Comparing popular and formal scientific writing using Bill Bryson and Simon Peyton Jones as example

By Gabriel Wurzer

October 18, 2012

1 Introduction

Writing papers has always been a miracle. The most inspiring person for me in this respect is Simon Peyton Jones, a researcher working for Microsoft Research who has shown 'How to write a good research paper' [1], among other very interesting guidelines [2][3]. In short, Peyton Jones is arguing for a radically changed research process, in which the paper is not the final act but the first and initial deliverable after an idea. According to Peyton Jones, this approach "*opens the way dialogue with others: reality check, critique, and collaboration*", as "*writing the paper is how you develop the idea in the first place*" [1]. Peyton Jones' writing is not formal at all when it comes to verbal style: He prefers the active voice, clings to hands-on examples wherever possible, and generally employs a very narrative language. However, from a structural point, his arguments are always very clearly and rigorously laid out, i.e. in:

- An *Abstract* given in accordance to Kent Beck's four-sentence structure [4]: (1.) *State the problem* (2.) *Say why it's an interesting problem* (3.) *Say what your solution achieves* (4.) *Say what follows from your solution.*
- An *Introduction* driving the whole paper, which *describes the problem* using an example, then produces a *list of contributions* that forward-reference the sections where they are described in full detail.
- A *Paper Body* which concentrates on the problem and the idea for a solution, stating full details of the approach. *Related work* is deferred until before the *Summary*, since the reader does neither know nor care to hear about any other research being related, until all relevant details of the paper have been laid out.

Having happily applied exactly this way of writing since almost ten years, I recently stumbled upon a popular science book called 'A Short History of Nearly Everything' by Bill Bryson, in which the author manages to describe the history and development of the world (*Big Bang to Humanity, Humanity to Modern Science*) in just 573 pages [5]. Both writers have a very similar style of writing: They use a direct, example-centric narrative style to develop the story, which they drive through an introduction (typically at the begin of each chapter, in Bryson's case). The interesting question for me was: How do we then discern popular scientific writing from formal (conference style) papers? As both persons are similar from writing style *but not from their structure*, a closer look at both styles may help understand what lies at the core of the separation between formal and popular scientific writing, if such a difference is discernible at all.

2 Method and Comparison Study

As background work, several papers by Simon Peyton Jones have been studied according to the structural propositions given his advice presentation [1]. Using the proposed form of the different parts of the papers, we searched Bryson's book for examples of similar structures (Section 2.1). Vice versa, we looked at the narrative and argumentative structures in selected chapters of Bryson's book and compared these to Peyton Jones' papers (Section 2.2).

2.1 Comparing popular to formal writing

An *abstract* summarizes the situation, a problem, the proposed solution and its effect. Clearly, a popular science book is not necessarily problem-centered - it is story-centered! Thus, similar structures exist only to a limited extent (13 of 30 chapters) and with a different intention: If it is that the situation, the problem, its solution and what followed is given in that order, the reason is that the author wants them to be told in a straightforward manner. Here is a contrasting example using the *what follows* part as first narrative element:

> At just the time that Henry Cavendish was completing experiments in London, four hundred miles away in Edinburgh another kind of concluding moment was about to take place with the death of James Hutton. This was bad news for Hutton, of course, but good news for science as it cleared the way for a man named John Playfair to rewrite Hutton's work without fear of embarrassment. *Bill Bryson, A Short History of Nearly Everything, p. 90*

So, we might conclude that an author of a popular scientific text has at least $4! = 24$ possibilities for varying the order of *situation, problem, solution* and *consequences*, while a formally written abstract has to adhere to the usual fixed order. This does, however, not mean that this constraint implies also a formal narrative style:

> [Situation] Even Haskell programs can occasionally go wrong. Programs calling *head* on an empty list, and incomplete patterns in functional definitions can cause program crashes, reporting little more than the precise location where *error* was ultimately called. [Problem] Being told that one application of the *head* function in your program went wrong, without knowing which use of *head* went wrong can be infuriating. [Solution] We present our work on adding the ability to get stack traces out of GHC, for example that our crashing *head* was used during the evaluation of *foo*, which was called during

> the evaluation of *bar*, during the evaluation of *main*. We provide a transformation that converts GHC Core programs into ones that pass a stack around, and a stack library that ensures bounded heap usage despite the highly recursive nature of Haskell. [Consequences] We call our extension to GHC *StackTrace*. *"Finding the needle - Stack Traces for GHC", Allwood, Peyton Jones and Eisenbach 2009 [6].*

A further typical feature of an abstract is that it reflects the title or at least subtitle of the paper. Both Peyton Jones' work (e.g. [7,8]) and Bryson's book adhere to this standard. In the latter case, however, much of what we would call *abstract* is intermingled with the *introduction* - for dramaturgical reasons; for example, contributions (or discoveries) are identified first, before describing the situation, the problem and the consequences.

> In 1953 Stanley Miller, a graduate student at the University of Chicago, took two flasks - one containing a little water to represent a primeval ocean, the other holding a mixture of methane, ammonia and hydrogen sulphide gases to represent the Earth's early atmosphere - connected them with rubber tubes and introduced some electrical sparks as a stand-in for lighting. After a few days, the water in the flasks had turned green and yellow in a heavy broth of amino acids, fatty acids, sugars and other organic compounds. 'If God didn't do it this way,' observed Miller's delighted supervisor, the Nobel laureate Harold Urey, 'He missed a good bet.' *Bill Bryson, A Short History of Nearly Everything, p. 350*

Likewise, the *related work* is part of the story - not always covering the subject in a broad technical sense, but really commenting on the evolution of the discipline (in which, as Bryson's book repeatedly tells us, credit for an insight is seldom given to the right person):

> In 1900, three scientists working separately in Europe rediscovered Mendel's work more or less simultaneously. It was only because one of them, a Dutchman named Hugo de Vries, seemed set to claim Mendel's insights as his own that a rival made it noisily clear that the credit really lay with the forgotten monk. *Bill Bryson, A Short History of Nearly Everything, p. 470*

The feeling of uninformedness that Peyton Jones holds against mentioning the related work right away ("the reader knows nothing about the problem yet; so your [...] description of various technical tradeoffs is absolutely incomprehensible") is exactly what makes Bryson's book fun to read: As relations between persons and research topics slowly evolve, the reader is drawn into the flow of events. This slow process of uncovering an aspect of the story might seem patronizing at first. However, it is quite common even in Peyton Jones' work, who uses illustrations and examples right away to get the reader interested, and only then describes the general ideas associated with them.

So what is the 'meat' in a book chapter? Clearly, one does not easily find a problem/solution/evidence cycle as proposed by Peyton Jones for establishing that a work that has been done is novel. The ideas reported are neither novel nor do they need to be proven - they are *reported*, that is all! This is the most obvious difference, as I see it, between popular and scientific writing. Here is another account of events in this fashion:

> It is a remarkable fact that well into the space age, most school textbooks divided the world of the living into just two categories - plant and animal. Micro-organisms hardly featured. Amoebas and similar single-celled organisms were treated as proto-animals and algae as proto-plants. Bacteria were usually lumped in with plants, too, even though everyone knew they didn't belong there. As far back as the late nineteenth century the German naturalist Ernst Haeckel had suggested that bacteria deserved to be placed in a separate kingdom, which he called Monera, but the idea didn't begin to catch on among biologists until the 1960s, and then only among some of them. *Bill Bryson, A Short History of Nearly Everything, pp. 373-374*

Identifying key ideas and concepts is vital for both forms of writing. However, in contrast to formal scientific writing, popular literature might also use humouristic elements (see citations earlier) that would seem offensive to the scientific field: It is, after all, entertainment that the reader is after.

2.2 Comparing formal to popular writing

In order to reverse the view and look also at the *entertaining sides of a paper*, we first have to study the structure of a popular scientific book chapter.

2.2.1 Titles

A chapter has a *title* describing the contents of the chapter, which already acts as a mini-abstract of what is going to come (e.g. "How to Build a Universe", "Welcome to the Solar System", "Elemental Matters"). Such a clear statement is needed because the chapter itself can start with a topic that is not at all related to the actual subject, thus giving some tension which is later to be resolved. In fact, we find that there are papers that are far more creative when it comes to titles than popular science books (also see [9] for a discussion on this topic):

- G. Storms, P. De Boeck, I. Van Mechelen and W. Ruts, "Not guppies, nor goldfish, but tumble dryers, Noriega, Jesse Jackson, panties, car crashes, bird books, and Stevie Wonder," in Memory and Cognition Volume 26, Number 1, pp. 143-145, 1998.
- G. Proulx and J. D. Sime, "To prevent 'panic' in an underground emergency: why not tell people the truth?," in Fire safety science: 3rd international symposium (G. Cox and B. Langford, eds.), (London), pp. 843–852, Elsevier Applied Science, 1991.

In the case of the example authors, we can see a certain resemblance in the way the work is advertised. While Bryson packages the whole book, the unit of work for Peyton Jones is the paper (not individual sections, which have a plain heading):

- **Bryson**. A Short History of Nearly Everything, Shakespeare: The World as a Stage, At Home: A Short History of Private Life, The Mother Tongue: English and How it Got That Way, Made in America: an Informal History of the English Language in the United States.
- **Peyton Jones**. Report on the programming language Haskell: A non-strict, purely functional language, Scrap Your Boilerplate: A Practical Design Pattern for Generic Programming, Tackling the awkward squad: monadic input/output, concurrency, exceptions, and foreign-language calls in Haskell, Type Classes: An Exploration of the Design Space, C–: A Portable Assembly Language That Supports Garbage Collection

2.2.2 Starting off

There are numerous ways in which a chapter introduction can take the reader by the hand and introduce the subject. Specifically for Bryson's work, two forms are common: (1.) The "*Once Upon A Time*" form, which describes the time and place some discovery was made, and (2.) the *definition* form, in which some scientific discipline, event or entity is put onto the stage.

> ["Once Upon a Time":] In 1787, someone in New Jersey - exactly who now seems to be forgotten - found an enormous thigh bone sticking out of a stream bank at a place called Woodbury Creek. *Bill Bryson, A Short History of Nearly Everything, p. 109*
>
> [Definition:] Thank goodness for the atmosphere. It keeps us warm. Without it, Earth would be a lifeless ball of ice with an average temperature of minus 50 degrees Celsius. *Bill Bryson, A Short History of Nearly Everything, p. 313*

The second form is the one scientists commonly employ in their texts (except, perhaps, for Historians).

2.2.3 Defense by Citation

As Peyton Jones states, failing to cite can ultimatively kill a paper [1]: Either the writer does not know that the idea he is describing is old (which is bad), or he does know and pretends it is his own work (very bad). In what Bryson writes, citations are used in an additional sense:

> Humboldt [...] observed that there are three stages in scientific discovery: first, people deny that it is true; then they deny that it is important; finally they credit the wrong person. *A Short History of Nearly Everything, p. 508*

The author is using a citation to back up his point. The source of the citation is important - it is both *credible* and *well-known*. Citing a well-known person does not improve what is being said at all; it only raises the bar for critics. But beware: Hiding behind "some say..." is very easily observable, and should only be used in cases where these sources are disclosed as well:

> Some say [the convection process] begins four 650 kilometers down, others more than 3,000 kilometers below us. The problem [...] is that there are two sets of data, from two different disciplines [...] *A Short History of Nearly Everything, p. 271*

2.2.4 Illustration

> Protons give an atom its identity, electrons its personality. *A Short History of Nearly Everything, p. 184*

It is obvious that complex concepts are easier to grasp when illustrated using an example. However, it doesn't end there: The example itself must be interesting, witty and memorable. In the above quote, an analogy to purely abstract concepts is made. The example itself is not self-sufficient - the reader will ask "*How is that?*". What is needed is a (preceding or succeeding) narrative that clarifies that hydrogen is an atom with one proton, hydrogen one with two and so on, and the number of electrons determines how reactive an atom will be. Leaving some level of uncompleteness is a good trick in keeping the reader interested, if one can fill in the necessary details in due course.

Another form of illustration, mostly for statistical data, is done by mapping to a known reference system:

> If you imagine the 4,500 million years years of Earth's history compressed into a normal earthly day, then life begins very early, about 4 a.m., with the rise of the first simple, single-celled organisms, but then advances no further for the next sixteen hours. *A Short History of Nearly Everything, p. 408*
>
> The most striking thing about our atmosphere is that there isn't very much of it. It extends upward for about 190 kilometers, which might seem reasonably bounteous when viewed from ground level, but if you shrank the Earth to the size of a standard desktop globe it would only be about the thickness of a couple of coats of varnish. *A Short History of Nearly Everything, p. 313*

Clearly, what is missing to some extent is graphical illustration - as in Stephen Hawking's popular scientific works such as "The Universe in a Nutshell" [10]. We find that this is a real pity, since much of the geniality found in the preceding quotes can be expressed also graphically, using the same analogies [11] and many more that cannot easily be verbalized [12]. Peyton Jones' contributions in this respect are also limited ("Find out how to draw pictures, and use them" [1]), which is understandable, since he comes from functional programming where most information can be expressed textually.

2.2.5 Controversy and Reinterpretation

It is definitively a good idea to combine controversy and a consideration of all relevant arguments and sources:

> Taxonomy is described sometimes as a science and sometimes as an art, but really it's a battleground. [...state of the art...] *A Short History of Nearly Everything, p. 437*
>
> In 1907, or so it has sometimes been written, Albert Einstein saw a workman fall off a roof and began to think about gravity. Alas, like many good stories this one appears to be apocryphal. According to Einstein himself, he was simply sitting in a chair when the problem of gravity occurred to him. *A Short History of Nearly Everything, p. 162*

As a popular science journalist, it is far easier to step back and look at the whole picture. In purely scientific writing, as in Peyton Jones' case, argumentation must likely focus on justification of novelty and applicability, in a far more directed manner:

> A sceptical reader might reasonably ask whether the Par monad merely solves a problem that is unique to Haskell, namely laziness. After all, many of the difculties identied in Section 2 relate to reasoning about strictness, a problem that simply does not arise in a call-by-value language. True enough – but the strict-vs-lazy debate is a larger question than we address here; what we show in this paper is how to integrate selective control over evaluation order into a lazy language, without throwing the baby out with the bathwater. *"A monad for deterministic parallelism", Marlow, Newton and Peyton Jones 2011 [13].*

Both forms, popular and formal scientific writing, do not criticize authors or works directly (i.e. "To make my work look good, I have to make other people's work look bad" [1]).

3 Conclusions

Popular scientific literature and research papers are fundamentally different:

- Popular science reports on *discoveries that lie in the past*,
- Research argues for work that is yet unprecedented - and thus *aimed at the future*.

This being clear, some elements of argumentation can be leveraged when writing a paper: The **conscious use of time** - i.e. a non-straightforward way of reporting on a sequence of events, can help establish a narrative that is engaging. Bold, **memorable catch-phrases** can summarize an idea far better than a lengthy argumentation line, even though both will have to be present to be credible. An **active selection of sources** from which work is cited does also play a huge role in this respect.

As many other researchers, Simon Peyton Jones has proven through his use of language to be *'more than an academic writer'*, which is why we have chosen him for a comparison with a popular scientific book authored by Bill Bryson.

Such a direct comparison might seem awkward, and admittedly: Whether it is a good idea or not is left to the reader. What it has shown us is that, maybe, the rather strict formal guidelines to which we as scientists adhere might need some creative exploitation to make reading papers as enjoyable as reading a good book.

References

[1] S. Peyton Jones, "How to write a good research paper," Talk at Vienna University of Technology, October 2004. Slides available from: research.microsoft.com/en-us/um/people/simonpj/papers/giving-a-talk/writing-a-paper-slides.pdf, Video available from: http://www.wit.at/events/peyton-jones/writing_a_paper.ram [accessed 16.10.2012]

[2] S. Peyton Jones, J. Launchbury and J. Hughes, "How to give a good research talk," in SIGPLAN Notices Volume 28, Number 11, November 1993. Slides available from: research.microsoft.com/en-us/um/people/simonpj/papers/giving-a-talk/giving-a-talk-slides.pdf, Video available from: http://www.wit.at/events/peyton-jones/giving_a_talk.ram [accessed 16.10.2012]

[3] S. Peyton Jones and A. Bundy, "How to write a good research proposal," Talk in Moscow, April 2006. Slides available from: research.microsoft.com/en-us/um/people/simonpj/papers/giving-a-talk/writing-a-proposal.zip [accessed 16.10.2012]

[4] R. E. Johnson, K. Beck, G. Booch, W. Cook, R. Gabriel and R. Wirfs-Brock, "How to get a paper accepted at OOPSLA (panel)," in Proceedings of the eighth annual conference on Object-oriented programming systems, languages, and applications, p. 435, 1993.

[5] B. Bryson, "A Short History of Nearly Everything", Black Swan Books, 2004.

[6] T. Allwood, S. Peyton Jones and S. Eisenbach, "Finding the needle: stack traces for GHC," in Proceedings of Haskell Symposium 2009, 2009. Available online from: research.microsoft.com/en-us/um/people/simonpj/papers/stack-trace/DebugTraces.pdf [accessed 16.10.2012]

[7] D. Vytiniotis, S. Peyton Jones and J. P. Magalhães, "Equality proofs and deferred type errors: A compiler pearl," in Proceedings of ICFP'12, 2012.

[8] M. Sulzmann, M. M. T. Chakravarty, S. Peyton Jones and K. Donnelly, "System F with Type Equality Coercions, " in The Third ACM SIGPLAN Workshop on Types in Language Design and Implementation (TLDI'07), 2007.

[9] I. Sagi and E. Yechiam, "Amusing titles in scientific journals and article citation, " in Journal of Information Science, Volume 34, Number 5, pp. 680-687, 2007.

[10] S. Hawking, “The Universe in a Nutshell”, Bantam Spectra Publishers, 2001.

[11] N. Vossoughian, ”Otto Neurath: The Language of the Global Polis“, NAI Publishers, 2011.

[12] E. R. Tufte, “Envisioning Information”, Graphics Press, 1990.

[13] S. Marlow, R. Newton and S. Peyton Jones “A monad for deterministic parallelism,” in Proceedings of the 4th ACM symposium on Haskell, Volume 42, Number 12, pp. 71-82, 2011.