

## Representing Parkosz’s alphabet in the Junicode font

Janusz S. Bien

### Abstract

The 15th century Latin manuscript containing a treatise by Jakub Parkosz was the very first proposal of Polish spelling. To account for all the phonemes of Polish some new letters were proposed, which are not available in present day fonts. This makes it difficult to quote the proposal when discussing the history of Polish spelling. A transliteration was designed which uses only the characters available in the Unicode standard, but it was rather cumbersome. Another approach, suggested by the present author and implemented by Peter S. Baker in his Junicode (version two) font, is to use so-called tag characters.

### 1 Introduction

A digression: when in 1990 I was writing my proposal of an extended font layout [3] and wanted to include some comments about the origin of the letters specific to the Polish language, I was very much surprised by the lack of published information on this topic. Now the situation is definitely better in one respect: the original sources have been digitized and are freely available to anybody. In my opinion the analysis of sources is still lacking many details, so some time ago I decided to try to answer my questions myself ☺.

The 15th century handwritten Latin treatise by Jakub Parkosz (called also Parkoszowic) was the very first proposal of Polish spelling. You can find the scans at (for example) [jsbien.github.io/Parkosz4IIIF/](https://jsbien.github.io/Parkosz4IIIF/). The best source of English language information on the treatise and its author seems to be [11]; in Polish, it is [8]. To account for all the phonemes of Polish, Parkosz created some completely new letters. As his proposal did not catch on, the letters haven’t become available in printer’s fonts. He also assigned some different meanings to the variants of handwritten letters, which also were not used in print. The almost complete repertoire of Parkosz’s letters is presented in Fig. 1.

Both the 1830 [2] and 1907 [10] editions of the treatise were typeset manually and undoubtedly special types had to be prepared for the missing characters. The 1985 edition [8] was typeset leaving empty space for Parkosz’s letters. The letters were drawn by hand, cut into pieces and carefully pasted into the empty spaces.

Two amateurish attempts have been made (by the author and his student) to create a font for Parkosz’s letters with FontForge. One was intended to reproduce the shape of the letters in the 1985 edition: [bitbucket.org/jsbien/parkosz-font-old](https://bitbucket.org/jsbien/parkosz-font-old). Another was intended to reproduce the shape of the letters in the 1907 edition: [github.com/jsbien/parkosz-font](https://github.com/jsbien/parkosz-font). However, their quality was not satisfactory, so for the electronic edition of the treatise [9] (see also [github.com/jsbien/Parkosz-traktat](https://github.com/jsbien/Parkosz-traktat)) a transliteration system was designed. The rationale for the transliteration decisions were presented in Polish in [5] and summarized in English in [6]; see also Appendix A. The transliteration covers all the characters used by Parkosz, not just the newly created ones.

As the transliteration uses regular Unicode characters but changes their meaning, some metadata is needed to distinguish the normal text from the transliteration. It would be convenient to make the metadata unnecessary. Theoretically it is possible with so-called variation sequences, but they have to be officially registered by the Unicode Consortium, which makes this approach impractical, at least at the present stage.

In [4] a brute force approach was proposed. It assumed that after creating a TrueType/OpenType font with appropriate ligatures a little used combining character, namely U+20E8 COMBINING TRIPLE UNDERDOT ☺, would serve as a kind of a private variant selector. At that time I was only vaguely aware of the usage of the so-called tag characters in Unicode.

In March 2022 Margaret Kibi ([marrus-sh](https://marrus-sh)) proposed using tag characters instead of the variable sequences in the Junicode font.<sup>1</sup> The proposal was supported by several font users and accepted by the font author. I hope this approach will become a kind of a *de facto* standard, as it shares many advantages with TrueType/OpenType features (cf. [1, pp. 12–13]). To make a long story short, using those features preserves the properties of the base character, while a character in the Private Use Area has no properties.

This approach was also applied to Parkosz’s letters as the implementation of my feature request<sup>2</sup> and this is the subject of the present paper. The characters accessed by variation sequences, tag characters or ligatures can have their own codepoints in the Private Use Area, but this is another topic.

<sup>1</sup> [github.com/psb1558/Junicod-font/discussions/122#discussioncomment-2416880](https://github.com/psb1558/Junicod-font/discussions/122#discussioncomment-2416880)

<sup>2</sup> [github.com/psb1558/Junicod-New/issues/27](https://github.com/psb1558/Junicod-New/issues/27)

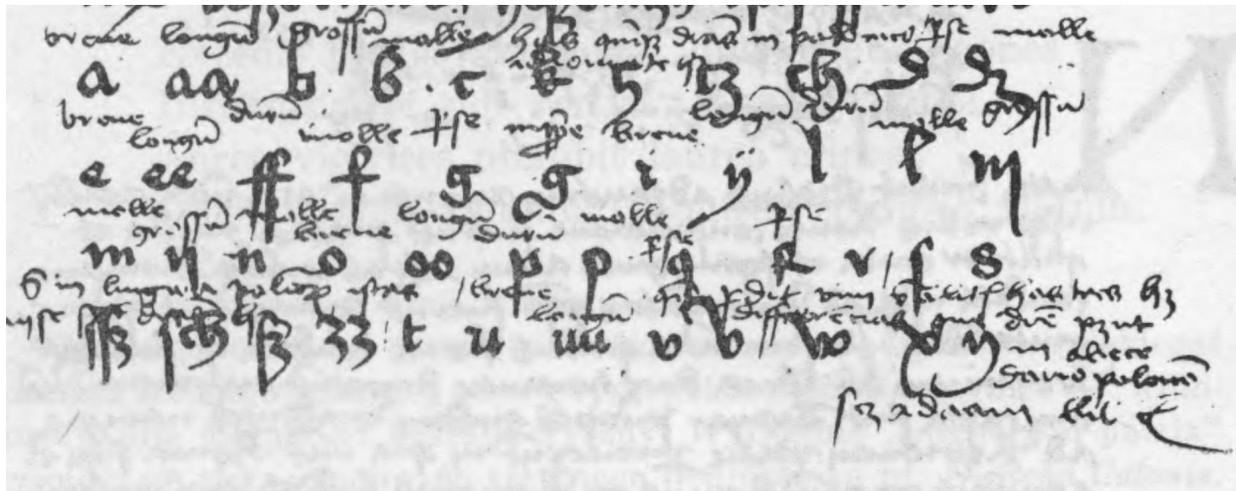


Figure 1: Parkosz's alphabet summary (with a few omissions)

Tag characters are in principle invisible, but for documentation purposes the tag sequences will be rendered here as, e.g., `[b][s]`; *p* stands for *Parkosz*<sup>3</sup> and is always the very first tag, *s* refers to the shape and means *square*. Other secondary tags used are `[r]` for *round*, `[h]` for *hook* and `[l]` for *loop*, `[s]` for *slashed*, `[b]` for *below*; `[d]` means *descender* or *dot*.

In a X<sub>q</sub>L<sup>A</sup>T<sub>E</sub>X source they can be written respectively as `\&\_\_p;`, `\&\_\_s;` etc. The same convention applies to X<sub>q</sub>T<sub>E</sub>X and even to Microsoft Word and OpenOffice [1, p. 38].

The input can and should be simplified by using special Emacs input methods or equivalent macros in other tools. On the other hand usually just isolated words will be quoted, so providing the tag characters explicitly is not excessively tedious.

For the reader's convenience the layout of this paper is similar to [6] and some figures are repeated here. Please consult [6] for more details.

## 2 New letters

### 2.1 *b grossum*

Called also *b durum* and *b quadratum*. Some occurrences in the manuscript are presented in Fig. 2. It was transliterated as U+0180 LATIN SMALL LETTER B WITH STROKE (`[b]`, Latin Extended-B block).

In Junicode the glyph is `[b]` and the input is `b[p][s]`.

### 2.2 *b molle*

Called also *b rotundum*. Some occurrences in the manuscript are presented in Fig. 3.

<sup>3</sup> The tag is available in Junicode since build 1.052beta of August 25, 2022.

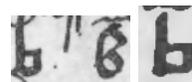


Figure 2: *b grossum* in the manuscript: p. [15] l. 28 (on the left), p. [8] l. 2



Figure 3: *b molle* in the manuscript: p. [15] l. 28 (on the right), p. [7] marginalia and l. 15.

It was transliterated as U+0253 LATIN SMALL LETTER B WITH HOOK (`[b]`, IPA Extensions block).

In Junicode the glyph is `[b]` and the input is `b[p][h]`.

### 2.3 *p durum*

Called also *p quadratum*. Some occurrences in the manuscript are presented in Fig. 4.

It was transliterated as U+1D7D LATIN SMALL LETTER P WITH STROKE (`[p]`, Latin Extended-C block).

In Junicode the glyph is `[p]` and the input is `p[p][s]`.

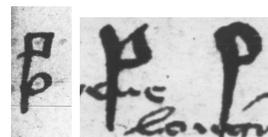


Figure 4: Letters *p durum* and *p molle* in the manuscript: p. [8] marginalia, p. [15] l. 32

## 2.4 *p molle*

Some occurrences in the manuscript are also presented in Fig. 4.

The letter was transliterated as U+01A5 LATIN SMALL LETTER P WITH HOOK ( $\text{p̃}$ , Latin Extended-B block).

In Junicode the glyph is  $\text{p̃}$  and the input is  $\text{p̃}$ .

## 2.5 *l molle*

The letter was transliterated as U+026C LATIN SMALL LETTER L WITH BELT ( $\text{ḷ}$ , IPA Extensions block),

In Junicode the glyph is  $\text{ḷ}$  and the input is  $\text{ḷ}$ .

## 2.6 *durum v*

The letter was transliterated as U+028B LATIN SMALL LETTER V WITH HOOK ( $\text{ṿ}$ , IPA Extensions block).

In Junicode the glyph is  $\text{ṿ}$  and the input is  $\text{ṿ}$ .

## 3 Adapted letters

### 3.1 *g improprie*

This is the letter *g* as written by Italians (*unco retorto versus dexteram partem sicut scribunt ipsum Italici*).

The letter was transliterated as U+A77F LATIN SMALL LETTER TURNED INSULAR G ( $\text{g̃}$ , Latin Extended-D block).

In Junicode the glyph is  $\text{g̃}$  and the input is  $\text{g̃}$ .

### 3.2 *grossum m*

This is the letter *m* as written at the end of words (*[...] spissum cum cauda, sicut in fine dictionum poni solet*). As discussed in [6, p. 48], the intended shape of the letter is not clear.

The letter was transliterated as U+0271 LATIN SMALL LETTER M WITH HOOK ( $\text{ṃ}$ , IPA Extensions block).

In Junicode the recommended input is  $\text{ṃ}$ ; it now renders  $\text{ṃ}$ , i.e., the MUFI (Medieval Font Unicode Initiative<sup>4</sup>) Private Use Area character M+F223<sup>5</sup> LATIN SMALL LETTER M WITH RIGHT DESCENDER. This may change in the future.

### 3.3 *grossum n*

Similar to *grossum m*, this is the letter *n* as written at the end of words. As discussed in [6, pp. 48–49], the intended shape of the letter is not clear; see also Fig. 5.

The letter was transliterated as U+0272 LATIN SMALL LETTER N WITH LEFT HOOK ( $\text{ṇ}$ , IPA Extensions block).

In Junicode the recommended input is  $\text{ṇ}$ ; it now renders  $\text{ṇ}$ , i.e., the MUFI character M+F228 LATIN SMALL LETTER N WITH RIGHT DESCENDER. This may change in the future.

## 4 Special use letters

To this category belong the letters with the standard shape, but with a non-standard (from the contemporary point of view) phoneme assigned to them by Parkosz. They are: *f molle*, *g per se*, *l durum*, *m molle*, *n molle*.

All three editions used for these just the standard letters, which is quite confusing for present-day readers, even if they are scholars. Therefore the decision was made to transliterate them as, respectively:

- U+1E1F LATIN SMALL LETTER F WITH DOT ABOVE ( $\text{ḟ}$ ),
- U+0121 LATIN SMALL LETTER G WITH DOT ABOVE ( $\text{ġ}$ ),
- U+1E37 LATIN SMALL LETTER L WITH DOT BELOW ( $\text{ḷ}$ ),
- U+1E43 LATIN SMALL LETTER M WITH DOT BELOW ( $\text{ṃ}$ ),
- U+1E47 LATIN SMALL LETTER N WITH DOT BELOW ( $\text{ṇ}$ ),

This part of my proposal can be considered an unnecessary complication, so should be treated as optional. Nevertheless in Junicode the characters can be entered as, respectively:  $\text{ḟ}$ ,  $\text{ġ}$ ,  $\text{ḷ}$ ,  $\text{ṃ}$  and  $\text{ṇ}$ .

If used in colored text, it would be desirable to render the dot in a different color to make clear this is an artificial addition (an idea of Jakub Wilk formulated long ago in a different context). I am aware that because of technical difficulties this is at present rather a dream.

## 5 Regular letters

There is no typographical problem with the letters listed below. We give their names using original early Latin spelling (*breue* meaning *breve*, i.e., ‘short’).

They are:  $\text{a}$  (*breue*),  $\text{c}$ ,  $\text{d}$  (*per se*),  $\text{c}$  (*breue*),  $\text{i}$  (*breue*),  $\text{y}$ ,  $\text{k}$ ,  $\text{o}$  (*breue*),  $\text{q}$  (*per se*),  $\text{r}$  and  $\text{R}$  (*per se*, see also section 8),  $\text{f}$  and  $\text{s}$ ,  $\text{t}$ ,  $\text{u}$  (*breue*),  $\text{v}$ ,  $\text{w}$  and  $\text{x}$ .

As for an unnamed variant of the letter  $\text{c}$ , it is not obvious how to interpret it. I assumed this is U+00E7 LATIN SMALL LETTER C WITH CEDILLA ( $\text{ç}$ , Latin-1 Supplement block) and transcribed as such.

<sup>4</sup> [mufi.info](http://mufi.info)

<sup>5</sup> For referencing MUFI codepoints I advocate the use of the M+ prefix.



Figure 5: Grossum *n* in the manuscript at the beginning of words: p. [8] l. 14



Figure 6: Nasal vowels in the manuscript

However in Junicode there is a special glyph for it, namely  $\overline{\text{c}}$  input as  $\text{c}\overline{\text{c}}$ .

In Junicode *long s*  $\overline{\text{f}}$  can be input as  $\text{s}\overline{\text{f}}$ .

## 6 Letters not listed in the alphabet summary

Most old Polish texts for the nasal vowel use U+A7C1 LATIN SMALL LETTER OLD POLISH O ( $\phi$ ) introduced to Unicode in version 14.0 [7]. However in Parkosz's treatise the vowel has the shape of U+00F8 LATIN SMALL LETTER O WITH STROKE  $\overline{\text{o}}$ , cf. the second and the third stroked letter in Fig. 6, so it was used in the transliteration. However in Junicode it can be input as  $\phi\overline{\text{o}}$  which makes its meaning clear.

The manuscript contains a single occurrence of U+2C65 LATIN SMALL LETTER A WITH STROKE ( $\overline{\text{a}}$ ), cf. the first stroked letter in Fig. 6, which could be a scribal mistake. Nevertheless it was transcribed just as  $\overline{\text{a}}$ .

Letter  $\overline{\text{z}}$  was omitted in the alphabet, probably by mistake; in the transliteration it was used without change, which is a simplification, as it looks more like  $\overline{\text{z}}$  U+0292 LATIN SMALL LETTER EZH.

The letter  $\overline{\text{h}}$  was also omitted probably by mistake.

## 7 Multigraphs

Some multigraphs in the manuscript are written as ligatures, but there is no need to represent them as such in the transliteration. On the other hand it would be desirable if the font rendered them as ligatures.

As it was mentioned earlier, some multigraphs consist of doubled vowels: *longum aa*, *longum ee*, *longum oo*, *longum uu*.

There is an open question whether the text would be more readable if *longum aa* was rendered as U+A733 LATIN SMALL LETTER AA ( $\overline{\text{a}}$ , Latin Ex-

tended-D) and *longum oo* as U+A74F LATIN SMALL LETTER OO ( $\overline{\text{o}}$ , also Latin Extended-D). For *longum ee* and *uu* the ligatures would have to be designed. The character M+E8C7 LATIN SMALL LIGATURE UU ( $\overline{\text{u}}$ ) would be rather misleading.

The other multigraphs are: *ch* (taken over from Latin), *cz*, *molle dz*,  $\overline{\text{fz}}$ , *fcb* and  $\overline{\text{fz}}$ , and *zz*.

## 8 Majuscules

Some of the examples are proper names, so they usually (not always) start with a majuscule (cf. the index in [9]).

The main Polish example is a verse, and the first words of some lines also start with a majuscule.

In consequence the transcription contained the following majuscules:  $\overline{\text{A}}$ ,  $\overline{\text{B}}$  (*B molle*, cf. Fig. 7),  $\overline{\text{C}}$ ,  $\overline{\text{G}}$  (*G per se*),  $\overline{\text{I}}$ ,  $\overline{\text{K}}$ ,  $\overline{\text{N}}$ ,  $\overline{\text{P}}$  (*P grossum*),  $\overline{\text{P}}$  (*P molle*),  $\overline{\text{Q}}$ ,  $\overline{\text{R}}$ ,  $\overline{\text{S}}$ ,  $\overline{\text{V}}$ ,  $\overline{\text{T}}$ ,  $\overline{\text{Z}}$ .

In Junicode we have now  $\overline{\text{P}}$  input as  $\text{P}\overline{\text{P}}$  and  $\overline{\text{P}}$  input as  $\text{P}\overline{\text{P}}$ ;  $\overline{\text{G}}$  can be input as  $\text{G}\overline{\text{G}}$ .

It is perhaps worth noting that the upper case of *g per se* is  $\overline{\text{I}}$ .

## 9 Concluding remarks

Adding Parkosz's letter to Junicode is in my opinion a large step forward in improving the editorial quality of publications concerning the treatise. Elsewhere I intend to present the difficulties caused by the lack of an appropriate font, including a misrepresentation of Parkosz's views.

It would be fun to have also a font simulating original handwriting, but the text would be unreadable for people without at least some knowledge of paleography; see, for example, the letters *d* and *x* shown in Fig. 8.

### A An excerpt: a word list

According to [11], this is a mnemonic verse; in that paper you can also find the English translation.

#### A.1 Pure Unicode transliteration

Some minor mistakes in [9] corrected.

Adaam bił bȳ! ca! kaal czas ąo chood daa! dzaa!  
 eřz ffitaa fięi i ęee ęe ęhaap kro! lis ęis mękaa  
 mękaa pięłki ęięki othoořz ęięęe pięchęo qras roffa

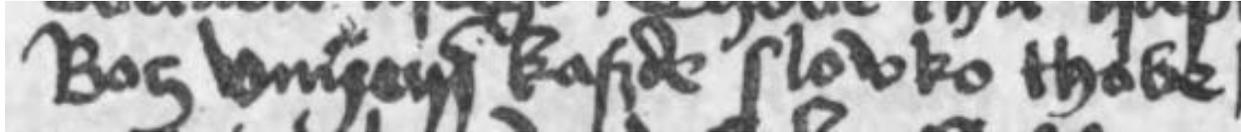


Figure 7: The manuscript p. 15 l. 24: Boç onyemj kafzde flooko thobe (transliteration)



Figure 8: Outlines (by Szymon Pilas): ç, d, ff, p grossum, w, x

rzoøffa rofuumj. faamj schaad fzaad! fzak zzaraa  
Zamjno to umee uuj wila wifaa! wfta xøodz  
yanczoc joczci jøøkaa

### A.2 Junicode transliteration

The characters accessed with tag sequences are set in italics.

Adaam bil był cal kaal czas *ça/ø* chood daa dzaal  
eż ffitaa *figi* i ee *je ghaan* krol *lis lis* mjkaa  
*mika nifki nifki* othoofz *pige pifchno* qvras  
roffa rzoøffa rofuumj. *faamj* schaad *fzaad!* fzak  
zzaraa *Zamjno* to *umee* uuj *wila wifaa!* wfta *xøodz*  
*yanczoc joczci jøøkaa*

### A.3 A sample transcription

Various publications use various transcriptions to quote the treatise. Here is the one used in [11], pp. 125–126].

Adam był bił cał kał; czas, ciało, chod dał  
dział; eż fyta figi i je je chan krol; lys lis myka,  
Mika nyski niski otoż pije pyszno kwas; rosa  
rzaśa, rozum; sam szad siadł; żak ziara za  
mną; to umie un wiła; wylał w usta ksiądz,  
jęcząc jęczy, jąka.

### References

- [1] P. Baker. Junicode — the font for medievalists. specimens and user manual for version 2, 2022. [github.com/psb1558/Junicod-font/](https://github.com/psb1558/Junicod-font/)
- [2] J.S. Bandtkie, E. Raczyński, eds. *Jacobi Parkossii de Żorawice antiquissimus de orthographia polonica libellus*. Wilh. Deckeri et Societatis, Posnaniae, 1830. [www.wbc.poznan.pl/publication/115430](http://www.wbc.poznan.pl/publication/115430)
- [3] J.S. Bień. On standards for Computer Modern font extensions. *TUGboat* 11(2):175–183, 1990. [tug.org/TUGboat/tb11-2/tb28bien.pdf](http://tug.org/TUGboat/tb11-2/tb28bien.pdf)
- [4] J.S. Bień. Standard Unicode i dawny język polski. *Acta Poligraphica* 14:7–28, 2019. [10.5281/zenodo.4058701](https://zenodo.org/record/4058701)
- [5] J.S. Bień. Traktat Parkosza. Eksperymentalna edycja elektroniczna. *Poznańskie Studia Polonistyczne. Seria Językoznawcza* 26(1):27–69, 2019. [10.14746/pspsj.2019.26.1.2](https://doi.org/10.14746/pspsj.2019.26.1.2)
- [6] J.S. Bień. Parkosz’s treatise from a typographic point of view. *Scripta & e-Scripta* 22:37–57, 2022. <https://www.ceeol.com/search/article-detail?id=1058326>. After fall 2023, also: <http://e-scripta.ilit.bas.bg/archives/year-2022/issue-22/yanush-s-bien-traktatt-na-parkosh-ot-tipografska-gledna-tochka>
- [7] D. Bunčić. Proposal to include the letter ‘Old Polish O’ in ISO/IEC 10646 and the Unicode Standard. l2/21-039, 2021. [www.unicode.org/L2/L2021/21039-old-polish-o.pdf](https://www.unicode.org/L2/L2021/21039-old-polish-o.pdf)
- [8] M. Kucała. *Jakuba Parkosza Traktat o ortografii polskiej*. Państwowe Wydawnictwo Naukowe, 1985. [ebuw.uw.edu.pl/publication/220504](http://ebuw.uw.edu.pl/publication/220504)
- [9] M. Kucała. Traktat o ortografii polskiej [Jakuba Parkosza]. Odczytanie, 2020. [10.5281/zenodo.3883863](https://zenodo.org/record/3883863)
- [10] J. Łoś. Jakóba syna Parkoszewego traktat o ortografii polskiej, 1907. [ebuw.uw.edu.pl/publication/238219](http://ebuw.uw.edu.pl/publication/238219)
- [11] R. Wójcik, W. Wydra. Jakub Parkoszowicz’s Polish Mnemonic Verse about Polish Orthography from the 15th Century. In *Culture of Memory in East Central Europe in the Late Middle Ages and the Early Modern Period*, R. Wójcik, ed., no. 30 in *Prace Biblioteki Uniwersyteckiej*, pp. 119–127. Biblioteka Uniwersytecka, 2008. Ciążeń, March 12–14, 2008. [hdl.handle.net/10593/6037](https://hdl.handle.net/10593/6037)

◇ Janusz S. Bień  
Warsaw, Poland  
[jsbien \(at\) uw.edu.pl](mailto:jsbien@uw.edu.pl)  
[sites.google.com/view/jsbien](https://sites.google.com/view/jsbien)  
ORCID 0000-0001-5006-8183