TYPESETTING POLISH TEXTS: CHARACTERISTIC FEATURES

(draft report)

TIP Ltd, Poland, November 1994

The aim of this report is to provide a help and some hints for those who are interested in adding Polish "look-and-feel" to the Polish versions of text processors. Polish quotations are not translated, as the text is chiefly meant for people familiar with the Polish language. Font considerations were stimulated by the invasion of digital fonts containing improperly designed Polish diacritical characters.

Diacritical characters. There are 18 Polish diacritical letters: ą, ć, ę, ł, ń, ó, ś, ź, ż, Ą, Ć, Ę, Ł, Ń, Ó, Ś, Ź, Ż. Sometimes, e.g., in header fonts, a variant of ż and Ż may be used with a cross in the middle of a letter rather than with a dot accent. Observe that crossed ƶ and Ƶ play the rôle of non-diacritical characters in certain decorative fonts.

REMARK 1. There are also several two-letter phonemes (ch, cz, dz, dż, dź, ni, rz, si, sz, zi) which, in general, should not be hyphenated. However, in certain context some of them are pronounced differently, namely, as two sounds; in such cases hyphenation may be possible.

REMARK 2. The following diacritical characters never begin a word: ą, ę, ń, Ą, Ę, Ń. The one-letter words, i.e., *a*, *i*, *o*, *u*, *w*, *z*, *A*, *I*, *O*, *U*, *W*, and *Z*, should not end the line of a text (especially capital ones), also when preceded by a brace or quotes. Such "hanging letters" are admissible in very narrow columns.

Numbers. The usual Arabic digits (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) are used in the Polish typesetting. In special cases Roman numerals (entirely either in lower- or uppercase letters) can be used. Text processors are supposed to recognize Roman numerals in order to avoid hyphenating them.

A comma should be used as a character marking the beginning of a decimal fraction, while the period should be used to separate groups of three digits, e.g., *3.345.678,123*. Nowadays, this custom seems to become obsolete, mainly due to the imported software technology (spread sheets, accounting programs, etc.). In order to avoid a possible ambiguity, using a tiny space as the separator of the groups of digits would be advisable, e.g., *3 345 678,123*.

A dash or a tiny space can be used as the separator of the groups of digits in telephone numbers, e.g., *tel. 56-78-77* or *tel. 56 78 77*. In mail codes (always preceding a city name), two leading digits should be separated with a dash from the trailing three digits, e.g., *80-331 Gdańsk*.

A slash may be used to denote a fraction, e.g., *1/2 km*. In addresses, a slash may be used to separate the number of a house from the number of a flat, e.g., *ul. Tatrzańska 6/1*.

The numbers can be preceded by the following characters: mathematical symbols (see below), (, [, „, « (or »—see below), - (dash), and §.

Punctuation. There is the list of punctuation characters which commonly appear in Polish texts:

a) usual punctuation: ., !, ?, ,, ;, :, *, ', ...;

b) quotes: „ (opening), ” (closing), « (opening), » (closing);

c) dashes: - (dash), – (en-dash), — (em-dash);

d) brackets: (,), [,];

REMARK 1. In general, punctuation marks are not separated from the neighbouring word with an extra space. For example, comma, dot, exclamation mark, etc., immediately follow the preceding character. The exception is an em-dash (see below). In certain cases, however, a tiny space may be used to improve the appearance of typesetting. Combinations of punctuation marks may also occur, e.g., ., (comma after abbreviation), ,– (used after the amount of money expressed in zlotys without groszes), !..., ?..., ...!, ...?, ?!, !!!.

REMARK 2. The usual Polish quotes are „ and ”. So called French quotes, « and », are used for quotations inside quotations. Sometimes » is used as opening quotes and « as closing ones, e.g., *„Byłem w hotelu «Polonia»” — powiedział* or *„Byłem w hotelu »Polonia«” — powiedział.* The first of these two forms is preferable. Rarely, a reversed closing quotes “ (German style) is used. Punctuation is normally put after the closing quote, e.g., *„1”, „2”*...

REMARK 3. The following punctuation characters may begin a sentence: „, « (or »), ..., (, — (em-dash, mainly in dialogs). The following punctuation characters may end a sentence: ., ..., !, ?, ”, » (or «), :,). The following characters may appear in the midst of a sentence: . (in abbreviations), ,, :, ;, ..., – (instead of an em-dash), —.

REMARK 4. The following characters may occur inside a word: - (dash—see below), . (in abbreviations), ' (apostrophe), /, e.g., *p.n.e.*, *d'Alembert, książka Greene'a*, *km/h, m/s*.

REMARK 5. The following characters should not begin the line of a text: ., ,, !, ?, /, :, ;,),], %, ‰. The following characters should not end the line of a text: „, (, [. See also Section “Diacritical characters,” remark 2.

REMARK 6. A dash - should be used only as a hyphen or as a character joining compound words, e.g., *biało-czerwony*. In such a case the proper hyphenation is:

biało-
-czerwony

i.e., the hyphen should be repeated at the beginning of the second line. Symbolic names, such as *K-202*, should rather not be hyphenated. It is not advisable to use a dash instead of an em-dash —. An em-dash plays the rôle of a break at the level of a sentence rather than of a word. It appears with normal spaces at both ends, e.g., *— Czy to jest prawda? — spytał Jan*. An en-dash – can either be used instead of an em-dash or for expressing a range of numbers, e.g., *123–125* denotes numbers from *123* to *125*; observe a tiny space added between the en-dash and the digits. In technical texts, however, symbol ÷ is preferable for expressing the range of quantities, in order to avoid intermixing a range symbol and a minus (see below).

REMARK 7. Square brackets surrounding an ellipsis character [...] are used for marking skipped fragments of a text.

Symbols. The following symbols may likely appear in scientific and/or technical texts:

a) units: [$], [zł] and other currency signs, [%], [‰];

b) reference marks: [*], [**], [***], [1], [2], [3], etc.;

c) basic mathematical symbols: [+], [−] (minus, not necessarily identical with an en-dash), [±], [×] (multiply), [·] (multiply), [/] (divide), [:] (divide), [÷] (range), [|] (absolute value), [{], [}], [<], [>], [⩽] (rather than [≤]), [⩾] (rather than [≥]), [∼] (proportionality), [≈] (approximate value), [←], [→], [∞] (infinity), [∫] (integral), [π];

d) basic technical symbols: [μ] (mu, prefix *micro*) [°] (degree), ['] (minutes), ["] (seconds or inches), [Ω] (ohm), [∅] (diameter);

e) other: [§] (section mark), [•] (bullet), [·] (centred period), [@], [&], [#].

REMARK 1. Units are usually put after a number, e.g., *15 $*, *100 zł*, *5 km*, *10,2 %*, etc.

REMARK 2. Rarely, chiefly in texts written in foreign languages, [†] (dagger) and [‡] (double dagger) are used for referencing.

REMARK 3. Straight single and straight double quotes, ['] and ["], should not be used as quotation marks. They should only be used as units, e.g., 54°24'57" (geographical position).

REMARK 4. Symbol [÷] is used in the meaning of "range" rather than "divide," especially in technical texts. For example, the formula ∅=15 ÷ 30 cm means: a diameter may vary from 15 cm up to 30 cm.

REMARK 5. One should be aware of the fact that the abbreviations of names of mathematical functions may be language-dependent, e.g., the Polish abbreviation for tangent function is *tg*, not *tan*, and the abbreviation for cotangent function is *ctg*, not *cot*.

REMARK 6. Section mark [§] is commonly used in formal agreements. At-sign [@] is presently used almost exclusively in email addresses. Ampersand [&] and hash sign [#] are occasionally used as elements of a logo of a firm. The latter is primarily used in musical notation (sharp).

Fonts. Polish diacritical characters should be designed with a great care. Not only their shape should be consistent with the overall appearance of a font (thickness of strokes, size of diacritical elements); also kerning is very important from the point of view of professional typesetting. *Courier Bold CE* and *Times New Roman CE* fonts, used in the examples which follow, come from the Polish Windows distribution. As the names suggest, they are meant for users from Central and Eastern Europe. One can doubt whether such fonts are useful for professional applications. Incidentally, standard fonts coming with Polish Windows (*Courier*, *Times New Roman*, *Arial*, etc.) have metrics differing from the metrics of fonts ***having exactly the same internal and external names*** which come with US Windows—disastrous consequences of such an incompatibility can easily be imagined.

THE LETTERS [ą], [ę], [Ą], and [Ę]. The size and shape of the "ogonek" should be consistent with other elements of a font. The ogonek should be attached as smoothly as possible to the contour of a letter. In the following example the ogoneks are apparently too small and too thin:

ę ą Ą Ę

Moreover, the ogonek in **ą** protrudes too much to the right. Such a protrusion may cause an undesirable collision with the neighbouring letters:

ap ąp ąg

Even worse, in both fonts the ogoneks are attached as separate elements to the letters:

ą ę Ą Ę ąę Ą Ę

which makes using these fonts in the context of cutting plotters nearly impossible.

THE LETTERS **ł** and **Ł**. These letters should usually be wider than **l** and **L**, respectively. Only in monospaced fonts, such as *Courier*, they can have the same width. In *Times New Roman CE* both **l** and **ł** are 0.278 em wide, and both **L** and **Ł** are 0.611 em wide. Moreover, there is no kern pair for either **ł** or **Ł**, while there are several kern pairs for **L**. Here are the results of such a state of the art:

ołw ŁW LW

A collision between **ł** and the neighbouring letters, and a gap between **Ł** and **W** can be observed.

Increasing the widths of **ł** and **Ł**, shifting a slash to the right by a small amount and aligning optically the crossing point with the top of lowercase letters would be advisable in this case.

THE LETTERS **ć**, **ń**, **ś**, **ź**, **Ć**, **Ń**, **Ś**, and **Ź**. The acute accents over the uppercase letters should be a bit flattened in order to avoid the collision between consecutive lines. The weight and slope of the accent should be consistent with other elements of a font, in particular with sloped ones. This apparently is not the case with both *Courier Bold CE* and *Times New Roman CE*:

kłó kłó

THE LETTERS **ż** and **Ż**. The dot accent should be exactly the same as the dot over **i** and **j**, i.e., it should be put at the same height and should have the same shape. In both *Courier Bold CE* and *Times New Roman CE* strange things happen with a dot accent:

iż iż

In the case of *Courier*, it is not obvious whether the dot of **ż** or the dot of **i** and **j** should be corrected.

FONT UNITS. The standard typographic unit used in Poland is didot point according to the Polish norm PN-70/P-55010 (1 didot point is equal to 1/2660 m, approximately 0.376 mm). Recently, however, "small points" (1/72.27 in, approximately 0.351 mm) and "big points" (1/72 in, approximately 0.353 mm) are becomming more and more popular due to the import of software technology.

CONCLUSION. So far, poorly designed (with respect to national characters) fonts dominate. The situation like this cannot last forever. Sooner or later font vendors will propose digital fonts suitable for high quality printing. It seems reasonable to consider producing better international fonts just now—within a few years one can expect a good market for such fonts.

Bogusław Jackowski