A stream of dancing lights, for all the world like the shimmering curtains of the aurora, blazed across the screen. They took up patterns that were held for a moment only to break apart and form again, in different shapes, or different colours; they looped and swayed, they sprayed apart, they burst into showers of radiance that suddenly swerved this way or that like a flock of birds changing direction in the sky. And as Lyra watched, she felt the same sense, as of trembling on the brink of understanding, that she remembered from the time when she was beginning to read the alethiometer.

— Philip Pullman, The Subtle Knife (Scholastic, 1997; USA, Knopf, 1997)

Introduction

Digital technologies have repeatedly redefined the paper world of books. Digital printing has overhauled the publishing processes, and the internet has revolutionised the way audiences and authors connect to share their enthusiasm and criticism. Now the digitization of books themselves, either for searching, browsing, and reading on a computer screen through services like Google Books, or for reading on dedicated devices like Amazon’s Kindle or the Sony Reader are threatening the established order.

But for this project we side-step these issues and concentrate instead on how the analytical power and display capabilities of computers may be used to enhance our understanding of book texts. We use the term “book texts” rather than the word “books” as we are not trying to build computer systems that might understand books, but rather we use the computer’s ability to treat books as an abstract sequence of words as the starting point for new analytical tools.

Who would use such tools? Anyone with an interest in books, be they authors, readers, publishers, agents, critics, academics, etc may find such tools useful, but we have designed our visualizations with fans and academic readers in mind. These readers form theories about the books that stand alongside the author’s own understanding and we hope that the abstract visualizations provided may help such an endeavour.

Background and Related Work

The statistical analysis of texts is an important area of work and is used widely in information retrieval (e.g. web search). It is also a mature area of research in its own right, and has been used in the past for things from author attribution to the ordering of works through time. For example in a letter published in 1882 Augustus De Morgan speculated about using statistical techniques to explore authorship questions around St Paul’s Epistles and the Epistle to the Hebrews [Lea76], while more recently Jockers, Witten, and Criddle used sophisticated statistical techniques to reassess the authorship of the Book of Mormon.
In contrast, the abstract visualization of book texts is not a large or a mature field of study, but there are notable and inspirational examples. The following sections list some of these (more on separate tab)

**Visualizations**

Our work focuses on the abstract visualization of children’s book series, and in particular the trilogy “His Dark Materials” by Philip Pullman. Pullman’s trilogy is made up of the three novels “The Northern Lights” (called “The Golden Compass” in the USA and in the movie adaptation), “The Subtle Knife”, and “The Amber Spyglass”. We choose this genre partly through personal passion and partly because of the range of potential enthusiastic readers. The best children’s book series (especially before they are completed) are read and discussed by child and adult readers and many of these readers develop their own theories which they share with their friends and with other readers online. Similarly academic interest is piqued and there are conferences and journals dedicated to the study of children’s literature (more on separate tab)

**Future Work**

There are several directions we’d like to take this work in now.

**User studies**

We need to take these visualizations out of the research lab and engage both the fans and the academics who are theorising about Pullman’s works. We should engage them with these tools and establish if the tools are useful, how they might be improved, and what other visualizations may be of value to the community.

**Infer.Net**

Throughout this work we took the view that computers were not adept at understanding books, but should just essentially count words and draw the results for people to interpret. However advances in machine learning, and especially toolkits enabling machine learning techniques to be applied quickly to new domains have led us to seek to apply Infer.Net to the analysis phase of the visualization.

**Other visualizations**

Inevitably building visualizations leads to 1,001 other ideas as to how the data may be visualized. We would like to add the ability to pivot (e.g. for one flowers bud to open another flower side-by-side). We would like to add animations so that the dynamic movement between visualizations or as a visualization is formed is part of the semantics of the visualization itself.

**Online version**

The visualizations we made are not available for public use either online or through downloading. This is partly because we have not spent time looking at the rights implications and partly because we have not engineered the code to the quality level required for public use. It would be great to get this to a level where people can try the visualizations we built for themselves without us present.

**Other Books**

It would be interesting to apply this work to other children’s book series, to see if the characteristic patterns revealed in the visualizations were different from author to author. We might also move from a reader’s perspective to a learner’s perspective and choose books which often appear on high-school syllabuses. But most intriguing would be to build visualizations that contrast the content and style of different author’s work.

**References**
The statistical analysis of texts is an important area of work and is used widely in information retrieval (e.g. web search). It is also a mature area of research in its own right, and has been used in the past for things from author attribution to the ordering of works through time. For example, in a letter published in 1882 Augustus De Morgan speculated about using statistical techniques to explore authorship questions around St Paul’s Epistles and the Epistle to the Hebrews [Lea76], while more recently Jockers, Witten, and Criddle used sophisticated statistical techniques to reassess the authorship of the Book of Mormon [JWC08]. In contrast, the abstract visualization of book texts is not a large or a mature field of study, but there are notable and inspirational examples. The following sections list some of these

**Clarence Larkin’s “Dispensational Charts”**
Data visualizations fall into two overlapping camps: exploration and communication. Larkin’s 1914-1918 Dispensational Charts are about communicating scripture and prophecy from The Bible. They diagram the structure of each topic (e.g. “The Heavens” or “The Second Coming”) and use flow, representational images, and references back to Bible passages to illuminate each topic.

Text Arc

The seminal work of abstract exploratory visualization of book texts is Brad Paley’s ”Text Arc”. TextArc is a screen based application Paley designed and implemented that takes a text and displays it twice. Firstly, line by line in tiny font around the edge of a giant ellipse. And then secondly word-by-word with each word anchored by invisible springs to the sentences in which it occurs.

Common words are removed (so called ‘stop words’) and the remaining words are rendered so that more common words use a larger font and are drawn on top of any less common words sharing the same screen area. Paley’s TextArc can be used to explore any text but he often demonstrates it using Alice in Wonderland and then at the centre, in big letters, is the word Alice as that occurs throughout the book. TextArc has many other features, including an elegant dynamic path sweeping through the work as the text is read through.

Text Arc was conceived as a tool to help academics and other readers analyse texts. Another outlet proved to be selling high quality printouts as a beautiful memento of one’s favourite texts. The application of book visualization to academic literary studies has been continued in work like Plaisant et al’s “Exploring Erotics in Emily Dickinson’s Correspondence”.

Partly because of the widespread availability of electronic versions of the text, partly because of its cultural significance, and partly because of the huge numbers of people who care about it The Bible has proved an intriguing source of visualizations.

Anh Dang’s “Gospel Spectrum”

While on NYU’s Interactive Telecommunications Programme Anh Dang built “Gospel Spectrum”, an interactive visualization exploring the gospel accounts of Christ’s life. Each episode in Christ’s life is represented as a coloured bar with the colours representing the different gospels and their length representing the number of verses spent on that episode. The resulting visualization allows one to see how Christ’s life unfolds through the gospels: which gospels concentrate on which parts of his life, and when the gospels come together to record an episode.

Linda Becker’s “In Translation”

Started at Central Saint Martin’s School of Art, Becker’s “In Translation” shows visually the structural similarities and differences between different language translations of the Tower of Babel story, for example showing the position allocated to each letter-combination. “In Translation” both enforces the message of The Tower of Babel Story by highlighting the differences between human languages, but also cuts across it by showing structural similarities.

Chris Harrison’s “Bible Visualizations”
Chris Harrison’s visualizations of The Bible follow two paths. Firstly Harrison took a set of textual cross references found in The Bible compiled by Lutheran Pastor Christoph Romhild and displayed the links visually, resulting in a beautiful picture that gives detail about which chapters contain most cross references that also impresses the viewer with the sheer number of cross references. The second set looks at proper nouns through The Bible and overlays them as a tag cloud. But rather than abstracting the positions of the nouns from their occurrence in the text they are placed at their ‘centre of mass’.

Steinweber and Koller’s “Similar Diversity”

The last Bible visualization we’ll touch on is Steinweber and Koller’s ‘Similar Diversity’. Like Harrison’s work Steinweber and Koller use arc-diagrams and other visual features, but rather than using them to explore the structure within The Bible Similar Diversity shows the similarities and differences between holy books of different religions.

Before moving on to describe our own visual explorations of the text of Pullman’s His Dark Materials trilogy there are four other interesting book visualization projects that are worth drawing attention to because of other potential features they make use of.

Ebany Spencer’s “Romancing Dimensions”

In her CSM MACD project ‘Romancing Dimensions’ Ebany Spencer attempts to use purely visual notations systems to retell Edwin Abbott Abbott’s “Flatlands” story. Though entirely paper based Spencer’s work uses three dimensions by using paper cut-outs to move some of her time-line representations of the work out from the background plane.

Tim Walter’s “textour”

Tim Walter’s textour (in German) has uses time and animation to show the structural elements of the book accruing as data is added or filtered.

Stephanie Posavec’s “Writing Without Words”

Stephanie Posavec’s beautiful visualizations of Jack Kerouac’s “On the Road” (and some other contrasting novels) are not the result of a computer analysis of the work but the result of careful, loving, and painstaking analysis by-hand of the text itself. Posavec produces several visualizations, from the spider-like Posavec diagrams which map the sentence lengths authors’ use (a line continues for the length of the first sentence, then turns ninety degrees and continues for the length of the second sentence, etc) through to the elegant ‘literary organism’ flower like structures.

IBM Research’s Visual Communications Lab’s “Many Eyes”

Many Eyes is a social visualization site. It is social in many ways: users upload data sets that are immediately shared with all the other Many Eyes members; anyone can use any of the provided visualization tools to visualize the data sets; these visualizations can be shared and discussed on the Many Eyes sites, or embedded into blog posts to foster conversation and analysis beyond the site. Many Eyes was conceived, designed, and built by IBM Research’s Visual Communications Lab. It was originally thought that most of the datasets and visualizations would be based on numeric data, and so the visualizations were tailored towards quantitative data. In fact the inventors were taken aback by the amount of textual data sets uploaded, including notably The Bible and political speeches, and they have written about the text based visualizations designed and added in response [WV08].

Our work focuses on the abstract visualization of children’s book series, and in particular the trilogy “His Dark Materials” by Philip Pullman. Pullman’s trilogy is made up of the three novels “The Northern Lights” (called “The Golden Compass” in the USA and in the movie adaptation), “The Subtle Knife”, and “The Amber Spyglass”. We choose this genre partly through personal passion and partly because of the range of potential enthusiastic readers. The best children’s book series (especially before they are completed) are read and discussed by both child and adult readers and many of these readers develop their own theories which they share with their friends and with other readers online. Similarly academic interest is piqued leading to conferences and journals dedicated to the study of children’s literature.
Design Ideas

In order to narrow the design space Linda Becker and I decided to focus on two questions:

1. How the language used about different characters contrasts and how it changes through the series;
2. How linguistic themes (like religious language) are used through the series.

We started with Linda sketching out how some visualizations might look (without using actual data).

In the first of these Linda looked at the distribution of words (e.g. characters names) throughout the text, using connecting arcs (among other ideas) to give a sense of the rhythm of related characters through the text.

The second set of sketches looks at the character word plots, what form they might take and what visual dimensions this would give us to plot differing data or to reinforce existing data.

Thirdly Linda tackled the notion of themes, and the sketches she produced show how we might plot themes progression through the books. These are the sketches that we have had least success moving into functioning visualizations since they rely on a more sophisticated notion of theme than looking at individual word positions may provide.

The last series of visualization sketches Linda produced looked at text. Instead of drawing structures based on the relationships between words we looked at drawing the structures with the words themselves. This proved quite playful. I had wanted the visualisations to be legible themselves as text, but some of the sketches jump to the opposite pole, for example rendering only the words of interest and leaving the surrounding text as measured space.

The two ideas that we built up into working visualizations are the flower-like structures showing the words occurring near the characters names (or other given words) and renderings of the whole text with the character names of interest highlighted with colours and arcs.

Character Flowers

The first of the visualization ideas that we implemented were the character flowers. Figure 1 shows the character flower for the word Lyra. Central to the flower is the word “lyra” itself, surrounded by a ‘lifebelt’ which shows, starting from the 12 O’clock position, the occurrences of the word “lyra” through the series, with each occurrence resulting in a thin red line.

We can see from the number of crowded red lines that “lyra” is a frequently occurring word, as we would expect, but that the second and third books contain episodes where she is not mentioned. Moving out from that each â€”budâ€™ represents a word. Here we are looking at all the words which immediately follow the word “lyra” in a sentence. Those words are arranged in order of the frequency with
which they appear after “lyra”, and the size of the bud reflects the frequency of the word overall (i.e. the number of chapters it occurs in, regardless of whether it occurs after the word “lyra”). The final measure is the distance from the centre that the bud is drawn. This reflects the probability that when the word occurs it occurs after “lyra”. So we see two buds placed near the centre at the start are two words that occur frequently after the word “lyra” and are unlikely to occur elsewhere. Indeed the two words are Lyra’s surnames: Silvertounge and Belacqua. Other words drawn towards the centre are evocative of Lyra’s personality: “joyfully”, “quelled”, “exulted”, “definitely”, “judged”, “raided”, … but two stand out as anomalous: “blushed” and “obediently”. Clicking on the bud brings up the sentences in which the word follows the word “lyra”. From these sentences we find that the terms are used when Lyra is in disguise. In some respects this shows that the visualization works â€“ the anomalies are indeed anomalies, but they are ones consciously placed by Pullman, rather than subconscious ones.

![Character Flower of the word “Lyra’s” (click for larger version)](image)

Characters names can also be used in their possessive sense, e.g. “lyra’s” and the character flower in Figure 2 shows the diagram for the words after “lyra’s”. These are mostly body parts (les, arms, hair, etc) and this style is born out in Pullman’s writing about the other characters.

**Whole Text**

These visualizations show the entire text of the three volumes that make up the trilogy. We were interested to see the rhythm of the characters occurrences in the whole text, especially two related characters. Figure 3 shows a fragment of the entire trilogy, with linked coloured disks over occurrences of Lyra and Will’s names. We can quickly see simple facts like Will’s absence from the first book, and more curious aspects like the periods of the second book where neither of them are mentioned (presumably the sections focussed on Mary Malone, Lord Asreil, or Mrs Coulter). Printed out this diagram is many feet long, and the text itself is (just) readable. This combination of text level detail and global pattern is particularly interesting. I was hoping that this visualization would highlight a poetic choice across the trilogy. Tolstoy starts and ends “Anna Karenina” at a railway station, and Pullman purposefully opens the first book with the word “Lyra” and ends the last book with the word “Lyra”. This should stand out as the visualization should start and end with a coloured disk. But it does not. In fact Pullman precedes the start of his book with a quote from Milton’s “Paradise Lost”, which stops the poetic symmetry coming out in the visualization.
The initial sketches were built in Adobe Illustrator. Having chosen our two initial candidates for implementation these were prototypes in Processing, a language aimed at designers new to programming. Later these prototypes were re-worked into C# and WPF. The texts themselves were drawn from the publishers Quark documents, saved to plain text, broken down into chapters, sentences, and words in C# and stored in a SQL Server 2008 database.
Union Cabinet Approvals – August 24 2016

i. On August 24th 2016, Union Cabinet headed by Narendra Modi has given its approval for MoUs with Myanmar, Cyprus, Fiji and South Africa in different fields.

ii. The Union cabinet approved an MoU with Myanmar for cooperation in the field of traditional systems of medicine.

iii. The Bill will regulate surrogacy in India by establishing National Surrogacy Board at the central level and State Surrogacy Boards and Appropriate Authorities in the State and Union Territories. The legislation will ensure effective regulation of surrogacy, prohibit commercial surrogacy and allow ethical surrogacy to the needy infertile couples.

August 24th individuals are observant by nature and pleasant side to their being. Browse through this article for a list of famous people born on 24th August and also know more about their personality traits.


Donald Trump has repeatedly attacked Hillary Clinton over the Clinton Foundation. This is Trump speaking Tuesday in Austin. A new Washington Post analysis of Hillary Clinton’s August schedule finds that she has raised at least $32 million over the last three weeks through a series of exclusive fundraisers. The events, held in wealthy areas such as Greenwich, Connecticut, Nantucket, Massachusetts, and Beverly Hills, California, frequently cost upwards of $50,000 a ticket.

Guantánamo Bay Prisoner Abu Zubaydah Argues for His Freedom. Aug 24, 2016. In news on Guantánamo Bay, prisoner Abu Zubaydah has argued for his release at a rare hearing held via video stream Tuesday.