# Perso-Arabic Orthography for the Polish Language

Gustek

August 23, 2025

### I. Introduction

This paper presents an adaptation of the Perso-Arabic script for the Polish language. Although based on a linguistically, historically and geographically foreign script, the following orthography is actually very "Slavic" in essence, the uniform marking of soft (*miękkie*) and hard (*twarde*) consonants being at the core of the writing system. Being based on an *abjad*, the soft/hard distinction is made by altering the consonant glyphs, whereas the Cyrillic script does so through different vowel glyphs (and the use of "soft/hard signs").

The orthography borrows heavily from Persian, especially in how it represents vowels; the way it distinguishes between soft and hard consonant can be (rightfully) likened to how Pashto indicates the retroflex equivalents to dental consonants. However, some technical constraints (namely the lack of glyphs in the Unicode standard) have prevented us from achieving consistency in how the soft version of a letter is derived.

We will first present the consonants (in the *abjadi* sense) signs, then the diacritical vowel marks and how vowel sounds can be combined. After this, we shall go over some exceptions created by the Latin orthography, present the process of loaning words from Arabic and provide some examples.

### II. Consonants

We present the orthographic consonants in the traditional order of the Perso-Arabic script, along with what letter (or digraph) their correspond to in the "standard" Latin Polish orthography) between strokes. Letters that are only found in foreign words are indicated by a star.

There are 39 letters and 12 soft variations, as follows: 1/a, 1/a,

There are 12 hard/soft pairs:  $\psi$  /b/ and  $\psi$  /bi/,  $\psi$  /p/ and  $\psi$  /pi/,  $\psi$  /t/ and  $\psi$  /ć/, and  $\psi$  /d/ and  $\psi$  /r/ and  $\psi$  /r/  $\psi$  /r/ and  $\psi$  /mi/,  $\psi$  /n/ and  $\psi$  /mi/,  $\psi$  /m/ and  $\psi$  /mi/,  $\psi$  /m/ and  $\psi$  /wi/.

Softness being indicated on *all* consonants allows for more etymological precision that the standard Latin orthography. For instance, words such as  $\dot{z}uraw$  (crane) have a soft ending, which is not displayed by the standard orthography, but is by this writing system:  $\dot{z}uraw$ .

 $<sup>^{1}</sup>$ Word-initially, | can be twritten as  $\sqrt{(alef\ z\ maddem)}$  when representing /a/ to avoid ambiguity. This letter is also found in loanwords.

<sup>&</sup>lt;sup>2</sup>The letter o (nun nosowy) is used to represent a nasal consonant, similarly to what was done in older Polish texts, e.g. zqb (tooth) being written as zomb.

 $<sup>^3 \</sup>mathrm{For}$  an explanation of the role of  $\ensuremath{\mathcal{Z}}$  (and conversely, of  $\ensuremath{^\dagger}$  and  $\ensuremath{^\dagger}$  ), see Section IV

#### III. Vowel marks

Whereas the /a/, /u/, /i/ and /y/ vowels are always explicitely written, /e/, /o/, /ó/ and partly /a/ and /e/ are indicated by three optional diacritic marks (shown here on the letter be):  $\dot{\psi}$  /e/, /a/ in some loanwords (fetche),  $\dot{\psi}$  /e/ (kiesre) and  $\dot{\psi}$  /o/, /ó/ (zemme). Another diacritical sign,  $\ddot{\psi}$  (szedde), indicates gemination.

The letters g (waw) and g (ig) cannot represent vowels word-initially, due to their dual consonant/vowel nature. When writing words starting with ig/i or ig/u/, an ig ale ig is placed before the corresponding letter, such as in ig ig (to go).

When writing words ending with the letter /o/, a waw is appended to indicate it, such as in szybko (fast).

Given that some vowels are not explicitly written, some words featuring vowel clusters could be misread. For instance, Europa (Europe) cannot be written as اورپا as that would be read as /uropa/. Instead, the special letter و is used, and the word is written as follows: اؤرپا. The same goes for clusters featuring a ja or an alef.

## IV. Exceptions

Some Polish words (of foreign origin) feature digraphs such as /ti/, /ri/ and /di/, in which the consonant is not soft, but hard. However, in this orthography,  $\ddot{z}$  and  $\ddot{z}$  indicate /ty/ and /ci/ rather than /ti/. In such cases, the special letter  $\ddot{z}$  ( $ja\ z\ chemzem$ ) is used Instead of the standard ja, e.g.  $\ddot{z}$  tiara (tiara).

The process of loaning words from Arabic is made pretty straightforward by the orthography. The only operation required is the transformation of  $t\bar{a}^{\,2}$   $marb\bar{u}tah$  into a te, e.g. رصحت (mercy) becomes رحمت. The fetche diacritic is read in most loanwords as /e/, except in high-prestige words, such as prophets' and saint's names, in which it is read as /a/, e.g. مُحْر /Mochammad/, عُمْر /Ali/, عُمْر /Omar/, etc.

# V. Examples

The letters have the following names and are ordered as follows:

Alef, be, bie, pe, pie, te, cie, se, dżym, cze, che, che, ce, dzym, dał, dział, zał, re, rze, zej, ziej, żej, syn, sin, szyn, sad, zad, ta, za, ejn, hejn, fe, fie, kof, kaf, gaf, łam, lam, mym, mim, nun, niun, nun nosowy, waw, wiaw, q, e, ę, ja, ja z hemzem.

The following is taken from the first poem of Adam Mickiewicz's *Crimean sonnets*, *Stepy akermańskie*:

Wpłynąłem na suchego przestwór oceanu, Wóz nurza się w zieloność i jak łódka brodzi; Śród fali łąk szumiących, śród kwiatów powodzi, Omijam koralowe ostrowy burzanu.

Już mrok zapada, nigdzie drogi ni kurhanu, Patrze w niebo, gwiazd szukam, przewodniczek łodzi; Tam z dala błyszczy obłok? tam jutrzeńka wschodzi? To błyszczy Dniestr, to weszła lampa Akermanu.

Stójmy!... Jak cicho!... Słyszę ciągnące żurawie, Których by nie dościgły źrenice sokoła; Słyszę, kędy się motyl kołysa na trawie,

Kędy wąż śliską piersią dotyka się zioła. W takiej ciszy — tak ucho natężam ciekawie, Że słyszałbym głos z Litwy. — Jedźmy, nikt nie woła!

وپلینىلم نا سوحگو پرستور الحأنو،
وز نورا ښۀ و ریاښټ ای یاک لدکا برډی؛
ښرد فالي لمک شوممىڅیح، ښرد کواتو پوډی،
امییام کرالوه استروی بوړانو.
یوژ مرک زاپادا، نیگډه درگی نی کورغانو،
پاتړۀ و ڼبو، گوازد شوکام، پرودنیچک لډی؛
تو بلیشچی دڼستر، تو وشلا پامپا آکرمانو.
تو بلیشچی دڼستر، تو وشلا پامپا آکرمانو.
کتریح بی ڼه دښتیگلی بررنیڅه سکلا؛
سلیشۀ، کمدی ښۀ متیل کلیسا نا تراؤه،
کدی ویژ ښلیسکؤ پرښؤ دتیکا ښۀ ږلا.
و تاکی ټیشی — تاک اوحو ناتیزام ټکاؤه،
ژه سلیشالیم گلس ز لیتوی. — یډمی، نیکت ڼه ولا!