



Experience from a Real-World Application of Micro-Typography with pdfT_EX

Hàn Th_ế Thành

River Valley Technologies hanthethanh@gmail.com

KEYWORDS micro-typography, margin kerning, font expansion, optimal page-breaking, semi-automatic pagination, automatic problem detection, paragraph stretching/shrinking.

ABSTRACT This article describes the experience from a real-world application of micro-typography with pdfT_EX. The project involved typesetting a study edition of the Bible in Czech, where there was a lot of further information apart from the original text: footnotes, references, further annotation etc. The design was not complex, but the typographic requirements were very strict. This posed a real challenge to typesetting everything without conflicts. The biggest problem was how to achieve perfect page-breaking without changing a single word of the text. The solution was to use a semi-automated method: manual breaking of each page, with auto-detection of problematic pages (like orphan/widow, too little/too much space between body and footnotes, etc.). When such a problem occurred, it had to be fixed by changing the length of one or more paragraphs on that page until the problem disappeared. Since we are not allowed to modify the text, the length of a paragraph could be changed only by changing the formatting of the paragraph (`\looseness`), in order to make it longer or shorter. This usually results in poor-looking paragraphs. However, with the aid of micro-typography provided by pdfT_EX, this was achieved without loss of quality.

1 Introduction

In 2000, a publisher in Czech Republic needed to re-publish a study edition of the Bible (only New Testament). The previous editions were typeset using Ventura. The publisher however was not very satisfied with the result, and wanted to achieve better result. They heard that a program called T_EX can do that job nicely, so they started looking for someone willing to typeset the Bible in T_EX. A T_EX fellow introduced me to the publisher. At that time the micro-typographic extensions of pdfT_EX were introduced and this was a perfect application for the new extensions. It had quite a good impact on pdfT_EX development, since during this project many issues were found and then were fixed or improved. This was perhaps the first serious application of micro-typography in pdfT_EX and was an important encouragement to me that the micro-typographic extensions can be really useful in practice.

In 2007 the publisher re-published the next edition of the Bible, and I did the typesetting again. Many things were easier than the previous time. However some problems remain difficult in principle and had to be solved with the “old tricks”. In this

article I would like to share the experience from the project, with the hope it might be interesting for the reader, and might be useful for someone who has to solve similar problems.

2 The workflow from the publisher's view

The workflow from the publisher's view (Figure 1) is rather simple and does not involve T_EX. The publisher maintains primary data in an in-house format which is not known to the typesetter. The data are exported to XML and sent to the typesetter. Whatever the typesetter does with the XML, the publisher needs not know: only the PDF result is relevant.

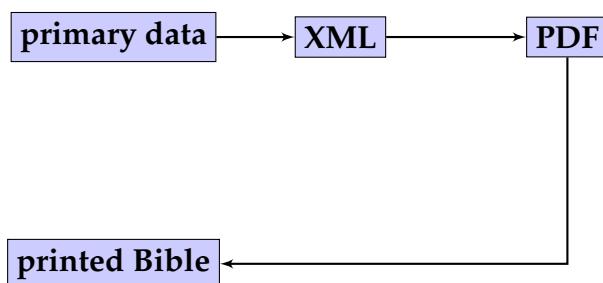


FIGURE 1. The workflow from the publisher's view

The choice of XML as the format to exchange data was a very good decision, since it frees the typesetter from knowing anything about the data format of the publisher. The publisher has their own tools to maintain and manipulate the data. The data format might be either very simple or complex, the tools might be either very reliable or buggy, but it is irrelevant to the typesetter. A DTD was developed for the project, and all what matters is whether the XML data from the publisher pass the validation against that DTD. If not, the problem is on the publisher's side and it is the responsibility of the publisher to check the data/tools and re-generate the XML data. If yes, then both sides know that the XML data are correct and if the PDF output is wrong, it is the typesetter's fault. So, the DTD acts as a contract for data exchange between the publisher and the typesetter. It helped to find out quickly who is responsible if a problem arises. Indeed, during the project many problems have been detected very early thanks to XML validation and that saved both sides from further communication problems.

3 The workflow from the typesetter's view

The workflow from the typesetter's view is more complicated (Figure 2).

When the typesetter receives the XML's, the following checks have to be done:

Validation: as mentioned above, validation is the first thing to do when the typesetter gets the XML data. If validation fails, the typesetter informs the publisher about the problem and waits for the updated data. The publisher is of course supposed

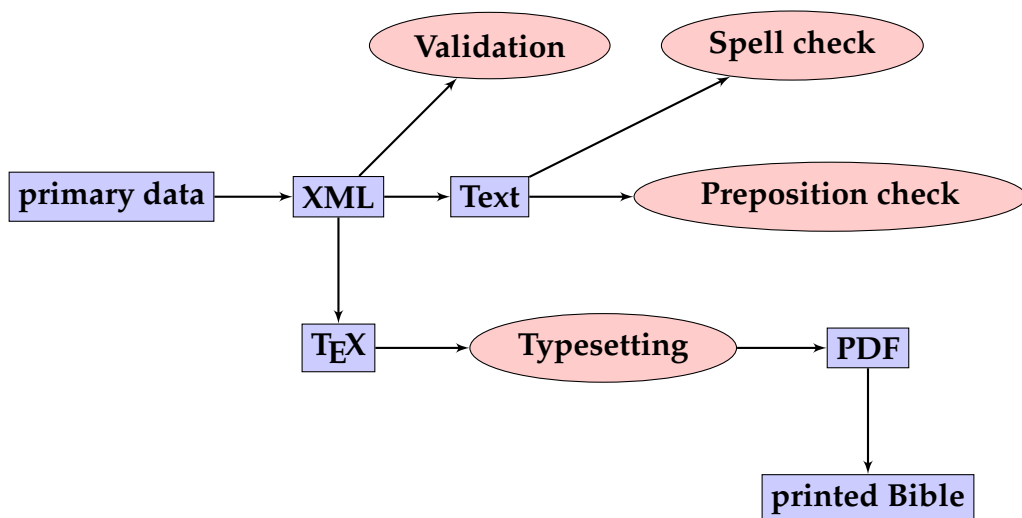


FIGURE 2. The workflow from the typesetter's view

to validate the XML data before sending them, but the typesetter should not rely on it. Instead, the typesetter must check and ensure the XML data are valid before processing further.

Text processing: the next step is to convert XML data to text to check for possible problems with the text contents:

Spell check: it might look a bit unlogical that spell checking is done here. Spell checking should be done on the publisher's side within the primary data format and not by the typesetter. However, during the typesetting a few misspelt words were found. Since it is not much work for the typesetter to do a spell check, it was added to the process just as a doubled check.

Preposition check: the czech language has many single-letter prepositions: **k**, **o**, **s**, **u**, **v**, **z**. Those prepositions should not end up at the end of a line. The conventional way to deal with those single-letter prepositions is to use a non-breakable space after them (~). The publisher is responsible for ensuring that tilde is used instead of space in such cases. However it happens from time to time that a tilde is missing. For that reason, another doubled check was added for this case.

If there is any problem in this stage, the publisher is informed about the problem and must send again the XML data.

Conversion to TeX: if the XML data pass the previous checks, they are converted to TeX, ready for typesetting.

All the checking steps above are done by some scripts, so anytime the XML data are updated, the typesetter can tell immediately if there is any problem and provide feedbacks to the publisher. If no problem is found, the TeX data are then used for the next stage: typesetting.

4 Typographic requirements

The layout of this typesetting is rather simple, as it can be seen in Figure 3.

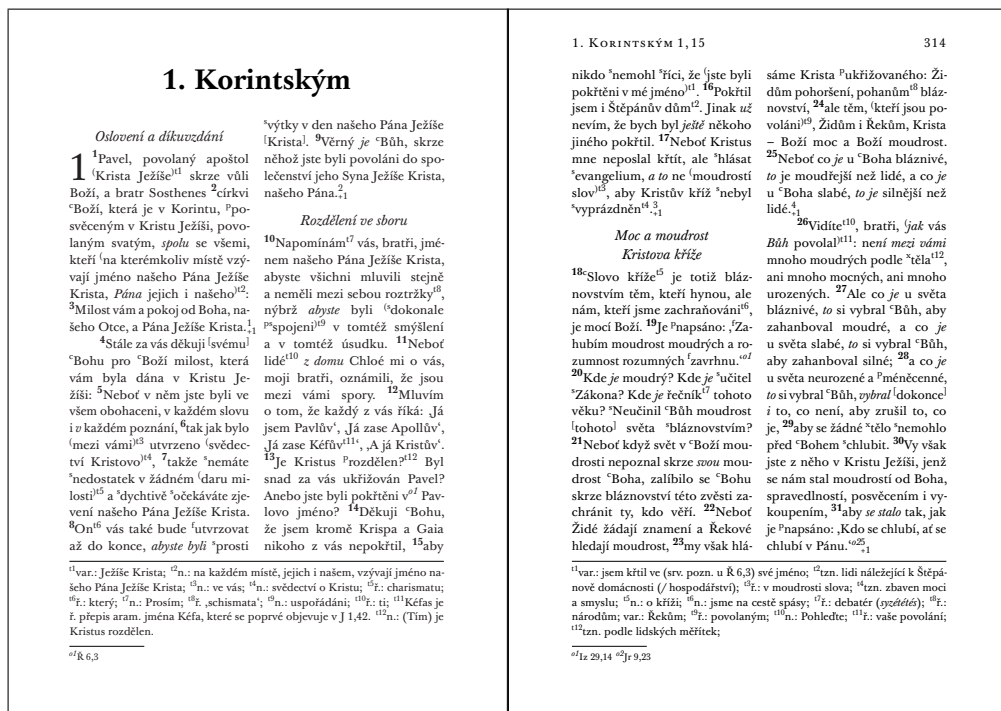


FIGURE 3. The sample layout

However, there are some requirements that are not that visible from the sample:

- Text contents must not be altered at all. As the XML data are converted to T_EX, such data must be used for typesetting without a single manual change.
- The line spacing is fixed (i. e. all lines must sit on a “grid”).
- The text contents are rich with many elements apart from the main text: subtitles, footnotes, references, various marking to aid the reader, etc.
- There must not be any typographic “sin”:
 - no widow,
 - no orphan,
 - chapter number must not end at page bottom (no “chapter orphan”),
 - not too much or too little gap between footnote and body,
 - the position where footnote ends is fixed,
 - always balanced columns.
- If there is any last-minute change in primary data, it must be reflected in the PDF quickly.

The conventional \TeX approach to solve problem with undesirable page breaks (to avoid widow/orphan) is to insert penalty at relevant places. However, this approach mostly leads to page breaking with very big gap between footnotes and body. To achieve optimal page breaking within that so restrictions, human decision and intervention is required.

5 How to achieve optimal page breaking?

When a document needs to be paginated optimally, the typesetter usually has to do two main things:

1. ensure that every problem is found (i. e. do not overlook any issue);
2. for a given issue, find an optimal way to solve the problem.

Solving a problematic case is usually done by changing the formatting of various elements around the problematic place. For example, the typesetter can change the gap around a figure, a paragraph, or a subtitle. Another example is to change the page dimensions (i. e. `\enlargethispage` in \LaTeX), or the gap between body and footnote, etc. In difficult cases where the above didn't work, the typesetter might have to change the formatting of a paragraph, or even the wording.

When we apply this approach to our case, it means that we must not miss anything from the issues mentioned in Section 4 and for each of them, find an optimal way to solve the problem.

However, in our case the possibility to change the formatting of an element is very limited: since grid typesetting is required, we cannot change spacing around any subtitle or paragraph. The gap between body and footnote cannot vary too much. Therefore, the only thing we can adjust is to change the formatting a particular paragraph, i. e. to make the formatting of a paragraph longer or shorter. Fortunately, \TeX provides a parameter to control this: `\looseness`. This is a \TeX parameter which means roughly the following: when `\looseness=1`, \TeX should try to format a paragraph to be one line longer than the default formatting. For example, if a paragraph is formatted to be 10 lines long by default, setting `\looseness=1` would cause the formatting of the paragraph to be 11 lines long.

Our approach to tune page breaking relies heavily on use of `\looseness`: when a problematic case is found, we change `\looseness` of relevant paragraphs to solve the problem. We can see this as a way to "stretch" or "shrink" a paragraph: stretching a paragraph is done by setting `\looseness=1`, and shrinking a paragraph is done by setting `\looseness=-1`. Setting `\looseness` does not always result in the desired effect: sometimes there is no way to make the formatting of a paragraph longer or shorter. So, a paragraph can be "stretchable" or "shrinkable", depending on whether changing `\looseness` gives the desired effect.

To illustrate how it works, let's look at an example. In Figure 4, there is a widow line. To avoid this widow line, the optimal way is to make some previous paragraph shorter so that the widow line would be moved to the left column. We do so by "shrinking" the second paragraph on the previous page, which means that we set `\looseness=-1` for that paragraph. The effect of this change is shown in Figure 5.

<p>1. KORINTSKÝM 2, 16</p> <p style="text-align: right;">316</p> <p style="text-align: center;"><i>Projevy nezralosti</i></p> <p>3 ¹A já, bratři, jsem k vám nemohl mluvit jako k duchovním, ale jako k ²tělesným, jako k nedospělým v Kristu. ³Dal jsem vám ⁴pit mléko, pokrm <i>jem vám nedat</i>, <i>neboť ten jste ještě nemohli snést</i>. Ale ani teď ještě nemůžete, ⁵neboť jste <i>stále ještě</i> ⁶tělesní. Vždyť pokud je mezi vámi žárlivost, rozbroj ⁷a rozdělení⁸, zdali nejste ⁹tělesní a ¹⁰nežijete jen po lidsku¹¹? ¹¹Když jeden říká: Já jsem Pavlův¹² a druhý: Já Apollův¹³, nejste <i>jako jiní lidé</i>¹⁴? ¹⁵Kdo je Apolos? A kdo je Pavel? Služebníci, skrze něž jste uvěřili, jak každému dal Pán. ¹⁶Já jsem zasadil, Apolos zalil, ale ¹⁷Bůh ¹⁸dával ¹⁹růst. ²⁰Tedy ani ten, kdo sází, ani ten, kdo zalévá, nejsou ²¹něčím <i>zvláštním</i>, ale Bůh, který dává růst. ²²Ten, kdo sází i ten, kdo zalévá, jsou jedno, každý však ²³dostane <i>svou vlastní mzdu</i> podle své²⁴ námahy. ²⁵Neboť jsme Boží spolupracovníci; ²⁶vy jste Boží pole, Boží stavba,²⁷ ²⁸podle ²⁹Boží³⁰ milosti, která mi byla dána, jsem jako</p> <p>³¹ř: nechodíte podle člověka; ³²var.: tělesní; ³³z.: vlastní; ³⁴z.: se odhaluje; ³⁵ř. přístavěl (na ten základ); podobně ve v. 10 a 12; ³⁶var.: zůstává; ³⁷n.: mezi vámi; ³⁸ř.: kterou; ³⁹zsn. podle měřiček tohoto světa;</p> <p>⁴⁰ř. 28,16</p>	<p>317</p> <p style="text-align: right;">1. KORINTSKÝM 4, 13</p> <p>hem bláznovstvím. Vždyť je ¹napsáno: On² chytá moudré v jejich chytráctví.³ ⁴20 A opět: Pán zná myšlenky moudrých a ví, že jsou marné.⁵ ⁶21 A tak ať se nikdo nechlubí lidmi. Vždyť všechno je vaše: ⁷22 ať Pavel nebo Apolos nebo Kéfas, ať svět nebo život nebo smrt, ať ⁸věci ⁹přítomné nebo budoucí – všechno je vaše, ¹⁰23 vy pak <i>jste</i> Kristovi a Kristus Boží.¹¹</p> <p style="text-align: center;"><i>Služebníci Kristovi</i></p> <p>4 ¹Ať o nás každý² smýšlí jako o služebnících Kristových a správcích Božích tajemství. ³A od správců se nakonec vyžaduje, aby byl každý shledán věrným. ⁴Pro mne je ⁵pramálo ⁶důležité, zda mě posuzujete vy nebo lidský soud⁷. Ale ani já sám sebe neposuzuji. ⁸Ničeho, ⁹co by <i>svědčilo proti mně</i>, si sice ¹⁰nejsem vědom, ale tím nejsem ¹¹ospravedlněn; ten, kdo mne posuzuje, je Pán. ¹²Proto nic nesude předčasně, dokud nepřijde Pán, který ¹³osvítí ¹⁴věci ¹⁵skryté ve tmě a ¹⁶zjeví úmysly srdcí. A tehdy se každému dostane pochvaly od</p> <p>¹⁷Boha.¹⁸ ¹⁹Toto jsem, bratři, vztáhl na sebe a <i>na</i> Apolla kvůli vám, abyste se na nás naučili ²⁰smýšlet! ²¹ne nad to, co je ²²napsáno, abyste se kvůli jednomu <i>učiteli</i> nenadávali jeden nad druhého. ²³Vždyť kdo ²⁴ti dává vyniknout²⁵? ²⁶Co z toho, co máš, ²⁷ji nedostal? A když jsi <i>to</i> dostal, proč se chlubit, jako bys <i>to</i> nedostal? ²⁸Už jste ²⁹nasyceni, už jste zbohatli, bez nás ³⁰jste ³¹začali ³²kralovat. ³³Kéž byste kralovali, <i>ale tak</i>, abychom i my ³⁴kralovali ³⁵spolu s vámi. ³⁶Zdá se mi, že nás, apoštolů, ³⁷Bůh postavil <i>jako</i> poslední, jako ³⁸odsouzené ³⁹na ⁴⁰smrt, neboť jsme se stali divadlem světu, an⁴¹đlům i lidem. ⁴²My jsme blázni kvůli Kristu, ale <i>vy jste</i> rozumní v Kristu; ⁴³my <i>jme</i> slabí, vy však silní; ⁴⁴vy slavní, my ⁴⁵bez ⁴⁶cti. ⁴⁷Až do této hodiny hladovíme a ⁴⁸žízňame, jsme ⁴⁹šora nazí, ⁵⁰dostáváme ⁵¹rány, jsme ⁵²bez ⁵³domova, ⁵⁴opotimne se ⁵⁵prací vlastních rukou. ⁵⁶Když ⁵⁷nám ⁵⁸spilají, ⁵⁹žehnáme, ⁶⁰když ⁶¹nás ⁶²pronásledují, ⁶³snášíme ⁶⁴to, ⁶⁵když ⁶⁶jme ⁶⁷hanění, ⁶⁸domlouváme. ⁶⁹Stali jsme se ja-</p> <p>¹ř.: Ten, který; ²ř.: člověk; ³ř.: den (tzn. zasedání tribunálu; srv. 3.13: zde den určený lidmi); ⁴ř.: té odlišuje (od ostatních);</p> <p>⁵ř. 5,13 ⁶ř. 29,11</p>
---	---

FIGURE 4. A problematic case: a widow line at the right page.

Of course this approach is very time-consuming and usually works only for short documents. Applying this approach to a book of a few hundred pages is not trivial and requires some tricks to speed up the process.

6 A semi-automatic approach

The way to tune page breaking in the previous paragraph can be improved in two places:

- problem detection: checking the result by eye is very tiresome and it is easy to overlook problems. It would be much better if we get a report of problems, saying for example *there is a widow line on page 317*.
- adjustment of paragraph formatting: we need an easy way to find which paragraph can be tweaked, i. e. to be formatted longer or shorter.

The workflow looks as in Figure 6.

To be able to refer to a paragraph, we number all paragraphs and typeset the number of each paragraph during tuning phase. In the final run they are left out.

6.1 Automatic problem detection

The method to detect problematic cases is to use a trick called absolute positioning. pdfT_EX allows recording the absolute position of a point using the primitive

<p>1. KORINTSKÝM 2, 16</p> <p style="text-align: right;">316</p> <p style="text-align: center;"><i>Projevy nezralosti</i></p> <p>3 ¹A já, bratři, jsem k vám nemohl mluvit jako k duchovním, ale jako k ²tělesným, jako k nedospělým v Kristu. ³Dal jsem vám ⁴pit míleko, pokrm <i>jem vám nedal, neboť ten</i> jste ještě nemohli <i>snést</i>. Ale ani teď ještě nemůžete, ⁵neboť jste <i>stále</i> ještě ⁶tělesní. Vždyť pokud je mezi vámi žárlivost, rozbroj ⁷a rozdělení¹, zdali nejste ⁸tělesní a ⁹nežijete jen po lidsku¹¹? ⁴Když jeden říká: Já jsem Pavlův¹ a druhý: Já Apollův¹, nejste <i>jako jiní lidé</i>?¹² ⁵Kdo je Apollós? A kdo je Pavel? Služebníci, skrze něž jste uvěřili, jak každému dal Pán. ⁶Já jsem zasadil, Apollós zalil, ale ⁷Bůh ⁸dával ⁹růst. ⁷Tedy ani ten, kdo sází, ani ten, kdo zalévá, nejsou <i>něčím zvláštním</i>, ale Bůh, který dává růst. ⁸Ten, kdo sází i ten, kdo zalévá, jsou jedno, každý však ⁹dostane <i>svou vlastní mzdu</i> podle své¹³ námahy. ⁹Neboť jsme Boží spolupracovníci; <i>vy jste Boží pole, Boží stavba</i>.¹⁴ ¹⁰Podle ¹⁵Boží¹ milosti, která mi byla dána, jsem jako</p> <p>¹¹Ě: nechodíte podle člověka; ¹²var.: tělesní; ¹³Ě: vlastní; ¹⁴Ě: se odhaluje; ¹⁵Ě: přistavěť (<i>na ten základ</i>); podobně v. 10 a 12; ¹⁶var.: zůstává; ¹⁷n.: mezi vámi; ¹⁸Ě: kterou; ¹⁹zsn. podle měřítka tohoto světa;</p> <p>¹¹1z 28,16</p>	<p>1. KORINTSKÝM 4, 13</p> <p style="text-align: right;">317</p> <p>psáno: „On¹¹ chytá moudré v jejich chytráctví.“¹² ²⁰A opět: Pán zná myšlenky moudrých a ví, že jsou marné.“²¹ ²¹A tak ať se nikdo nepoloží jiný základ nežli ten, který je již položen,¹ a tím je Ježíš Kristus. ¹²Jestliže někdo na <i>(tomto)</i> základě staví ze zlata, stříbra, drahých kamenů, dřeva, sena <i>nebo slámy</i>,¹³ dílo každého se stane zjevným. Ten den <i>je ukáze</i>, neboť se zjeví¹⁴ v ohni, a oheň vyzkouší dílo každého, jaké je. ¹⁴Jestliže někdo <i>na tomto základě</i> vystaví¹⁵ dílo a ono mu zůstane¹⁶, dostane odměnu, ¹⁵jestliže <i>mu</i> jeho dílo shoří, ¹⁶utrpí ¹⁷škodu; sám se sice zachrání, ale jako skrze oheň. ¹⁶Nevíte, že jste Boží svatyně a že ve¹⁷ vás bydlí ¹⁸„Duch Boží“? ¹⁷Ničili někdo ¹⁸Boží svatyni, ¹⁹zničí ²⁰Bůh jej. Neboť ²¹Boží svatyně je svatá, ²²a tou <i>matyní</i>¹⁸ jste vy.¹⁹ ¹⁸At nikdo neklame sám sebe. Jestliže si někdo mezi vámi myslí, ¹⁹že ²⁰je v tomto věku¹⁹ moudrý, ať se stane bláznem, aby se stal moudrým. ¹⁹Neboť moudrost tohoto světa je před ²⁰Bohem bláznovstvím. Vždyť je ²¹na</p> <p style="text-align: center;"><i>Služebníci Kristovi</i></p> <p>4 ¹At o nás každý² smýšlí jako o služebnících Božích tajemství. ²A od správců se nakonec vyzaduje, aby byl každý shledán věrným. ³Pro mne je ⁴„pramáto“ důležité, zda mě posuzujete vy nebo lidský soud³. Ale ani já sám sebe neposuzuji. ⁴Ničeho, <i>co by svědčilo proti</i> mně, si sice ⁵nejsem vědom, ale tím nejsem ⁶ospravedlněn; ten, kdo mne posuzuje, je Pán. ⁷Proto nic nesudte předčasně, dokud nepřijde Pán, který ⁸osvítí i věci ⁹skryté ve tmě a ¹⁰zjeví úmysly srdcí. A tehdy se každému dostane ¹¹pochovaly od ¹²Boha.¹¹</p> <p>¹Ě: Ten, který; ²Ě: člověk; ³Ě: den (tzn. zasedání tribunálu; srv. 3:13; zde den určený lidmi); ⁴Ě: tč odlišuje (od ostatních); ⁵může zn. ty, kteří nesou nečistotu světa (Lv 16,10); ⁶Jb 5,13 ⁷22,94,11</p>
--	--

FIGURE 5. Solving the problem by making a paragraph shorter.

\pdfsavepos and its friends. To see how it can help to find problematic cases, let's have a close look at widow detection. A widow line is the last line of a paragraph that ends up at the beginning of a page or column. We detect widow lines as follows:

1. For each paragraph, record the absolute position of the last line of the paragraph. It can be done by appending the primitive \pdfsavepos to all relevant paragraphs.
2. For each page, record the absolute position where the page body starts.
3. For each page, compare the absolute position of the page start against the absolute position of the last line of each paragraph. If they are close enough, then a widow line is found.

In the example in the previous section, we have the following lines recorded:

```
\bodytoppos{317}{28146752}
\endparpos{11}{317}{27423153}
```

The first line was generated for the body top of page 317, the second line was generated for the last line of the paragraph number 11 on page 317. The last numbers in those commands are the absolute positions (in scaled point) of corresponding elements. Since they are close enough (28146752 - 27423153 = 723599 = 11pt), this indicates that there is a widow line from paragraph 11 on page 317.

Other problematic cases are detected in a similar manner.

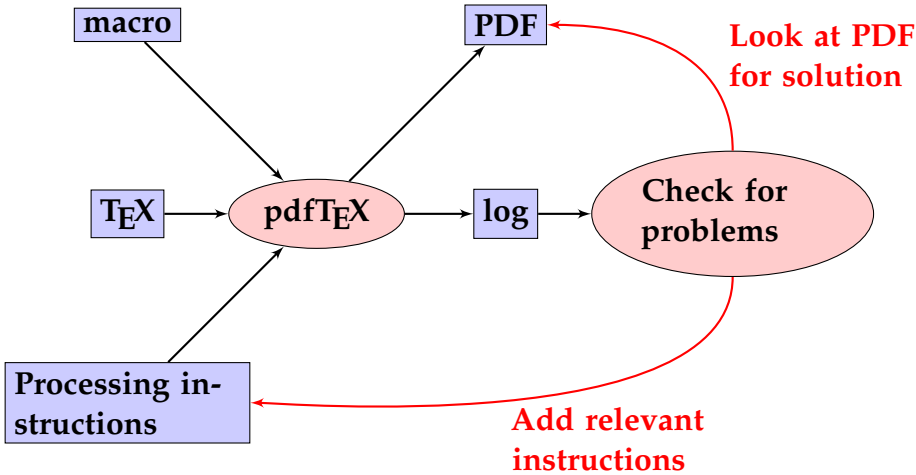


FIGURE 6. The process to tune page breaking

6.2 Adjustment of paragraph formatting

Once a problematic case is detected, we need to tweak some paragraph to solve the problem. Let's continue with the previous example. To make the widow line disappear, we have two choices:

- move the widow line to the left column, i. e. to “shrink” one of the previous paragraph by one line; or
- move another line from the left column to the right column, so that there are 2 lines of the paragraph 11 on the right column. In this case, we need to “stretch” one of the previous paragraphs.

The question here is to find which paragraph can be “shrunk” or “stretched”. Since we already number all paragraphs, this turned out to be simple: together with the paragraph number, we also show whether the paragraph can be “stretched” or “shrunk”. If a paragraph can be stretched, a sign +1 is shown below the paragraph number. If it can be shrunk, -1 is shown too. For example, the paragraph containing the widow line in our example is paragraph number 11, and it can be stretched (+1 is shown) but cannot be shrunk (-1 is not shown). Usually most of paragraphs can be stretched, but only a few can be shrunk.

Having all these numbers displayed, finding a suitable paragraph to tweak is quite easy. In our example, we see that paragraph number 9 can be shrunk, so we add a command saying “shrink the paragraph number 9 by one line” and rerun pdfT_EX. The result is shown in Figure 5.

This tuning process must be done as the last step, since any change in the data might require retuning page breaking.

7 Use of micro-typography

7.1 Margin kerning

Margin kerning is heavily used in this project. This is essential to make the margins look smooth, since too often it happens that non-letter elements ends up at the beginning or the end of a line. Those elements include:

- hyphen (very frequent, due to narrow columns),
- punctuations,
- footnote marks and references,
- verse number,
- various text marks.

How to apply margin kerning to hyphen, punctuations and other characters is not difficult and is described well in pdfTeX manual, microtype manual and various articles. However it is not clear how to apply margin kerning to elements that are not a character from the main font, so it is worthwhile to take a close look at such an example.

Suppose we want to apply margin kerning to the verse numbers. Those numbers often end up at the left margin. Since those verse numbers are typeset as superscript at a smaller size than the body, their shape actually take less “ink” than a regular character. So when they end up at the left margin, they will cause the visual effect that the actual line looks slightly shorter than other lines. If we leave out margin kerning for them, the left margin would look a bit uneven. However, we cannot apply margin kerning to those verse numbers in the same way as for hyphen or punctuations, since a verse number is typeset as a (raised) hbox, and margin kerning applies only to characters.

We make use of the fact that the amount of marginal kerning for all verse numbers are approximately the same. So, our trick is to insert before each verse number a “virtual” character. Such a character has zero dimensions and no “shape”. This can be done using a special virtual font, where each character definition is empty. Since the character is “empty”, it is not visible but we can apply margin kerning to such a character and therefore to the element after it, namely the verse number in this case. Other superscript elements are handled in a similar manner.

About 50% of lines has right margin kerning and 20% of lines has left margin kerning.

7.2 Font expansion

Font expansion is even more important than margin kerning for this task. Since we rely mainly on tweaking `\looseness` to stretch or shrink a paragraph, it is important that there are as many stretchable/shrinkable paragraphs as possible. Font expansion gives us exactly what is needed in this case: it gives more room to line-breaking, therefore makes a paragraph more likely stretchable or shrinkable.

Also, when a paragraph is stretched or shrunk without font expansion, usually it looks very ugly due to loose spacing between words. Font expansion is critical to compensate this: it makes a stretched or shrunk paragraph look reasonable.

Font expansion was used at the limit $\pm 2\%$, using `autoexpand` and without tuning expansion factor of individual characters. Going further than this limit is not recommended, since the effect of font expansion might be visible in some cases.

8 The result

The job took about 3 days to tune page breaking for about 500 pages. There are only 2 places in the book where there is an orphan line in the left column, since it is impossible to avoid them without breaking other rules. The total number of processing instructions (to stretch or shrink a paragraph) is about 250. About 15 paragraphs were shrunk, the rest was stretched. Almost all requirements are fulfilled, and the publisher was satisfied with the result.

9 What is still imperfect

Despite our effort, the typesetting was not perfect. We will take a quick look at issues that could be improved.

Hyphen at page breaks: it was not required by the publisher to prohibit hyphenation at page breaks, however it is better if we could do so. It is possible to detect such a problem automatically. Unfortunately it is not easy to fix such problem without changing the text at all. Due to lack of time, this issue was left unsolved.

Subtitles break at undesired places: Some subtitles are longer than one line and hence must be broken into 2 lines. However, sometimes they are broken at undesired place. This cannot be auto-detected, and cannot be fixed without understanding the actual subtitle. So, this issue must be fixed in the primary data by the publisher.

Last line in paragraph too short: we did put no restrictions on the length of the last line in a paragraph. Due to frequent stretching/shrinking paragraphs, there are many cases when the last line of a paragraph contains a single word (or even worse, a fragment from a hyphenated word). One could argue that this is tolerable for narrow columns, but it is better if we could have avoided it.

10 Conclusions

- T_EX is still a very powerful typesetting system if properly used. It is not suitable for every application, however in certain domains it is really hard to beat T_EX, despite the fact that it is a 30-years old program.
- The micro-typographic extensions of pdfT_EX can be very helpful in extreme cases, where we need to give a little extra flexibility to paragraph formatting.
- High-quality typesetting requires human decision and intervention.