

Computer tools for preparing stylesheets

Paul Beverley

Archive Publications, paul@archivepub.co.uk

Abstract: When editing technical and scientific documents, a stylesheet is an essential tool for maintaining consistency, both internally within the document and with the publisher's house style. Paul Beverley has provided some free computer tools, in the form of Word macros, that will analyse a document and highlight potential inconsistencies. This allows the stylesheet to be populated early in the editing process, thereby saving time and allowing the editor to spot inconsistencies that might otherwise have been missed. The tools cover inconsistencies with hyphenation, spelling (including proper nouns) and accents, and certain aspects of punctuation and capitalisation.

Keywords: consistency, hyphenation, macros, spelling, stylesheets

This essay describes the use of some free computer programs that I have created to help editors with the process of creating a stylesheet, an essential tool for maintaining consistency, both internally within the document and with the publisher's house style. It is important to make these style decisions early in the editing process, and so using the computer's power to analyse your document is a very time-efficient way to assess the document's current (in) consistency.

The creation of this software has its origins in a situation that you might recognise. When editing a book, I noticed that in chapter 1 the author had mainly used 'nonlinear' but occasionally 'non-linear', so I applied 'nonlinear' consistently. Unfortunately, the frequency of 'non-linear' increased in chapter 2, but I continued 'correcting' it. By chapter 3, I could see that the author was now consistently using 'non-linear'. I then had to decide whether to persevere with 'nonlinear', or go right back through the first two chapters.

I realised then that making the effort to check this sort of thing *before* I started to read would save time in the long run. So over the past few years, I have developed a number of computer tools that check for several forms of inconsistency, and I make these tools available free of charge to fellow editors, in the form of Word macros, from: www.archivepub.co.uk/macros.

These are computer tools and so they can only follow rigid rules and point up discrepancies that might possibly be inconsistencies; it's up to you as the editor with an understanding of the vagaries of the English language to interpret the findings.

The analysis is done using five main macros, plus a number of smaller ones, covering possible inconsistencies with: (a) hyphenation, (b) the spelling of proper nouns, (c)

spelling generally, (d) the use of accents and (e) a range of other issues including punctuation and capitalisation.

Hyphenation issues

With the 'non(-)linear' example in mind, Table 1 shows an extract from a Word file that my macro *HyphenAlyse* generated for one book that I was editing.

Table 1: Part of a hyphenation frequency list produced by the *HyphenAlyse* macro

non-circular...1		
non-collinear...2	non collinear...2	
non-commutative...3		
non-commuting...2		
non-conforming...5		nonconforming...3
non-conservative...4		
non-equilibrium...4	non equilibrium...1	
		nonetheless...2
non-euclidean...1		
non-existent...1		
non-existing...1		
non-isothermal...1		
non-issue...3		
non-linear...5		nonlinear...207
non-linearity...1		nonlinearity...18
non-negative...6		
non-negativity...2		

First, the macro has listed *all* the hyphenated words in the whole book (Table 1 only shows a small extract) and has given me the frequency of each, but also if they appear as single words and/or as two separate words. In addition, it has also found the frequency of all the words that begin with anti, hyper, inter, non, etc, whether they are hyphenated or not.

I can look through this list and see what the author has done, but from the extract in Table 1, you can see that the macro tries to draw my attention to those words that appear both as a single word, and in hyphenated form (eg nonlinear/non-linear), because those are the words most likely to represent inconsistencies in the text.

By looking through this list I am able to see the author's (in)consistency as regards hyphenation and make decisions about whether some, all or none of the words using, say, the 'non' prefix should be hyphenated. I can then list these as part of a word list in my stylesheet – see Table 2. And having made these sorts of decisions before starting to edit chapter 1, time-wasting problems over hyphenation can be avoided.

Table 2: An extract from a word list, forming part of a stylesheet

hyper<word> – NONE are hyphenated
inter<word> – NONE are hyphenated except “inter-element”
iso<word> – ALL are hyphenated except “isoparametric”
non<word> – ALL are hyphenated except “nonlinear”
pre-image
pre-multiply
pseudo<word> – ALL are hyphenated
quasi<word> – ALL are hyphenated

Proper noun inconsistencies

The names of people and places are ripe for misspellings, and Word’s spell-checker is almost powerless to help. However, in one 150,000-word book, I found one mention of Siroishi, near the beginning, and then one of Sirioshi near the end. When I asked the author, “Is this the same person?” he was amazed that I had spotted the inverted vowel order. I didn’t tell him, but it wasn’t me that spotted it – it was the computer, specifically my macro *ProperNounAlyse*.

What this macro does is first to create a frequency list of all the words in the book that start with a capital letter, as shown in Table 3. As you can see, these first few words aren’t proper nouns. This is of course because the word at the start of every sentence has an initial capital, whether it’s a proper noun or not. But that’s OK because the macro then looks through the entire list, comparing the different words and using various different tests to find pairs of words, such as Siroishi/Sirioshi, that might possibly be variations of one another. It then highlights these word pairs, to catch my attention as I look down the list.

Table 3: The start of an example proper noun frequency list

Abbreviations ... 1
ABECAS ... 3
Academic ... 3
Academy ... 2
Accelerating ... 2
Access ... 2
Accessed ... 60
Accessible ... 2
Accident ... 5

Table 4 shows the sort of proper nouns pairs that I pulled out of the list. It’s true that some of these might in fact be entirely different people or places, and not misspellings, but at least I’m alerted and can check them.

Table 4: Some possible proper noun inconsistencies detected by *ProperNounAlyse*

Allain ... 1
Allen ... 1
Blas ... 1
Blass ... 1
Cresswell ... 3
Creswell ... 12
Ferret ... 1
Ferrett ... 1
Frances ... 1
Francis ... 1
Korcynski ... 2
Korczynski ... 2

Spelling inconsistencies

I have various macros to help with spelling, but the first two are only of any use if I’m not told what spelling conventions to use. To find out if the book predominantly uses UK or US English the *UKUSCount* macro will count how many specific UK-only and US-only words there are (eg labour/labor, pedalling/pedaling). Then if I am going to use UK spellings, the *IZISCount* macro gives me the frequency of -is-/-ys- as against -iz-/-yz- words.

Again for UK English, once I’ve made my is/iz decision, either *IStoIZ* or *IZtoIS* can be used to correct those words that need changing (or just highlight them if I prefer). But note that *IZtoIS* will not to change ‘prize’ into ‘prise’, and *IStoIZ* won’t give me ‘arize’. Each macro carries its own list of exceptions, and these lists are held as Word files, so you can add any specialist words for your own field of expertise.

The main spelling macro, *SpellingToolkit*, is really several spelling macros in one, and I can choose which bits to use and how I want to apply them for any given job. The fundamental aim is to combine the speed of Word’s spell-checking with the editor’s knowledge of the book’s context. So the macro starts by using Word’s spelling system to produce a single alphabetic list of all the different words that it *thinks* are errors. I can then look through the list and decide which words really are spelling errors – or might be, in context.

There are then various different ways to use this list of spelling errors: for example, I can get *SpellingToolkit* to change those words that are definitely errors and to highlight others that depend on the context. Users of my *FRedit* macro can do the same sort of thing by adding the words to their *FRedit* list. (For those not familiar with it, *FRedit* is a macro that does a sequence of global F&Rs, automatically, from a list that I give it, so I can make all those changes on each new chapter in a matter of seconds.)

Varying use of accents

My tools were written for my own use with the English language, but I gather (not being a linguist) that some of them can be used with documents written in other languages. What's more, at the suggestion of a group of editors that I met in Spain, I wrote the *AccentAlyse* macro, which looks for possible inconsistencies such as those shown in Table 5.

Table 5: A set of word pairs showing different accent use

Ängquist	1
Angquist	2
CIGRÉ	1
CIGRE	47
Dennetiere	1
Denetière	8
Dube	4
Dubé	4
edition	1
Édition	1
facade	1
façade	3
Gerin-Lajoie	1
Gérin-Lajoie	2
Saint-Germain-des-Prés	1
Saint-Germain-des-Près	1

Other inconsistencies

Finally, *DocAlyse* produces a frequency list of a whole range of different consistency issues, such as capitalisation (eg Chapter 6/chapter 6, Section 4.3/section 4.3), alternative spellings (eg focus(s)ed, co(-)operate, learnt/learned, etc),

serial (or not) commas, punctuation (e.g. p6/p.6/p 6/p. 6 or Eqn 2.1/eqn 2.1/Equn. 2.1/eqn. 2.1/Equation 2.1), and many others. Of course, some of these variations (possible inconsistencies) might not occur in a particular document, but the macro only lists those that do represent possible inconsistencies.

Other useful tools

One other important related macro is *MultifileWord*. This is used when your document arrives as a number of separate chapter files. Clearly, to do the analyses above, you need one big file of all the text from the whole document. *MultifileWord* allows you to look into a folder and combine the text from all (or a selection) of those files.

With very big files, some of the analyses can take quite a long time to run, so one way to speed things up is to use *MultifileText* instead of *MultifileWord*. This alternative macro creates a pure text version of the combined file which, since there is no formatting information, will be smaller and easier for Word to handle.

Worth the effort?

Using these macros does take time and effort – both initially to learn how to install and use them, and also to apply them to any given job, so only you can decide if any or all are worthwhile. Certainly, the longer the job the bigger the pay-back from the preparation time.

But regardless of the time saved, these tools enable me to produce a better standard of work (a higher level of consistency), which seems to me a good investment.

These macros are all available, including instructions, in the form of a book from my website (<http://www.archivepub.co.uk/TheBook>). If you are new to macros, you might like to try the 'Proofreader's Pack' (equally applicable to editors), which gives you very simple instructions for loading and installing a set of macros (including most of those mentioned here). These will get you off to a good start (www.archivepub.co.uk/documents/ProofingPack.zip).

And if you have other ideas for programs that could make you more efficient and effective, do please let me know – I'm adding more macros to my list all the time.

Paul Beverley has been an editor and proofreader of technical documents for over ten years. He's partly retired now, but doesn't want to stop altogether because he enjoys his work far too much.