The Asian Journal of TEX, Volume 2, No. 1, April 2008
Article revision 2008/2/18

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

Hàn Thế 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 pdfTEX. 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 pdfTEX, 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 TEX can do that job nicely, so they started looking for someone willing to typeset the Bible in TEX. A TEX fellow introduced me to the publisher. At that time the micro-typographic extensions of pdfTEX were introduced and this was a perfect application for the new extensions. It had quite a good impact on pdfTEX 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 pdfTEX 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 TEX. 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.

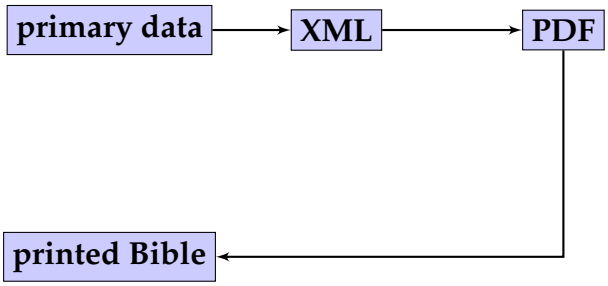


FIGURE 1. The workflow from the publisher's view

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

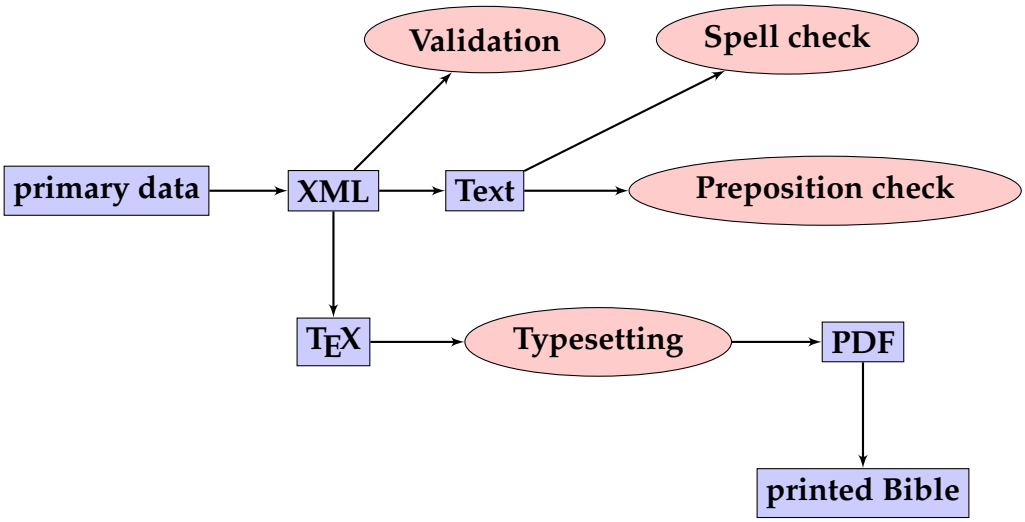


FIGURE 2. The workflow from the typesetter's view

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 looks 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 missing. For that reason, another doubled check was added for this case.

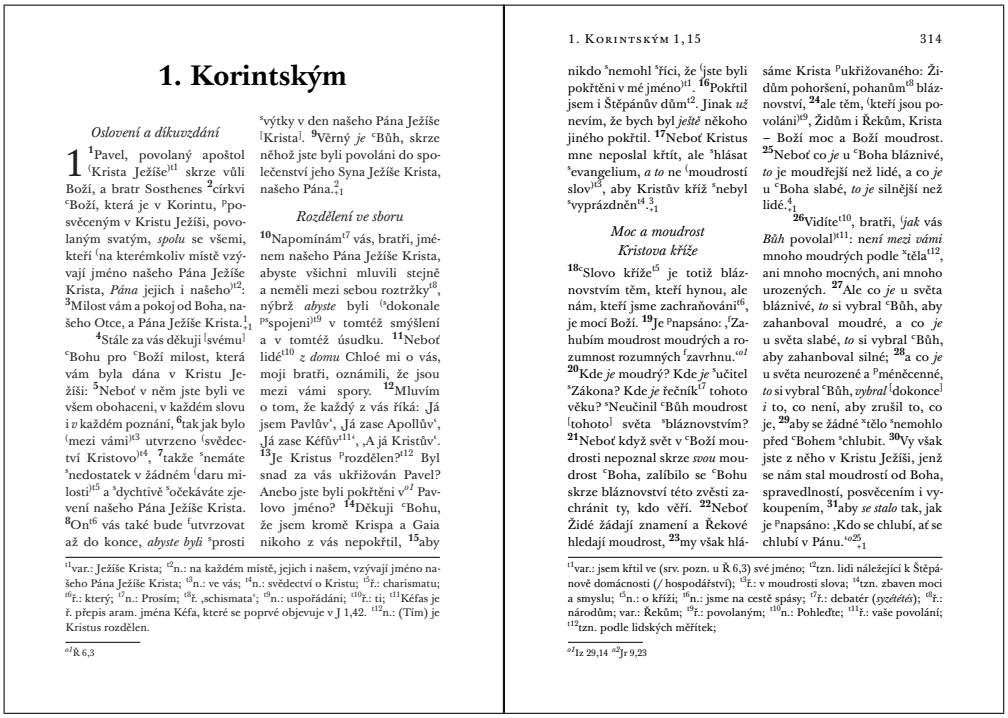
If there is any problem in this stage, the publisher is informed about the problem and must send again the XML data.

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.



1. Korintským 1,15 314

1. Korintským

Oslovení a díkuvzdání

1 [1]Pavel, povolaný apoštol [ʲKrista Ježíše][ʲt1] skrze vůli Boží, a bratr Sosthenes [2]církvi [ᶜ]Boží, která je v Korintu, [ᴾ]posvěceným v Kristu Ježíši, povolaným svatým, *spolu* se všemi, kteří [ʲ]na kterémkoliv místě vzývají jméno našeho Pána Ježíše Krista, *Pána* jejich i našeho[ʲt2]: [3]Milost vám a pokoj od Boha, našeho Otce, a Pána Ježíše Krista.[1]₊₁

[4]Stále za vás děkuji [svému] [ᶜ]Bohu pro [ᶜ]Boží milost, která vám byla dána v Kristu Ježíši: [5]Neboť v něm jste byli ve všem obohaceni, v každém slovu i *v* každém poznání, [6]tak jak bylo [mezi vámi][ʲt3] utvrzeno [svědectví Kristovo][ʲt4], [7]takže [ˢ]nemáte [ˢ]nedostatek v žádném [ʲ]daru milosti][ʲt5] a [ˢ]dychtivě [ˢ]očekáváte zjevení našeho Pána Ježíše Krista. [8]On[ʲt6] vás také bude [ʲ]utvrzovat až do konce, *abyste byli* [ˢ]prosti [ˢ]výtky v den našeho Pána Ježíše [Krista]. [9]Věrný *je* [ᶜ]Bůh, skrze něhož jste byli povoláni do společenství jeho Syna Ježíše Krista, našeho Pána.[2]₊₁

Rozdělení ve sboru

[10]Napomínám[ʲt7] vás, bratři, jménem našeho Pána Ježíše Krista, abyste všichni mluvili stejně a neměli mezi sebou roztržky[ʲt8], nýbrž *abyste* byli [ᵈ]dokonale [ᴾˢ]spojení][ʲt9] v tomtéž smýšlení a v tomtéž úsudku. [11]Neboť lidé[ʲt10] *z domu* Chloé mi o vás, moji bratři, oznámili, že jsou mezi vámi spory. [12]Mluvím o tom, že každý z vás říká: ‚Já jsem Pavlův', ‚Já zase Apollův', ‚Já zase Kéfův[ʲt11]', ‚A já Kristův'. [13]Je Kristus [ᴾ]rozdělen?[ʲt12] Byl snad za vás ukřižován Pavel? Anebo jste byli pokřtěni v[ᵒ1] Pavlovo jméno? [14]Děkuji [ᶜ]Bohu, že jsem kromě Krispa a Gaia nikoho z vás nepokřtil, [15]aby nikdo [ˢ]nemohl [ˢ]říci, že [ʲ]jste byli pokřtěni v mé jméno)[ʲt1]. [16]Pokřtil jsem i Štěpánův dům[ʲt2]. Jinak *už* nevím, že bych byl *ještě* někoho jiného pokřtil. [17]Neboť Kristus mne neposlal křtít, ale [ˢ]hlásat [ˢ]evangelium, *a to* ne [ʲ]moudrostí slov[ʲt3], aby Kristův kříž [ˢ]nebyl [ˢ]vyprázdněn[ʲt4].[3]₊₁

Moc a moudrost Kristova kříže

[18][ᶜ]Slovo kříže[ʲt5] je totiž bláznovstvím těm, kteří hynou, ale nám, kteří jsme zachraňováni[ʲt6], je mocí Boží. [19]Je [ᴾ]napsáno: [ʲ]Zahubím moudrost moudrých a rozumnost rozumných [ʲ]zavrhnu.'[ᵒ1] [20]Kde *je* moudrý? Kde *je* [ˢ]učitel [ˢ]Zákona? Kde *je* řečník[ʲ7] tohoto věku? [ˢ]Neučinil [ᶜ]Bůh moudrost [tohoto] světa [ˢ]bláznovstvím? [21]Neboť když svět v [ᶜ]Boží moudrosti nepoznal skrze *svou* moudrost [ᶜ]Boha, zalíbilo se [ᶜ]Bohu skrze bláznovství této zvěsti zachránit ty, kdo věří. [22]Neboť Židé žádají znamení a Řekové hledají moudrost, [23]my však hlásáme Krista [ᴾ]ukřižovaného: Židům pohoršení, pohanům[ʲt8] bláznovství, [24]ale těm, [ʲ]kteří jsou povoláni][ʲt9], Židům i Řekům, Krista – Boží moc a Boží moudrost. [25]Neboť co *je* u [ᶜ]Boha bláznivé, *to* je moudřejší než lidé, a co *je* u [ᶜ]Boha slabé, *to je* silnější než lidé.[4]₊₁

[26]Vidíte[ʲt10], bratři, [ʲ]*jak* vás *Bůh* povolal][ʲt11]: *není mezi vámi* mnoho moudrých podle [ˣ]těla[ʲt12], ani mnoho mocných, ani mnoho urozených. [27]Ale co *je* u světa bláznivé, *to* si vybral [ᶜ]Bůh, aby zahanboval moudré, a co *je* u světa slabé, *to* si vybral [ᶜ]Bůh, aby zahanboval silné; [28]a co *je* u světa neurozené a [ᴾ]méněcenné, *to* si vybral [ᶜ]Bůh, *vybral* [ʲ]dokonce] *i* to, co není, aby zrušil to, co je, [29]aby se žádné [ˣ]tělo [ˢ]nemohlo před [ᶜ]Bohem [ˢ]chlubit. [30]Vy však jste z něho v Kristu Ježíši, jenž se nám stal moudrostí od Boha, spravedlností, posvěcením i vykoupením, [31]aby *se stalo* tak, jak je [ᴾ]napsáno: ‚Kdo se chlubí, ať se chlubí v Pánu.'[ᵒ25]₊₁

[ʲt1]var.: Ježíše Krista; [ʲt2]n.: na každém místě, jejich i našem, vzývají jméno našeho Pána Ježíše Krista; [ʲt3]n.: ve vás; [ʲt4]n.: svědectví o Kristu; [ʲt5]ř.: charismatu; [ʲt6]ř.: který; [ʲt7]n.: Prosím; [ʲt8]ř. ‚schismata'; [ʲt9]n.: uspořádáni; [ʲt10]ř.: ti; [ʲt11]Kéfas je ř. přepis aram. jména Kéfa, které se poprvé objevuje v J 1,42. [ʲt12]n.: (Tím) je Kristus rozdělen.

[ᵒ1]Ř 6,3

[ʲt1]var.: jsem křtil ve (srv. pozn. u Ř 6,3) své jméno; [ʲt2]tzn. lidi náležející k Štěpánově domácnosti (/ hospodářství); [ʲt3]ř.: v moudrosti slova; [ʲt4]tzn. zbaven moci a smyslu; [ʲt5]n.: o kříži; [ʲt6]n.: jsme na cestě spásy; [ʲt7]ř.: debatér (*zyzáčtěs*); [ʲt8]ř.: národům; var.: Řekům; [ʲt9]ř.: povoláným; [ʲt10]n.: Pohleďte; [ʲt11]ř.: vaše povolání; [ʲt12]tzn. podle lidských měřítek;

[ᵒ1]Iz 29,14 [ᵒ25]Jr 9,23

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 TEX, 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.

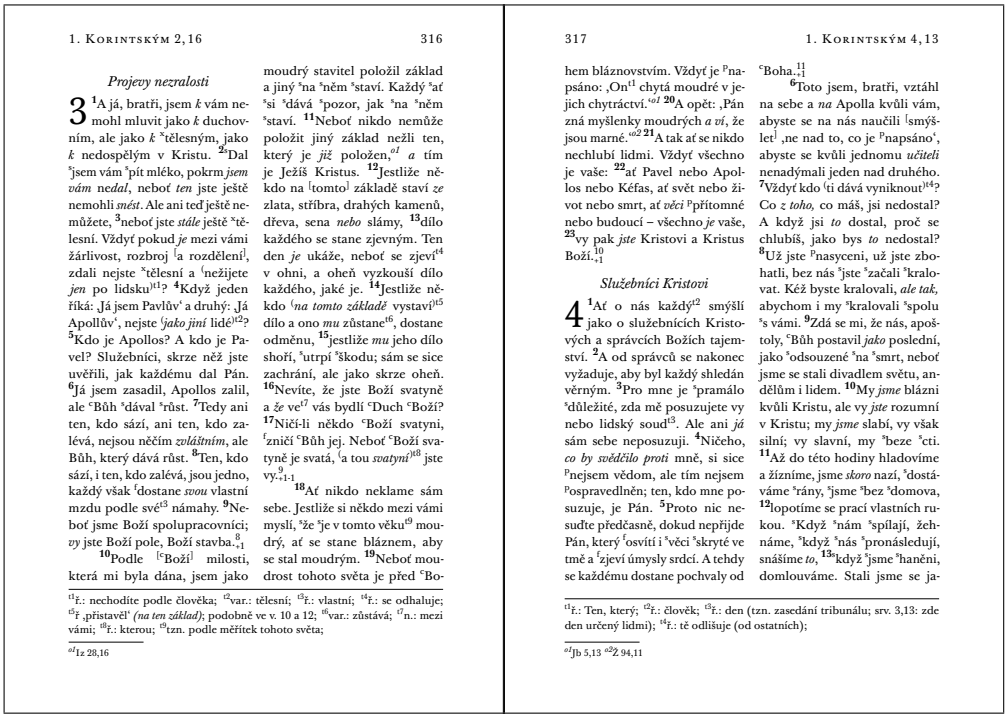
Solving a problematic case is usually done by changing the formatting of various elements around the problematic place. For example, the typesetter can change the gap around a figure, a paragraph, or a subtitle. Another example is to change the page dimensions (i. e. `\enlargethispage` in LATEX), or 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.



1. Korintským 2,16 316

Projevy nezralosti

3 [1]A já, bratři, jsem *k* vám nemohl mluvit jako *k* duchovním, ale jako *k* [x]tělesným, jako *k* nedospělým v Kristu. [2]Dal [s]jsem vám [s]pít mléko, pokrm *jsem vám nedal*, neboť *ten* jste ještě nemohli *snést*. Ale ani teď ještě nemůžete, [3]neboť jste *stále* ještě [x]tělesní. Vždyť pokud *je* mezi vámi žárlivost, rozbroj [l]a rozdělení[l], zdali nejste [x]tělesní a [l]nežijete *jen* po lidsku)[t1]? [4]Když jeden říká: ‚Já jsem Pavlův' a druhý: ‚Já Apollův', nejste (*jako jiní* lidé)[t2]? [5]Kdo je Apollos? A kdo je Pavel? Služebníci, skrze něž jste uvěřili, jak každému dal Pán. [6]Já jsem zasadil, Apollos zalil, ale [c]Bůh [s]dával [s]růst. [7]Tedy ani ten, kdo sází, ani ten, kdo zalévá, nejsou něčím *zvláštním*, ale Bůh, který dává růst. [8]Ten, kdo sází, i ten, kdo zalévá, jsou jedno, každý však [f]dostane *svou* vlastní mzdu podle své[t3] námahy. [9]Neboť jsme Boží spolupracovníci; *vy* jste Boží pole, Boží stavba.[9]+1

[10]Podle [c]Boží[l] milosti, která mi byla dána, jsem jako moudrý stavitel položil základ a jiný [s]na [s]něm [s]staví. Každý [s]ať [s]si [s]dává [s]pozor, jak [s]na [s]něm [s]staví. [11]Neboť nikdo nemůže položit jiný základ než ten, který je *již* položen,[o1] *a* tím je Ježíš Kristus. [12]Jestliže někdo na [l]tomto[l] základě staví *ze* zlata, stříbra, drahých kamenů, dřeva, sena *nebo* slámy, [13]dílo každého se stane zjevným. Ten den *je* ukáže, neboť se zjeví[t4] v ohni, a oheň vyzkouší dílo každého, jaké je. [14]Jestliže někdo [l]*na tomto základě* vystaví)[t5] dílo a ono *mu* zůstane[t6], dostane odměnu, [15]jestliže *mu* jeho dílo shoří, [s]utrpí [s]škodu; sám se sice zachrání, ale jako skrze oheň. [16]Nevíte, že jste Boží svatyně a *že* ve[t7] vás bydlí [c]Duch [c]Boží? [17]Ničí-li někdo [c]Boží svatyni, [l]zničí [c]Bůh jej. Neboť [c]Boží svatyně je svatá, ([a]a tou *svatyní*)[t8] jste vy.[9]+1-1

[18]Ať nikdo neklame sám sebe. Jestliže si někdo mezi vámi myslí, [s]že [s]je v tomto věku[t9] moudrý, ať se stane bláznem, aby se stal moudrým. [19]Neboť moudrost tohoto světa je před [c]Bohem bláznovstvím. Vždyť je [p]napsáno: ‚On[t1] chytá moudré v jejich chytrácství.'[o1] [20]A opět: ‚Pán zná myšlenky moudrých *a ví*, že jsou marné.'[o2] [21]A tak ať se nikdo nechlubí lidmi. Vždyť všechno je vaše: [22]ať Pavel nebo Apollos nebo Kéfas, ať svět nebo život nebo smrt, ať *věci* [p]přítomné nebo budoucí – všechno *je* vaše, [23]vy pak *jste* Kristovi a Kristus Boží.[10]+1

[t1]ř.: nechodíte podle člověka; [t2]var.: tělesní; [t3]ř.: vlastní; [t4]ř.: se odhaluje; [t5]ř. ‚přistavěl' *(na ten základ)*; podobně ve v. 10 a 12; [t6]var.: zůstává; [t7]n.: mezi vámi; [t8]ř.: kterou; [t9]tzn. podle měřítek tohoto světa;

[o1]Iz 28,16

317 1. Korintským 4,13

Služebníci Kristovi

4 [1]Ať o nás každý[t2] smýšlí jako o služebnících Kristových a správcích Božích tajemství. [2]A od správců se nakonec vyžaduje, aby byl každý shledán věrným. [3]Pro mne je [s]pramálo [s]důležité, zda mě posuzujete vy nebo lidský soud[t3]. Ale ani *já* sám sebe neposuzuji. [4]Ničeho, *co by svědčilo proti* mně, si sice [p]nejsem vědom, ale tím nejsem [p]ospravedlněn; ten, kdo mne posuzuje, je Pán. [5]Proto nic nesuďte předčasně, dokud nepřijde Pán, který [f]osvítí i [s]věci [s]skryté ve tmě a [f]zjeví úmysly srdcí. A tehdy se každému dostane pochvaly od [c]Boha.[11]+1

[6]Toto jsem, bratři, vztáhl na sebe a *na* Apolla kvůli vám, abyste se na nás naučili [l]smýšlet[l] ‚ne nad to, co je [p]napsáno', abyste se kvůli jednomu *učiteli* nenadýmali jeden nad druhého. [7]Vždyť kdo [l]ti dává vyniknout)[t4]? Co *z toho,* co máš, jsi nedostal? A když jsi *to* dostal, proč se chlubíš, jako bys *to* nedostal? [8]Už jste [p]nasyceni, už jste zbohatli, bez nás [s]jste [s]začali [s]kralovat. Kéž byste kralovali, *ale tak,* abychom i my [s]kralovali [s]spolu [s]s vámi. [9]Zdá se mi, že nás, apoštoly, [c]Bůh postavil *jako* poslední, jako [s]odsouzené [s]na [s]smrt, neboť jsme se stali divadlem světu, andělům i lidem. [10]My *jsme* blázni kvůli Kristu, ale vy *jste* rozumní v Kristu; my *jsme* slabí, vy však silní; vy slavní, my [s]beze [s]cti. [11]Až do této hodiny hladovíme a žízníme, jsme *skoro* nazí, [s]dostáváme [s]rány, [s]jsme [s]bez [s]domova, [12]lopotíme se prací vlastních rukou. [s]Když [s]nám [s]spílají, žehnáme, [s]když [s]nás [s]pronásledují, snášíme *to,* [13]když [s]jsme [s]haněni, domlouváme. Stali jsme se ja-

[t1]ř.: Ten, který; [t2]ř.: člověk; [t3]ř.: den (tzn. zasedání tribunálu; srv. 3,13: zde den určený lidmi); [t4]ř.: tě odlišuje (od ostatních);

[o1]Jb 5,13 [o2]Ž 94,11

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:

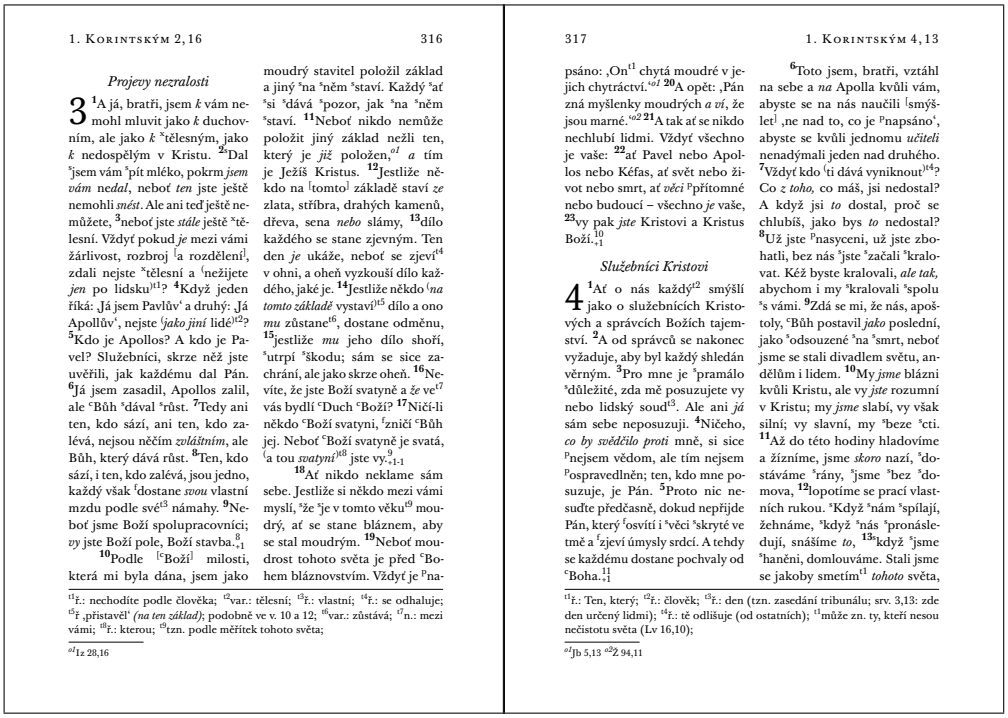
- problem detection: checking the result by eye is very tiresome and it is easy to overlook problems. It would be much better if we get a report of problems, saying for example *there is a widow line on page 317*.
- adjustment of paragraph formatting: we need an easy way to find which paragraph can be 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. pdfTEX allows recording the absolute position of a point using the primitive



1. Korintským 2,16 316

Projevy nezralosti

3 [1]A já, bratři, jsem *k* vám nemohl mluvit jako *k* duchovním, ale jako *k* [x]tělesným, jako *k* nedospělým v Kristu. [2]Dal [s]jsem vám [p]pít mléko, pokrm *jsem vám nedal*, neboť *ten* jste ještě nemohli *snést*. Ale ani teď ještě nemůžete, [3]neboť jste *stále* ještě [x]tělesní. Vždyť pokud *je* mezi vámi žárlivost, rozbroj [l]a rozdělení[l], zdali nejste [x]tělesní a [l]nežijete *jen* po lidsku)[t1]? [4]Když jeden říká: ‚Já jsem Pavlův' a druhý: ‚Já Apollův', nejste (*jako jiní* lidé)[t2]? [5]Kdo je Apollos? A kdo je Pavel? Služebníci, skrze něž jste uvěřili, jak každému dal Pán. [6]Já jsem zasadil, Apollos zalil, ale [c]Bůh [s]dával [s]růst. [7]Tedy ani ten, kdo sází, ani ten, kdo zalévá, nejsou něčím *zvláštním*, ale Bůh, který dává růst. [8]Ten, kdo sází, i ten, kdo zalévá, jsou jedno, každý však [f]dostane *svou* vlastní mzdu podle své[t3] námahy. [9]Neboť jsme Boží spolupracovníci; *vy* jste Boží pole, Boží stavba.[9]+1

[10]Podle [c][Boží][l] milosti, která mi byla dána, jsem jako moudrý stavitel položil základ a jiný [s]na [s]něm [s]staví. Každý [s]ať [s]si [s]dává [s]pozor, jak [s]na [s]něm [s]staví. [11]Neboť nikdo nemůže položit jiný základ nežli ten, který je *již* položen,[o1] *a* tím je Ježíš Kristus. [12]Jestliže někdo na [l]tomto[l] základě staví *ze* zlata, stříbra, drahých kamenů, dřeva, sena *nebo* slámy, [13]dílo každého se stane zjevným. Ten den *je* ukáže, neboť se zjeví[t4] v ohni, a oheň vyzkouší dílo každého, jaké je. [14]Jestliže někdo (*na tomto základě* vystaví)[t5] dílo a ono *mu* zůstane[t6], dostane odměnu, [15]jestliže *mu* jeho dílo shoří, [s]utrpí [s]škodu; sám se sice zachrání, ale jako skrze oheň. [16]Nevíte, že jste Boží svatyně a že ve[t7] vás bydlí [c]Duch [c]Boží? [17]Ničí-li někdo [c]Boží svatyni, [f]zničí [c]Bůh jej. Neboť [c]Boží svatyně je svatá, [l]a tou *svatyní*[l][t8] jste vy.[9]-1-1

[18]Ať nikdo neklame sám sebe. Jestliže si někdo mezi vámi myslí, [s]že [s]je v tomto věku[t9] moudrý, ať se stane bláznem, aby se stal moudrým. [19]Neboť moudrost tohoto světa je před [c]Bohem bláznovstvím. Vždyť je [p]napsáno: ‚On[t1] chytá moudré v jejich chytrácství.'[o1] [20]A opět: ‚Pán zná myšlenky moudrých *a ví*, že jsou marné.'[o2] [21]A tak ať se nikdo nechlubí lidmi. Vždyť všechno je vaše: [22]ať Pavel nebo Apollos nebo Kéfas, ať svět nebo život nebo smrt, ať *věci* [p]přítomné nebo budoucí – všechno *je* vaše, [23]vy pak *jste* Kristovi a Kristus Boží.[10]+1

[t1]ř.: nechodíte podle člověka; [t2]var.: tělesní; [t3]ř.: vlastní; [t4]ř.: se odhaluje; [t5]ř. ‚přistavěl' *(na ten základ)*; podobně ve v. 10 a 12; [t6]var.: zůstává; [t7]n.: mezi vámi; [t8]ř.: kterou; [t9]tzn. podle měřítek tohoto světa;

[o1]Iz 28,16

317 1. Korintským 4,13

Služebníci Kristovi

4 [1]Ať o nás každý[t2] smýšlí jako o služebnících Kristových a správcích Božích tajemství. [2]A od správců se nakonec vyžaduje, aby byl každý shledán věrným. [3]Pro mne je [s]pramálo [s]důležité, zda mě posuzujete vy nebo lidský soud[t3]. Ale ani *já* sám sebe neposuzuji. [4]Ničeho, *co by svědčilo proti* mně, si sice [p]nejsem vědom, ale tím nejsem [p]ospravedlněn; ten, kdo mne posuzuje, je Pán. [5]Proto nic nesuďte předčasně, dokud nepřijde Pán, který [f]osvítí i [s]věci [s]skryté ve tmě a [f]zjeví úmysly srdcí. A tehdy se každému dostane pochvaly od [c]Boha.[11]+1

[6]Toto jsem, bratři, vztáhl na sebe a *na* Apolla kvůli vám, abyste se na nás naučili [l]smýšlet[l] ‚ne nad to, co je [p]napsáno', abyste se kvůli jednomu *učiteli* nenadýmali jeden nad druhého. [7]Vždyť kdo [l]ti dává vyniknout[l][t4]? Co *z toho,* co máš, jsi nedostal? A když jsi *to* dostal, proč se chlubíš, jako bys *to* nedostal? [8]Už jste [p]nasyceni, už jste zbohatli, bez nás [c]jste [s]začali [s]kralovat. Kéž byste kralovali, *ale tak,* abychom i my [s]kralovali [s]spolu [s]s vámi. [9]Zdá se mi, že nás, apoštoly, [c]Bůh postavil *jako* poslední, jako [s]odsouzené [s]na [s]smrt, neboť jsme se stali divadlem světu, andělům i lidem. [10]My *jsme* bláznivi kvůli Kristu, ale vy *jste* rozumní v Kristu; my *jsme* slabí, vy však silní; vy slavní, my [s]beze [s]cti. [11]Až do této hodiny hladovíme a žízníme, jsme *skoro* nazí, [s]dostáváme [s]rány, [s]jsme [s]bez [s]domova, [12]lopotíme se prací vlastních rukou. [s]Když [s]nám [s]spílají, žehnáme, [s]když [s]nás [s]pronásledují, snášíme *to*, [13]s když [s]jsme [s]haněni, domlouváme. Stali jsme se jakoby smetím[t1] *tohoto* světa,

[t1]ř.: Ten, který; [t2]ř.: člověk; [t3]ř.: den (tzn. zasedání tribunálu; srv. 3,13: zde den určený lidmi); [t4]ř.: tě odlišuje (od ostatních); [t1]může zn. ty, kteří nesou nečistotu světa (Lv 16,10);

[o1]Jb 5,13 [o2]Ž 94,11

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 ($28146752 - 27423153 = 723599 = 11\text{pt}$), this indicates that there is a widow line from paragraph 11 on page 317.

Other problematic cases are detected in a similar manner.

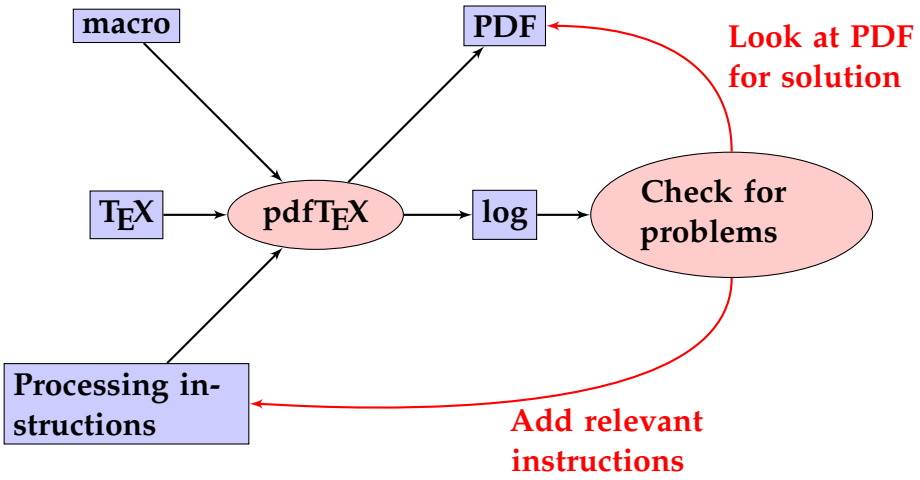


FIGURE 6. The process to tune page breaking

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 pdfTEX. 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 ±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

- TEX 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 TEX, despite the fact that it is a 30-years old program.
- The micro-typographic extensions of pdfTEX 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.