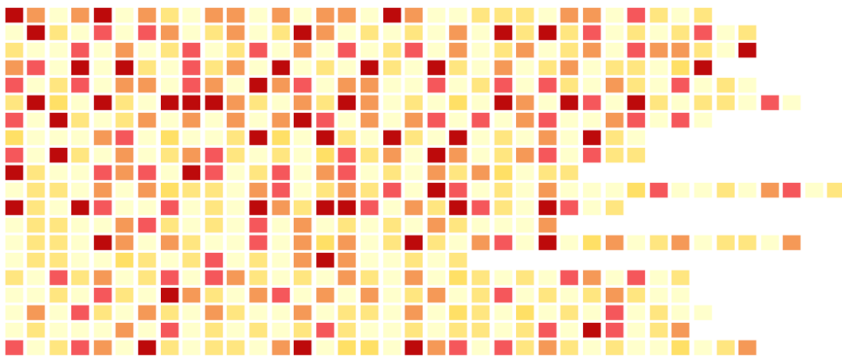


Figure 9.4. Visualisation of the texture of “The Glacier”

A central objective of visualisation, and more, generally, of data analyses, is to arrive at patterns which can stimulate interpretation via more focused forms of close reading. Graphical representations of individual texts can elucidate specific patterns in the use of literary devices, and such patterns can encourage scholars to explore whether or not these formal features support and reinforce the meaning of the text. The aim to produce such interpretable patterns simultaneously implies a predicament. Texts can be processed and visualised in an endless number of ways, and it is often difficult to know beforehand if particular manipulations can actually result in patterns that can foster a hermeneutic engagement. Some manipulations may open up new interpretative possibilities, while other manipulations may prove unproductive. As the precise effects of statistical operations are often unforeseen, the discovery of patterns typically contains an erratic and an aleatory element. The production of interpretable graphics often demands various cycles of trials and of errors.

Literary texts are invariably highly complex and multifaceted phenomena, and through the process of visualisation, researchers can reduce such complexity. Graphic renditions of quantitative data mostly aim to clarify the main characteristics of multivariate data collections by rendering these in a succinct format. Such abstractions or summarisations unavoidably imply a loss of information, as they are created by privileging a limited set of dimensions at the expense of certain other dimensions. There can subsequently be a more concentrated examination of

the dimensions that remain. Visualisations are generally abstractions which simplify complex data sets and which obscure much of the ambiguities and uncertainties that may interfere in these data.

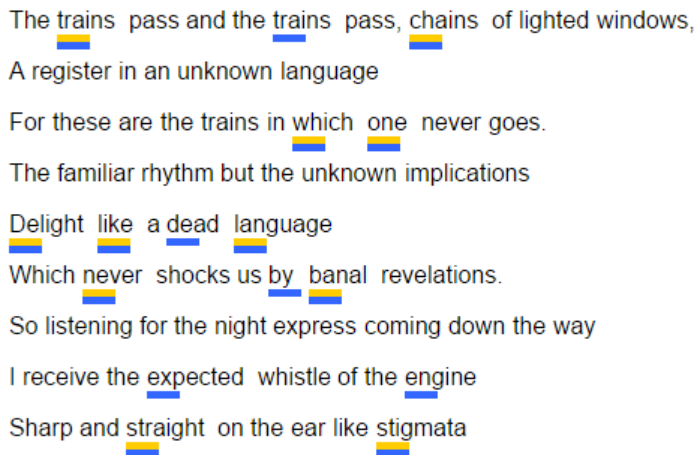
While such simplifications may be useful, literary scholars may also benefit from methods which can expose the nature of such reductions and which can effectively indicate the ambiguities and the contradictions that exist within the objects that are studied. Kouw et al. stress, more generally, that ambiguity and uncertainty can frequently form a source of new knowledge. Ambiguities can serve as a disruptive force which can illuminate the ways in which prevailing conceptualisations misrepresent an actual situation. When scholars are attentive to uncertainty, this “opens up potentials otherwise veiled”, as problems of uncertainty and ambiguity may provoke new types of solutions. While computer applications are often concerned with disambiguation and with reducing complexity and irregularity, literary informatics can also benefit from applications that support processes of reambiguation, and from visualisations that can effectively demonstrate that texts are more complicated and more varied than was originally assumed. Digital methods can stimulate interpretation by generating compelling quandaries which may force scholars to read texts in a different light. Scholars do not necessarily aim to solve all the difficulties that inhere in a poem, and they are frequently intent on demonstrating that texts can yield to many different, and potentially contrastive, readings.

At present, however, there are no unequivocal rules or guidelines for creating visualisations that can productively expose complexity and ambiguity. In natural languages, speakers can often use qualified language to indicate that specific assertions lack convincing support. Levels of ambiguity or of uncertainty can be made explicit, for instance, via the choice of specific verbs and adverbs. When data sets include ambiguous data, it is pivotal, analogously, to qualify these visualisations. Leland Wilkinson stresses that data sets may be incomplete, indeterminate, biased or downright erroneous, and graphs which do not explicitly indicate the presence of uncertain or faulty values are obviously deceptive. Wilkinson notes that when a data set contains systematic indications of the confidence levels of variables these can be used in a variety of ways. Such quantifications of the levels of uncertainty can be used to determine the degree of transparency, the textures or the shapes of the icons and of glyphs which are used in a visualisation. Such visual aids may “guide, qualify, or soften our judgments of uncertain data”.⁵⁹⁴

In studies that concentrate on data about literary texts, three causes of ambiguity can be distinguished. A first form of ambiguity is generated by a conflict between the connotations of literary devices and the thematic concerns of the texts in which these devices are used. A number of such cases of ambiguity have been discussed in Chapter 8. Poetry written in light verse is generally experienced as

⁵⁹⁴ Leland Wilkinson, *The Grammar of Graphics*, p. 450.

buoyant or as cheerful, but, as was shown, MacNeice often used this verse form in poems which address morose and gloomy subjects. Visualisations which juxtapose phenomena which seem incompatible or contrastive can often pose stimulating quandaries.



The trains pass and the trains pass, chains of lighted windows,
 A register in an unknown language
 For these are the trains in which one never goes.
 The familiar rhythm but the unknown implications
 Delight like a dead language
 Which never shocks us by banal revelations.
 So listening for the night express coming down the way
 I receive the expected whistle of the engine
 Sharp and straight on the ear like stigmata

Figure 9.5. Two alternative methods for detecting alliteration

The fact that many of the data have been generated via algorithms forms a second cause of ambiguity. Using a term proposed by Mons and Velterop, the annotations that have been produced automatically may be referred to as “hypothetical data”.⁵⁹⁵ The algorithms that have been used to create these data are critically based on a theory about the manner in which particular literary phenomena may be detected. These data remain hypothetical until they have been vetted by human scholars. If manual interventions in the results of the algorithms are registered in metadata, visualisations may also represent the difference between hypothetical data and data which are endorsed manually. Alternatively, the hypothetical nature of data about literary texts may potentially be emphasised by visualising the results of different implementations of algorithms. In Figure 9.5, the colours beneath the words indicate the implications of two different approaches towards the recognition of alliteration in MacNeice’s poem “A Contact”. The first algorithm, whose results are shown in yellow, considers repetitions of consonants in stressed syllables. The second algorithm asserts that all repeated sounds at the beginning of both stressed and unstressed syllables can produce alliteration. Visualisations such as these effectively demonstrate the notion that idiosyncratic

⁵⁹⁵ Barend Mons & Jan Velterop, “Nano-Publication in the E-Science Era”, p. 8.

decisions about the logic that is used within an algorithm can have direct implications on the eventual visualisations. They may simultaneously expose cases in which two different implementations of an algorithm both result in relevant cases, underscoring the complexity both of the heuristic activity that is modelled and of the text that is analysed.

The ambiguity of data sets can also be connected, thirdly, to the inevitable subjectivity of observations. Even when specific assertions which were generated algorithmically are approved by one particular scholar, other scholars may still disagree fervidly with these observations. In the case of subjective, or observer-dependent, observations, it is generally impossible to quantify the level of certainty. Whereas visualisations typically reflect one particular author's preferred manner of simplifying, static graphs misleadingly evoke a sense of closure, and they consequently acquire an aura of objectivity. Rieder and Röhle observe that there is a tendency among viewers to automatically accept visual information as evidence.⁵⁹⁶ While arguments which are presented in a textual resource often disclose the larger process that has been followed to arrive at a particular conclusion, static graphics often lack information about their genesis. As it is difficult for viewers to identify the potential errors that have been made in the production of the static visual, it is correspondingly difficult to refute the findings that are presented.

In the digital realm, the seeming closure of data visualisations can be counteracted through the construction of interactive and elastic graphs. When other scholars are empowered to alter the parameters that govern the graphical rendition, this may effectively stress the notion that diagrams are also arbitrary and contingent. Interactive diagrams can illustrate the notion that different points of view also result in different diagrams. Interactivity strongly weakens a scholar's ability to communicate fixed and authoritative statements about texts, however, adding to the deferral of the interpretative burden. Interactive visualisations primarily help scholars explore the characteristics of a data set, to assess some of the assumptions that underlie the data, and to form new ideas about the objects that are depicted.

9.5. Conclusion

Louis MacNeice viewed verbal communication as a superior mode of expression. In his poem "To Posterity", MacNeice expresses his anxiety that this manner of communicating may eventually be replaced by "other, less difficult, media". He fears that, when experiences are no longer "framed in words", birds will "be always wingless birds". In digital humanities research, scholars increasingly make use of visualisations, and, these graphic displays, to a large extent, are indeed "less

⁵⁹⁶ Bernhard Rieder & Theo Röhle, "Digital Methods: Five Challenges", in: David Berry (ed.), *Understanding Digital Humanities*, Basingstoke: Palgrave Macmillan 2012, p. 74.

difficult”, in the sense that they offer less refined possibilities for the communication of complex ideas. It is not feasible, for instance, to convey a sustained argument. Although visualisations are generally created for a particular purpose and from a particular theoretical perspective, they do not explicitly sanction or reject specific views. Furthermore, there are currently no widely accepted conventions for the visual expression of subjectivity and ambiguity. For this reason, it does not seem probable that visualisations will fully supplant discursive and argumentative texts. A visualisation is primarily a rhetorical device which can support an argument, and not an argumentation in itself.

Graphical displays do support a new form of studying texts, however. Data visualisations form compelling examples of the form of research which Jerome McGann described as “performative” and “deformative criticism”. McGann surmises that computational analyses may reinforce hermeneutic processes via the creation of transformed versions of texts. Deformative operations, such as the elimination of specific word types, a reversal of the order in which lines are presented or the creation of frequency lists can expose unexpected perspectives on familiar texts, and these new vistas may subsequently spawn new ideas. Aspects such as “the structural forms of words, phrases, and higher morphemic and phonemic units” are often viewed as “preinterpretative and precritical”,⁵⁹⁷ but a rearrangement of these components can fruitfully “release or expose the poem’s possibilities of meaning”.⁵⁹⁸ Visualisation derived from data about the literary techniques can be likewise valuable for literary critics, as they allow for a highly systematic scrutiny of the language that is used within a text. Works of literature are often very rich repositories of ideas and of sonic patterns. Using the graphical modality, the nature of text corpora may be illuminated via the production of distribution graphs, via filtering on the basis of secondary data, and via the creation of clusters. These basic operations result in patterns which may provoke specific ideas and interpretations.

Importantly, visualisations can be created at any level of aggregation. Next to visualisations of corpora in their entirety, scholars can also produce abstracted renditions of individual text fragments. Visualisations of separate texts may help to investigate the more precise ways in which words, sounds and grammatical categories collaborate to produce a meaning. The macro-level and the micro-level patently complement each other, and an exclusive focus either on the whole or on the units seems incomplete. A recognition of the deficiency of such a singular and unadaptable approach is also manifest in the appellation of one particular visualisation tool which can provide support for the process of close reading, namely, the Myopia application. The name of the tool notably implies an inability to perceive objects which are far away. It may be argued that, if close reading is

⁵⁹⁷ Jerome McGann, *Radiant Textuality: Literature after the World Wide Web*, p. 115.

⁵⁹⁸ *Ibid.*, p. 108.

myopic, distant reading is a hyperopic form of engagement. In the latter form of reading, the focus is predominantly on the broader patterns, at the expense of particularities of individual texts. Unlike healthy eyes, eyes affected by myopia or by hyperopia are deprived of the capacity to focus variably. In the case of such defects, technology such as glasses or contact lenses may intervene to restore the ability to focus both on details which are nearby and on objects which are distant. Literary critics are arguably myopic by default, but they can similarly use technology to study the phenomena which can only be observed from a distance. Both perspectives are important for a solid understanding of literary works. Digital methods, importantly, enable scholars to switch between myopic and hyperopic reading, and to concentrate both on patterns and on particulars.