TEX and Copyediting

SK Venkatesan[a] and CV Rajagopal[b]

[a]TNQ *Books and Journals Pvt Ltd, Dr Vikram Sarabhai Instronic Estate, Kottivakkam, Chennai 600041, India*

[b]*River Valley Technologies,* JWRA *34, Jagathy, Trivandrum 695014, India*

20 Jul 2014

1   Introduction

There can be many a slip between the cup and the lip in the publishing process. The manuscript that arrives in a modern publisher's office, usually as a LATEX or a MS word file, gets transformed bit by bit into a central XML form and then it is typeset into its final PDF form. It is a bit like smelting and purifying iron from in its raw form and molding it into the final finished product. Copyediting is a crucial step in the process and is receiving increasing attention now as the copyediting changes are being clearly indicated in the proofing process to the author.

Copyediting involves a broad range of activity: the accurate conversion of the initial input to XML; ensuring consistency of usage within the manuscript, correcting basic language and grammar, applying the finer aspects of the publisher's style, and placing XML hooks to ensure finer typographic aspects are taken care. The XML keeps the link alive between the present print-led world and future worlds such as HTML5. Copyeditors and XML form the bridge between these two worlds. However, there exist a lot of different ways in which TEX can be misused to make life difficult for a copyeditor [1] but we have come a long way from the earlier days when the technology was still underdeveloped [2]. LATEX's own secret little macros and TEX4ht have also made it easier to form this bridge between the two worlds.

Just as in all professions copyeditors also come from a long lineage of tradition. Copyediting tries to filter-out what it deems imperfections and inconsistencies in the manuscript and also ensures that author-reader communication is improved. Each publisher has an in-house style guide that has been refined over many years and forms the basis for copyediting. Our experience with different publishers has established that it is possible to design a generic set of TEX macros that can be used in the spirit of BibLATEX macros.

It should be mentioned here that these set of macros are not designed to replace copyeditors but to make it easier for them to take care of mundane aspects of copyediting in a systematic way, so that they will be able concentrate

on improving the crucial author-reader semantic communication aspects. Despite market trends that go in the reverse direction, the role of copyediting has never been more important in the present world with varied rendering devices, with different aspect ratios and modern semantic capabilities.

It should be mentioned here that these sets of macros are not designed to replace copyeditors but to make it easier for them to take care of mundane aspects of copyediting in a systematic way, so that they will be able to concentrate on improving the crucial author-reader semantic communication aspects. Despite market trends that go in the reverse direction, the role of copyediting has never been more important in the present world with varied rendering devices, with different aspect ratios and modern semantic capabilities.

2   Copyediting macros

Copyediting involves quite a broad spectrum of activity. At one end of the spectrum it improves semantic communication between the author and the reader. At the other end of the spectrum it reinforces certain stylistic and typographic conventions of the publisher. Semantic aspects are much beyond the capability of ordinary TeX macros, so it is at the latter end of the spectrum that most of this effort will be focussed.

We first attempt to list the copyediting process into various modular components:

1. Localization — British-American-Australian-Canadian

2. Close-up, Hyphenation, and Spaced words

3. Latin abbreviations

4. Acronyms and Abbreviations

5. Itemization, nonlocal lists and labels

6. Parenthetical and serial commas

7. Non-local tokenization in language through Abbreviations and pronouns.

3   Localization — British-American-Canadian-Australian

There are many sub-categories in British-American-Australian-Canadian variations:

3.1   DG (Am) versus DGE (Au, Br, Ca)

In American spelling, words like Acknowledge**e**ment and Judg**e**ment loose the e and become Acknowledgment and Judgment.

3.2 S (Am, Ca) versus Z (Br, Au)

In American and Candian spelling prefers ize, while Australian and British use ise spelling as in words like apolog**ize**/apolog**ise** or as in author**ize**/author**ise**. However, the rule is different for yze/yse patterns as in words like anal**yze**/anal**yse** although American prefer z but the rest use the British s.

3.3 C (Am, Ca) and S (Br, Au)

In words like defen**s**e/defen**c**e, offen**s**e/offen**c**e, the American and Candian prefer s instead of c

3.4 G (Am) and GUE (Au, Br, Ca)

In words like dialog/dialog**ue**, catalog/catalog**ue** Americans prefer to drop the ue

3.5 OR (Am) and OUR (Au, Br, Ca)

In words like color/colo**ur**, favor/favo**ur** Americans do away with u while teh rest keep the British spelling.

3.6 ER (Am) and RE (Au, Br, Ca)

In words like cent**er**/cent**re**, calib**er**/calib**re** Americans prefer the er spelling while the rest follow teh British spelling.

3.7 L (Am) and LL (Au, Br, Ca)

In words like cance**l**ed/cance**ll**ed, mode**l**ed/mode**ll**ed American prefer the single l spelling while rest prefer double l.

3.8 Others

There are also many other patterns and differences that doesn't fall into the above set of regular expression patterns and so these can only be handled by a word list with their language mapping table.

We use a very simple macro to care of all of this complexity: \vara{color} to take care of British-American-Australian-Canadian. The switch to particular language spelling can made by using:

```
\usepackge[lang=uk]{copyediting}
```

in the preamble. Both \vara{color} and \vara{colour} would produce the same output: colour, so the author's original need not be changed. The other options for language switch in this context: lang=uk,ca,au. The default language for the package is British spelling. The exceptions when one wants to force a

particular use in a particular instance one should use: `\vara*{analog}` as this starred macro will leave the input unchanged as analog.

4   Close-up, Hyphenation, and Spaced words

Although American spellings use less hyphenation, the modern preference for closed prefixes has a few exceptions:

1. if the root word is a proper noun or a number (post-Depression, pre-2001)

2. for double prefix (non-self-governing)

3. if the prefix precedes a proper open compound then ndash is used (pre–Civil War)

4. if two instances of the letter i or the letter a are adjacent (anti-intellectual, extra-action), or another combination of letters that could hamper reading (pro-labor)

5. for a double prefix (anti-antibody)

6. for a repeated prefix with implicit use (over- and understimulation)

However, many house styles have their own preferences, which can be dealt with the starred macros. We use the macro: `\hyp{anti}{body}` to hyphenate a compound word and for a close-up word we use: `\closeup{anti}{body}` for compound words that occur as two separate words: `\sword{Civil}{War}` You might wonder what use is such macros in a LaTeX file? They give visibility to the corrections the copyeditor makes and offers hooks to produce a global inventory of various changes while at the same time making it convenient to make switches on a global scale.

5   Latin abbreviations

Latin abbreviations such as:

cf.	compare
et. al.	and others
etc.	and so forth
e.g.	for example
i.e.	that is
NB	note
viz.	namely

is quite straight-forward to handle using macros: `\lat{et al.}` where the stylistic aspect will be taken care by global switches such as:

```
\usepackge{copyediting}[lat=0,abbr=italic]
```

The default `lat=0` leaves the text as it is and italic sets it to italic style. The other option `lat=1` removes all the dots and `lat=2` sets the value to its English equivalent shown above.

6 Acronyms and Abbreviations

Depending on if the initial letter abbreviations are spoken together as a word, as in AIDS (Acquired Immune Deficieny Syndrome), the term acronym is used but we will not make this distinction here and treat them as one and the same. A simple macro: `\ac{AIDS}` is good enough and the default global switch will ensure that it is expanded correctly the first time. The mapping between the acronym and its expansion is declared the first time as:

```
\newacro{AIDS}{Acquired Immune Deficieny Syndrome}
```

However many standard acronyms would be available by default from the package and only new acronyms need to be added this way. This can be checked during compilation.

7 Itemizations and nonlocal lists and labels

In many cases where there are only a few instances of a list we tend to use like in this example:

First, this is an endangered species;

Second, humans find them delicious;

Third, they are only found on this island.

In this example we could have as well have used, first, second, third, instead of `*ly`, making that a global option. It is also possible that this can be changed into a standard arabic numeral list: `1)... 2)... 3)...` etc. In order to keep the possibility to make such changes with a simple switch one can use macros:

```
\begin{eitem}
  \item this is a endangered species;
  \item humans find them delicious;
  \item they are only found on this island.
\end{eitem}
```

If we run LaTeX the third and final time then there is an option to change the last item in the list to lastly but that's a global switch:

```
\usepackage{copyediting}[eitem=0,last]
```

where the `eitem=0` is the default option that causes firstly, secondly... and last indicates that the last item should be lastly. If `eitem=1` is set then ly drops out and for `eitem=2,3,..`, it switches to standard enumerated and bulleted list.

eitem=0, last=true

Firstly, this is an endangered species;

Secondly, humans find them delicious;

Lastly, they are only found on this island.

eitem=1, last=true

First, this is an endangered species;

Second, humans find them delicious;

Last, they are only found on this island.

eitem=2

1. this is an endangered species;

2. humans find them delicious;

3. they are only found on this island.

eitem=3

- this is an endangered species;

- humans find them delicious;

- they are only found on this island.

eitem=4, last=true

This option and the succeeding one will make the list in paragraph mode instead of the usual vertical list. Also, the semicolon at the end of each item and the and connector at the end of penultimate item will be automatically added.

> Firstly, this is an endangered species; secondly, humans find them delicious and lastly, they are only found on this island.

eitem=5, last=true

> First, this is an endangered species; second, humans find them delicious and last, they are only found on this island.

8 Parenthetical and serial commas

Many long sentences are difficult to read and can be communicated better with parenthetical constructs or footnotes rather than commas. It would be nice to have switches that can make this change. For example:

> The enthusiastic young ducks flying in front of the group `\pc{led by the sugecious older ones at the back, make a lot of noise and turbulence}` which are used by older ones at the back to warm their heart and the wings.

would outputs to:

> The enthusiastic young ducks flying in front of the group, led by the spacious older ones at the back, make a lot of noise and turbulence, which are used by older ones at the back to warm their heart and the wings.

Depending on global switch `pc=0,1,2,3` or `4` we have the option of choosing parenthetical comma, parenthesis, emdash, a footnote or a sidenote.

See the output when `pc=1` (in parentheses):

> The enthusiastic young ducks flying in front of the group (led by the spacious older ones at the back) make a lot of noise and turbulence, which are used by older ones at the back to warm their heart and the wings.

See the output when `pc=2` (between emdashes):

> The enthusiastic young ducks flying in front of the group — led by the spacious older ones at the back — make a lot of noise and turbulence, which are used by older ones at the back to warm their heart and the wings.

See the output when `pc=3` (as a footnote):

> The enthusiastic young ducks flying in front of the group[1] make a lot of noise and turbulence, which are used by older ones at the back to warm their heart and the wings.

See the output when `pc=4` (as a marginpar):

> The enthusiastic young ducks flying in front of the group make a lot of noise and turbulence, which are used by older ones at the back to warm their heart and the wings.

led by the spacious older ones at the back

---

[1] led by the spacious older ones at the back

8.0.1 Elist

For a list of items such as: Suddenly warblers, tits, wrens, and hummingbirds started singing in chorus from the bushes. . . we turn them into:

> Suddenly `\elist{warblers,tits,wrens,hummingbirds}` started singing in chorus from the bushes. . .

This will be tranformed into:

> Suddenly warblers, tits, wrens and hummingbirds started singing in chorus from the bushes. . .

This macro will help bring consistency across the document regarding the placement of comma before and after and in the last item and in ensuring proper white-space after the comma.

9 Non-local tokenization in language through Abbreviations and pronouns

In a sequence of minimization operation, in typical news column the copyeditor addresses:

> His Holyness, the Prince of Mangoistan addressed a gathering of ordinary mangoes in the capital New Mango. The Prince of Mangoistan pointed out the serious threat of foreign insects in the country. He further pointed out the precautionary methods taken such as the use of organic insect repellants like the neem leaves and cow-dung to keep the country free of foreign pests. . .

His Holyness the Prince of Mangoistan shrinks to The Prince of Mangoistan and then finally to He. This copyediting operation can be denoted using:

```
\definetoken{mango}{His Holyness, the Prince of Mangoistan}
    {The Prince of Mangoistan}{He}
```

at the first instance and then just `\Token{mango}` at the later instances. The `\Token{mango}` macro would be quite useful to even just indicate what the important pronouns link to in a paragraph. However, not all pronouns have corresponding original objects as in the case of it in It is raining.

10 Interactive proofing

The above set of macros bring certain level of transparency and consistency to the copyediting process. Using additional macros, this also has the potential to convey further the key aspects of copyediting to the author using menus and dashboards, bringing certain interctive aspect to the proofing process.

11 Conclusion

We have made an attempt at bringing together many copyediting aspects as LaTeX macros. This involves some amount of drastic simplification and abstraction that may or may not work in all cases. The starred macros could be used in those cases where one needs to escape the global switch. The non-local linkages work just as in the case of bibliography links by multiple compilation of LaTeX that passes information through auxiliary files. Of course, this is only a small step towards the Himalayan task of climbing the semantic hill through LaTeX macros as envisaged by SenseTeX [3].

Acknowledgements

We would like to thank Lorna O'Brien for important inputs on English language and its varied usages across countries and publishers. Of course, this work would not have been possible without the constant encouragement of Mariam Ram, TNQ and C.V. Radhakrishanan, River Valley Technologies.

References

[1] E. Gregorio (2005) Horrors in LaTeX: How to misuse LaTeX and make a copy editor unhappy, *TUGboat* **26**(3), 273–279.

[2] P. Flynn (1993) TeX and SGML: A Recipe for Disaster? *TUGboat* **14**(3), 227–230.

[3] S.K. Venkatesan (2005) Moving from bytes to words to semantics. *TUGboat* **26**(2), 165–168.