16th century Latin printed brevigraphs in Unicode: a computer resource Janusz S. Bień

(retired)

jsbien@uw.edu.pl

Abstract

The content of a public git repository is presented. It contains the brevigraphs, i.e. a specific forms of scribal abbreviations found in the two editions of Stanisław Zaborowski's Latin treatise on Polish spelling entitled *Ortographia seu modus recte scribendi et legendi Polonicum idioma quam utilissimus*. The brevigraphs are encoded in Unicode following the principles of typemic transcription.

Introduction

The very first computational approach to Latin abbreviation seems to be Olaf Pluta's *Abbreviationes* TM: A Database of Medieval Latin Abbreviations awarded the 1993 German-Austrian Academic Software Prize (Deutsch-Österreichischer Hochschul-Software-Preis) for outstanding software in the humanities. In 2015 a new version, called Abbreviationes TM Professional, was released, which uses Unicode. To use the software a paid license is required.

The repository presented here is an open resource. Everybody can use it for any pupose, modify it and distribute modified version etc.

3.2 Indexes for DjVu documents

From the technical point of view the indexes are just simple CSV files (with the semicolon as the separator). The indexes can have other uses beside being browsed with djview4poliqarp.

	parkosz - DjView dla Poliqarpa	_
<u>Plik W</u> idok <u>Przejdź Index Pomoc</u>		
Serwer: https://szukajwslownikach.uw.edu.pl • • • • • • • • • • • • • • • • • • •		 ZaborowskiABr.csv: 2,508 item(s) Find: dura durar_R * durarum
Kwerenda: • Szukaj nouice naich & latis male: gaibi chnon talë fonü fcribüt p s duplex fic Proffoz. i.omniü. Regularius tñ talis fonus p s aliarü litteratz duratz fignari debetfic 1 ftaŭt & tercius fonus mollis: g etiã mo gnari debet. gnari debet. @Dew & V. pi in latinis dictiõibus nõ em eftlittera, elius docuit, vult & non wlt.vulpes & turatz fignari por cpo mi folettverü in bis vulgaribus e filibus cal pere fonatz eft bi c durütideo more aliarus duratz fignari p ficro vnico fugius polito fic ĉop.ĉastejo el regularins. Jn bis vero vulgaribus clo.i.teloneñ.oÿca s.pis ce bula.i.cepe.cvöe.i.alienű z filibul 3 poft c nõ tebet ponitnec aliquo puncto fignarita bie	alphabetilatini, q etiam patet in his vulgaribus pan.i.dñs. lan.i. Ioannes. Ecce hui? littere a quatuor notabiles funt foni aliquans do & fignificatū dictionū variantes. Accipe exm de his vltimis fonis in hoc Polonico üercimak. In quo fi a cum puncto inferiori fcribatur erūt due dictiões latine fignificantes tere vel terit papa uer. Si cum fimplici a erit vna dictio, que latine dicii tribulū alias üercoch.fic droga fi fcribat p a cū pūcto inferiori dicii vialatine. fi vero fine puncto fignificabit idē qd' preciofa latine.o etiam in primo vulgari eft durū, & licet plures fint foni huius littere a, Po lonis tamen prefertim illiteratis, hi quatuor fufficiūt, qui funt ma gis principales ad quos alij reducunt. "C in vulgari in fuo, pprio fonolegi debet, fecus eft in latinis dis ctionibus, vbi c cū a vel o vel u cofillabicū loco k poni folet & eit fonū habet. In dictiõibus yo vulgaribus c loco k nūcip poni folet & eit fonū habet. In dictiõibus yo vulgaribus c loco k nūcip poni debet,	durarū * durarum durarū * durarum durarū * durarum durarū * durarum ourarū * durarum ourar * durarum ourar * durarum ourar * durarum ourar * durum durū * durum



1 Printed texts

Typecases and *typème*; typoglyphs and typochars. Cf. Fig. 9.

2 Unicode

2.1 Basics

Unicode characters are not the same as user-perceived characters. The notions of *typeme* and *textel* was proposed to remedy this.

In theory Unicode can be easily extended by interested communities by agreeing on the use of private characters. Medieval Unicode Font Initiative is a good example. However the private characters are crippled because they are missing the properties provided by the Unicode Character Database. Even if the properties are provided, it is impossible or prohibitively difficult to make programs to use them.

One of the main Unicode design principle is the distinction between the characters (some abstract objects) and glyphs (their visual images). Another important principle is that Unicode encodes characters, not glyphs. It is difficult to apply this principle in practise, because the difference is not always clear.

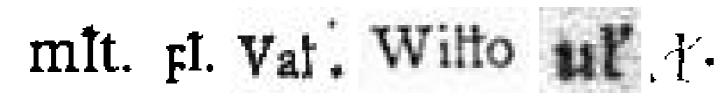


Figure 1: Different glyphs of **U+A749** 'LATIN SMALL LETTER L WITH HIGH STROKE' — the representative glyph is **†**

Theoretically there is a method to circumvent this principles by using so called character variation sequences. The problem is that every usage of variation sequences has to be officially registered by the Unicode Consortium. Up to now no proposal concerning the Latin script has been accepted.

In March 2022 Margaret Kibi (marrus-sh) proposed to use tag characters instead of the variant sequences in the Junicode font; the tags themselves are invisible but can be be visualized for documentation purposes, see below. The proposal was supported by several font users and accepted by the font author. I hope it will become a kind of a *de facto* standard.

In consequence the important glyph variant of the character discussed above can be encoded as l. yielding l' (it can be named *l with high stroke ending with flourish*).

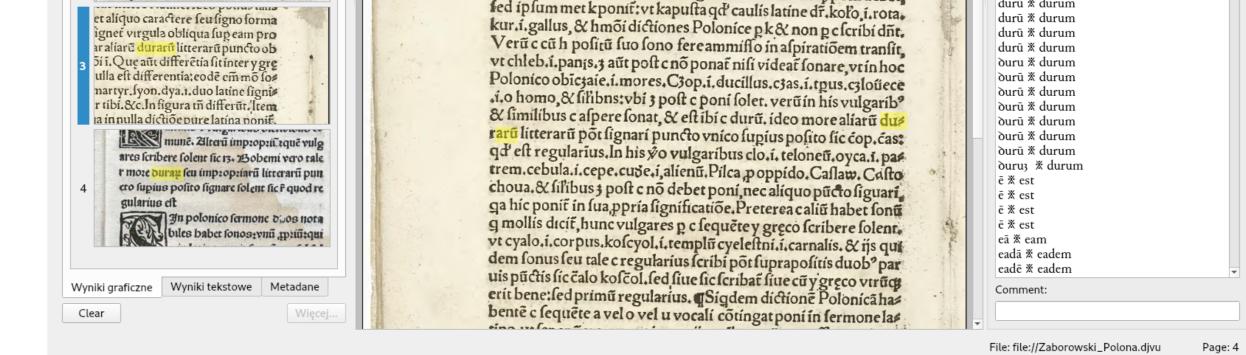
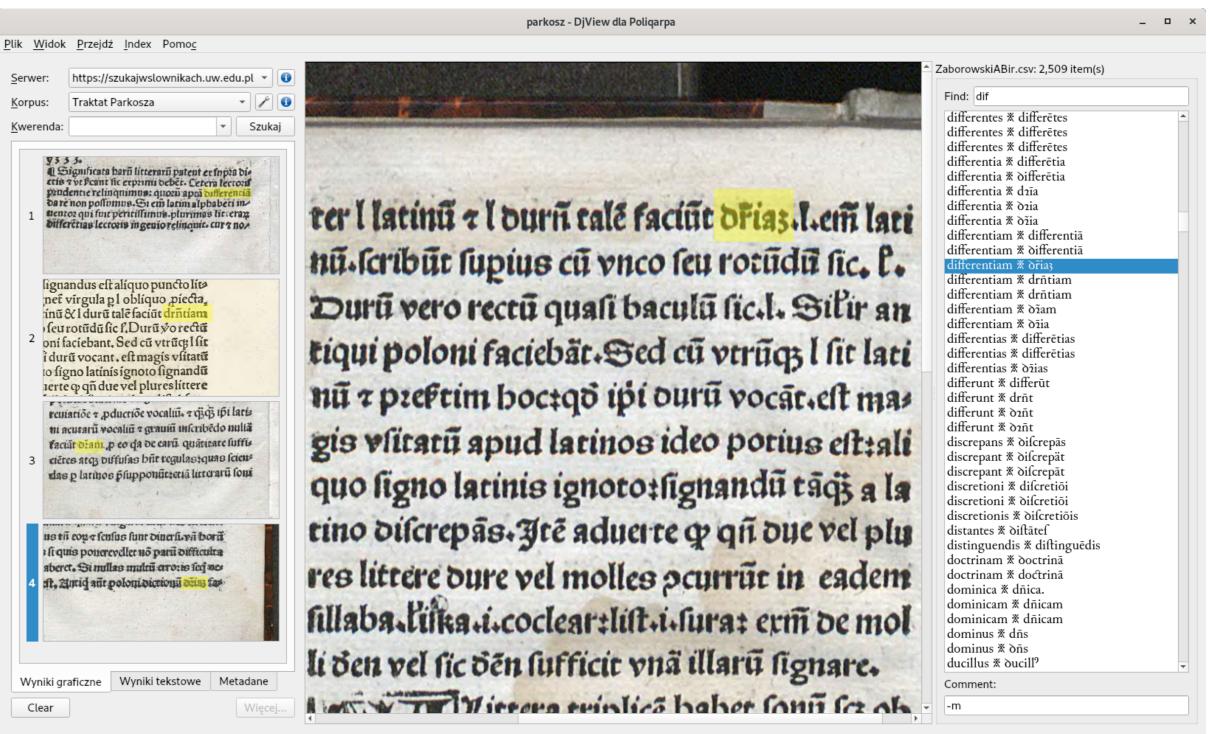


Figure 6: A primary index: the entry is an abbreviation



2.2 Examples of encoding problems

vi vi regla popfus

Figure 2: From the left: vel, vel, regula, populus (Zaborowski's treatise)

'LATIN SMALL LETTER L WITH HIGH STROKE' (1) not obvious, Junicode 1 [] [] (I': l with high stroke ending with flourish).

alit fimilit

Figure 3: From the left: *aliter*, *similiter* (Zaborowski's treatise)

U+035B 'COMBINING ZIGZAG ABOVE'? U+F1C7 'MUFI COMBINING ABBREVIATION MARK ZIGZAG ABOVE ANGLE FORM' and U+F1C8 'MUFI COMBINING ABBREVIATION MARK ZIGZAG ABOVE CURLY FORM'? Junicode: ² (주도보) and ³ (주로보)?

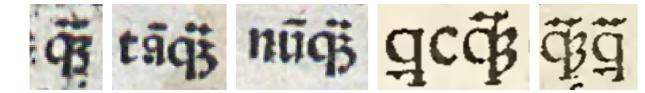


Figure 4: From the left: quam, tantquam, nunquam, quicquam, quamquam (Zaborowski's treatise)

U+A76B 'LATIN SMALL LETTER ET', U+E8BF 'MUFI LATIN SMALL LETTER Q LIGATED WITH FINAL ET', U+A757 'LATIN SMALL LETTER Q WITH STROKE THROUGH DESCENDER'

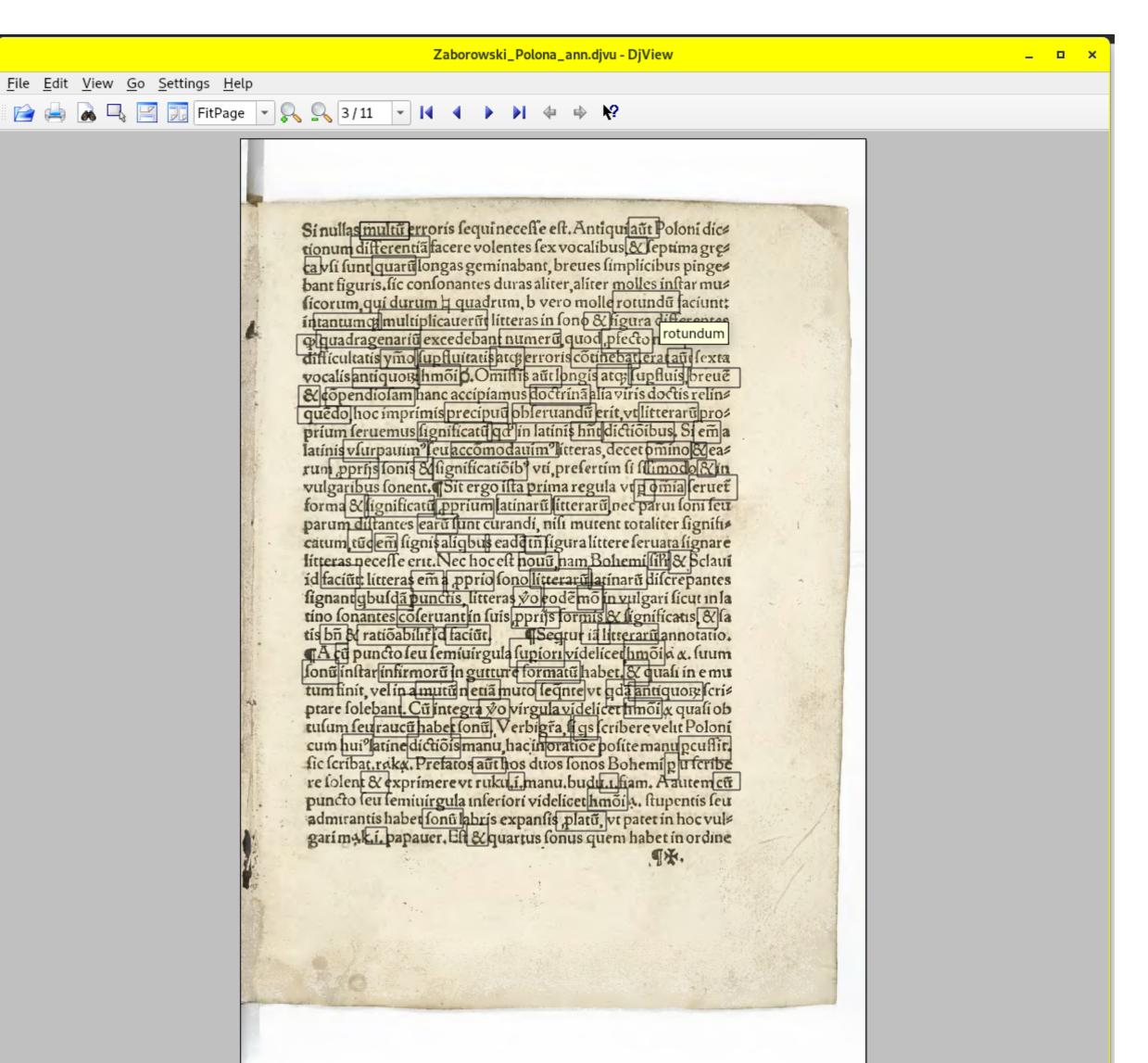
U+0305 'combining overline'? U+0304 'combining macron'? U+0303 'combining tilde'? U+1DD3 'combining latin small letter flattened open a above'?

3 DjVu

3.1 The format

The format was developed in 1999-2001 for serving scans and underlying text layer over Internet. The open source viewer djview4 is still actively maintained. A typical DjVu document internally contains a dictionary of glyphs (actually connected components).

Figure 7: A secondary index: the entry is the abbreviated word



traktatParkosza_041-054_faksymile_Transkribus.djvu - DjView	_ _ ×	
Edit <u>V</u> iew <u>Go</u> <u>S</u> ettings <u>H</u> elp		
🚔 🙈 🖳 📴 🗊 Fittwidth 🔹 🛼 💁 1 🔹 🕨 4 🔹 🕨 4 🍁 🌿	DjView Shapes	
	File Help	
lesus Christus		
3	o nomo ousnow-uses przenizenize a okacional ruszys ko- bi ska Czerwonego antonicze zdywała a ska z Sieńska do kacional skala skala skala skala skala skala skala skala	t Szlachta, 174
	DLSKA Czerwonego Antiguitas zażywalac a tak a Sigers in iswa nazywal bę. sZLA: HERBY zowia fię inaczey ARMA, z tey rac	
	2 fwoie Swint o changemail, rynkrautkiem lie piec	zętowali, partim też, iż na
Pugna L mun h ipsam	Tarczach malowane nofili Oreże, lub wyrznięte. 2	tad dotychezas na znak Ry-
dgin a commence of the state	WA bete set as forwar Gale v Offshich estimated and a state of the set of the	
di neritoria.	acve, GELONI, AGATTRSOWIF, Szlacheckwo (mo-	noty, zowią lię od Pierście-
Pine late and letturner	3 rem: nim febie farbuiąc Twarze y Ciało, według niow, Sygnetow, ktorych zażywali Starożytni y te Ciż Brytańczykowie, y Hollaudowie na Ramieniach z Herbami śwemi.	erozoieyfi, do pieczętowania
Patr	fobie, wedtug Lipfinfza THRACES, Zichość fwoię TARCZE albo POLAHI	
utur medine Hind lest quod	um oznávmisli fujern kolorem werthe nemienio- TARCZE, albo Pola na HERBY bywaia w tryan	ant w fundament v olera
inco + ritas: Sacius, reipublice	Carnym bywaią dyflyngwowane. 1. ZLOTY Kolor, y ZOLTY znśrzy Światlość, bo niadze Scuta, pulgo Szkudy z Włofka, ktorych za	
ore insudance and a second address of address of address of address of a second address of a second address of a	4 ieft od przezroczyfiego powietrza, według /jidora zowie przy takża ukrzątew formy brz żadnego an	seute, zowie fie dneile, jakie 3 3 5 5 5 5 5 5 5 5 3 3 3
	neratow, znaczy, dum Soll,, albo Nudzielę, znaczy też za Numy Pompiltufza Krola Rzynifkiego ipadło z	
	trza, y z Wielzczkow interpreta yi, znaczyło Ævi cyl, ktoraby ten Poklerz konferwowała wiecznie.	
redicamentiskiic ibus, impossibile est aliquid orum / iduest k international internation	Gr hardem defectunt y Xieltwo Mazowieckie Koronie le Marka Bożka, złożył go, y dwunaftu Kapłanow	
	5 Koiatowicza. Z ktorych precipui byli Słuccy, Kijowicy, W Pukierze bolic im każąc, według Owiatujza.	
noraliter.) Und	Niektorzy Clarefcunt podziś dzień Natałum, Virtuun dore, jako to: WISZNIOWIECCY, CZARTORYSCY, (derent Kleynot, Miałło trzech Bofonow, olim za	LICOD de L. L. CL
ariishonorib atetreorum gesta inspicere	WORONFECCY Xiata Stucking Succingent LLO O ragiew AureaFlamma, albo Złoty Płomień nazwa	na, z materyi iedwabney zło- W W W W W W W W W W W
adit aude	s sZLACHECTWA było, Szabli dobywać pod czas tew tkaney, w Kościele S. Dyonizego konterwowa Francuzom, przeciw Niewiernym: Lecz że też y n	Chrześcian podtymie wos
contist of the second s	6 + nofić: Dwory także y Rezydencye Szlachty Politiey, iowali znakiem, zniknęta ta Choragiew Złotego	
tonus. Oumpl. Aureavictricestophubit taurea crines, Ducit portipa comum,	winowsyce do nich umknawfzy, maia te pro Afriti, ieni bydž nie moga, chyba arciztowani, weotog Kon- FORMY kwadratowey TARCZE zowia fiç	Toffers an Grade sind WWWWWWW 1 1 1 1 1
sequitur confusio victum. Hackeciam saluberrima ratione exigente et quasi	sheine set Damt, Pitzee &c. to will winnemit to me. kwadrzaffe, CLIPFL, byli TARC7F, zrzejha jaka	
sequitur confusio/victum. Hacteciam (satuberrima ratione) (exigente) et quasi x, necesitate, inducente, nedum (apud) fideles, verumeeiam (et/circa (gentiles,), ut efert/Aristotiles, / quotienscumque, quempiam, pro/comunitate (certantem) con ;	Miaflo Gdafúk, Eiblag, Toruń, poczeli czerwonego zażywać Wo	
efert/Aristotiles, quotienscumque quempiam prorcomunitate certantem con-	towaniu (woich rożnych Inflrumentow, 7	
ngit, ponere, vitam, ipsum honorifice sepultum in suis superstitibus, puta tiis et filiabus, fovere, indesinenter favoribus ratum atquergratum	w HERBACH KOLORÝ o	o ZNACZA?
Iliis ettifiliabus, iroverei indesinenteri Iravoribus ratum atque/gratum	PCLA HERBOWNE albo Złotym, albo Srebryy	
omunitati extitit. Claret ergo patriam defendenti gloriam attingere miuti.	ab alexandro. STEMMA zas uformowało frę od Greek e- Carnym bywaia dyffyngwowane.	m ³ Sucondary Latherondan ³
Nos itaque monitu talium inducti, non quod intendamus laudi, iquoniam hoci ipsum	to tell Auronos, wienics, store takze ex Herrors, to tell	
olius, Dei est, sed inspicientes ad nutum, ex quo reipublice, agitur comodum).	Dikcyonarzu u. Językow. STEMMATA redy albo HERBY refr od przezroczyfiego powietrza, według Ijudora	
ensuimus ex causis infra dicendis paternum idioma, quod notabiliter traximus x Polonorum lingua, fore caracteribus latinis cum paucis differenciis appositis	tiuta, fa Znaki, Znamiona, ktoremi Szlachetni od Niefzla- Ll	m 3888888888888888888888888888888888888
R. Polonorum lingua, tore caracteribus latinis cum paucis differenciis appositis	y TARCZE zowią fię Tiffera po Greeka, toiett	22222222222
scribendum,) ut in hoc, quantum ad presens attinet, contra cacetem id est malum	FARCZE, zrzeżbą iską na tobie, od flowka olypice,	
this ufficientem scripture polonice/ aborantes/ videamur patriam comodose	Y 9 Yanna manufacti DELT & hula Tarma from nut	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8

Figure 5: djview4: the scan and the underlying text, djview-shapes: shapes similar to the letter C

P3/11 2000x3008 300dpi x=1332 y=2471

Figure 8: An index converted to the tooltips

4 Final remarks

Perhaps in some future Zaborowski's treatise will be encoded also in TEI XML, but for the present purpose the tools and resources described are fully adequate.



Figure 9: A ligature type and its printed image (Daniel Ullrich, Threedots, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=855947)