Plain English in Two Australian Organisations: 
Readability and Style Analysis

Baden Eunson

Abstract

Samples of text were obtained from the annual reports of two Australian organizations- the Bank of New South Wales (a.k.a. Westpac) and the Commonwealth Scientific and Industrial Organisation (CSIRO). Over historical time, these samplings comprise diachronic corpora or corpuses. Readability scores are considered, taking into account the strengths and weaknesses of such scores. Both corpora are measured for readability, with lower scores being used as a proxy for plain English style. Conclusions are reached about the genre of annual reports and also about the success or otherwise of the organizations in communicating with different audiences.

Keywords: annual reports; readability scores; diachronic analysis; corpora; sampling; passive voice.

Introduction

This paper examines publications of two Australian organizations to determine whether this philosophy of plainness of style has affected the writing style of the creators of those publications. The publications examined are the annual reports of the Bank of New South Wales (Westpac) from 1851 to 2001 and the Commonwealth Scientific and Industrial Organisation (CSIRO) from 1916 to 2001. The original research design encompassed readability analysis and corpus linguistics Multi-Dimensional Analysis (Biber, 1988; McEnery and Hardie, 2012). This paper will be confined to readability analysis.

Annual reports as texts: background and considerations of genre

As a type of document annual reports are very attractive sources of style variations over time. Perhaps all publications produced by organisations, they are the most likely to be chosen for purposes of research: they exist, by their very nature they represent a record over historic time, and they often purport to reveal the culture and nature of the organisation being reported on.

Several caveats need to be noted, however:

1. Merely because a publication exists does not mean that it is typical or capable of revealing insights into the activities of groups of individuals over time. Toynbee warned of the tendency for the potter to become the slave of his clay, or for historians to only concentrate on studying those societies and civilisations whose records were extant (Toynbee, 1972:30–34). Thus, organisational and management scholars speak routinely of ‘organisational culture,’ which may be a mind-set or system of beliefs and behaviour expressed through rituals, stories, material symbols and spoken as well as written language (e.g., Driskill and Brenton, 2005, Chapter 7; Eunson, 2012, Chapter 16), and annual reports
1. The obfuscation hypothesis. This hypothesis is that writers of annual reports will present ‘good news’ in easier to read writing, and mask ‘bad news’ with more difficult writing (Merkle-Davies, Brennan and McLeay, 2011) – that is, they will use ‘impression management’ and ‘spin’ techniques in style used in reports.

2. The Pollyanna hypothesis. This hypothesis is that the language used in annual reports will be positive regardless of the financial state of the company (Guillarmon-Saorin et al, 2009).

3. The readability-performance hypothesis. This hypothesis is that companies that perform well have more readable annual reports while companies that do not perform well have reports that are not as readable. (Subramanian, Insley and Blackwell, Rodney, 1993; Laksmana, Tietz and Yang, in press).

This is achieved through various writing techniques, as follows. One particular aspect among the discursive efforts in annual reports has received wide interest, viz. that of agency: it has been shown that sentences with active verbs and animate subjects tend to refer to the management
team if the news is good, as in ‘we have met our target’, whereas abstracts, nominalisations and passives abound if the performance has been poor, as in ‘pricing pressures and unusually hot weather combined to severely erode margins’ (Kohut and Segars 1992). Subramanian, Insley and Blackwell (1993) showed that writers who report profit write in clearer language with shorter sentences, whereas losses are found at the bottom of longwinded paragraphs (Bülow, 2011).

**Corpus design**

A corpus (Latin = body) is a collection of texts. As Meyer puts it, ‘For the purposes of this book, then, a corpus will be considered a collection of texts or parts of texts upon which some general linguistic analysis can be conducted’ (Meyer, 2002: xi). Meyer also suggests that corpus linguistics is not a discrete area of enquiry like sociolinguistics or psycholinguistics, but rather just a methodology for all areas of linguistic enquiry.

McEnery and Hardy give a bit more information on this process:

> Corpus construction. The process of designing a corpus, collecting texts, encoding the corpus, assembling and storing the relevant metadata, marking up the texts where necessary and possibly adding linguistic notation (2012: p. 241).

Historically, theoretical linguists such as Noam Chomsky have been indifferent or hostile towards corpus approaches, as early corpora were fairly primitive, and did not always produce insights into linguistic patterns (Fillmore, 1992). Fillmore humorously summed up the dichotomy as:

> …when the corpus linguist asks the theoretician (or ‘armchair linguist’) ‘Why should I think that what you tell me is true?’, the generative grammarian replies back ‘Why should I think that what you tell me is interesting?’ (1992:35).

Corpora can be synchronic (relating to the study of a language or languages as they exist at a particular moment in time, without reference to how they might change over time) or diachronic (relating to the study of a language or languages as they change over time. A diachronic corpus samples text across a span of time or from multiple time periods; see McEnery and Hardie, 2012).

**Plain English style**

Part of this project was to find out whether writers and editors of annual reports had been influenced by the ‘Plain English’ and ‘Plain Language’ concepts that have been salient now for several decades (see, e.g., Cutts, 2009).

Plain style can be meaningfully defined as comprising language that is characterised by:

- Shorter rather than longer sentences;
- An emphasis on simple sentences, less emphasis on compound and complex sentences, and least emphasis upon compound-complex sentences;
- A preference for active rather than passive voice;
- A preference for Anglo-Saxon words (which tend to be shorter) rather than Latinate or French words (which tend to be longer);
- Possible use of information document design such as layout, colour, bullet points and diagrams (Eunson, 1996).
A central aspect of such style is readability statistics, in that easy readability can be seen as a proxy for plain English style.

**Readability: What is it?**

When people refer to whether handwriting can be understood, they are talking about legibility. When people talk about whether the ideas expressed in writing can be understood, they are talking about readability. Readability can, within certain limits, be quantified, providing a score or statistic analysing the writing in terms of:

- Sentence length,
- Word length,
- Syllabic structure of words used,
- Punctuation.

Some readability tests break down the scores to also give a passive verb index for written passages. Passive verb constructions can lengthen sentences and make them harder to understand (ACTIVE: I gave you the book; PASSIVE: You were given the book by me). Most word processors now incorporate at least one type of readability score.

There are over 200 formulas or tests currently available (DuBay 2004). Here are a few of them:

- The FOG (Frequency of Gobbledygook) Index was developed by Gunning (1968) and measures the number of syllables and complex words (words that contain three or more syllables) in the selected text.

- The SMOG (Simple Measure of Gobbledygook) Index was developed by McLaughlin (1969) and estimates the years of education needed to understand a piece of writing using a simplified formula based on measuring words of three or more syllables in three 10-sentence samples. SMOG is widely used in checking health messages (see, e.g., Hedman, 2008).

- The Flesch Reading Ease test rates texts from 100 (very easy) to 0 (extremely difficult) (Flesch, 1949). Generally speaking, when we talk about high readability scores for a text or sample of text, we are talking about difficulty. Of course, with the Flesch Reading Ease score, the opposite is true (i.e. the higher the score, the easier it is to understand).

- The Flesch-Kincaid Grade test rates texts according to levels of years of education (based on American school grade levels 1–12) from 1 to 16.3 (or 4.3 years beyond the standard 1–12 year sequence) (Kincaid et al., 1975). Originally intended to rate technical documents, this is probably now the most widely used formula.

- The Dale-Chall formula is based on sentence length and a list of 3000 words that are familiar to most fourth grade students. That is, the longer the sentences and the more unfamiliar the words, the greater the chance of a passage having low readability (Dale and Chall, 1995).

- The CLEAR (Colors Label Ease for Adult Readers) developed by McLaughlin (2009), which attempts to colour-code text, based upon lists of difficult words. This tends to score higher in terms of readability years than other scores.

- The Lexile framework, developed by the US MetaMetrics corporation, gives a text a rating in terms of word difficulty and sentence length, on scales from 0 to 1700. Entire books are
sampled in slices, and then rated within a hierarchy of 100-unit divisions of increasing complexity. The theory is that such a hierarchy will motivate readers to tackle texts that are more complex. The Australian Council of Educational Research has worked with the US version to produce an Australian hierarchy (Scholastic, 1999).

The comparability of some of these readability scales can be seen in Table 1, where scores are given for Rachel Carson’s *Silent Spring*.

<table>
<thead>
<tr>
<th>FOG INDEX</th>
<th>FLESCH READING EASE</th>
<th>FLESCH-KINCAID GRADE</th>
<th>LEXILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.2</td>
<td>41.4</td>
<td>13.2</td>
<td>1340L</td>
</tr>
</tbody>
</table>

Table 1. Rachel Carson’s *Silent Spring* evaluated across four readability tests (Sources: Fog Index, Flesch Reading Ease and Flesch-Kincaid grades, retrieved from Amazon.com; Lexile ratings from MetaMetrics, Inc., www.lexile.com )

**Readability in Australia**

Regarding the current situation in Australia in terms of reading ability by years of education, the authors of a document on medical writing point out that:

[Based on the 1996 ABS Literacy survey] it can be seen that (at the time of the last survey) about 44 percent of Australians had prose literacy at or below the 6th grade level (PsychAssessment.com.au, 2006).

This may also have some interesting socio-economic implications, as pointed out by an Australian Bureau of Statistics report:

The median weekly income for people assessed with the highest level of prose literacy was $890 compared to $298 for those assessed at the lowest level. (Australian Bureau of Statistics, 2007).

**Readability statistics: Advantages**

Readability test scores and tools for measuring texts have proved very useful for many writers. They can help to refine writing to by providing a rough proxy for a more plain English style. Some of the many advantages of using readability tests and tools include:

They seem to provide a precise and scientific means of matching text to the capabilities of the reader or audience.

They make the writer more aware of the semantic (word choice) and syntactic (sentence construction) aspects of their own writing, and thus improve the writer’s self-editing capabilities.

They have helped to improve the general level of communication in fields such as business, medicine, law, and academia as writers in those fields apply readability concepts to their writing. That is, stockholders understand financial documents more easily, relatives of patients can more easily understand medical forms, citizens can understand regulations more easily, and academics can write more influential articles (Hartley et al, 2002; Hochhauser, 1999; Johnson, 2004; McKinley, 1998; Sawyer, Laran and Xu, 2008; Smith and Richardson, 1999).
They may have helped save lives and prevent injuries, as US research has shown that installation instructions on child safety seats in vehicles and online information about domestic violence were written at levels well above the average year 8 level recommended by the US government (DuBay, 2007; Yick, 2008).

Readability statistics: Disadvantages

Nevertheless, readability test scores have attracted a fair amount of criticism. Some of the disadvantages pointed out by the critics of using readability tests and tools include:

Most tests only take into account that which can be most easily measured: semantic (numbers of syllables of words or whether a word appears in a ‘familiar words’ collection), and syntactic (sentence length) factors. Other factors, such as content (propositions, organisation, coherence), structure (chapters, headings, navigation), and design (typography, format, illustrations) are not easily measured, and are thus ignored, even though they may play a critical role in whether a reader comprehends the meaning of a text.

Most readability tests necessarily concentrate on the text, not the reader, and not all readers are the same. Some readers, for example, may have prior experience in the subject matter of the text (Mikk, 2001), which means that they will comprehend the text content more effectively.

Similarly, the motivation patterns and intensity of the readers are not factored into most readability measures. As Fry (2002) observes, ‘What high school teacher has not seen a student with reading test scores near fourth-grade level sit for hours and read a drivers’ licence manual that has a readability of ninth grade?’ (Fry, 2002: 289).

Readability statistics: Diachronic problems?

As touched upon previously, text analyses of these kinds are usually either synchronic (relating to the study of a language or languages as they exist at a particular moment in time) or diachronic (relating to the study of a language or languages as they change over time). A diachronic corpus samples text across a span of time or from multiple time periods. The texts from other times may or may not be so radically different that they cannot be compared to those of another, much later period. If they are different kinds of text, probably they cannot be compared, but if the differences are of degree, then probably they can be compared. Thus, Biber, Conrad and Reppen (1998) have compared medical documents with drama scripts from 1650 AD to 1990 AD, and found meaningful linguistic differences of personal and impersonal style, which can be measured with such variables as pronouns, passive voice, conjuncts (such as ‘however’, ‘All told’, etc.), and adverbial subordinators (such as ‘because’, ‘while’, ‘when’, and ‘after’). Stajner and Mitkov (2011) use readability formulae to measure language changes in texts over a shorter period (1961–1991), while Heffernan et al (2010) looked at the influence of American spelling on Canadian spelling over a century.

‘Register’ refers to the type of language used in particular situations or for particular audiences. If diachronic analysis using readability formulas was contested, a case would have to be made that the styles and registers of other times were so radically different as to confound
comparisons. Yet, if we look at the lowest Flesch-Kincaid readability score (1906) for the Bank of New South Wales in Fig. 3 (below), we see this type of register:

It will be the duty of this Meeting to elect a Director in the room of the Hon. Reginald James Black, M.L.C., who retires by rotation, but is eligible for re-election, and has given the requisite notice that he is a candidate.

The meeting will also have to elect two Auditors in the room of Messrs. William Harrington Palmer and Frederick Wilson Uther, who now retire, the latter being eligible for re-election.

The Dividend will be payable at the Head Office to-morrow, and at the Branches on receipt of advice.

Fig. 1. Text from the 1906 sample of the Bank of New South Wales annual report.

This text has a Flesch-Kincaid score of 7.0, and, to the extent that readability is a robust proxy for clarity, plain English and comprehensibility, then the text is understandable. After all, Upton Sinclair’s *The Jungle* and John Galsworthy’s *The Man of Property* were published in 1906 and are still read today without any major comprehension barriers. Again, 1851, the year of the sample Bank of New South Wales text given in Fig. 2, was when Herman Melville’s *Moby Dick* and Elizabeth Gaskell’s *Cranford* were published – texts quite comprehensible and popular today.

**Readability statistics: Yes or no?**

There are thus advantages and disadvantages when considering readability as an aspect of style, but, since the cons do seem to outweigh the pros, the point of bothering with such scores seems arguable. On balance, they are worth considering, but only if the potential weaknesses of the scores is kept in mind.

The formulas have their defenders. Readability consultant Bill DuBay calls them ‘good enough’ and adds ‘they’ve been extremely beneficial to millions of readers’ (Bialik 2008). Weaver and Kintsch (1991, p. 242) take a more world-weary, but pragmatic view of things:

> The theoretical case against readability formula has been argued many times… Their merits are undeniable: at best, these formulas provide a scandalous oversimplification, more frequently a serious distortion. Nevertheless, as we have just shown, they continue to be used. It is the only game in town (emphasis added), and as long as modern research described earlier in this chapter does not provide viable alternatives, there is not much else to do.

**Method: selection and change in corpora**

The corpora built were those based on the annual reports of Westpac and CSIRO, from the private and public sectors respectively. Westpac was chosen because of its historical longevity. Established as the Bank of New South Wales in 1817 and as Westpac from 1982, it provides a potential corpus of a private sector organisation’s documents unparalleled in post-European settlement Australia (Holder, 1970; Van den Broek, 2011). The CSIRO (Commonwealth Scientific and Industrial Research Organisation) was chosen for similar reasons: established in World War I in 1917 as the Advisory Council of Science and Industry, it too provides a unique potential corpus of a public sector organisation’s documents (Collis, 2002).
The Bank of New South Wales’ annual reports from 1817 to the late 1830s were hand-written, and were mainly brief financial statements with little in the way of narrative text. They were typewritten from the mid-1830s onwards and it is not until 1851 that the typewritten text includes narrative discussion. I chose to begin the corpus from that time. Over time, particularly from the 1960s onwards, greater attention was paid to the annual reports as not simply information tools for shareholders but as professionally produced marketing tools for a number of stakeholders. This is consistent with global trends (Hopwood, 1996; Clarke, 1997). In particular, the chairman’s address emerges as a discrete and important part of the reports, together with the rest of the document, which tend to comprise not only numerical data but also narrative text describing operations and programs.

The CSIRO’s reports are typewritten and printed from the outset, including narrative text, and used primarily to describe the year’s activities in different scientific and technological departments or divisions. As with Westpac, there is a change in approach and information and document design over time, and there is also a change in organisational voice directed towards audiences. Whereas in earlier decades it seems that the reports are couched in a register or style influenced by scientific writing and addressed by their writers to scientific peers inside the organisation, in later decades the register changes to a more general one, addressing politicians and the general public, who can affect the flow of funding to a public sector organisation.

Samples or slices of corpora should be reasonably large, at least 500 words (Meyer, 2002), so I collected 1000 word samples of the Bank of New South Wales or Westpac annual reports and the CSIRO annual reports. As mentioned earlier, there are sampling problems caused by the changes in register and document design over historic time, which makes sampling difficult, but not impossible. Here, for example, are samples of the Westpac 1851 and 2001 reports:

<table>
<thead>
<tr>
<th>1851 SAMPLE</th>
<th>2001 SAMPLE</th>
</tr>
</thead>
</table>
| It will be gratifying to the Proprietors to learn that the establishment of branches of this Bank at Moreton Bay and Melbourne has been effected under the most favourable circumstances, with every prospect of success. The Moreton Bay branch has been in operation since the beginning of November last, and that at Melbourne is expected to commence business on the 9th instant. The separate accounts of these establishments are laid on the table for the information of the Proprietors. | Getting Real: Chief Executive Officer's Review  
People don't like banks. This is an unarguable fact that we can ill afford to ignore. That's why we have set ourselves a clear challenge: to front up and tackle the issues that our customers are most concerned about. We owe them a clear explanation and, most importantly, practical cost effective solutions. To restore trust we must demonstrate a genuine empathy for their needs. We must ensure that the revolution of our times works for all our customers and to the benefit of the whole community. Our commitment to meeting this challenge demands fresh thinking and decisive action. |

Fig. 2. Samples of 1851 and 2001 Westpac (Bank of New South Wales) corpora

Apart from the obvious historical differences in style, it needs to be noted that the 2001 text is part of a separate area of the annual report, the ‘Chief Executive Officer’s Review’, known in various examples of the report genre as the ‘Manager’s Report’, the ‘CEO Statement’ or the ‘CEO’s Letter’. Following world practice, these tended to emerge in reports from the 1940s.
onwards. This text or register tends to be more personal in style than the rest of the report (note the use of the pronoun ‘we’), and this needs to be accounted for in the sampling. In both Westpac and CSIRO corpora, I attempted to do this by sampling 500 words from the personal style of the CEO’s Review and 500 words from the rest of the narrative. While this may bias the results towards a personal style, such possible biases are partly negated by being replicated over enough time. There are no major unaccountable jumps in data or the graphic rendition of the data (see below) to suggest that this is a critical methodological weakness. In fact, the emergence of such different sub-registers within the report seems to have made these documents more accessible and readable.

Nevertheless, the disturbing thought remains that personal style (low readability scores) may be over-represented, and that in fact the totality of the annual reports scored even higher than is shown in Fig. 3 – perhaps skewing into the 20s and 30s on the Flesch-Kincaid scale. This would make the documents even less transparent, even less plain English in style and even less readable than the current simple curves show.

I did not attempt at this stage to test the Pollyanna, obfuscation or readability performance hypotheses, partly because (a) such testing would be irrelevant for a public sector organisation such as the CSIRO, and (b) it was not always possible to get an objective accounting of profit outcomes from the Westpac reports over historic time. However, using proxies for plain English style such as readability scores and passive verb indexes, it was possible to reach conclusions about success or failure in clarity of communication in the documents of both organisations.

I applied a number of readability indexes using Microsoft Word and the Readability Plus software package to yield data on Flesch and Flesch-Kincaid scales. Microsoft Word was also used to calculate passive verb indexes.

There is also a problem with data that is becoming slightly stale. The data was gathered in five-year-period samples, so the next year following this data set would have been 2006, and indeed 2011. However, given that most of the research took place before 2006, any samplings after that date might be problematic, because editors of the reports from both organisations had completed questionnaires before 2006, and thus might well have behaved atypically, because of Hawthorne effects, observer effects, and Heisenbergian uncertainty principle effects (Levitt and List, 2011; Reynolds and Schoech, 2012; Brown and Brown, 2011) and thus might produce compromised data.

Discussion

Bearing in mind the potential weaknesses of readability scores, the following can be observed.

The first years in the 1850s, the Westpac register is complex, but then the language seems to become less complex, reaching remarkably low scores at the turn of the century. It then begins to ascend rapidly, and stay at considerable levels of difficulty until the mid-1990s, when the text becomes easier to read, coinciding with time that the Plain English Plain Language concept was becoming popular in different genres of non-fiction writing. The spokespersons for Westpac and CSIRO both claimed to have been influenced by this concept (Fig. 3).
To the extent that the rationale of readability scores is correct, Westpac annual reports would have been read with difficulty by average readers for most of the 150 years sampled, although there may be a case for saying that shareholders as a stakeholder group may be more able to grasp complex text (see ABS data on reading skill and income).

CSIRO readability starts out high, and tends to remain high, only declining in the mid-1990s – a decline similar to that of Westpac. This would make sense if the authors or editors were influenced by scientific technical styles of writing, and if they perceived that their audience comprised other scientists and technologists rather than lay people (citizens, politicians) who might influence funding of the organisation.

The passive verb index is fairly low for Westpac (Fig. 4) for much of the sample period, but begins a decline slightly before the readability scores begin to decline, which may be a sign of increasing clarity of expression.

The decline of the passive verb index for CSIRO is spectacular, going from 54% in 1916 to 8% in 1996, with a slight uptick in 2001 (although that may be an artefact, or at least not statistically significant). This would be consistent with the change in approach over time, with the
The annual report is a major tool used by organisations to communicate with various publics. These include not only shareholders but also political and administrative actors, and the general public. Over the time considered, annual reports have evolved from simple financial reports with no narrative text to documents that not only feature substantial amounts of narrative expositional text but also opinion texts by chief executives. To the extent that readability scores are accurate and useful, it appears that the bank’s best days were in the latter part of the nineteenth century and the first few years of the twentieth, after which Flesch-Kincaid readability scores virtually doubled (Flesch scores mirror these trends with high early scores and later low scores). In fact, for much of the twentieth century, the bank’s annual reports are as poorly readable as those of a scientific and technical organisation. It may well be that the audiences for these documents were capable of decoding such demanding text, but it did not send out a good message about plain English text making basic concepts transparent and understandable by most, if not all, in the community. With the dramatic growth in stock ownership amongst the wider community, there may also have been a need to communicate more clearly with that new audience.

At the time of the strengthening of the Plain English Movement in the late 1990s, there is correlation, and perhaps causation, evident in the final few data points that an ethos of text simplification was in the minds of the creators of such documents. This was confirmed, or at least alleged, in interviews with current editors of both annual reports series, and can be seen anecdotally in the register of the Westpac Chief Executive Officer’s Review, ‘Getting Real’, Fig. 2 above. There may also be a factor of writing tasks being relinquished by technical specialists – in the bulk of narrative exposition separate from a CEO’s report – and taken up by professional writers, working with such technical specialists.

The diachronic passive verb index of Westpac roughly mirrors its diachronic Flesch-Kincaid score, while the diachronic passive verb index of CSIRO is nothing short of astonishing. It seems to reflect a real change in style, and does not appear to be merely a statistical artefact. Scientific writing is often characterised by heavy use of passives, Graeco-Roman lexis and nominalisations, but the debate about whether this was desirable (for example, was the use of the passive merely a way of masking identity and responsibility in discourse on scientific/technical processes?) did not really get under way – in print, at least – until the 1960s and 1970s (Weil et al, 1963; Wilkinson, 1992). Perhaps it was the need to communicate with small audiences of lay decision-makers and funding controllers that motivated writers and editors from the 1930s on to use more active voice, which would have had the usual effect of making discourse more direct and even shorter.

This certainly bears further investigation, as do the obfuscation, Pollyanna and readability-performance hypotheses in annual reports as a genre, as well as the general issues of organisational culture and discourse production raised earlier.

The methodological question of whether 50/50 sampling of CEO reports and general expositional text leads to a downward skewing of results, i.e., that the real reports are even more unreadable than Fig. 3 might suggest, must also be explored, by trying out new sampling models.
REFERENCES


Collis, Brad, 2002, Fields of Discovery: Australia’s CSIRO, Allen and Unwin, Crow’s Nest, NSW.


Eunson, Baden, 1996, Writing in Plain English, John Wiley and Sons, Brisbane.


Hochhauser, Mark 1999, ‘Informed consent and patients’ rights documents: a right, a rite or a re-write?’, *Ethics and Behavior*, vol. 9, no. 1, pp. 9–12.


