

On the Etymology of Diacritics in General and the Origin of the Czech Diacritics in Particular

Daniel Bunčić (Cologne)

This paper demonstrates that diacritics have their own etymologies, just like words. They are created according to certain mechanisms (based on iconicity, from letters, from deletion marks, or from disambiguation marks), and when they are borrowed from one language to another, both their form and their function stay the same or change according to general rules and as necessitated by the concrete borrowing situation.

However, the authors of hypotheses about the origin of the Czech diacritics do not take this into account. Therefore, all the hypotheses put forward so far about possible sources of the dot (the later háček) and the stroke (čárka) are evaluated anew (viz., marks in earlier Czech texts, the *i* tittle, Hebrew, Glagolitic, the Greek acute, the Latin apex, and Old Irish). It turns out that the seemingly most far-fetched idea is by far the most plausible: The Czech diacritics were borrowed from Old Irish.

Keywords: graphematics, orthography, Jan Hus, De Orthographia Bohemica, háček, caron, čárka, acute accent, apex, overdot, Old Irish, Glagolitic, Hebrew

“Everything is the way it is
because it got that way.”
Kenneth E. Boulding¹⁾

1. Introduction

The word *etymology* in the title of this article might look inappropriate. After all, *words* have etymologies, and morphemes too – but diacritics? But think of letters. We know that the Latin letter ⟨A⟩ goes back to Greek alpha ⟨Α⟩, which is derived from Phoenician ⟨𐤀⟩, which was called *ʾālep* /ʔa:lɛp/ ‘ox’, because the letter originally depicted an ox head ⟨𐤀⟩. We can follow the history of this letter’s shape, understand how the original ox head was simplified and rotated until it came out as ⟨A⟩. We can also follow the history of the letter’s function: In Phoenician, it was used as a consonant letter for /ʔ/, and this sound can be explained by the “acrophonic principle” (despite the objections by Gelb 1963, 251), i.e. by the fact that *ʾālep* ‘ox’ starts with the sound /ʔ/. Apart from that, the Phoenicians sometimes used ⟨𐤀⟩ as a ‘mater lectionis’ to signify a vowel. When the Greeks adapted the Phoenician letters to the Greek language, they systematized this use (cf. Gelb 1963, 181f.), turning the symbol into a vowel letter for /a/, because the Greek language had no phoneme /ʔ/ and its structure needed the systematic representation of vowels more than Semitic languages do. We can also explain on the basis of historical phonology why ⟨A⟩ for /a/ also came to stand, for example, for /æ/, /ɛ̃/, and /ə/ in Modern English (cf. *bat*, *late*,

1) Boulding (1953), an economist writing about growth, attributes the phrase to the biologist D’Arcy W. Thompson, referring to his 1917 book *On growth and form*, but seems to have coined it himself (cf. Branch 2015).

atop). This history of the letter ⟨A⟩ therefore fulfils the requirements of any etymology: to plausibly explain the historical changes of both form (signifier) and meaning (signified) on the basis of general principles.

Regarding diacritical marks, however, it is often assumed that they do not need to be explained in this way. After all, to the modern linguist it seems straightforward that, if an alphabet does not have enough letters for all the phonemes of a language, one can simply attach some kind of mark to some letters (e.g. ⟨è⟩ for [ɛ] in French, ⟨č⟩ for [tʃ] in Czech, ⟨ş⟩ for [ʃ] in Turkish, etc.), which can “also be used entirely unsystematically” (“auch ganz unsystemat[isch] verwendet werden”, Glück 2016, 144). One aim of this paper is to show that this notion is wrong, that all diacritics have an etymology just like words and letters, and that in order to trace their origin one has to have a plausible explanation, based on general principles, for all changes in both form and function.

The other aim of this paper is to find the source of the two Czech diacritics that were proposed in the 15th-century treatise *De Orthographia Bohemica*: the dot (·), which later turned into the háček (ˇ), and the acute-like stroke (´). A great number of possible inspirations have been suggested, but most of these suggestions are based on similarities as superficial as the ones on which medieval word etymologies were based, considering “dots and strokes above the letters – whatever their function may have been” (“Punkte und Striche über den Buchstaben – mag ihre Funktion wie auch immer geartet gewesen sein”, Stejskal 1971, 270) enough evidence to assume a connection. Some hypotheses even seem to be wishful thinking in much the same way as the false derivation of the name of the Slavs from **slava* ‘glory’. If a scientific method is introduced into this area of linguistics, i.e. if the history of diacritics has to be plausible for both the signifier and the signified, it turns out that the only possible source for the diacritics is one that at first glance seems the most far-fetched.²⁾

After giving a short overview of the content of the *Orthographia Bohemica* and the circumstances of its genesis (section 2), I will first provide a general definition of what diacritics are (section 3) and then examine how they are created (section 4) and how they are borrowed (section 5) before evaluating all the hypotheses about the Czech diacritics’ origin put forward so far (section 6), which will reveal which of them is the most plausible (section 7).

2. *De Orthographia Bohemica*

2.1 Content of the treatise

Until the 15th century, Czech was written in a spelling called ‘digraph orthography’ [spřežkový pravopis]. Then, at the beginning of the 15th century, the treatise *De Orthographia Bohemica* ‘About Czech orthography’ written by the church reformer Jan Hus (John Huss) introduced the modern diacritical orthography [diakritický pravopis].

However, the preceding sentence includes at least four doubtful statements: (1) the treatise actually has no title, and “De Orthographia Bohemica” is just a label conventionally given to it by modern scholars; (2) the text is anonymous, and its common attribution to Jan Hus is not substantiated by firm evidence; (3) different scholars date the text to

2) These two ideas have already been the subject of Bunčić (2020b) and Bunčić (2020c), respectively. The present paper is partly based on these two papers (which were written in German and published in Germany) but contains many new arguments and a lot of additional material as well as recent literature not yet incorporated in the preceding papers – first and foremost the new edition of *Orthographia Bohemica* by Voleková (2019).

anything between 1406 and the 1440s; and (4) it is not at all clear whether the author of the treatise was also the inventor of the diacritic orthography. We will come back to these issues in subsection 2.2.

It has been noted in many places that the treatise provides an extraordinarily exact phonemic analysis and description of the pronunciation of 15th-century Czech (e.g. Glück 2005). The author of the treatise disapproves of the use of digraphs and trigraphs to represent Czech phonemes, which were current at the time, and advocates a one-to-one representation. In order to achieve this, he introduces two diacritical marks. One of them is a dot (˙) on consonant letters (č đ ě ň ř š ț ž) to indicate a pronunciation different from the one in Medieval Latin:

si non ponitur punctus rotundus super litera aliqua ex iam dictis, tunc debet pronunciari more Latinorum. Sed si ponitur punctus desuper, tunc ad ydionia Boemicum debet flecti. (<i>Orthographia Bohemica</i> , fol. 36r, Voleková 2019, 50) ³	if no round dot is placed above one of the mentioned letters, it is to be pronounced the Latin way; but if a dot is placed on top, then it has to be changed according to the Czech language.
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

The dot (˙) later turned into the modern háček (ˇ) (or caron; Schröpfer 1968, 27–28; printed as an apostrophe on letters with ascenders: (č đ ě ň ř š ț ž)). Note that in the orthographic system advocated in this treatise the dot has nothing to do with palatalization but is a mere marker of a non-Latin pronunciation. Consequently, the author places the dot over the ‘hard’ *l rather than the soft *l because obviously the latter was more similar to Latin l, whereas the former was probably pronounced as a velarized /l/. Since ‘hard’ *l and ‘soft’ *l later merged into /l/ in Czech, the háček could then be reinterpreted as a marker of palatalization, so that in the modern Slovak orthography (already in Bernolák 1790) (ḷ) represents ‘soft’ /l/. Note that very early on the function of the dot was extended to also mark /g/ (which basically only appeared in foreign words) as (ḡ) to distinguish it from (g) otherwise representing /j/ (e.g. (angel ḡbriel) for /anjel gabriel/ in a manuscript from 1417, Voleková 2018, 137).

The other diacritic described in the treatise is the stroke (˘) for long vowels ((á é í ó ú ý)), which the author calls “gracilis virgula” (“thin slash”; *Orthographia Bohemica*, fol. 41r, Voleková 2019, 74). Nowadays it is called *čárka* ‘stroke’ in Czech, and since in typography it is more or less identical to the acute accent it is often called *acute* in English. However, because of its different linguistic function we ought to distinguish the Czech stroke from the Greek or French acute.

The author of the treatise does not simply give a neutral description of how these two diacritics can be used. Instead, he argues vigorously for using them, lamenting that the Czechs have to use foreign letters that were designed for the Latin language, whereas the “Jews” and the “Slavs” have their own alphabets (fol. 35r f., Voleková 2019, 46–47), ridiculing the use of digraphs like (ff) for /f/ and (zz) for /z/ because doubling a letter cannot change its pronunciation (fol. 41r, Voleková 2019, 76–77) and in the end imploring the readers to observe at least some of his rules even if they cannot follow all of them (fol.

3) I quote the *Orthographia Bohemica* from the latest edition by Voleková (2019), which is “without doubt more reliable than the preceding ones” (“senza dubbio più affidabile rispetto alle precedenti”, Tomelleri 2020, 206). However, the page number of the original manuscript is always given as well to make it easy to use Schröpfer’s (1968) or other editions of the treatise as well. The English translation by Marcela Koupišová and David Livingstone from Voleková (2019) is used where referenced; if no reference follows the translation, it is mine.

41v f., Voleková 2019, 78–79). Moreover, throughout the text he gives reasons for all his choices. The text should therefore be regarded as argumentative rather than just descriptive (cf. Schröpfer 1968, 13).

The use of the two diacritics argued for in the *Orthographia Bohemica* is attested in Czech manuscripts from 1414 (Šlosar 2010, 204) and, less consistently, in 1413 (Voleková 2018, 134f.). In handwriting, the system is generally accepted until the mid-15th century but then the manuscripts become less consistent again (Voleková 2019, xxxviii f., lxiv f.), and the full system of the treatise is not implemented completely until the 18th century (Porák 1983, 104–105; Čejka 1999, 28–30; Berger 2012, 264–265; Bunčič 2016). In printing, however, after inconsistent beginnings (Berger 2012, 261–264; Porák 1983), in the course of the 16th century the so-called Brethren Orthography [bratrský pravopis] prevails as the uncontested spelling norm for Middle Czech. It is based on the principles of *Orthographia Bohemica*, though with some deviations necessitated by typographic restrictions: a) there is no space above capital letters and letters with ascenders, so that ⟨ř⟩ has to be replaced with ⟨ŕ⟩, ⟨č⟩ with ⟨cz⟩, etc.; b) the tittle on the ⟨i⟩ already has the shape of a stroke in the blackletter typefaces used for Czech until the early 19th century (e.g. [i⁴] in Schwabacher, [i] in Fraktur; cf. Voleková 2018, 135), so that instead of placing a stroke on the ⟨i⟩ to indicate /i:/, the ⟨i⟩ has to be doubled, which results in ⟨ij⟩, which in turn is subsequently simplified to ⟨j⟩ (see the quotation from Palacký in 2.2 below as a late specimen of this orthography). This spelling system was used in the Kralice Bible [Bible kralická] of 1579–1593, which came to be viewed as the epitome of good Czech (and was praised even by Catholics, who for a long time were officially forbidden to read this Protestant translation; cf. Berger 2012, 264). In the 1840s (cf. Berger 2010), Schwabacher was replaced with Roman type, which – along with modern printing technology – made it possible to implement the principles of *Orthographia Bohemica* consistently, e.g. by using ⟨š⟩, ⟨č⟩ and ⟨i⟩, which yielded modern Czech orthography.

2.2 Circumstances of the treatise's genesis

Jan Hus's authorship was claimed already in František Palacký's first short note about his discovery of the text, which is quoted here in full:

5) M. Jana z Husince pogeďnánj latinské, o české dobropjsennoŝti od něho uŝtanovené. Z o tomto dŝležitém pogeďnánj, kteréž ŝe také v Třeboni našlo, hodně geŝt abychom buďaucně mjŝněgŝj zprávu dáli. (Palacký 1827, 134)

5) Latin treatise by Master Jan of Husinec [i.e. Hus] about the Czech orthography established by him. About this important treatise, which was also found in Třeboň, it also seems in order that we give a more extensive account in the future.

No arguments for this attribution were given, but many have accepted it “without reservations” (“ohne Bedenken”, Miklas 1989, 18). However, there is still no proof that the treatise was written by Jan Hus (cf. Vykypěl 2020, 97–98) – even though his authorship is not entirely implausible, given the erudition of the author, the style of the treatise and the fact that the diacritical spelling was first used by the Hussites (Schröpfer 1968, 14; Stejskal

4) I follow Fuhrhop – Buchmann's (2009) proposal to distinguish between graphemic and graphetic material by using (...) only for graphemes and enclosing graphetic units (i.e. specific characters' shapes, variants of graphemes, etc.) in [...] (cf. the well-established tradition of using /.../ vs. [...] for phonemic vs. phonetic material).

1971, 265–269; Voleková 2019, xxxvii).⁵ Nonetheless, there are alternative suggestions that cannot be ruled out (e.g. Jan of Holešov or simply an unknown author, Voleková 2019, xxxvif.), and the fact that none of the few preserved Czech texts that Hus evidentially wrote himself consistently implements the proposals of the treatise (Bermel 2007, 86; cf. also Vidmanová 1982, 88) seems a bit disturbing. As to the time of writing of the treatise, many Czech scholars prefer 1412, but this date is based on the assumption that it was written by Jan Hus, who was burned at the stake in Konstanz in 1415 (Voleková 2019, XXXVI, LVII), which makes the argument a bit circular. František Ryšánek, who did not presuppose any specific authorship (cf. Vykypěl 2020, 97), instead dated the treatise to the 1430s–1440s. The first document to be written in the diacritical orthography is commonly assumed to be the *Horčičkův sborník*, a collection of writings by Jan Hus, of 1414 (Voleková 2019, xxxvif., LXIII). However, Voleková (2018, 136f.) has shown that both diacritics advocated in the *Orthographia Bohemica* were already used, though inconsistently, in texts written toward the end of the 14th century. If the diacritics in these texts were actually added later and if the early dating of the treatise is correct, then it is probable that the author of the treatise is also the inventor of the diacritical orthography, which he described in the treatise in 1412 and which was then implemented in the following years. If, however, these dates diverge, i.e. if either the diacritics in the texts from the second half of the 14th century were really written by the original scribes or if the treatise was only written in the 1430s or 1440s (or both), then it becomes more probable that the person who invented the Czech diacritic orthography and the person who wrote the treatise about it were different people.

But does not the author of the treatise say that he invented the diacritics himself? One can indeed read it like this; for example, Voleková (2019, 47) translates the second sentence of the text (fol. 35r) as “I therefore decided (and I believe, usefully) to shorten somewhat the Latin alphabet for its use in Czech writing, to supply what has been missing and explain the difference between letters.” However, the Latin original only has *placuit* ‘it pleased’, not *placuit mihi* ‘it pleased me’, i.e. one could also translate this as “It was therefore decided (and I believe, usefully) to shorten...”. Certainly, the author later uses the first person in statements like “posui *c* cum *h* et vocavi illam literam *che* {ché} et nolui novam imponere” (fol. 35v, “I have put *c* with *h* and called the letter *che* {ché}. I did not want to introduce a new one”, Voleková 2019, 48–49), but the verb *ponere* ‘to put’ is extremely polysemous and does not have to refer to the act of inventing a grapheme; it can also be translated simply as ‘write’; in this case, this is certainly the intended meaning, because the spelling ⟨ch⟩ for /x/ had been around long before the introduction of the diacritical orthography.

The treatise therefore remains anonymous. Moreover, even if Jan Hus wrote the text, he might not have invented the Czech diacritical orthography. Or he might have invented it, but someone else might have written the treatise about it. Therefore, following the well-established practice with anonymous texts, the author of the treatise *De Orthographia Bohemica* and the inventor of the system described in it will be called *Orthographus Bohemicus* in this article. Where necessary, I will distinguish between the author of the treatise and the inventor of the diacritical orthography.

5) Schröpfer’s (1968, 14) argument that the Hussites and later also the Catholics attributed the Czech diacritics to Jan Hus, however, is rather weak if one considers how widespread misattributions are, e.g. of “Hier stehe ich; ich kann nicht anders” (“Here I stand; I can do no other”) to Martin Luther, of “Eppur si muove” (“And yet it moves”) to Galileo Galilei, of “La Garde meurt mais ne se rend pas” (“The Guard dies but does not surrender”) to Pierre Cambronne (cf. Rasper 2017), or indeed of the term *nabodeničko* to Jan Hus (cf. Pleskalová 2007, 501).

3. Diacritics: functions and forms

A diacritic can be defined as a “differentiating mark added to a grapheme to form a new grapheme which is distinctive within the specific writing system” (Bunčić 2020a). This definition excludes disambiguation marks such as the *i* tittle, the breve-like mark some people place above ⟨u⟩ in handwriting to distinguish it from ⟨n⟩, the horizontal line often placed above cursive Cyrillic ⟨т⟩ and/or below ⟨ш⟩, or the slash that can help distinguish zero from capital O, because none of them creates a new grapheme. It also excludes marks that are graphemes of their own (despite being placed above or below another grapheme), e.g. the vowel signs in Arabic and Hebrew (because it makes less sense to say that, e.g. in Hebrew, hiriq ⟨.⟩ and patah ⟨.⟩ create the new graphemes ⟨װ⟩ for /bi/ and ⟨ױ⟩ for /ba/, which are separate from ⟨ב⟩ for /b/, than it makes to say that ⟨װ⟩ consists of two graphemes, ⟨ב⟩ signifying /b/ and hiriq ⟨.⟩ signifying /i/). Functionally, these marks are not diacritics (although some classify them as such; e.g. Coulmas 1999, 127; Thümmel 2002, 162).

A complex grapheme created by adding a diacritic can signify: a) a different phoneme than the original grapheme without the diacritic (e.g. Czech ⟨š⟩ signifying /ʃ/ as opposed to ⟨s⟩ signifying /s/); b) suprasegmental information like stress (e.g. in Italian ⟨città⟩ /t͡ʃiˈtːa/ ‘city’) or tone (e.g. in Ancient Greek ⟨οἴκοι⟩ /oĩˈkoĩ/ ‘at home’ vs. ⟨οἴκοι⟩ /oĩˈkoĩ/ ‘houses’); or c) the distinction of homophones (e.g. in French ⟨où⟩ ‘where’ vs. ⟨ou⟩ ‘or’ or Dutch ⟨één⟩ ‘one’ vs. ⟨een⟩ ‘a’).

As to the form of diacritics, they can be superscribed, i.e. placed above the grapheme they modify (e.g. ⟨ä⟩, ⟨č⟩, ⟨é⟩, ⟨ñ⟩, ⟨í⟩, ⟨á⟩, ⟨ˆ⟩), subscribed, i.e. placed under it (e.g. ⟨à⟩, ⟨ç⟩, ⟨ʒ⟩, ⟨ø⟩, ⟨↔⟩), inscribed, i.e. written inside or through it (e.g. ⟨đ⟩, ⟨ł⟩, ⟨ø⟩, ⟨κ⟩, ⟨ⱥ⟩), or adscribed, i.e. written next to it (e.g. ⟨σ⟩, ⟨l⟩, ⟨ḍ⟩, ⟨ʹΩ⟩, ⟨ḷ̥⟩).

A special case of a diacritic is a diacritical element, which is not detached from the basic grapheme, e.g. in ⟨η⟩ vs. ⟨n⟩, Cyrillic ⟨ш⟩ vs. ⟨ш̣⟩ or ⟨ѣ⟩ vs. ⟨ѣ̣⟩. The division between diacritical marks and diacritical elements is not clear-cut, however, because some of the ‘classical’ diacritics touch the basic grapheme as well, e.g. the cedilla in ⟨ç⟩ or ⟨ş⟩, the stroke in ⟨ø⟩ or ⟨đ⟩, and especially the ogonek in ⟨ą⟩ or ⟨ę⟩, which in handwriting is always drawn together with the basic letter, without lifting the pen.

Sometimes letters can have the same function as diacritics. A clear case of such a ‘diacritical letter’ is ⟨h⟩ in Italian, which is almost exclusively used to signify that ⟨g⟩ has to be pronounced /g/ rather than /d͡ʒ/ or ⟨c⟩ has to be pronounced /k/ rather than /t͡ʃ/ despite a following front vowel, e.g. in ⟨spaghetti⟩ /spaˈɡetːi/ or ⟨gnocchi⟩ /ˈɲɔkːi/. The only other place where the letter ⟨h⟩ is used in Italian are four forms of *avere* ‘to have’, where *ho* (1SG), *hai* (2SG), *ha* (3SG) and *hanno* (3PL) are distinguished from homonymous *o* ‘or’, *ai* ‘to the’, *a* ‘to’ und *anno* ‘year’ by the presence of the mute ⟨h⟩. Interestingly enough, over the centuries these four forms were often spelled with a grave accent: ⟨ò⟩, ⟨àì⟩, ⟨à⟩, ⟨anno⟩ (cf. *TLIO*), which clearly demonstrates the equivalence of the Italian ⟨h⟩ with a diacritical mark. Similarly, the ⟨u⟩ in French ⟨guerre⟩ /gɛʁ/ ‘war’ or the ⟨i⟩ in Polish ⟨ciasto⟩ /ˈt͡ɕastɔ/ ‘dough’ can be viewed as ‘diacritical letters’, although in these languages the same letters can also have other, non-diacritical functions. The ‘soft sign’ ⟨ь⟩ in the East Slavic languages (e.g. in ⟨уголь⟩ [ˈuːgɔlʲ] ‘coal’ vs. ⟨угол⟩ [ˈuːgɔl] ‘corner’) is another example of a ‘diacritical letter’. The ‘hard sign’ ⟨ѣ⟩ in Russian and the apostrophe in Ukrainian

and Belarusian can also be interpreted as ‘diacritical letters’, but I would argue that all of them represent /j/ (cf. Russian ⟨объятие⟩ [Λ'bʲjætʲiː] ‘hug’ vs. ⟨обязанный⟩ [Λ'bʲazənːiː] ‘obliged’ or Ukrainian ⟨з’ясувати⟩ [zʲjasuˈvatʲ] ‘to clarify’ vs. ⟨зябра⟩ [ˈzʲabɾa] ‘gills’).⁶⁾

4. How diacritics are created

4.1 Iconic diacritics

Iconicity means that there is some kind of formal resemblance between the signifier and the signified. The most well-known examples in spoken languages are onomatopoeic words like English *cuckoo*, Czech *kukačka*, German *Kuckuck*, etc. Sign languages tend to have a larger number of iconic signs than spoken languages because of the fact that their medium is visual three-dimensional space rather than sound. Similarly, diacritics, whose medium is visual two-dimensional space (on the paper or the screen), can be iconic by virtue of resembling what they signify.

Such iconicity can be assumed for the original Ancient Greek accent marks, with the acute accent (´) marking a rising tone, the grave accent (`) a falling tone, and the circumflex (ˆ) a rising-falling tone. Of course, the sound we produce in our larynx does not really go up or down anywhere in space; but this iconicity is based on the widespread spatial metaphor of pitch (see Grutschus 2009, 147–152, on its development). The Greek and Latin names of these diacritics, ὀξεῖα / *acutus* ‘sharp’, βαρεῖα / *gravis* ‘heavy’, and περισπωμένη / *circumflexus* ‘curved around’ are based on another metaphor of pitch, the haptic metaphor (*Ibidem*, 136–140).

Iconicity is probably also at the base of the macron (¯) (Greek μακρά ‘long’) indicating vowel length and the breve (˘) (Greek βραχεῖα, Latin *breve* ‘short’) indicating shortness (and, by extension, non-syllabicity). Either the shape of the breve itself indicates how the pen drawing it only shortly touches ‘the ground’ (as any circle touches any tangent at a point with a length of zero), or the breve might actually have evolved from a point or a shorter line just like the háček derives from an original point; cf. the symbols of the Morse signs *long* (–) and *short* (·).




Another iconic diacritic originally invented for Ancient Greek is the diaeresis (¨) (from Greek διαίρεσις ‘separation’), which indicates that two vowels are not to be read as a diphthong but separately. Apparently, the two dots represent the two separate syllables.

4.2 Letter diacritics

If diacritics are not iconic, they often arise from other characters, for example letters or parts of a letter. Maybe the most famous example of such diacritics are the Ancient Greek breathing marks, which derived from the Greek letter *eta* (archaic *hēta*) (Η): the rough breathing (´) (δασὺ πνεῦμα, Latin *spiritus asper*), which signifies a [h] before the vowel, derives from the left half (F) of *eta*, and the smooth breathing (˘) (ψιλὸν πνεῦμα, Latin *spiritus lenis*), which marks the absence of a [h] (and maybe [ʔ] instead?), from the right half (I).⁷⁾

6) Note that the traditional view of the ‘hard sign’ is that it marks the preceding consonant as being ‘hard’, i.e. non-palatalized. However, there is no empirical evidence that this is indeed the case, and it seems rather unlikely that in a language in which every consonant before [ɛ] and [i] is automatically palatalized a consonant before [j] or [i] should be able to avoid this palatalization.

7) Originally Aristophanes of Byzantium introduced these marks at the turn of the second century BCE only for disambiguating homonyms, e.g. ὄρος / ‘oros’ ‘mountain’ and ὄρος / ‘horos’ ‘border’ (cf. Sturtevant 1937,

The so-called *e caudata* ⟨e⟩ (‘tailed e’), which was used in Latin since the 7th century, is another example of a letter diacritic. Its development was aptly described by Robert (1895, 634). It started out as the ⟨æ⟩ ligature, which was used for *ae* in classical Latin. In the Middle Ages, its *a* part could have an ‘open’ shape, so that it looked like this: . The *a* here is hardly more than a bow, and this bow then moved towards the bottom of the *e*, yielding the shape . This downward movement might have happened as a consequence of the fact that medieval scribes often represented *ae* by a simple ⟨e⟩, and the cauda could be added only later, sometimes even by a reader of the document (Stotz 1996, 83, §67.4). Often it turned out very thin, as in . Of course, for such a belated correction there would be no space next to the ⟨e⟩, so that the cauda representing *a* could for technical reasons only be added below the ⟨e⟩. Later the cauda moved further to the middle of the character, which enabled scribes to write it together with the basic letter without lifting the pen.

The ring in Czech ⟨ů⟩ derives from an ⟨o⟩. In Czech, **ō* first turned into a diphthong, which was initially spelled ⟨uo⟩ and then written vertically as ⟨ů⟩. Since the diphthong merged with /u:/, the ring has been a reminder of the fact that ⟨ů⟩, unlike ⟨ú⟩, alternates with ⟨o⟩. The ring in Scandinavian ⟨å⟩, however, derives from an ⟨a⟩. The sound nowadays spelled ⟨å⟩ used to be [a:], which was initially (and until 1948 in Danish) spelled ⟨aa⟩. (It later developed into [ɔ], [ɔ] and even [o].) Written vertically, just like in Czech, this turned into ⟨å⟩, which was then simplified to ⟨å⟩. Consequently, the ring in Czech and the ring in Scandinavian look exactly the same but have completely different etymologies (deriving from ⟨o⟩ and ⟨a⟩, respectively). In this sense, they are homonyms.

Another case of homonymy are the umlaut dots, which in modern typography have the same shape as the diaeresis treated above, but a different function and etymology. They derive from ⟨e⟩, and in German, ⟨ä⟩, ⟨ö⟩, and ⟨ü⟩ can still be replaced with ⟨ae⟩, ⟨oe⟩, and ⟨ue⟩ even today (e.g. in internet addresses that do not permit characters outside the ASCII code). The vertical placement of these letters yielded |ä|, |ö|, and |ü|, which can be seen in German prints as late as the early 20th century. In the old German cursive known as Kurrent (and often called Sütterlin), however, the ⟨e⟩ is nothing more than two vertical lines next to each other: |e|. Consequently, |œ|, |œ|, and |œ| could be written as |e|, |e|, and |e|, which then yielded |e|, |e|, and |e| with two little strokes above the letters, and this was also done in printed blackletter: |ä|, |ö|, and |ü|. In roman type, the two strokes were simplified to two dots: |ä|, |ö|, and |ü|.

Since the Middle Ages, ⟨n⟩ and ⟨m⟩ were often written above the preceding letter to save space, and their form was soon reduced to a wavy line (˘). This nasal mark (which was sometimes also simplified to a straight line (˘)) was purely optional, so that |tādē|, |tādēm|, |tandē| und |tandēm| were mutually exchangeable realizations of Latin ⟨tandem⟩ ‘at last’ occurring in free variation. However, in Spanish, long /n:/, which could be spelled |nn| or |ñ|, was palatalized to /ɲ/ (just like /l:/, which was spelled |ll|, was palatalized to /ʎ/

116). Later, however, the marks were placed on all vowels at the beginning of words. Although the consistent use of the rough breathing mark made the smooth breathing mark redundant, it was kept for many centuries even after /h/ had vanished from the language, up to the year 1982, when both breathing marks were abolished.

8) Scan from the facsimile in Steffens (1929, plate 32, line 29) of manuscript K. k. V. 16 of Cambridge University Library, a copy of Beda’s church history written probably in Echternach around 737 in the Insular script.

9) Scan from the facsimile in Steffens (1929, plate 38, line 7) of charter I.13 of the St Gall Monastery Archive, a deed of foundation for the St Gall Monastery written in 757 in the Merovingian script.

10) Scan from Baker (n.d.) from an 11th-century manuscript by Eadwig Basan of Canterbury.

and later simplified to /j/). As soon as /j/ was consistently spelled ⟨ñ⟩ and the nasal mark was not used in other contexts anymore, the tilde (˘) had turned into a diacritical mark. Something similar happened in Portuguese, where vowels followed by /m/ and /n/ could turn into nasal vowels, e.g. ⟨bom⟩ /bõ/ ‘good’. In certain positions the tilde has to be used instead of ⟨m⟩ or ⟨n⟩, e.g. ⟨nação⟩ /nẽˈsẽw̃/ ‘nation’ or ⟨nações⟩ /nẽˈsõj̃/ ‘nations’. The tilde indicating nasal vowels has also been borrowed into other languages, e.g. into Kashubian (where ⟨ã⟩ for /ã/ contrasts with ⟨ą⟩ for /ɔ/), and into the International Phonetic Alphabet.

The cedilla is a prominent example of a diacritic derived from a part of a letter. It emerged in Spain, where during the Middle Ages the local variant of the Latin alphabet was the so-called Visigothic script. One of its characteristic features was the *zeta* [or *zeda*] *copetuda* ‘tufted z’, a tailed ⟨z⟩ with an additional top hook: |ʒ|. This *copete* ‘tuft’ grew over time and was eventually reinterpreted as the letter ⟨c⟩, so that the rest of the original ⟨z⟩ was only a *zedilla* ‘little z’ under the main letter: ⟨ç⟩ (Menéndez Pidal 1954, 212–221). As soon as this grapheme was interpreted as a variant of ⟨c⟩ and not as a form of ⟨z⟩, a new diacritic was born: the cedilla (¸). According to Kramer (1996, 587), Italian ⟨ç⟩ has a different etymology: When ⟨c⟩ and ⟨z⟩ were not sufficient to distinguish the results of various palatalizations (especially [tʃ], [ts], and [dʒ]), the Greek letter zeta (ζ) was borrowed into Italian, which was then reinterpreted as a variant of ⟨c⟩ in a similar way as Spanish |ʒ|, and ultimately printers used the same sort for it, ⟨ç⟩. If this is true, the Spanish and Italian cedilla are another pair of homonymous diacritics, although their functions are so similar that they can synchronically be viewed as the same diacritic.

4.3 Deletion diacritics

Another kind of characters that can develop into a diacritic are deletion marks. The most common way of marking something as deleted is striking it out. This can be seen in many characters, e.g. Polish ⟨ł⟩ for /w/ (< /ł/), traditional Lower Sorbian ⟨fch⟩ for /ʃ/ (vs. ⟨fch⟩ for /ɛ/) or Old Norse ⟨ǰ⟩ for /ŋ/ (vs. ⟨g⟩ for /g/; e.g. in the 12th-century *First Grammatical Treatise*, p. 88, cf. Haugen 1972, 26–27, 57). The main idea of this change of function is that a letter is needed for a sound for which the alphabet used does not really have a letter; therefore one uses a letter representing a similar sound but strikes it out at once to indicate that this is not exactly the right letter. In contrast to a real mistake, however, no correction is possible, so no ‘correct’ letter is placed next to the deleted letter. Readers knowing the language are capable of taking the hint and pronouncing a sound resembling the one of the ‘deleted’ letter.

Since this ‘deletion’ is, after all, merely a metaphor and it is important for the ‘deleted’ letter to still be recognizable (and over time it also becomes important that the stroke through the letter can be written smoothly together with it), many of these deletion dashes are not drawn through the middle of the character but only through a part of it, e.g. in Serbo-Croatian ⟨Đ đ⟩ (rather than *(Đđ) or *(Dd)) for /ǰz/ or Maltese ⟨Ħ ħ⟩ (rather than *(Ħħ) or *(Hh)) for /ħ/.

Historically, scribes did not want to spoil a beautiful manuscript by striking out a wrong letter (“mis~~s~~take”) or painting it over (“mis█take”). Instead, they placed a punctum delens (deletion dot) above or below the wrong letter (“mis[•]take” or “mis_•take”). This did not stick out on the page, but when readers arrived at the passage, they knew to skip the letter with the dot.

In Old Irish, this punctum delens was turned into a diacritical mark, the so-called *ponc séimhithe* [ˈpʷɪŋk ˈʃeːvʲiħə] ‘lenition dot’ ⟨ ˙ ⟩. At the beginning it was placed only on ⟨ f ⟩ and ⟨ s ⟩, of which the former completely disappeared (after being pronounced as [w] for some time) and the latter was sometimes pronounced [h] and sometimes deleted (e.g. in the family name *McEntegart*, which goes back to *mac int śacairt* /mak iNt aɣəRdʲ/ ‘son of the priest’, with *śacairt* < Latin *sacerdos* ‘priest’; cf. McManus 1996, 659). This made the use of the deletion dot even more compelling than in the cases mentioned above, in which there still is a sound pronounced, albeit the wrong one. Later, the diacritic was by analogy also used to mark other lenited consonants: ⟨ p ⟩ [f], ⟨ b ⟩ [w], ⟨ t ⟩ [θ > h], ⟨ d ⟩ [ð > ɣ], ⟨ c ⟩ [x], ⟨ ǵ ⟩ [ɣ], and ⟨ m ⟩ [w̃ > w].

An interesting case is the Old Polish nasal vowel /ã/ (from Proto-Slavic *ę and *ǫ), which was written as ⟨ ą ⟩ in medieval manuscripts. This character can almost certainly be derived from ⟨ ǫ ⟩. When this nasal vowel was long, it was often doubled, as in ⟨ rǫkǫǫ ⟩ ‘by hand’. At the beginning of the 16th century, the quantitative opposition /ã/ : /ã:/ was gradually replaced by the qualitative opposition /ɛ̃/ : /ɛ̃:/, so that Zaborowski (1983, 56, 78, 91) proposed in his orthographical treatise to distinguish between an *a* with a short stroke at the top (“semivirgula superior”) and an *a* with a long stroke (“integra virgula”). Consequently, the word ‘by hand’ appears as |**Raka**| in the printed treatise (Zaborowski 1983, 71). However, Hieronymus Vietor improved this solution by writing /ɛ̃/ < /ã:/ as ⟨ ą̇ ⟩ (like Zaborowski) and /ɛ̃/ < /ã/ as ⟨ ę ⟩ (cf. Bunčić 2012, 235; e.g. |**reęa**|). In roman type, ⟨ ę ⟩ was often replaced with the already existing sort ⟨ ę̇ ⟩ (see 4.2 above), and later ⟨ ą̇ ⟩ was analogically written as ⟨ ą̇ ⟩ (yielding today’s ⟨ ręka ⟩).¹¹

4.4 Diacritics from disambiguation marks

Disambiguation marks (German *Unterscheidungszeichen* (Bunčić 2014), Czech *znaménko zdůrazňovací* (Voleková 2018)) can become necessary due to cursivization and aesthetic principles that increase the uniformity of a script and thus the similarity of its graphemes as a calligraphic / typographic ideal at the expense of clarity. For example, various medieval book hands often emphasized the vertical stems, especially Gothic hands like textualis (see fig. 1): ⟨ minimum ⟩ → |**mmmm**|. This led to the use of a disambiguation mark on ⟨ i ⟩ to make the text readable despite the Gothic aesthetics.¹² (In Czech texts, it appears sporadically since the first half of the 14th century; cf. Voleková 2018, 136.) At the beginning, this mark was a stroke, which might be interpreted as a repetition of the basic shape of ⟨ i ⟩: |**iiiiiiii**|. Later it could be reduced to a dot: |**iiiiiiii**|. The cursive that was taught in German schools until 1941 would not be readable without the (obligatory!) disambiguation marks on ⟨ i ⟩ and ⟨ u ⟩: Compare *|**uuuuuuuuuu**| with |**iiiiiiiiiiii**|.

11) Kamusella (2019), seemingly without looking into any of the relevant prints of the early 16th century or any medieval manuscript, claims that ⟨ ą̇ ⟩ was derived from Cyrillic ⟨ ą̇ ⟩ (the letter for Proto-Slavic *ę), arguing that this explains why the letter for /ɛ̃/ is ⟨ ą̇ ⟩ rather than ⟨ ǫ̇ ⟩. Apart from the extant sources clearly showing a different development, this fails to comply with the fact that /ɛ̃/ developed from /ã:/ and that as late as the early 16th century “both nasal vowels resembled nasal *a*, save that one was somewhat farther front and the other farther back” (Stieber 1973, 79, §56).

12) Another way of dealing with the problem was the replacement of letters in especially difficult cases. Thus, in English, words like ⟨ wunder ⟩ or ⟨ sum(e) ⟩, originally spelled |**uuuuder**| (with a real *double u*) and |**sum(e)**|, were made more readable by replacing ⟨ u ⟩ with ⟨ o ⟩: |**uouuder**| and |**sum(e)**| (Millward 1996, 160). The pronunciation of ⟨ o ⟩ as /ʌ/ in such words still betrays its origin in a short *u*.

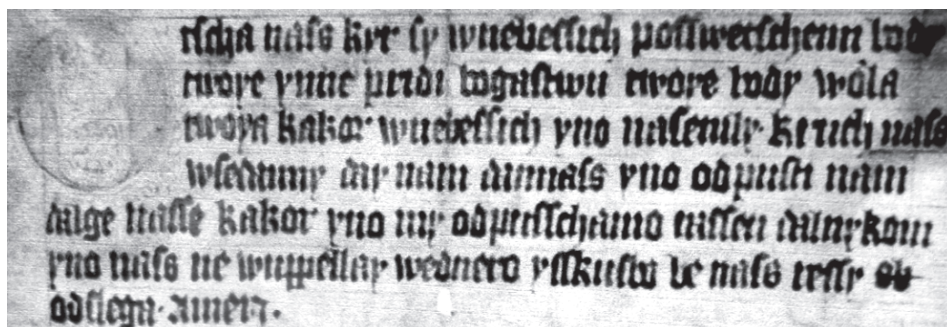


Fig. 1. Textualis (in the Slovenian Rateče Manuscript, late 14th century)

Of course, the *i* tittle is not a diacritic (and neither is the breve-like disambiguation mark on the $\langle \dot{z} \rangle$), because $\langle \dot{z} \rangle$ is not a different grapheme from the old $\langle z \rangle$ without the dot. However, when the Latin orthography for Turkish was introduced in 1928/1929, this dot was given a distinctive function within the new Turkish writing system by differentiating $\langle \dot{i} \rangle$ with a dot (with the capital letter $\langle \dot{I} \rangle$) for /i/ from $\langle i \rangle$ without a dot (with the capital letter $\langle I \rangle$) for /u/.

The Arabic script is characterized by strong cursivization, which led to the graphical merger of many letters. Thus, both Phoenician zayin ($\langle \tau \rangle$) and resh ($\langle \rho \rangle$) (cf. Hebrew $\langle \tau \rangle$ and $\langle \rho \rangle$) turned out as $\langle \rho \rangle$. In order to ensure readability, a dot was then used to distinguish $\langle \dot{\rho} \rangle$ (z) from $\langle \rho \rangle$ (r). However, once the dot was established as the only feature distinguishing these two letters, it could be used for other distinctions as well. Thus, the letter $\langle \dot{\rho} \rangle$, which derives from Phoenician heth ($\langle \aleph \rangle$) (cf. Hebrew $\langle \aleph \rangle$), signified two Arabic sounds, /x/ and /h/. By using the dot as a diacritic, a distinction between $\langle \dot{\rho} \rangle$ (h) for /x/ and $\langle \rho \rangle$ (h) for /h/ was made. Although synchronically there is no difference now between the diacritic dots that originally emerged as a disambiguation mark and those that were analogically created from scratch as diacritics, one can still tell the difference from the fact that the ‘new’ oppositions have something in common in the signified, like the two back fricatives /x/ and /h/, whereas the functions of the original disambiguation marks that arose from pure visual similarity are random.

4.5 Arbitrary creations?

The four paths of the emergence of diacritics presented in 4.1–4.4 are generalizations from findings in my investigation of a sizable number of diacritics in various languages and scripts (Bunčić 2013–2020; Bredel – Bunčić 2020). Nonetheless, I cannot exclude that there are other paths in some writing systems I did not examine. The more interesting question, however, is: Can diacritics also be created ‘without an etymology’, in a completely arbitrary way? My hypothesis is that they cannot.

An example of a diacritic that is often claimed to be arbitrary is the stroke supposedly added to $\langle C \rangle$ during the Roman classical period to get $\langle G \rangle$ (cf. e.g. Balázs 1958, 253; Diring 1962, 166f.; Friedrich 1966, 112; Haarmann 1991, 296; Glück 2016). In the earliest

period of written sources in Latin, the Romans used the letter ⟨C⟩ to represent both /g/ and /k/ because they had inherited their alphabet from the Etruscans, who did not have a voiced / voiceless opposition.¹³⁾

However, this ‘diacritical’ G-bar is a myth, which was already debunked by Hempl (1899). As he plausibly deduces, the letter ⟨G⟩ has to be traced back to Greek zeta ⟨Ζζ⟩ – which explains why its place in the alphabet is neither after C (as a variant of it, like ⟨J⟩ after ⟨I⟩ or ⟨V⟩ after ⟨U⟩) nor at the end (like other Greek letters that were added later: ⟨X⟩, ⟨Y⟩, ⟨Z⟩) but in the exact location of zeta:

Greek (classical)	A	B	Γ	Δ	E	†F	Z	H	Θ	I	...
Old Italic	Α	Β	Ɔ	Ɔ	Ɔ	Ɔ	Ɔ	Ɔ	⊗	Ι	...
Latin (classical)	A	B	C	D	E	F	G	H	—	I	...

The parallel development of gamma to ⟨C⟩ and of zeta (in its original Western Greek and Old Italic shape, which still resembled Phoenician zayin ⟨𐤆⟩ and did not yet have the vertical line tilted as in classical Greek ⟨Ζ⟩) to ⟨G⟩ is demonstrated by Hempl (1899, 31) using shapes of these letters attested in inscriptions in chronological order:



Consequently, the bar distinguishing ⟨G⟩ from ⟨C⟩ is not a diacritic (or at least it was not created as such), just as the leg distinguishing ⟨R⟩ from ⟨P⟩ or the additional bar distinguishing ⟨E⟩ from ⟨F⟩ is not a diacritic.

There is a diacritical mark from the classical Roman period which is less well-known because it is never used in modern text editions: the so-called *apex*, which was a widespread mark indicating the long vowels /a:/, /e:/, /o:/, and /u:/ in Latin texts from the late second century BCE into the third century CE (Oliver 1966, 131; long /i:/ was usually represented by the *i longa* ⟨I⟩; see fig. 2). Although the shape of the apex is more or less identical to the Greek acute accent (´), it is probably not related to it, both because it indicates vowel length rather than stress and because of the historical development of its shape in Latin texts.

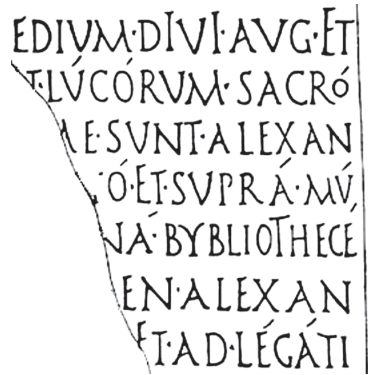


Fig. 2. Roman inscription (1st c. BCE).¹⁴⁾

13) Originally, the Etruscans used both kappa ⟨𐌕⟩ and koppa ⟨𐌗⟩ alongside gamma ⟨𐌆⟩, but all these letters represented /k/, so this unnecessary variation was later reduced by abolishing ⟨𐌕⟩ and ⟨𐌗⟩. The most archaic Latin inscriptions similarly use ⟨K⟩ for /k/ before ⟨A⟩, ⟨Q⟩ for /k/ before ⟨O⟩ and ⟨U⟩, and ⟨C⟩ for /k/ in other positions, but then ⟨C⟩ is generalized (and ⟨Q⟩ kept only in the combination ⟨QU⟩ for /kw/ and ⟨K⟩ in the word *kalendae* ‘first day of the month’).

14) Clipping from a sketch uploaded to *Wikimedia Commons* by Tomisti [on-line] <<https://commons.wikimedia.org/w/?oldid=141578342>> [cit. 02-09-2022], based on Keil (1923, 127–129).

While the macron used in modern text editions¹⁵ is iconic by virtue of its length extending in the direction of writing (see 4.1), an iconic connection between the apex (´) and its meaning is hard to establish. However, the inscriptions and the few preserved papyri show the apex not only in its ‘canonical’ form |´| but also in many other shapes (cf. Christiansen 1889, 4–5; Oliver 1966, 149–150), e.g. |´|, |ʔ|, |´|, |ʃ|, or |´|, and there are even a few attestations of the macron-like horizontal shape |´| (e.g. <POLYCLITŌ> and <SŌTACVVS> in the inscription CIL 8979 of the early imperial era, Henzen et al. 1882, 1191). If this variant |´| was the oldest one, the apex could be explained as iconic.

However, according to Balázs (1958, 157f.) and Oliver (1966, 150–151), the oldest variant, attested as early as 104 BCE, is the “sickle-shaped” one, i.e. a right half-circle |´|. Moreover, the original function of the mark seems not to have been the indication of vowel length (at least not directly) but the indication of the doubling of a letter (vowel or consonant). Oliver (1966, 152–158) has shown convincingly that the apex was thus originally identical to a diacritic known as *sicilicus* ‘little sickle’, which was placed exclusively over consonants. Consequently, for an etymology of the apex we have to look for a motivation of the relation between the semicircle shape |´| and letter gemination. Unfortunately, the *sicilicus* seems to have been used mainly in “books” (“libri”) before 204 BCE (Fontaine 2006, 106–107) but not in inscriptions, where at the beginning single letters could represent geminated consonants, before in the course of the second century BCE the doubling of consonant letters and the apex for long vowels took hold. However, books, i.e. papyri, have not been preserved from this early period. Therefore, we can only speculate about the earliest history of this diacritic. An explanation of the circular shape might be a kind of iconic repetition sign like |↻| in the sense of ‘read again’. Another might be that geminated letters may originally have been written above each other (just as with Scandinavian <aa> turning into <â> and then <å>, see 4.2) and that a simplified form of a frequently used letter (<ð>? <þ>?) might then have been analogically transferred to all geminated letters.

However, these are just speculations. We probably have to accept that we will never know exactly how the apex was motivated. But just like with the many words whose etymology is unclear, disputed or completely unknown, we can assume that this diacritic does have some etymology, that it did not arise out of thin air. Or more generally: Diacritics with unclear etymologies do not falsify the hypothesis that diacritics always have some motivation for their shape and their function.

The hypothesis thus seems to be true for all the diacritics created before the era of the printing press. In the Modern Era, however, technical restrictions could force printers to pick an already existing letter with a diacritic from their letter-case. We have already seen that when the Polish orthography was transferred to roman type at the turn of the 17th century, the letter <ę> with its ‘deletion diacritic’ was replaced with the *e* caudata <ę> because it was a similar-looking character that simply happened to be in the letter-case, thus creating the *ogonek* as a diacritic indicating nasality.

Another example comes from the so-called Hodža-Hattala Reform of the Slovak standard language (Slovak *Hodžovsko-hattalovská reforma*) in the middle of the 19th century. In Western Slovak dialects, the reflex of **ō* was simply long [o:], whereas Central Slovak had a diphthong [uo]. Consequently, Anton Bernolák and others advocated spelling this phoneme <ó>, whereas Ľudovít Štúr spelled it <uo>. As a compromise, a sort already in

15) Christiansen (1889, 4) claims that it was introduced to Latin editions by Wilhelm Paul Corssen in the 19th century. The mark itself, however, was already used in antiquity (cf. the words cited by Oliver 1966, 134, fn. 14, from a Greek manuscript of the second century CE).

the letter-case which contained ⟨o⟩ with a different diacritic that was traditionally also associated with vowel length was found, so that the phoneme has since then been spelled ⟨ô⟩ (Hattala 1852, 3). In the meantime, the orthoepic norm has also been unified, so that the standard pronunciation of ⟨ô⟩ now is [uɔ]. In the manuscript era, one might instead have invented an ⟨o⟩ with a little ⟨u⟩ above (⟨^uo⟩ or similar), which might then over time have undergone some kind of simplification (e.g. ⟨ō⟩).

Probably the most extreme case is the Turkmen alphabet approved by the parliament of Turkmenistan on 12 April 1993, which was designed to make it possible to write Turkmen with the character set of code page 437 (the US character set of the original IBM PC). Thus, not only was ⟨ñ⟩ with tilde (which signifies /ɲ/ in Spanish, Basque, Quechua, Guaraní, and many other languages) used to spell /ŋ/, and ⟨ÿ⟩ with diaeresis to spell /j/ – just because these characters happened to be included in the character set; it was also decided to use ⟨ç⟩ (as a small letter) and ⟨Ş⟩ (as the corresponding capital letter) for /ʃ/, ⟨{⟩ (the top half of the integral symbol ⟨∫⟩ designed to span rows) and ⟨£⟩ as small and capital letter for /z/, and ⟨Ÿ⟩ as the capital variant of ⟨ÿ⟩ (because ⟨ÿ̈⟩ was unfortunately not included in code page 437). (In 1999, this rather unsuccessful alphabet was replaced by one with ⟨ž⟩ for /z/, ⟨ñ̈⟩ for /ŋ/, ⟨ş̈⟩ for /ʃ/, and ⟨ÿ̈⟩ for /j/, which is now in widespread use.)

While the technical restrictions that led to the decisions for such diacritics are part of their history, one might indeed say that these diacritics have ‘no etymology’, just like phantasy words such as *Kodak* (Kiplinger 1962), *Thalys* (de Kemmeter 2021), or *Catan* (Teuber 2020, 203–204), which were invented by free association, without reference to any pre-existing word or morpheme. However, it is important to note that such more or less arbitrary ‘inventions’ of diacritics entail three preconditions: 1. a technical restriction that does not allow for designing a diacritic according to the methods outlined in 4.1–4.4; 2. a culture in which the use of diacritics for designing new orthographies is already well-established; 3. a technical environment that already contains a certain set of diacritics to choose from. None of these preconditions is met by the diacritics proposed in the *Orthographia Bohemica*.

5. How diacritics are borrowed

All the etymologies in section 4 concerned diacritics that were created directly for a given language. More frequently, however, diacritics are borrowed from a different language. The basic mechanisms of borrowing are the same as with the borrowing of alphabets or words: Which donor language a diacritic is borrowed from depends on the areal, social, cultural, economic, and political relations between the societies using the languages (5.1); after the moment of borrowing, the form of a diacritic can change either in the donor language or in the recipient language (5.2); and a diacritic can be borrowed with only a certain aspect of the meaning/function it has in the donor language (5.3). I will demonstrate these mechanisms with a few examples.

5.1 Borrowing relations

As pointed out above, the cedilla was ‘invented’ independently in Old Spanish and Italian. From Spanish (where it is not used anymore nowadays, having been replaced with ⟨z⟩, the origin of ⟨ç⟩), ⟨ç⟩ was taken over into neighbouring languages like Portuguese, Catalan, Occitan, and French. Similarly, from Italian (where it is not used anymore nowadays either) it was borrowed into neighbouring Friulian, Croatian, and Albanian. We know that

loanwords can be borrowed not only from a neighbouring language but also from a remoter language that the speakers have come into contact with for socio-political reasons; similarly, an orthography with its choice of diacritics “is not just a writing technique” (“является не только техникой письма”), as Jakovlev (1930, 36) put it, but reflects “an alphabet’s ideology” (“идеологии алфавита”). In the 19th century, due to the French Revolution and Napoleon’s reforms, French became a dominant language in Europe, and it is from there that ⟨ç⟩ was adopted to Romanian and, in the 1920s, to the Turkic languages of the Soviet Union (the so-called New Turkic Alphabet or Jaṅalif, first introduced in 1927 to Tatar) as well as to Turkish (introduced in 1928), and from Turkish to Kurdish (proposed in 1932). Note that Jaṅalif and the Turkish alphabet differ in many respects (e.g. Turkish ⟨ö⟩ vs. Jaṅalif ⟨ə⟩, ⟨ğ⟩ vs. ⟨q⟩, ⟨j⟩ vs. ⟨z⟩, ⟨ı⟩ vs. ⟨b⟩), including the sound values of ⟨c⟩ and ⟨ç⟩: While in Turkish ⟨c⟩ represents [d͡ʒ] and ⟨ç⟩ represents [t͡ʃ], in Jaṅalif it is the other way around. The modern Latin alphabets of Azerbaijani (since 1991), Turkmen (since 1999), and Tatar (introduced in 2001, banned by a Russian law in 2002) use ⟨ç⟩ for [t͡ʃ] as in Turkish. But all these languages agree in borrowing French ⟨ç⟩.

Note also that in Croatian, ⟨ç⟩ borrowed from Italian during the Renaissance was used only in the coastal region, where Italian influence was strong, whereas the orthographies of the hinterland were based on the Hungarian model (Marti 2012, 282–284). During the 19th-century Slavic National Revival, the cedilla as well as some digraphs were replaced with combinations with the háček borrowed from Czech, which also spread over most of the other Slavic languages using the Latin alphabet (cf. Moiseenko 1997).

In the Northwest European periphery, we can see a medieval borrowing chain with the Latin apex, which was preserved as a length mark in Old Irish (cf. Balázs 1958, 264).¹⁶ Around the end of the 11th century, the apex also began to be used in early Middle English (Sisam 1953, 185–191), most systematically in the 12th-century *Ormulum* (Anderson – Britton 1999, 300, 325) and certainly modelled on Irish (Balázs 1958, 264). The 12th-century *First Grammatical Treatise* (FGT) then introduced it to Old Norse, most probably borrowing it directly from Irish (Haugen 1972, 55), at a time of massive Scandinavian presence both in Ireland and in England.¹⁷ (Apart from that, chronicles report that when the Norse settled Iceland, they found Irish monks already living there, who left “Irish books” behind when they were forced to leave; cf. Grønlie 2006, 4, 17–18.)

16) As mentioned in 4.5, the apex had fallen out of use in Latin texts as early as the 3rd century CE, and Irish manuscripts in the Latin alphabet have only been preserved from the 8th century onwards. However, it is very probable that the Irish language was written in Latin letters long before the oldest preserved texts. For once, the Irish had used the Ogam alphabet to inscribe stones at least since the 4th century, and Ogam seems to be a secondary encoding of a different alphabet (cf. Fjellhammer Seim 2007, 159–160); and second, the Irish orthography was so refined even in the earliest preserved texts from around 700 that “Latin-letter literacy in Irish must have existed all through the Ogam period, at the very least to record native Irish names in ecclesiastical documents” (Mac Eoin 1998, 126). Consequently, we can assume that the apex was borrowed directly from Latin texts into Irish.

17) The First Grammarian also marked nasal vowels with a dot (e.g. ⟨â⟩ for /ā/, ⟨ô⟩ for /ō/, ⟨ô̇⟩ for /õ/, etc.; cf. *Ibidem*, 38–39), whose origin has so far not been established (*Ibidem*, 55). It is not impossible that the dot was also borrowed from Irish, where it marked lenited consonants (see 4.3). Having no use for a dot on consonant letters but needing a way to distinguish a whole range of vowels (since all nine vowels could, in principle, be nasal, *Ibidem*, 38–39), the First Grammarian might have used the dot but slightly changed its function (see 5.3 below) from marking a special class of consonants to marking a special class of vowels. However, it is equally possible that he invented this diacritic himself as a deletion diacritic (see 4.3), given that the punctum delens was widespread among medieval scribes all over Europe.

Another example, which demonstrates chronological rather than geographical circumstances of borrowing, are the French accent marks. Until the 18th century, French was written without accent marks. Only in the third edition of the dictionary of the Académie Française, published in 1740, were the Greek accent marks acute, grave, and circumflex officially adopted to the French orthography. It is no coincidence that this happened when the Age of Enlightenment and the French classicist form of Baroque made inspirations from classical antiquity fashionable.

5.2 Changes of form

An example of the change of shape of a diacritic is the Czech dot (ˇ), which was borrowed into Polish to form the grapheme (ż) – and originally also (ś), (ć), and (dź) – by Stanisław Zaborowski (1983) at the beginning of the 16th century. While in Czech itself the dot changed its shape to the modern háček (ˇ), it remained unperturbed in Polish, and is still printed as a dot to this day (though in handwriting, (ż) is often replaced with [z], with a bar through the middle). Additionally, since Standard Polish has three rows of sibilants (dental: /s z ts dz/; postalveolar: /ʃ ʒ tʃ dʒ/; alveolopalatal: /ç z tç dç/), whereas Czech (like all other Slavic languages) has only two, Zaborowski extended the idea of the diacritic dot to mark the third row by two dots: (š) for /ç/, (ž) for /ʒ/, (č) for /tç/, and (ď) for /dç/ (Zaborowski 1983; cf. Bunčić 2012, 229). From the beginning, these two dots had a tendency to be smeared in print and to be connected in handwriting, which at first made them appear as [š̄], [ž̄], etc. and eventually resulted in modern (ś), (ź), (ć), and (dź) (cf. Bunčić 2012, 232f.).

The cedilla also changed its shape after having been borrowed from French into Romanian: Instead of the connected ‘little z’ tail used in French (ç) and also Turkish (ş) and Gagauz (ṭ), it turned into a detached mark in the shape of a comma: (ș) and (ț). Consequently, although the two marks are historically identical, their forms have to be distinguished, so that, for example, they were given separate Unicode values in September 1999 (in Unicode version 3.0). It is not quite clear whether the Romanian change from cedilla to comma was technically induced (because people used the comma on the typewriter due to the lack of separate keys for (ș) and (ț)) or a consequence of a handwriting practice. In Latvian, which uses the cedilla (in the shape of a comma, as in Romanian) to signify palatals, the diacritic was also needed with the letter (ģ) for /j/, and because of this letter’s descender it had to be placed above it: (ģ̆) (cf. the chart of the International Phonetic Alphabet, which stipulates at the bottom of the page: “Some diacritics may be placed above a symbol with a descender, e.g. ģ̆”).

A well-known example of a diacritic with varying shape is the Greek circumflex (περισπωμένη) (ˆ), which is often cursivized to [̂] (also in modern text editions), whereas in the European languages that borrowed it it usually has the angular form [^].

5.3 Changes of function

The circumflex is also a good example of functional changes. Originally, it was invented to indicate the rising-falling tone in Ancient Greek wherever words differed only by tone (cf. the example *οἴκοι* vs. *οἴκοι* in section 3 above). Later, it was used on all words with this tone, irrespective of whether the word would have been ambiguous without the circumflex. Then the tone differences were lost in the Greek language, so that the marking of the stressed syllable and disambiguation were left as the sole function of the accent marks. However, since the circumflex tone could only occur in long syllables, the diacritic

was also an indicator of long vowels. (While long vowels could have other accents than the circumflex, a vowel marked with a circumflex could not be short.) This function was what inspired German scribes to use the circumflex as a length mark, e.g. in the Old High German Tatian (a translation of the gospel harmony *Diatesseron*) from around 830 (e.g. *Thaz uuas in imo lib inti thaz lib [!] uuas lioht manno* ‘In him was life; and the life was the light of men’, John 1:4; Sievers 1892, 13).

In humanist Latin the circumflex was also used as a length mark, e.g. in contractions like *amâsse* for *amavisse* ‘to have loved’ (Hale 1995, 24), whereas the acute marked only stress. Thus, Erasmus of Rotterdam (1528, 4v) could write, “Aio dômi nâtum, audis accentû acutû in priore voce, in posteriore circûflexû” (“I say *dômi nâtum* ‘born at home’, you hear the acute in the former word and the circumflex in the latter”), to clarify that he had not meant *dominâtum* ‘dominated’, referring to the two stresses that the two words have: on a short syllable in the word *domī* ‘at home’ and on a long syllable in the word *nâtum* ‘born’.

When the Greek accent marks were introduced into French in 1740, the circumflex also marked long vowels (which after the loss of vowel quantity differ in their quality as [a], [o], [ε] from the originally short vowels [a], [ɔ], [ə]), e.g. in *grâce* or *fantôme* and especially in words with compensatory lengthening after the loss of a consonant as in *fenêtre* ‘window’ < *fenestre* (< Latin *fenestra*) or *bête* ‘beast’ < *beste* (< Latin *bēstia*). Note that the circumflex on the French vocative particle *ô* ‘O’ (e.g. in *Ô ma noble et fidèle amie* ‘O my noble and faithful friend’, Camus 1959, 929) corresponds exactly to the circumflex in the Greek vocative particle *ὦ* (as in *ὦ Ἀχιλλεῦ* ‘O Achilles’ [Iliad, I.74] or *ὦ Θεόφιλε* ‘O Theophilus’ [Acts 1:1]). Even Hattala’s ⟨ô⟩ introduced in 1852 (mentioned in 4.5) was meant to signify either a long vowel [o:] or a diphthong [uɔ], which is long by definition.

The disambiguating function of the Greek accent marks, which can also be seen in Greek pairs like the (orthotone) demonstrative *ὅ* ‘this’ vs. the (proclitic) article *ὁ* ‘the’, where the two words differ only by the absence vs. presence of an accent mark, was exploited in many Western European languages, e.g. in humanist Latin *quòd* ‘because’ vs. *quod* ‘which’ (relative pronoun), in French *à* ‘to’ vs. *a* ‘has’, or in Italian *è* ‘is’ vs. *e* ‘and’.

The *e* caudata ⟨ę⟩ (see 4.2), which was very popular in medieval Latin texts, was borrowed into Old Norse in the 12th century by the author of the *FGT*. The First Grammarian came back to the initial function of ⟨ę⟩ to substitute for ⟨æ⟩, which originally signified an open [ɛ] or [æ], and used it to distinguish between ⟨ę⟩ for Old Norse open /ɛ/ and ⟨e⟩ for closed /e/. By analogy, he also used the cauda to create ⟨o⟩ for open /ɔ/ in contrast to ⟨o⟩ for closed /o/. At the end of the 16th century, as described above (see 4.3), the *cauda* turned into a diacritic signifying nasality in Polish (i.e. ⟨ę⟩ for /ɛ̃/) and was subsequently also transferred to ⟨ą⟩ for /ɔ̃/. The mark’s modern (Polish) name *ogonek* means ‘(little) tail’, just like Latin *cauda*. From Polish, the *ogonek* was borrowed into Lithuanian, whose four nasal vowels came to be spelled as ⟨ą⟩, ⟨ę⟩, ⟨į⟩, and ⟨ų⟩. Nowadays, however, they are simply long oral vowels, but they have kept the spelling with the diacritic, which is still called *nosinė* ‘nasalization mark’ in Lithuanian.

So we have seen that the form and function of diacritics can develop in quite idiosyncratic ways – just as the development of the form and meaning of words can take strange and intriguing courses (e.g. Greek *ἐλέφας* ‘elephant’ > ... > Czech *velbloud* ‘camel’ and occasionally end up at the opposite end of the starting point, Czech *kriminalista* ‘criminalist’ with Polish *kryminalista* ‘criminal’, or German *abdecken*, which can mean both ‘cover’

and ‘uncover’, depending on context; cf. Bunčić 2004, 207, 209). However, all these developments can be explained by their situational conditions, and so can the developments of diacritics. They are not arbitrary.

6. Hypotheses about the source of the Czech dot and stroke

After having laid out the theory about how diacritics generally evolve, let us now concentrate on the treatise *De Orthographia Bohemica*. We will review various possible inspirations for the creation of the two Czech diacritics, ⟨˙⟩ for a non-Latin pronunciation of consonants and ⟨˘⟩ for long vowels. (Overviews are given by Schröpfer 1968, 23–30; Pleskalová 2007, 501; Šlosar 2010, 200–204; and Voleková 2019, xxxiii–xxxv, lviii–lx.)

6.1 Czech texts before the 15th century

It is often pointed out that the two diacritics had been in use even before the time of writing of the *Orthographia Bohemica* (e.g. Schröpfer 1968, 24–26; Pleskalová 2007, 501; Voleková 2019, xxxiii, lviii). This can mean two things: a) that the system laid out in the *Orthographia* had already been implemented before it was written down in the preserved text of the *Orthographia*; b) that the *Orthographia* systematized the use of two diacritics which had already been used sporadically with the same function.

Interpretation a) does not really change anything in our question about the origin of the diacritical system. Irrespective of whether the system was first described and then used or the other way around, we still want to know where its inventor got the idea for it. In fact, it seems rather probable that the inventor would at least first test his idea by writing down at least a short Czech text in the new orthography before taking the trouble to formulate the rules and argue for them in an extensive treatise. There might even have been some development in the system before it was recorded in the treatise. So if, for example, the diacritics used in the fragment of the Gospel of Mark pointed out by Flajšhans (1891, 92) and Voleková (2018, 137) can be proven to have been written by the scribe himself in 1370–1390 (and not added later by someone else), or if the *Orthographia Bohemica* really has to be dated to the middle of the 15th century (cf. Voleková 2019, xxxvi, lxii), this merely changes the chronological order of events. However, if the system was already used in the 14th century, Jan Hus – who enrolled at the University of Prague around 1389 and received his Masters degree in 1396 (Gillet 1863, 52) – is less likely to have been its inventor, and the author of the *Orthographia Bohemica* might very well have been describing a system he had not invented himself.

However, some authors lean towards interpretation b), claiming that the *Orthographia Bohemica* does not contain any new ideas because the dot and the stroke had already been around in Czech manuscripts for a long time, though they had not been used as systematically as the treatise then recommended. This would mean that they might have emerged independently from each other and much earlier.

Which evidence is given for this claim? For example, Kvítková (2010, 76) reports exactly two cases in the Hanuš fragments [Hanušovy zlomky] of the Chronicle of Dalimil [Kronika tak řečeného Dalimila] from the first half of the 14th century in which a dot is placed above ⟨y⟩: ⟨giezdity⟩ for jezdit ‘to ride’ and ⟨tluftoty⟩ for tlustosti ‘thickness (GEN. SG)’. She interprets this dot as a diacritic to indicate the palatalization of the preceding consonant (which is nowadays indicated by using ⟨i⟩ vs. ⟨y⟩). In contrast to this, she claims, the non-palatalizing /i/ could be written as ⟨ÿ⟩ with a háček. An example of this

is the Hradec Manuscript [Hradecký rukopis] from the mid-14th century (which is also referred to by Schröpfer 1968, 101). However, ⟨ŷ⟩ occurs exactly once in this manuscript, namely in ⟨Snŷmyz⟩ for *s nimiž* ‘with which (INSTR.PL)’ (Patera 1881, xx, 40–41) – where ⟨n⟩ is actually palatalized, which contradicts Kvítková’s theory. Apart from that, we once find the spelling ⟨ij⟩ in ⟨milofrdenftwij⟩ for *milosrdenství* ‘mercifulness (LOC.SG)’ (*Ibidem*, xx, 210–211).

With all these examples of ‘diacritics’ on ⟨y⟩, we have to take into account that in the Middle Ages (and up to the 19th century) ⟨y⟩ was generally considered to be a variant of ⟨i⟩ – either as the Greek form of ⟨i⟩ (cf. its French name *i grec*, Polish *igrek*, Dutch *Griekse ij*, Latin *y graeca*, etc.) or as a double ⟨i⟩ in the form of ⟨ij⟩, which is practically identical to ⟨ŷ⟩ (cf. cursive |ȝ| and the equivalence of Dutch ⟨ij⟩ to Afrikaans ⟨y⟩, e.g. *pījp* ≡ *pyp* ‘pipe’, both going back to older ⟨ŷ⟩). Consequently, when the i tittle was introduced as a disambiguation mark (see 4.4), placing one or two dots (or sometimes strokes, see section 2) also on the ⟨y⟩ (to form |ȝ|, |ŷ|, |ȝ̇| or |ŷ̇|) became a widespread practice (cf. Voleková 2018, 235; see fig. 3¹⁸). Clearly, all these marks are no more diacritics than the dot on the ⟨i⟩.

Note also that in the *Orthographia Bohemica* there is no diacritical dot on ⟨y⟩, so it would be hard to see a |ȝ| as a predecessor of the Czech diacritical system.

A different kind of evidence are cases like *wsekirzie* ‘in the axe’ in the Clementine Psalter from the mid-14th century (Žaltář klementinský, Ps 73/74:6; Gebauer 1878, 219), *nepr̄siblizie* ‘not come nigh’ (Ps 31/32:6), *otewr̄zeff* ‘thou openest’ (Ps 144/145:16) and *w̄sicz̄kny* ‘all’ (Ps 6:9/8) from the same manuscript (Patera 1890, xx), *czyrw̄ow skařez̄dych* ‘of ugly worms’ from the late 14th-century Ráj duše manuscript (fol. 22r; *Ibidem*), *wid̄yel* in a fragment of the Gospel of Mark dated to 1370–1390 (Voleková 2018, 137), or the isolated stroke on the personal name *Noóstup* in a Latin charter from 1188 (Friedrich 1904–1907, 294; Pleskalová 1998, 137; Voleková 2018, 136). In all these cases, the diacritics are clearly redundant because the phonemes ⟨r̄/⟩, ⟨t̄j̄/⟩, ⟨j̄/⟩, and ⟨o:̄/⟩ are already indicated by a digraph (in these cases, ⟨rz̄⟩/⟨r̄f̄⟩, ⟨cz̄⟩, ⟨dȳ⟩, and ⟨oō⟩). As Gebauer (1878, 220) himself concedes, “Někdy možná, že takováto znaménka dostala se do rukopisu staršího teprve později, vlivem pravopisu Husova” (“Sometimes it is possible that such a mark was added to an older manuscript only later, under the influence of Hus’s orthography”), i.e. the diacritics might have been added in the 15th century or even later by a reader who already knew the diacritical orthography. Although Gebauer is optimistic that “mostly this is not so” (“většinou není tomu tak”, *Ibidem*), one may have doubts about this, and



Fig. 3. Štítiny (1463), 67r

18) There are dots on all the ⟨y⟩’s, which sometimes, e.g. in *lopoty* in line 4, look like strokes, and two dots on the ⟨y⟩ in line 5 – all of this irrespective of whether ⟨y⟩ corresponds to modern ⟨y⟩, ⟨ŷ⟩, ⟨i⟩, or ⟨i̇⟩, or whether it indicates palatalization or /j/ as in *sobye* for modern *sobě*.

Patera (1890, 27) often assumes that a dot is “posterior, as in many other places” (“pozdější, jako i na mnoha jiných místech”). Gebauer’s alleged spelling *wsekiřzie*, for example, is represented in Patera’s edition (1890, 129) as “w fekirzie”, with a footnote declaring: “Nad *r* tučná tečka aneb skvrna” (“Above the *r* there is a fat dot or stain”). Gebauer’s interpretation of the blotch above *skarzedych* shown in fig. 4 as a háček, or indeed any diacritic, also seems overly optimistic, especially given the otherwise total absence of diacritics in the whole manuscript.

Some of the marks that were probably added by readers in the Clementine Psalter (Patera 1890, xx) are not redundant because the original scribe did not mark the special pronunciation at all, e.g. *bud* ‘be!’

(Ps 9:36/10:15), *uhanyl* ‘shamed’ (Ps 13:6, Patera 1890, xx; represented as *uhanyl* *Ibidem*, 35), *množství* ‘wealth’ (Ps 43:13/44:12, *Ibidem* xx; *množství* *Ibidem*, 81), *doňawadz* ‘as long as’ (Ps 103/104:33, *Ibidem*, xx; *donawadz* *Ibidem*, 181). Note that next to such dots that are in accordance with the *Orthographia Bohemica*, there are many others in the same manuscript that are not, e.g. *rúczie* ‘suddenly’ (Ps 6:10/11), *chudeho* ‘poor (GEN.SG)’ (Ps 13/14:6), *kralowství* ‘kingdoms’ (Ps 45:7/46:6), *fwřchka* ‘above’ (Ps 77/78:23), *žlořtwi* ‘wicked (NOM.PL)’ (Ps 103/104:35, all from Patera 1890, xx), *feizé* ‘burneth’ (maybe for *feize*, Ps 45:10/46:9, *Ibidem*; *feize*, *Ibidem*, 85). All in all, the assumption that the scribe of the Clementine Psalter already used the dot of the *Orthographia Bohemica*, though inconsistently, is unsubstantiated. Instead, we have to agree with Patera (1890, xixf.) that the dots were placed “either due to the carelessness of the scribe himself or by later readers” (“bud neopatrností samého pisaře aneb od pozdějších čtenářů”).

Another piece of evidence presented by Gebauer (1878, 219) is *že neosta* in the Hradec Manuscript from the last third of the 14th century. However, a look at the page shows a dot that, in view of the ink colour differing from the colour of both the ⟨z⟩ and the following ⟨e⟩, might very well have been added later (see fig. 5), and Patera (1881, 16; fol. 6^b, not 5^b as indicated by Gebauer 1878, 219) decided not to include the dot in his edition, representing these words as “Ze neofla”. Similarly, Gebauer (*Ibidem*) cites from the 1376 manuscript XVII.A.6 in the Czech National Library (the so-called *Klementinský štitenský*

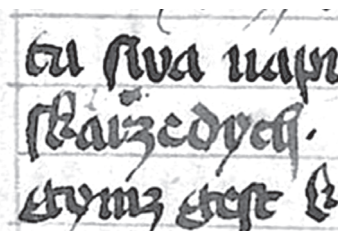


Fig. 4. Ráj duše, 22r

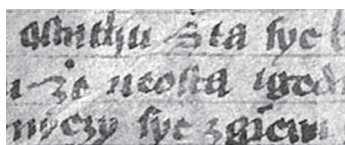


Fig. 5. Hradec Manuscript, 6v

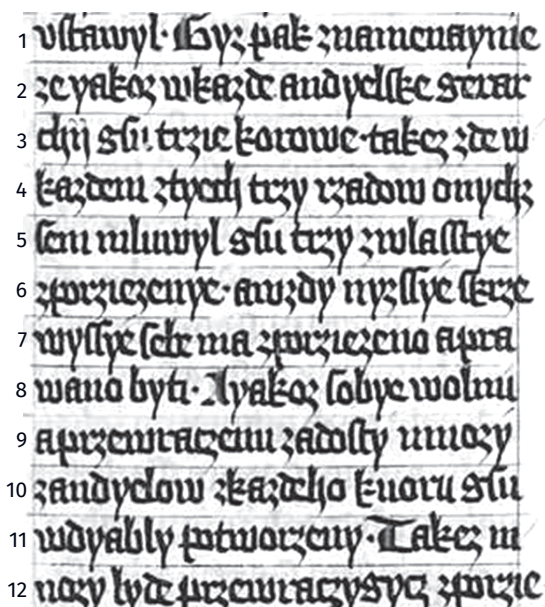


Fig. 6. Klementinský štitenský sborník, 67r

sborník with texts by Tomáš Štítný ze Štítného): “wolnú zadošťý” (actually the text reads *fobyewolnú a przewraczenu zadošťý* ‘by their whimsical and perverse desire’, fol. 67r), *nyzffýe skrze wyffýe* ‘the lower ones by the higher ones’ (*Ibidem*) and *wydúcz* ‘seeing’ (fol. 74v). The first two of these examples can be seen in fig. 6, lines 6–9. If we compare these faint strokes with the much bolder i tittles in *gerar|chij* ‘hierarchies’ (lines 2f.) and the periods (lines 1, 3, 6, 8, 11), it seems at least possible that the strokes might not have been written by the writer of the text but added by a later reader, who also added hyphens at the ends of lines 3, 11 and 12 to connect the words *wkazdem* ‘in every’, *m-nozy* ‘many’, and *zporzie-zenye* ‘order’. Note also that in *nyzffýe skrze wyffýe* the diacritics are, again, redundant, because the long /i:/ is already indicated by the digraph ⟨ye⟩. If these words had been written by a scribe who was already familiar with the function of the length mark, we should expect him to write either ⟨ý⟩ or ⟨ye⟩, but not ⟨ýe⟩. This spelling can be explained much more plausibly if we assume that the 14th-century scribe spelled the endings ⟨ye⟩ and a reader in the 15th century, who was already used to seeing them spelled ⟨ý⟩, added the stroke. The same reader, obviously, also added strokes in quite a few other places in this particular manuscript.

To conclude, there is no robust evidence of the use of dots to indicate non-Latin pronunciations of consonants and of strokes to indicate vowel length before the turn of the 15th century.

6.2 The i tittle and other disambiguation marks

An interesting hypothesis about the diacritical dot is that it might have its origin in the i tittle (Schröpfer 1968, 26). Indeed, this would have been an ingenious idea: The inventor of the diacritical system would have associated the raising of the tongue body in the Czech palatalized consonants with the articulation of [i], and then he would have indicated this [i]-like secondary articulation by placing a dot as a *pars pro toto* for ⟨i⟩ on top of the palatalized consonants.

This theory, however, has two flaws. First, the diacritical dot as defined in the *Orthographia Bohemica* was not a mark of palatalization. It marked any consonant whose pronunciation differed from Latin, including ⟨l̃⟩ for /l/. Second, at the beginning of the 15th century the i tittle had just started to be used, at first in the shape of a stroke, not a dot, and usually inconsistently. Consequently, the dot was not linked closely enough to the letter ⟨i⟩ to be used as a recognizable *pars pro toto*. Note also that disambiguating dots were also used on ⟨ý⟩ and, according to Voleková (2018, 138), on ⟨č⟩ (which could stand for /tʃs/ or /tʃ/, the dot simply helping with distinguishing it from ⟨t⟩, which could look rather similar, e.g. |ṭ| vs. |t| in *textualis*). This further blurred any possible association between the dot and the letter ⟨i⟩.

But might the disambiguation mark on ⟨č⟩ have been an inspiration for the Czech diacritical dot? After all, the Arabic disambiguation dots were very successfully developed into an extensive system of diacritics (see 4.4). However, the difference is that the Arabic disambiguation marks still exist; the idea that the dot could create a new grapheme was only available after the marks had become the only distinction between otherwise not just similar but identical letters. In the Czech case, the diacritical dot replaced the disambiguation dot (with problems arising when writers mixed up the two systems, as Voleková 2018, 138–139, has shown), which means that the shapes of ⟨c⟩ and ⟨t⟩ had not fully merged and it was

still possible to distinguish the two graphemes without using the dot.¹⁹⁾ Consequently, the dot was only used sporadically (maybe whenever the writer, looking at the ⟨c⟩ he had just written, decided that it looked too similar to ⟨t⟩ to be readable without the dot) and not by everybody because it was not the only distinctive feature. Therefore, the developmental stage necessary for introducing additional dots as diacritics had not been reached. But if it had, the dot on ⟨ċ⟩ could not have been reused to distinguish its pronunciation /tj/ from ⟨c⟩ for /ts/ because ⟨c⟩ would then have been identical to ⟨t⟩. While the extension of the idea to ⟨š⟩ vs. ⟨s⟩ etc. would have been feasible in this scenario, the contrast between /tj/ and /ts/ would have had to be expressed by something like ⟨ċ⟩ vs. ⟨ċ⟩ (vs. ⟨c⟩ for /t/), just like in the Persian version of the Arabic script, where the disambiguation mark on ⟨z⟩ that distinguishes it from ⟨r⟩ could of course not be omitted, so that two more diacritical dots were added to create the additional letter ⟨żż⟩ for /z/.

6.3 Hebrew

Since the Hebrew letter šin ⟨w⟩ is mentioned in the *Orthographia Bohemica* itself (fol. 35r), it is often claimed that the šin dot distinguishing ⟨ẇ (š)⟩ for /ʃ/ from ⟨w (š)⟩ for /s/ was an inspiration for the Orthographus Bohemicus to use a dot to distinguish consonants (e.g. Balázs 1958, 280; Šlosar 2010, 201; Kamusella 2019).

However, this presupposes a familiarity with the Masoretic text of the Hebrew Bible (including vowel signs) as we know it today, which was not at all self-evident at the beginning of the 15th century; Hebrew studies were introduced at the University of Prague only in 1541 (Segert 1966, 70). Note also Segert's (1981/1984/1985, 349) verdict that "Hus was certainly not able to read a Hebrew text". As a matter of fact, all the *Orthographia Bohemica* says about the Hebrew alphabet is: "Iudei habent unam literam, que dicitur *ches*, et aliam, que dicitur *šyn*" (fol. 35r, "the Jews have one letter called *ches* and another one called *šyn* {šin}" , Voleková 2019, 46–47). Neither the Bible nor the šin dots nor indeed the Hebrew language are mentioned, nor does the treatise draw the shape of the Hebrew letters or cite example words. Consequently, all the author of the treatise needed to have seen to make this statement (and to mention the name "gymel" for ⟨g⟩ in fol. 37v) was a list of the names of the Hebrew letters. As a matter of fact, this list was familiar to every medieval scholar: It is contained, for example, in the Book of Psalms (which every monk learned by heart in its entirety). Concretely, it is part of Psalm 119 (118), which in the Hebrew original is an acrostic with each stanza starting with a different letter, in alphabetical order. Therefore, even the Latin Vulgate includes the names of the Hebrew letters at the beginning of each stanza (gimel starts the third stanza, verses 17–24; het the eighth stanza, verses 57–64; and šin the twenty-first stanza, verses 161–168).²⁰⁾ So, according to Ockham's razor, we should not base any argument on the unsubstantiated assumption that the author of the *Orthographia Bohemica* had to have "a certain knowledge of [...] the Hebrew language" ("jisté znalosti i o [...] hebrejštině", Voleková 2019, xxxv). He just had to know his Psalms.

19) A merger occurred only later, in the German cursive of the 17th–20th century, but not between ⟨c⟩ and ⟨t⟩. Instead, ⟨c⟩ merged with ⟨i⟩ and ⟨n⟩ with ⟨u⟩, so that they had to be distinguished by obligatory disambiguation marks: |c| vs. |i| and |n| vs. |u|. Note that in this context ⟨c⟩ was distinguished from ⟨i⟩ by the *absence* of a dot. The ⟨t⟩ remained distinct as |t|.

20) Additionally, lists of the Greek and Hebrew letters were often given as a kind of appendix in Bible manuscripts, e.g. in manuscript XI.A.14 of the Czech National Library (Biblia XI.A.14, fol. 242r).

The list of letter names, however, implies an opposition between šin (װ (š)) for /ʃ/ and samekh (׀ (s)) for /s/ (Ps 119:113). The same opposition is also mentioned in the “essential book of the whole Middle Ages” (“Grundbuch des ganzen Mittelalters”, Curtius 1948, 489), Isidore of Seville’s *Etymologiae* (or *Origines*) from the 7th century, which all medieval scholars had read. In book IX (chapter 2, verse 18; Lindsay 1911: no page numbers; <https://archive.org/details/isidori01isiduoft/page/n349/mode/2up>) of the *Etymologiae*, two people called Saba in the Latin translation of Genesis 10:7 are distinguished as one being originally spelled with samekh and one with šin. (In the King James Bible of 1611, for example, they are accordingly called *Seba* and *Sheba* (and modern Czech Bible translations also give them as *Seba* and *Šeba*.) Saint Jerome, the translator of the Latin Bible, who was also “the foremost authority in Hebrew matters for mediaeval Christian Bible scholars” (Segert 1981/1984/1985, 350), explains in his *Liber de nominibus Hebraicis* ‘Book about the Hebrew names’ that samekh “might almost be described by our letter S” (“quasi per S nostram litteram discrebatur”, Hieronymus 1845, col. 827f.), whereas šin is characterized by “a certain hushing sound not of our language” (“stridor quidam non nostri sermonis”, *Ibidem*). This is also how these two letters are used, for example, in Yiddish (a language that must have been ubiquitous in 14th/15th-century Prague), where /s/ is always spelled (׀ (s)), and (װ (š)) exclusively spells /ʃ/. No šin dots are necessary in the Hebrew alphabet to make this distinction. It is only due to certain sound changes in the history of the Hebrew language that in some cases (װ (š)) came to be pronounced like (׀ (s)), which is usually left unmarked (just like the pronunciation of the middle ⟨s⟩ in Czech *husitismus* ‘Hussitism’ as /z/ is not marked) except in the Masoretic Text of the Hebrew Bible. No Hebrew dots are mentioned in the *Orthographia Bohemica*, and *ffyn* is cited as a letter – not a variant of a letter! – for /ʃ/, in the same breath as Slavic *ša* (i.e. Glagolitic (ш) or Cyrillic (ш)) and, as Schröpfer (1968, 59) remarked, without mentioning its “twin brother” (“Zwillingsbruder”) *šin*. A century later, János Sylvester (2006, 35) has a three-way distinction between *samek*, *schin dextrum* ‘right šin’ and *schin sinistrum* ‘left šin’ in his Hungarian Grammar of 1539, and he later became a professor of Hebrew (cf. Balázs 1958, 279), but in the first half of the 15th century information about the two pronunciations of šin would have been hard to come by for a Christian scholar, since Jerome’s information on this “is ambiguous” (Segert 1981/1984/1985, 350). So we can conclude that the treatise refers to the opposition between (װ) and (׀), not between (װ) and (װ̣).

But even if the inventor of the Czech diacritic orthography had known the šin dots – would the orthography have turned out the way it did if he had borrowed them? The šin dots are used to disambiguate a letter that has two pronunciations. If this idea had been borrowed into Czech, we should expect ⟨s⟩ to be able to signify both /s/ and /ʃ/, with the possibility of being disambiguated into ⟨ṣ⟩ for /s/ vs. ⟨ṣ̣⟩ for /ʃ/, and similarly ⟨ṭ⟩ for /t/ vs. ⟨ṭ̣⟩ for /c/, etc. The opposition between ⟨s⟩ and ⟨ṣ̣⟩, ⟨ṭ⟩ and ⟨ṭ̣⟩, etc. that we see in the *Orthographia Bohemica* differs from the alleged Hebrew model both in its form (left / right vs. middle) and in its function (disambiguation vs. grapheme creation).

Another model often mentioned in the literature (e.g. Balázs 1958, 280f.; Schröpfer 1968, 24–25; Šlosar 2010, 201f.) is the Hebrew dagesh. The dagesh is a diacritic which is placed inside a Hebrew letter to indicate gemination (e.g. in ⟨הָמָּה⟩ (lāmma)) ‘why’ with ⟨מָּ) for *mm*). When it is used with the six letters for plosives, it indicates that they are pronounced as plosives even in positions where they would otherwise turn into fricatives (see fig. 7). This happens, for example, at the end of the letter name ⟨הֶתְ) (het), where the final *t* historically turned into [θ], and the spelling *ches* in the *Orthographia Bohemica* (35r)

indicates the Ashkenazic pronunciation of [θ] as [s].²¹ The dagesh can prevent this change, e.g. in $\langle \text{קָטַבְתָּ} \rangle$ (*katavt*) ‘you have written’, where both the initial *kav* $\langle \text{כ} \rangle$ and the final *tav* $\langle \text{ת} \rangle$ are marked by a dagesh to keep their original pronunciation as plosives [k] and [t] instead of the fricatives [x] and [θ].

However, this function and the function of the dot in Czech “are basically incommensurable” (Segert 1981/1984/1985, 347). Therefore, if the dagesh had been borrowed into Czech, we should expect the dot to have a completely different function: It should then mark the letters for /d/, /t/, /s/, etc. to indicate that they keep their original pronunciation, whereas the absence of the dot should allow their pronunciation to change into /j/, /c/, /ʃ/, etc. Furthermore, the form should also be different. We should expect the dot to be placed inside, not above, the letter, just as in Hebrew. Fig. 8 illustrates what Czech with a dagesh would look like.

In Hebrew, the dot is used for various other purposes as well, most importantly to mark vowels, depending on position (e.g. a dot below a letter stands for /i/, a dot above for /o/, two dots side by side under a letter for /e/, etc.; and, most interestingly for the application to an alphabet: the letter *vav* $\langle \text{ו} \rangle$ with a dot above $\langle \text{ו} \rangle$ signifies /o/, whereas with a middle dot $\langle \text{ו} \rangle$ it signifies /u/ and without these dots in a vocalized text it represents the consonant /v/. This raises another problem: If the Hebrew dots had been the inspiration for the Czech diacritic orthography, then they should also have been used with multiple functions both for consonants and vowels, i.e. also to mark vowel quantity.²²

כ [b]	כ [v]
ג [g]	ג [ɣ]
ד [d]	ד [ð]
כ [k]	כ [x]
פ [p]	פ [f]
ת [t]	ת [θ]

Fig. 7.
Hebrew dagesh

c [ts]	c [tʃ]
d [d]	d [j]
f [f]	f [ʃ]
n [n]	n [ɲ]
r [r]	r [ɹ]
s [s]	s [ʃ]
t [t]	t [c]
z [z]	z [ʒ]

Fig. 8.
Czech dagesh

6.4 Glagolitic

The *Orthographus Bohemicus* also mentions the “Slavic” letters *chir* and *ffa* (fol. 35v, Voleková 2019, 46–47), which seems to imply some familiarity with either the Cyrillic or the Glagolitic alphabet. Again, this does not necessarily mean that the author had seen or even read texts in one of these alphabets himself. At least four Glagolitic (and probably as many Cyrillic) abecedaria in 14th–15th-century manuscripts from the Czech lands have been preserved to the 21st century (Čermák 2012), so quite a few must have been around at the time. They usually give the names of the letters in the Latin alphabet (e.g. *Biblia XI.A.14*, fol. 242v, with *cher* and *ffa*).

Nonetheless, the hypothesis of the Glagolitic origin of the Czech diacritics is probably the most popular one. It was first put forward by Antonín Frinta (1940), who hypothesized that the inspiration for the Czech diacritical dot were the Glagolitic manuscripts produced by Croatian Benedictine monks at the Emmaus Monastery in Prague, who had

21) Schröpfer (1968, 24) mentions the use of Ashkenazic forms instead of the traditional Bible forms (in this case, *chet* or *cheth*) as characteristic of Jan Hus; cf. also Segert (1981/1984/1985, 350).

22) Kamusella (2019) claims that this was originally the case and that the dot later “evolved into several different New Czech diacritics, namely, the acute accent [‘], háček (‘little hook’) [ˇ], and overring [˘]”. However, it is well-documented in a multitude of texts from the 15th century onward that the stroke (‘) was clearly separate from the dot even in the earliest texts containing these diacritics and, of course, in the *Orthographia Bohemica*, which distinguishes between *gracilis virgula* ‘thin slash’ and *punctus rotundus* ‘round dot’, and that the ring (˘) emerged from an (o). So this thesis cannot be upheld.

Apart from the completely different function, there is also a formal difference between the Glagolitic *yer* dots and the Czech diacritic: Since the former actually represent the štapić, which is a full letter usually placed *after* the preceding letter, the point is also not centred above the letter but placed above the right top corner of the letter.²³⁾

The author of the *Orthographia Bohemica* had an astonishing knowledge of language history, e.g. pointing out in fol. 38r that Czechs once spoke [g] instead of [h] (Voleková 2019, 58–59). But even a reader who does not know anything about the history of the reduced vowels will understand quickly that the dots in the Glagolitic manuscripts are placed above word-final consonants and in consonant clusters. Since the vast majority of these dots appear above completely normal non-palatalized consonants, the idea that they mark a non-Latin pronunciation, as Frinta (1940, 254) wants to suggest, could not occur to a medieval reader. Note also that in one of Frinta's selected examples, ⟨ϣⱱ (naš)), while the dot is in a place where the Orthographus Bohemicus would indeed have placed a dot as well (⟨naš)), it does not change the pronunciation of ⟨ϣ (s) to /ʃ/ but is placed above a letter that already indicates /ʃ/ without any diacritic, namely the Slavic ša ⟨ϣ (š)), which happens to be explicitly mentioned in the *Orthographia Bohemica* as a letter in the 'Slavic alphabet' that, in contrast to Latin ⟨s), does not need modification (fol. 35v, Voleková 2019, 48–49).

To sum up, the dots in the Glagolitic texts from the Emmaus Monastery have a function that is easy to grasp for any reader and at the same time so different from the function of the Czech diacritical dot that there can be no connection between them. The facts that the positions of the dots differ and that there is no trace of a length mark in the Glagolitic texts only add to the improbability of this connection.

Various students of the *Orthographia Bohemica* have suggested that the Glagolitic alphabet might have been an important model, if not for the diacritics as such, then for the general idea of each phoneme corresponding to exactly one grapheme (cf. Vykypěl 2020, 96–97). This question is not the subject of the present paper, but, apart from the fact that the Glagolitic alphabet is not ideal in this respect (*Ibidem*, 95), I would like to remark that he did not need any other model because the Latin alphabet certainly provided one grapheme for every phoneme of Medieval Latin (in which the vowel quantities of Classical Latin, for example, were not heeded anymore; note also that there are no minimal pairs distinguishing the two pronunciations of ⟨c) [k ~ t̪s], and the pronunciations of ⟨g) [g ~ j] led to respellings, cf. Vidmanová 1969, 297). Where the alphabet had even more than one grapheme for a Medieval Latin phoneme, scribes tended to simplify the orthography, cf. spellings from the *Orthographia Bohemica* like ⟨litere) for *litterae* (fol. 35r), ⟨gencium) for *gentium* (fol. 35v), ⟨lingwa) for *lingua* (*Ibidem*), etc. At the beginning of his treatise the Orthographus Bohemicus says that the Latin alphabet is insufficient for Czech, Greek, 'Jewish', German, "and other languages" (fol. 35r, Voleková 2019, 46–47); he seems to take for granted, however, that it is sufficient for Latin.

23) In the example from the Emmaus Bible quoted above, there are also three dots on the letter ⟨e): ⟨ḗḡḡḡ (ēnoš)), ⟨ḗḡḡḡ (ēnoš)) und ⟨ḡḡḡ (noē)). These dots are much rarer and have a completely different function from the *yer* dots: They mark an ⟨e) as being syllable-initial but nonetheless to be pronounced without a prothetic [j]. This dot also has a different etymology, going back to the Greek breathings, which were also used in Cyrillic and some Glagolitic texts. For example, they are already used in the shape of dots in the 11th-century Glagolitic Codex Assemanianus, e.g. "ḡḡḡḡḡ" for ⟨ḡḡḡḡḡ (iōānъ)) 'John', fol. 6v). Accordingly, in contrast to the *yer* dots, these breathing dots are centred above the vowel letters.

Another reason for assuming the Orthographus Bohemicus's familiarity with the Glagolitic alphabet is the fact that he has given the letters names that make up sentences, e.g. *a bude celé čeledi dáno d'edictvie* (35r, Voleková 2019, 44) 'And may the whole family be given the heritage', like Cyril's *az buky vědě* 'I know the letters'. However, Schröpfer (1968, 92–95) has shown that the idea could easily have been taken from Saint Jerome, who makes sentences for memorizing the order of the Hebrew letters.

6.5 The Greek acute accent

According to Pleskalová (2007, 501), a potential origin of the length marks might be Greek words in Latin texts in which sometimes the Greek accent marks were preserved, so that the Greek acute accent could occur above Latin letters. Apart from the fact that this usage must have been extremely marginal (I have not been able to find an example), the function is clearly different: The acute accent marked the stressed syllable and could be placed on both long and short vowels. If such Greek loanwords had been the inspiration for the Orthographus Bohemicus, he would not have used the acute accent to mark length but the circumflex, which only occurs on long vowels – as the 9th-century Germans did (see 5.3).

6.6 The Latin apex

The apex, as already mentioned in 4.5, was a Latin diacritic that was used in inscriptions between 104 BCE and 200 CE and in papyri up to the 3rd century CE (Kramer 1991, 142). It is often identified with the Czech stroke (e.g. Balázs 1958, 282), because its classical shape (since the time of Augustus) is more or less the same, and its function is to mark long vowels. So in contrast to most of the marks mentioned so far, it is both formally and functionally identical to one of the marks of the *Orthographia Bohemica*.

However, the question here is: Could a Czech of the turn of the 15th century have known about it? Books from the time when the apex was used have not survived except for a few fragmentary papyruses. In the late Middle Ages, texts by classical authors like Cicero or Pliny were only known from medieval copies (which is why in the 15th century the Carolingian minuscule was chosen as the source of the small letters of 'roman type', or '(littera) antiqua'), and these copies did not use the apex. Since the Czech lands had never been part of the Roman Empire, there were of course no Roman inscriptions there. But even if the Orthographus Bohemicus had travelled to Italy or some other country where he could see such ancient inscriptions, he would have been in the same position as the modern scholars up to the 19th century, who, if they did not take the often very thin apices to be fissures in the marble, were confronted with a great variety of shapes and their inconsistent use, so that the apex's "meaning and function was long hidden in the dark" ("vis ac potestas diu in tenebris latuit", Christiansen 1889, 4).

However, the apex was mentioned first by Quintilian in the 1st century CE and then a bit later by Quintus Terentius Scaurus, and their ideas were summarized in the 7th century, long after the apex had been in active use, by Isidore of Seville in his *Etymologiae* (see 6.4):

In dubiis quoque verbis consuetudo veterum erat ut, cum eadem littera alium intellectum correpta, alium producta haberet, longae syllabae apicem adponebant; utputa ‘populus’ arborem significaret, an hominum multitudinem, apice distinguebatur. (*Etymologiae*: I.xxvii.29, Lindsay 1911)

NB: *apex* ‘summit, top; conical cap of the flamens; helmet, crown’; *pōpulus* ‘poplar’; *pōpulus* ‘people’.

In ambiguous words, when a letter had a different meaning when short than when long, it was the custom of the ancients to add an *apex* to a long syllable; for example, when *populus* signifies a tree and not a multitude of people it is distinguished by an *apex*.

If, however, the Orthographus Bohemicus had borrowed his length mark from the ancient Romans, we can assume that he would have mentioned an authority like Isidore, which would have given his ideas a lot of additional support, because in the 15th century an *argumentum ab auctoritate* was still the prime form of evidence. The Orthographus Bohemicus argues vehemently for the diacritical orthography, e.g. admonishing the reader, “do not write double *zz* with vowels – for even if you wrote *z* a hundred times, they would not make a sound different or more than *z*” (“noli ponere duplex *zz* cum vocalibus, quia si centum *z* pones, non facient plus in sono vel aliter quam *z*”, fol. 41r, Voleková 2019, 76). It is quite unthinkable that he would have foregone the chance to point out that what he was proposing had already been good practice among the ancient Romans. Instead, he talks about ‘Jewish’ and ‘Slavic’ letters, but never once mentions Isidore or any other authority. Furthermore, if the Czech length mark was based on the Roman apex, the treatise would certainly call it by the name *apex* instead of circumscribing its shape as *gracilis virgula* ‘thin slash’.

Another problem with this hypothesis is that Isidore does not say what the apex looks like or where it is placed. As pointed out above, a Czech of the early 15th century was unlikely to come across a Latin text that actually uses an apex, and all Isidore says is *apicem adponebant* ‘they added a summit / top / hat / crown’. It would be an enormous coincidence if from this formulation the Orthographus Bohemicus correctly guessed the exact shape and position that the apex had in 2nd-century inscriptions and papyri. (Christiansen 1889, 6, reports that Justus Lipsius in the 16th century assumed that the mark discussed by Quintilian and Scaurus had the shape of a macron (ˉ).) The only way the Orthographus Bohemicus could have achieved this result would have been to take the inspiration from Isidore to go looking for ancient Roman inscriptions, for example, in Italy, examine them more closely than anyone had done for centuries – and then not to say a word about this.

Furthermore, Isidore explicitly says that the apex should be placed only on *dubia verba* ‘ambiguous words’. Quintilian had even explicitly expressed his disgust at the over-use of the apex: “longis syllabis omnibus adponere apicem ineptissimum est” (*Institutio Oratoria* I.7.2, cf. Kramer 1991, 142; “placing an apex on *all* long syllables is extremely silly”). If the Orthographus Bohemicus had really borrowed the apex, we would now have a Czech orthography in which the length mark is only used for disambiguation (much the way it is used in Serbian and Croatian, where people sometimes write ⟨sâm⟩ ‘by oneself’ to distinguish it from *sam* ‘am’ or ⟨lingvistâ⟩ ‘of the linguists’ to distinguish this genitive plural from the genitive singular *lingvista* ‘of the linguist’).

6.7 Old Irish

When working on diacritics in the context of Volume 5 *Grapholinguistics* (edited by Martin Neef, Rüdiger Weingarten, and Said Sahel) of the “Dictionaries of Linguistics and Communication Science” (Bunčić 2013–2020, 2020a, Bredel – Bunčić 2020), I noticed that

almost every language has a different combination of diacritical marks but that the only two European languages whose inventory of diacritics consists of the dot and the stroke are Old Irish and Old Czech (see fig. 9). There are no other pairs of languages in which such an identity of inventories happens by chance. The only other pair of languages with more than one diacritic that have the same diacritics is Norwegian and Danish, where it is well-known that the Norwegian orthography is entirely based on Danish. So if the fact that Old Irish has the same two diacritics as Old Czech is pure coincidence, it is the only such coincidence in all the European languages. Therefore, Old Irish merits a closer look.

Modern Irish (Gaelic) has an extremely complex and ‘deep’ orthography, even more so than English. For example, the Irish name of Dublin is spelled (Baile Átha Cliath) but pronounced [bl̪iːˈkl̪iə]. In Old Irish, orthography was still considerably ‘shallower’, but by no means phonological. Among the phenomena that complicate the relationship between morphology and phonology and thus make a simple orthographic representation difficult is the so-called lenition (cf. Thurneysen 1909, §§119, 128–131), which under certain circumstances can turn [p] into [f], [t] into [θ], etc. The *ponc séimhithe* ‘lenition dot’ (´), which came to be used to mark these lenited consonants, was already mentioned in 4.3. In the texts of the Old Irish period, it was only used consistently with four consonant letters: ⟨f̃⟩ for [Ø] (< [w]), ⟨s̃⟩ for [h], ⟨m̃⟩ for [w] (< [w̃]), and ⟨ñ⟩ for [n] (corresponding to ⟨n⟩ for some kind of ‘more fortis’ sound commonly transcribed as [N]). Lenited plosives were either spelled with digraphs including a ‘diacritic’ ⟨h⟩ to indicate fricativization, i.e. ⟨ph⟩ for [f], ⟨bh⟩ for [w]/[v], ⟨th⟩ for [h] (< [θ]), ⟨dh⟩ for [ɣ] (< [ð]), ⟨ch⟩ for [x], and ⟨gh⟩ for [ɣ], or lenition could also not be marked at all. Only later was the consistent use of the lenition dot extended to the plosives (⟨p̃⟩, ⟨b̃⟩, ⟨t̃⟩, ⟨d̃⟩, ⟨c̃⟩, ⟨g̃⟩). This dot has survived to this day in some Irish texts written in Gaelic type, but especially in roman type it is generally replaced with the ⟨h⟩ digraphs (including ⟨fh⟩, ⟨sh⟩, ⟨mh⟩, ⟨nh⟩ for Old Irish ⟨f̃⟩, ⟨s̃⟩, ⟨m̃⟩, ⟨ñ⟩; so the Irish name of Dublin and of the lenition dot given in roman type above are to be spelled |Baile Átha Cliath| and |ponc séimhithe| when written in Gaelic type).

While the Irish ‘fortis’ consonants generally resemble the Latin sounds corresponding to these letters (⟨f̃⟩ for [f], ⟨s̃⟩ for [s], etc.), lenition always leads to sounds that do not correspond to the common function of the letter (with the possible exception of ⟨ñ⟩ [n]), although many of them are not as such foreign to the Latin language. Consequently, someone who does not understand the intricacies of Irish consonant alternations would perceive the dot as simply signifying that the sound to be pronounced is different from the sound the letter is usually associated with. Even if someone read Old Irish texts without even knowing how they were pronounced, it would be easy to figure out that the dots meant that the letter under them somehow had to be pronounced differently. This is exactly the function of the dot in the *Orthographia Bohemica*: to mark a sound that is somehow different from the sound normally associated with the letter (irrespective of whether this different sound was caused by palatalization, as in the case of ⟨s̃⟩ for [ʃ], or by velarization, as in the case of ⟨l̃⟩ for [ɫ]).

The other diacritic used in Old Irish is the so-called *síneadh fada* [ˈʃiːn̪ə ˈfʲadʲə] ‘length mark’ (˘), which was used to mark long vowels. As we have seen in 5.1, it was borrowed from the Latin apex with the same shape and function. In contrast to the Latin grammarians’ recommendations about the apex, however, the Irish length mark was used to mark all the long vowels rather consistently, not just to avoid ambiguity. Note that apart from marking vowel length, the stroke also helps with determining which vowel letters are

	acute/stroke (V)	stroke (C)	dot	grave	circumflex	umlaut/diaeresis	háček	macron	breve	ring	tilde	double acute	dash	cedilla/comma	ogonek
	´	˘	·	ˋ	ˆ	¨	ˇ	ˉ	˘	◌◌	˜	˝	/	¸	˛
Kashubian	ó	ń	ż	ò	ô	ë	–	–	–	–	ã	–	ł	–	ą
Polish	ó	ć	ż	–	–	–	–	–	–	–	–	–	ł	–	ą
Sorbian	ó	ć	–	–	–	–	š	–	–	–	–	–	ł	–	–
Catalan	é	–	ł	à	–	ü	–	–	–	–	–	–	ç	–	–
Old Norse	á	–	è	–	–	–	–	–	–	–	–	–	–	–	ø
Old Irish	á	–	š	–	–	–	–	–	–	–	–	–	–	–	–
Old Czech	á	–	š	–	–	–	–	–	–	–	–	–	–	–	–
Portuguese	á	–	–	ã	ô	ü	–	–	–	–	õ	–	–	ç	–
French	é	–	–	à	ô	ë	–	–	–	–	–	–	–	ç	–
Welsh	ŵ	–	–	ŵ	ŵ	ï	–	–	–	–	–	–	–	–	–
Italian	è	–	–	ù	–	–	–	–	–	–	–	–	–	–	–
Slovak	á	–	–	–	ô	ä	š	–	–	–	–	–	–	–	–
West Frisian	ú	–	–	–	ô	–	–	–	–	–	–	–	–	–	–
Spanish	á	–	–	–	–	ü	–	–	–	–	ñ	–	–	–	–
Hungarian	á	–	–	–	–	ö	–	–	–	–	–	ő	–	–	–
Icelandic	á	–	–	–	–	ö	–	–	–	–	–	–	ð	–	–
Luxembourgish	é	–	–	–	–	ë	–	–	–	–	–	–	–	–	–
Czech	á	–	–	–	–	–	š	–	–	ů	–	–	–	–	–
Northern Sami	á	–	–	–	–	–	š	–	–	–	–	–	č	–	–
Orm's English	á	–	–	–	–	–	–	–	ă	–	–	–	ț	–	–
Dutch	é	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Irish (roman)	á	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Serbo-Croatian	–	ć	–	–	–	–	š	–	–	–	–	–	đ	–	–
Lithuanian	–	–	ė	–	–	–	š	ū	–	–	–	–	–	–	ą
Maltese	–	–	ġ	–	–	–	–	–	–	–	–	–	ħ	–	–
Breton	–	–	–	ù	ô	ü	–	–	–	–	ñ	–	–	–	–
Romanian	–	–	–	–	î	–	–	–	ă	–	–	–	–	ș	–
Estonian	–	–	–	–	–	ö	š	–	–	–	õ	–	–	–	–
Turkish	–	–	–	–	–	ö	–	–	ğ	–	–	–	–	ç	–
Swedish	–	–	–	–	–	ö	–	–	–	å	–	–	–	–	–
Albanian	–	–	–	–	–	ë	–	–	–	–	–	–	–	ç	–
Finnish	–	–	–	–	–	ö	–	–	–	–	–	–	–	–	–
German	–	–	–	–	–	ö	–	–	–	–	–	–	–	–	–
Latvian	–	–	–	–	–	–	š	–	–	–	–	–	–	ņ	–
Slovenian	–	–	–	–	–	–	š	–	–	–	–	–	–	–	–
Danish	–	–	–	–	–	–	–	–	–	å	–	–	ø	–	–
Norwegian	–	–	–	–	–	–	–	–	–	å	–	–	ø	–	–
Basque	–	–	–	–	–	–	–	–	–	–	ñ	–	–	ç	–
English	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Fig. 9. Inventories of diacritics in European languages

to be pronounced as vowels at all and which are used to determine whether the adjacent consonants are palatalized or not, e.g. ⟨*déanamh*⟩ /'d̪eːn̪əw/ 'doing' vs. ⟨*Seán*⟩ /'ʃaːn̪/ 'John' or ⟨*móin*⟩ /m̪ˠoːn̪i/ 'sod, turf' vs. ⟨*croíleacán*⟩ /'kr̪iːl̪əkaːn̪/ 'core'.

How might someone who did not know Irish and came across these strokes in a manuscript perceive them? A modern Western European might be inclined to read the 'acutes' as stress marks because this is what they are used for in many modern languages and transcription systems. However, the fact that these strokes occur on many monosyllables and sometimes more than once within a word (see the last example above) would contradict this interpretation. Moreover, someone coming from the background of Latin education would be prejudiced by the importance of vowel length for Latin versification (where the syllables of a dactyl are defined as long–short–short, not stressed–unstressed–unstressed), whereas word stress in Latin could be predicted from syllable structure and vowel length. Furthermore, if the person seeing these marks was a Czech, then there would be no other way but to interpret them as length marks. Since the Czech language has no phonological word stress (stress generally falling on the first syllable of a phonological word) but does have distinctive vowel length as a rather prominent feature, a Czech would have guessed the stroke to be a length mark even if it had actually marked something else (e.g. nasality, creaky voice, or whatever).

Consequently, Old Irish had exactly the same two diacritics that were introduced by the Orthographus Bohemicus, and a medieval Czech scholar could be expected to arrive at the result we see in the *Orthographia Bohemica* even by guessing at their function without any knowledge of Irish. The question is: How could the Czech scholar come to see Old Irish texts at all? How could they have travelled from Hibernia to Bohemia?

The idea that the Orthographus Bohemicus might have borrowed his diacritics from Irish was already formulated by Schröpfer (1968, 27, 30, 99–102). However, ever since then, if it was mentioned at all, Bohemists have invariably dismissed it as far-fetched (e.g. Stejskal 1971, 270f.; Mareš 1973, 90; 1975, 70; Miklas 1989, 21; Pleskalová 2007, 501; Pleskalová p.c.; Voleková 2019, xxxv, lx).

And far-fetched it is, literally: Dublin is roughly 1,500 km away from Prague as the crow flies. And if you are not a crow, you have to take a boat across two seas on the way. In the Middle Ages, the journey would have taken at least two months, probably more. As to language contact: there is not a single Irish loanword in Czech, nor is there a Czech loanword in Irish.²⁴⁾

Nonetheless, people did travel between the Czech lands and the British Isles. For example, Jan Hus's close confidant Jerome of Prague (who was burned roughly eleven months after Hus, at the same spot) went to Oxford in 1402 to study there and to copy writings by John Wyclif so that they could also be read in Prague. In 1407 he returned to Oxford to study there once more. There, Jerome might easily have met fellow students

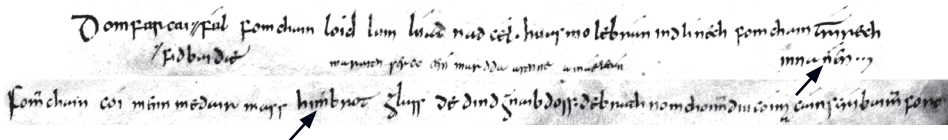
24) An exception, of course, are proper names and country-specific terms that are used when talking about the respective country or its people; e.g. the Irish Wikipedia article about the Czech Republic contains words like *koruna* ([on-line]. <<https://ga.wikipedia.org/w/?oldid=1040052>> [cit. 02-09-2022]), and in the Czech Wikipedia article about Ireland we can find sentences like “Taoiseachem je od června 2020 Micheál Martin ze strany Fianna Fáil” (“Since June 2020, the taoiseach has been Micheál Martin from the party Fianna Fáil” ([on-line] <<https://cs.wikipedia.org/w/?oldid=21629006>> [cit. 02-09-2022])). An Irish name also borne by Czechs is *Oskar*, which became popular throughout Europe in the 18th century due to James Macpherson's *Poems of Ossian*.

from Ireland. In general, the contacts between England and Bohemia were surprisingly close at the time, since in 1382 King Richard II of England had married Anne of Bohemia [Anna Česká], the eldest daughter of Charles IV (cf. Heller 1835, 7).

Even more important for our question is the Hiberno-Scottish mission in the 6th–8th centuries, which was renewed in the 11th–12th centuries. In this process, many monasteries were founded in Central European towns, which were still exclusively run by Irish monks around 1400, among them Regensburg (Ratisbon), Vienna, Erfurt, Würzburg, Leipzig, Munich, Nuremberg, Prachatice, and, of course, St. Gall(en), named after one of the most prominent Irish missionaries, Saint Gall(us) (who died around 646; the Czech form of the name *Gallus* is *Havel*). His skull has been in Prague since 1353, in St. Gall Church [Kostel svatého Havla], where Jan Hus gave sermons in 1402 (Hejda n.d.).²⁵

If a student nowadays wants to learn Old Irish, most of the original texts she will read do not belong to libraries in Ireland. Instead, they were preserved in the Irish monasteries on the continent, some of which (Prachatice, Regensburg, Vienna) had close relations to Prague (and even to Jan Hus, who probably went to school in Prachatice; cf. Schröpfer 1968, 100f.). The Irish monks had taken the manuscripts with them to Central Europe, partly because they needed them for their mission and partly to save them from the frequent Viking attacks in Ireland. Some of these texts are still preserved today; at the turn of the 15th century, when the manuscripts were six centuries younger and when the monasteries were still peopled by Irish monks, there must have been many more of them.

Most of the preserved manuscripts are not completely written in Irish but are in fact Latin manuscripts with Irish glosses and sometimes other texts added between the lines or in the margins. An example of such a text is the St. Gall Priscian (Codex Sangallensis 904). It is a copy made around 845 of the famous Latin Grammar written by Priscian around 500 CE. Of course, the St. Gall Priscian is in Latin, but it contains a lot of inter-linear explanations of Latin words in Old Irish (just like three other preserved Irish copies of Priscian, cf. Thurneysen 1909, 6). Additionally, in the bottom margins of pp. 203 and 204 we find the following lines:



Domfarcai fidbaidae fál. fomchain lóid luin
lúad nad cé. huas mo lebrán indlínech.
fomchain trírech innaén ...

Fommchain cói menn medair mass.
himbrot glass de dindgnaib doss. debrath
nomchoimmdiu cóima. cáinscríbaimm
foróida r<oss>.

(Bauer et al. 2017)

A hedge of trees surrounds me: a blackbird's
lay sings to me – praise which I will not hide ...

Above my booklet, the lined one, the trilling
of the birds sings to me. In a gray mantle the
cuckoo's beautiful chant sings to me from the
tops of bushes: may the Lord be kind to me!
I write well under the greenwood. (*Ibidem*)

25) When on 4 November 2016, at the conference *Intellectuals, identities and ideas in the cultural space of Central Europe in the 20th and 21st century – and their history*, I gave a talk about this subject at Charles University in Prague, chance would have it that I could demonstrate how near Irish culture is by pointing my finger at St. Gall Church, which was visible through the window of the auditorium, on the opposite side of the street.

Priscian's grammar was a text that an author of an orthographic treatise, of course, had to have read. If in reading it he should have come across glosses and marginalia like these (Schröpfer 1968, 30), the strokes above the vowels and the dots above the consonants would have been clearly visible (the latter are marked by arrows in the facsimile and printed bold in the transcript).

For Celtology, the fact that most of the Old Irish manuscripts contain Irish text only in the form of interlinear and marginal notes means that hardly any longer Old Irish texts have been preserved. For us, however, it means that these manuscripts were not useless for people who did not know Irish. Consequently, the Orthographus Bohemicus could easily have been reading the main text and might then have noticed that there were glosses in a language he did not understand, which contained dots and strokes above the letters.

6.8 Old Norse or early Middle English

As we saw in 5.1, Irish was not the only language that used the Latin apex as a length mark; it was also used in Old Norse and in some English texts from the eleventh and twelfth centuries. However, Old Norse texts have mainly been preserved in Scandinavia, and early Middle English texts in England. There were no English or Scandinavian monasteries in or near Bohemia in the 14th/15th century. Even those Czechs who visited England (like Jerome of Prague, see 6.7) would not have stumbled across length marks in English texts, because by the time of Chaucer (who died in 1400) the apex was not used anymore. As to Old Norse, the apex was also used much more sporadically there than it was in Old Irish, because Old Norse makes ample use of scribal abbreviations, which tend to conflict with other supralinear marks. More importantly, Old Norse and early Middle English texts contain only the apex but no dot, so that we would still need a different explanation for the dot.

The only exception is the *First Grammatical Treatise*, which also uses the dot, but on vowels rather than consonants, and this use has not been adopted in any other text. The treatise is preserved in a single mid-14th-century manuscript, the *Codex Wormianus*, which was written in Iceland and seems to have remained on the island until it came to Copenhagen in the 17th century (Benediktsson 1972, 16–19), but once there probably were more copies of it (and a 12th-century original), and it cannot be excluded that one of them was somehow seen by a Czech scholar. In principle, it is also feasible that this Czech, seeing both strokes and dots on vowels (and sometimes both together, as in ⟨mǫna⟩ 'mother' or ⟨rūnar⟩ 'runes', cf. Benediktsson 1972, 130, 220–223), decided to use the dot to mark the extra consonant phonemes he needed in Czech. However, the additional assumptions needed for this connection make it extremely unlikely that the Czech diacritics were borrowed from the *FGT*.

7. Conclusion

The rather extensive sections 4 and 5 have hopefully convinced you that every diacritic, wherever we can reconstruct its history, has an etymology in which the development of both its form and its function follows general rules of linguistic development. On the basis of this finding, I have analysed all the published hypotheses about the origin of the two Czech diacritics that I am aware of. As is often the case when dealing with history, it is hardly possible to prove or disprove any of these hypotheses. However, we can evaluate how well a hypothesis agrees with our knowledge about how diacritics develop. The criteria for this evaluation are: a) whether the diacritics have the same form (including positioning);

b) whether they have the same function (or the function expected to be given to them in the borrowing process); c) whether the source can explain the origin of both dot and stroke or if it necessitates another explanation for the other diacritic; d) whether the source was available to a learned Czech in the late 14th and early 15th century.

As fig. 10 shows, most hypotheses fulfil only some of the criteria. The only one that fulfils all is the hypothesis that the Czech system of diacritics described by the *Orthographus Bohemicus* was borrowed from Old Irish.

A last objection to this solution of the riddle might be: Why then is Old Irish not mentioned in the *Orthographia Bohemica*? There are two possible answers to this. First, it is possible that the inventor (or, rather, importer!) of the Czech diacritics and the author of the treatise were two different people; then the author might not even have known where the original idea came from. Second, even if the author of the treatise knew that the diacritics had been borrowed from Old Irish, the purpose of his treatise was not to give an accurate historical account of the emergence of the orthography. This is quite different from a text like the Monk Xrabr's *O pismenъхъ* [On the letters], whose aim is exactly that: to explain how the 'Slavs' received their already established alphabet (and to stress the role of Saint Cyril and divine inspiration). By contrast, the purpose of the treatise *De Orthographia Bohemica* was to describe how the Czech language could be spelled accurately and to convince the readers to really use this system. All the languages the author mentions help with this argument: Latin, Greek, and Hebrew as the three holy languages that each had their own letters; 'Slavic' as the ancestral language of the Czechs, which was highly valued not only by Charles IV and which also had its own alphabet(s); and German as the language of the powerful neighbours, who also had to find their own solutions to cope with the insufficiencies of the Latin alphabet, so that it would be no disgrace if the Czechs had to do so too. Irish would, for most of the readers, have been a language they neither knew nor cared about, so there was no point mentioning it. None of the languages

	dot		stroke		explains both	source available
	form	function	form	function		
i tittle	+	-			-	+
mark on (č)	+	-			-	+
šin dots	-	-			-	+
dagesh	-	-			-	+
Glagolitic	-	-			-	+
Greek acute	+	-			-	?
Latin apex			+	+	-	-
Old Irish	+	+	+	+	+	+
Old Norse			+	+	-	?
FGT	+	-	+	+	+	-
English			+	+	-	?

Fig. 10. Evaluation of hypotheses

is cited to tell us where the idea for the dot or the stroke was taken from, because this question was not the subject of the treatise. But it was the subject of the present article, and I hope to have answered it.

References:

- Anderson – Britton 1999: ANDERSON, J. – BRITTON, D.: The orthography and phonology of the *Ormulum*. *English Language and Linguistics* 3, 1999, 2, 299–334.
- Baker (n.d.): BAKER, P. S.: A note about *e caudata*. In: Baker, P. S.: *Junicode* [on-line]. <<http://junicode.sourceforge.net/eacadata.html>> [cit. 20-09-2018].
- Balázs 1958: BALÁZS, J.: Zur Frage der Typologie europäischer Schriftsysteme mit lateinischen Buchstaben. *Studia Slavica Academiae Scientiarum Hungaricae* 4, 1958, 3–4, 251–292.
- Bauer et al. 2017: BAUER, B. – HOFMAN, R. – MORAN, P.: *St Gall Priscian Glosses*, v2.0 [on-line]. <<http://stgallpriscian.ie/index.php?id=11430&an=1>> [cit. 31-08-2022].
- Benediktsson 1972: BENEDIKTSSON, H. [ed.]: *The First Grammatical Treatise. Introduction, text, notes, translation, vocabulary, facsimiles*. Reykjavík 1972.
- Berger 2010: BERGER, T.: Argumentace reformátorů českého pravopisu v letech 1780–1850. In: Čornejová, M. – Rychnovská, L. – Zemanová, J. [edd.]: *Dějiny českého pravopisu (do r. 1902)*. Sborník příspěvků z mezinárodní konference. Host, Brno 2010, 343–359.
- Berger 2012: BERGER, T.: Religion and diacritics. The case of Czech orthography. In: Baddeley, S. – Voeste, A. [edd.]: *Orthographies in Early Modern Europe*. De Gruyter Mouton, Berlin 2012, 255–268.
- Bermel 2007: BERMEL, N.: *Linguistic authority, language ideology, and metaphor. The Czech orthography wars*. De Gruyter Mouton, Berlin 2007.
- Bernolák 1790: BERNOLÁK, A.: *Grammatica slavica*. Landerer, Posonii 1790.
- Biblia XI.A.14: Biblia [manuscript XI.A.14 of the Czech National Library, 1417–1436]. In: *Manuscriptorium* [on-line]. <https://manuscriptorium.com/apps/?direct=record&pid=AIPDIG-NK_CR_XI_A_14_4D6NEGD-cs> [cit. 02-09-2022].
- Boulding 1953: BOULDING, K. E.: Toward a general theory of growth. *Canadian Journal of Economics and Political Science / Revue canadienne d'Economique et de Science politique* 19, 1953, 3, 326–340.
- Branch 2015: BRANCH, G.: “Everything is the way it is ...”. In: *National Center for Science Education*, 12 Feb 2015 [on-line]. <<https://ncse.ngo/everything-way-it>> [cit. 18-08-2022].
- Bredel – Bunčić 2020: BREDEL, U. – BUNČIĆ, D.: Apostroph. In: Neef, M. – Weingarten, R. – Sahel, S. [edd.]: *Schriftlinguistik* (Wörterbücher zur Sprach- und Kommunikationswissenschaft 5) [on-line]. <https://www.degruyter.com/database/WSK/entry/wsk_id_wsk_artikel_artikel_14048/html> [cit. 02-09-2022].
- Bunčić 2004: БУНЧИЦ, Д.: Энантисемия внутриязыковая и межъязыковая как проблема коммуникации. In: Grabska, M. [ed.]: *“Słowa, słowa, słowa”... w komunikacji językowej II*. Gdańsk 2004, 207–212.
- Bunčić 2012: BUNČIĆ, D.: The standardization of Polish orthography in the sixteenth century. In: Baddeley, S. – Voeste, A. [edd.]: *Orthographies in Early Modern Europe*. De Gruyter Mouton, Berlin 2012, 219–257.
- Bunčić 2013–2020: BUNČIĆ, D.: Adhak; Anusvāra; Apex (2); Avagrāha; Bindī; Cedille; Chandra; Chandrabindu; Chvostik; Dagesch; Durchstreichung; Gravis; Háček; Mappig; Nasalstrich; Nukta; Ogonek; Raphe; Tilde (1); Tippī; Virāma; Visarga; Zirkumflex. In: Neef, M. – Weingarten, R. – Sahel, S. [edd.]: *Schriftlinguistik* (Wörterbücher zur Sprach- und Kommunikationswissenschaft 5) [on-line]. <<https://doi.org/10.1515/wsk>> [cit. 02-09-2022].

- Bunčić 2014: BUNČIĆ, D.: Unterscheidungszeichen. In: Neef, M. – Weingarten, R. – Sahel, S. [edd.]: *Schriftlinguistik* (Wörterbücher zur Sprach- und Kommunikationswissenschaft 5) [on-line]. <https://www.degruyter.com/database/WSK/entry/wsk_id_wsk_artikel_artikel_13690/html> [cit. 02-09-2022].
- Bunčić 2016: BUNČIĆ, D.: Czech (16th–18th centuries): diamesic diorthographia. In: Bunčić, D. et al.: *Biscriptality. A sociolinguistic typology*. Winter, Heidelberg 2016, 140–143.
- Bunčić 2020a: BUNČIĆ, D.: Diakritisches Zeichen. In: Neef, M. – Weingarten, R. – Sahel, S. [edd.]: *Schriftlinguistik* (Wörterbücher zur Sprach- und Kommunikationswissenschaft 5) [on-line]. <https://www.degruyter.com/database/WSK/entry/wsk_id_wsk_artikel_artikel_14748/html> [cit. 02-09-2022].
- Bunčić 2020b: BUNČIĆ, D.: Wie entstehen diakritische Zeichen? In: Lederer, I. – Meyer, A.-M. – Schlund, K. [edd.]: *Linguistische Beiträge zur Slavistik. XXVI. und XXVII. JungslavistInnen-Treffen, 6. bis 8. September 2017 in Bamberg und 12. bis 14. September 2018 in Heidelberg*. Peter Lang, Frankfurt am Main 2020, 13–31.
- Bunčić 2020c: BUNČIĆ, D.: Jan Hus und die Traditionslinien europäischer diakritischer Zeichen. In: Bochmann, K. [ed.]: *Sprechen und Schreiben. Schriftsysteme und ihre linguistischen, kulturellen und politischen Dimensionen*. S. Hirzel, Stuttgart – Leipzig 2020, 52–68.
- Camus 1959: CAMUS, A.: *Les possédés. Pièce en trois parties adaptée du roman de Dostoïevski*. Gallimard, Paris 1959.
- Čejka 1999: ČEJKA, M.: Srovnání Devotyho opisu Loutny české Adama Michny z Otradovic s původním tiskem. In: Zand, G. – Holý, J. [edd.]: *Tschechisches Barock. Sprache, Literatur, Kultur*. Peter Lang, Frankfurt am Main 1999, 21–32.
- Čermák 2012: ČERMÁK, V.: Hlaholská abecedaria v českém prostředí. In: Nejedlý, P. – Vajdlová, M. [edd.]: *Cesty slov*. Praha 2012, 36–42.
- Christiansen 1889: CHRISTIANSEN, J.: *De apicibus et i longis inscriptionum latinarum*. C. F. Delff, Husum 1889.
- Codex Assemanianus: *Асеманиево евангелие. Старобългарски глаголически паметник от X век. Факсимилно издание*. София 1981.
- Coulmas 1999: COULMAS, F.: Diacritic. In: Coulmas, F.: *The Blackwell Encyclopedia of Writing Systems*. Blackwell, Malden (MA) 1999, 126–127.
- Curtius 1948: CURTIUS, E. R.: *Europäische Literatur und lateinisches Mittelalter*. 3rd edn. Francke, Bern 1948.
- Diringer 1962: DIRINGER, D.: *Writing*. Praeger, New York 1962.
- Emmaus Bible: Bible Hlaholská [manuscript XVII.A.1 of the Czech National Library, 1416]. In: *Manuscriptorium* [on-line]. <https://www.manuscriptorium.com/apps/index.php?direct=record&pid=RTRTR1-NKCR_XVII_A_1___3636BJ9-xx#search> [cit. 02-09-2022].
- Erasmus 1528: ERASMUS ROTERODAMUS, D.: *De recta latini graecique sermonis pronuntiatione*. Parisiis 1528 [on-line]. <<http://digital.ub.uni-duesseldorf.de/ihd/content/pageview/328038>> [cit. 02-09-2022].
- Fjellhammer Seim 2007: FJELLHAMMER SEIM, K.: Runologie. In: Haugen, O. E. [ed.]: *Altnordische Philologie. Norwegen und Island*. De Gruyter Mouton, Berlin 2007, 147–222.
- Flajšhans 1891: FLAJŠHANS, V.: Klasobraní po rukopisích. XXIX. Staročeský zlomek evangelia sv. Marka. *Listy filologické* 18, 1891, 1/2, 92–97.
- Fontaine 2006: FONTAINE, M.: “Sicilicissitat” (Plautus, “Menaechmi” 12) and early geminate writing in Latin (with an appendix on “Men.” 13). *Mnemosyne* 59, 2006, 1, 95–110.
- Friedrich 1904–1907: FRIEDRICH, G. [ed.]: *Codex diplomaticus et epistolaris Regni Bohemiae*. Vol. 1. Wiesner, Pragae 1904–1907.
- Friedrich 1966: FRIEDRICH, J.: *Geschichte der Schrift unter besonderer Berücksichtigung ihrer geistigen Entwicklung*. Winter, Heidelberg 1966.

- Frinta 1940: FRINTA, A.: O předloze Husovy reformy pravopisné. *Listy filologické* 67, 1940, 3/4, 252–255.
- Fuhrhop – Buchmann 2009: FUHRHOP, N. – BUCHMANN, F.: Die Längenhierarchie. Zum Bau der graphematischen Silbe. *Linguistische Berichte* 218, 2009, 127–155.
- Gebauer 1878: GEBAUER, J.: Ku kvantitě ve staročeských Zlomcích Epických. *Listy filologické* 5, 1878, 2–4, 219–227.
- Gelb 1963: GELB, I. J.: *A study of writing*. (Revised edition.) University of Chicago Press, Chicago – London 1963.
- Gillet 1863: GILLET, E. H.: *The life and times of John Huss; or, The Bohemian Reformation of the fifteenth century*. Vol. 1. Gould and Lincoln, Boston 1863.
- Glück 2005: GLÜCK, H.: Jan Hus als Phonologe und Orthographiereformer. In: von Erdmann, E. et al. [edd.]: *Tusculum slavicum. Festschrift für Peter Thiergen*. Pano, Zürich 2005, 583–593.
- Glück 2016: GLÜCK, H.: Diakritikon. In: Glück, H. – Rödel, M. [edd.]: *Metzler Lexikon Sprache*. Metzler, Stuttgart 2016, 144.
- Grønlie 2006: GRØNLIE, S. [trad.]: *Íslendingabók. Kristni saga / The Book of the Icelanders. The story of the conversion*. University College, London 2006.
- Grutschus 2009: GRUTSCHUS, A.: *Strategien der Musikbeschreibung. Eine diachrone Analyse französischer Toneigenschaftenbezeichnungen*. Berlin 2009.
- Haarmann 1991: HAARMANN, H.: *Universalgeschichte der Schrift*. 2nd edn. Campus, Frankfurt 1991.
- Hale 1995: HALE, J. K.: Observations on Milton's accents. *Renaissance and Reformation* 19, 1995, 3, 23–34.
- Hattala 1852: [HATTALA, M.]: *Krátká mluvnica slovenská*. Prešporok 1852.
- Haugen 1972: HAUGEN, E. [ed.]: *First Grammatical Treatise. The earliest Germanic phonology. An edition, translation and commentary*. 2nd edn. Longman, London 1972.
- Hejda (n.d.): HEJDA, J.: Historie. Kostel sv. Havla na Starém městě pražském. In: *Římskokatolická farnost u kostela Matky Boží před Týnem* [on-line]. <<http://www.tyn.cz/?stranka=historie>> [cit. 04-02-2018].
- Heller 1835: HELLER, L.: *Hieronymus von Prag*. Friedrich Aschenfeldt, Lübeck 1835.
- Hempl 1899: HEMPL, G.: The origin of the Latin letters G and Z. *Transactions and Proceedings of the American Philological Association* 30, 1899, 24–41.
- Henzen et al. 1882: HENZEN, G. et al. [edd.]: *Inscriptiones urbis Romae Latinae* (Corpus Inscriptionum Latinarum 6, 2). Georgius Reimer, Berolini 1882.
- Hieronymus 1845: HIERONYMUS, S. E.: *Opera omnia*. Vol. 2. (Patrologiae cursus completus, ed. by Jacques Paul Migne, series 1, vol. 23.) Vrayet, Parisiis 1845.
- Hradec Manuscript: Hradecký rukopis [manuscript XXIII.G.92 of the Czech National Library, 1366–1400]. In: *Manuscriptorium* [on-line]. <https://manuscriptorium.com/apps/?direct=record&pid=AIPD1G-NKCR__XXIII_G_92_35XWCJB-cs> [cit. 02-09-2022].
- IPA chart: International Phonetic Association: *Full IPA chart. Reproduction of The International Phonetic Alphabet* [on-line]. <<https://www.internationalphoneticassociation.org/content/full-ipa-chart>> [cit. 18-08-2022].
- Jakovlev 1930: ЯКОВЛЕВ, Н. Ф.: За латинизацию русского алфавита. *Культура и письменность востока* 6, 1930, 27–43.
- Kamusella 2019: KAMUSELLA, T.: Is hot air mightier than states? The big Central European history of a little tail (ogonek). *New Easter Europe*, 12 Dec 2019 [on-line]. <<https://neweasterneurope.eu/2019/12/12/is-hot-air-mightier-than-states/>> [cit. 16-01-2023].
- Keil 1923: KEIL, J.: 7. Inschriften. In: Österreichisches Archäologisches Institut [ed.]: *Forschungen in Ephesos*. Vol. 3. Wien 1923, 91–168.

- de Kemmeter 2021: DE KEMMETER, F.: 25th anniversary of Thalys: How did it start? In: *RailTech*, 2 June 2021 [on-line]. <<https://www.railtech.com/policy/2021/06/02/25th-anniversary-of-thalys-how-did-it-start/>> [cit. 02-09-2022].
- Kiplinger 1962: KIPLINGER, W. M. [ed.]: The story behind ... Kodak trademark. *Changing Times: The Kiplinger Magazine* 16, 1962, 4, 40.
- Kralice Bible: *Bibl české. Djl druhý*. [Kralice] 1580. (Facsimile [on-line]. <<https://nbn-resolving.org/urn:nbn:de:bvb:12-bsb10861203-8>> [cit. 02-09-2022].)
- Kramer 1991: KRAMER, J.: Die Verwendung des Apex und P. Vindob. L 1 c. *Zeitschrift für Papyrologie und Epigraphik* 88, 1991, 141–150.
- Kramer 1996: KRAMER, J.: Verschriftungsarten und -tendenzen in der Romania. In: Holtus, G. et al. [edd.]: *Lexikon der Romanistischen Linguistik*. Bd. II, 1. Niemeyer, Tübingen 1996, 584–597.
- Kvítková 2010: KVÍTKOVÁ, N.: Grafická stránka nejstarších zápisů kroniky tak řečeného Dalimila. In: Čornejová, M. – Rychnovská, L. – Zemanová, J. [edd.]: *Dějiny českého pravopisu (do r. 1902)*. Sborník příspěvků z mezinárodní konference. Host, Brno 2010, 75–79.
- Lindsay 1911: LINDSAY, W. M. [ed.]: *Isidori Hispalensis episcopi Etymologiarum sive Originum libri XX*. E typographeo Clarendoniano, Oxonii 1911.
- Mac Eoin 1998: MAC EOIN, G.: Literacy and cultural change in early Ireland. In: Ehler, Ch. – Schaefer, U. [edd.]: *Verschriftung und Verschriftlichung. Aspekte des Medienwechsels in verschiedenen Kulturen und Epochen*. Narr, Tübingen 1998, 99–131.
- Mareš 1973: MAREŠ, F. W.: Kyrillo-methodianische Wurzeln der tschechischen diakritischen Orthographie. *Anzeiger der Österreichischen Akademie der Wissenschaften – Philosophisch-historische Klasse* 110, 1973, 4, 81–99.
- Mareš 1975: MAREŠ, F. V.: Emauzské prameny českého diakritického pravopisu. In: Petr, J. – Šabouk, S. [edd.]: *Z tradic slovanské kultury v Čechách. Sázava a Emauzy v dějinách české kultury*. Universita Karlova, Praha 1975, 169–172.
- Marti 2012: MARTI, R.: On the creation of Croatian. The development of Croatian Latin orthography in the 16th century. In: Baddeley, S. – Voeste, A. [edd.]: *Orthographies in Early Modern Europe*. De Gruyter Mouton, Berlin 2012, 269–320.
- McManus 1996: MCMANUS, D.: Irish. In: Daniels, P. T. – Bright, W. [edd.]: *The world's writing systems*. Oxford University Press, New York 1996, 656–660.
- Menéndez Pidal 1954: MENÉNDEZ PIDAL, R.: *Cantar de mio Cid*. Espasa, Madrid 1954.
- Miklas 1989: MIKLAS, H.: Slavische Sprachreformatoren in West und Ost. Der Fall des Tschechen Jan Hus und des Bulgaren in serbischen Diensten Konstantin von Kostenec. *Die Welt der Slaven* 34, 1989, 1, 18–31.
- Millward 1996: MILLWARD, C. M.: *A biography of the English language*. 2nd edn. Heinle, Boston (MA) 1996.
- Moiseenko 1997: МОИСЕЕНКО, В. Е.: Из истории славянского диакритического правописания. *Studia Slavica Academiae Scientiarum Hungaricae* 42, 1997, 1–2, 99–107.
- Oliver 1966: OLIVER, R. P.: Apex and sicilicus. *American Journal of Philology* 87, 1966, 2, 129–170.
- Pacnerová 1996: PACNEROVÁ, L.: Česká varianta charvátské hranaté hláholice. *Slovo* 44–46, 1996, 45–62.
- Palacký 1827: PALACKÝ, F.: Liternej zprávy. *Časopis společnosti vlastenského Museum v Čechách* 1, 1827, 1, 132–140.
- Paris et al. 1852: PARIS, L. – KOPITAR, J. – SILVESTRE, J. B.: *Évangélique slave dit Texte du sacre, de la Bibliothèque de Reims*. Didron, Paris 1852.
- Patera 1881: PATERA, A. [ed.]: *Hradecký rukopis*. Matice česká, Praha 1881.
- Patera 1890: PATERA, A. [ed.]: *Žaltář klementinský*. Matice česká, Praha 1890.
- Pleskalová 1998: PLESKALOVÁ, J.: *Tvoření nejstarších českých osobních jmen*. Masarykova univerzita, Brno 1998.

- Pleskalová 2007: PLESKALOVÁ, J.: Pravopis. Od počátků do začátku národního obrození. In: Pleskalová, J. et al. [edd.]: *Kapitoly z dějin české jazykovědné bohemistiky*. Academia, Praha 2007, 499–515.
- Porák 1983: PORÁK, J.: *Humanistická čeština. Hláskosloví a pravopis*. Univerzita Karlova, Praha 1983.
- Ráj duše: Ráj duše [manuscript XVII.A.19 of the Czech National Library, 1383–1400]. In: *Manuscriptorium* [on-line]. <https://manuscriptorium.com/apps/?direct=record&pid=AIPDIG-NKCR__XVII_A_19__1A2YHZ6-cs> [cit. 02-09-2022].
- Rasper 2017: RASPER, M.: „No Sports“ hat Churchill nie gesagt. *Das Buch der falschen Zitate*. Ecowin, Salzburg 2017.
- Rateče Manuscript: Rateški rokopis [manuscript, 1362–1390]. In: *Wikipedia* [on-line]. <<https://sl.wikipedia.org/w/?oldid=4064047>> [cit. 02-09-2022].
- Robert 1895: ROBERT, U.: Note sur l'origine de l'ę cédillé dans les manuscrits. In: Delisle, L. [ed.], *Mélanges Julien Havet. Recueil de travaux d'érudition dédiés à la mémoire de Julien Havet (1853–1893)*. Ernest Leroux, Paris 1895, 633–637.
- Ryšánek 1948: RYŠÁNEK, F.: Další zlomek českého hlaholského Komeštorá. In: Kurz, J. et al. [edd.]: *Slovanské studie. Sbíрка statí, věnovaných relátu univ. prof. dr. Josefu Vajsovi k uctění jeho životního díla*. Vyšehrad, Praha 1948, 207–225.
- Schröpfer 1968: SCHRÖPFER, J.: *Hussens Traktat „Orthographia Bohemica“*. Die Herkunft des diakritischen Systems in der Schreibung slavischer Sprachen und die älteste zusammenhängende Beschreibung slavischer Laute. Harrassowitz, Wiesbaden 1968.
- Segert 1966: SEGERT, S.: Husův zájem o hebrejský jazyk. In: Říčan, R. – Flegl, M. [edd.]: *Husův sborník: Soubor prací k 550. výročí M. Jana Husa*. Praha 1966, 69–72.
- Segert 1981/1984/1985: SEGERT, S.: Jan Hus's reform of the Czech orthography: A Hebraist's contribution. *Byzantine Studies / Etudes Byzantines*, 8/11/12, 1981/1984/1985, 343–350.
- Sievers 1892: SIEVERS, E. [ed.]: *Tatian: Lateinisch und altdeutsch mit ausführlichem Glossar*. 2nd edn. Schöningh, Paderborn 1892.
- Sisam 1953: SISAM, K.: *Studies in the history of Old English literature*. Clarendon, Oxford 1953.
- Šlosar 2010: ŠLOSAR, D.: Diakritický pravopis. In: Čornejová, M. – Rychnovská, L. – Zemanová, J. [edd.]: *Dějiny českého pravopisu (do r. 1902)*. Sborník příspěvků z mezinárodní konference. Host, Brno 2010, 200–205.
- Steffens 1929: STEFFENS, F.: *Lateinische Paläographie. 125 Tafeln in Lichtdruck mit gegenüberstehender Transkription nebst Erläuterungen und einer systematischen Darstellung der Entwicklung der lateinischen Schrift*. 2nd edn. Walter de Gruyter, Berlin 1929.
- Stejskal 1971: STEJSKAL, K.: Die *Orthographia* Hussens und das slavische Kloster in Prag-Emaus. *Mediaevalia Bohemica* 70, 1971, 3, 265–277.
- Stieber 1973: STIEBER, Z.: *A historical phonology of the Polish language*. Heidelberg 1973.
- Štítýný 1376: ŠTÍTÝNÝ ZE ŠTÍTNEHO, T.: Knížky šestery o obecných věcech křesťanských [“Klementinský štítýnský sborník”, manuscript XVII.A.6 of the Czech National Library, 1376]. In: *Manuscriptorium* [on-line]. <https://manuscriptorium.com/apps/?direct=record&pid=AIPDIG-NKCR__XVII_A_6__0PIACLA-cs> [cit. 02-09-2022].
- Štítýný 1463: ŠTÍTÝNÝ ZE ŠTÍTNEHO, T.: Knihy naučení křesťanského [manuscript XVII.C.18 of the Czech National Library, 1463]. In: *Manuscriptorium* [on-line]. <https://manuscriptorium.com/apps/?direct=record&pid=AIPDIG-NKCR__XVII_C_18__2KXRQEB-cs> [cit. 02-09-2022].
- Stotz 1996: STOTZ, P.: *Handbuch zur lateinischen Sprache des Mittelalters (Handbuch der Altertumswissenschaft, Abt. 3, Teil 5)*. Vol. 3 (Lautlehre). Beck, München 1996.
- Sturtevant 1937: STURTEVANT, E. H.: The smooth breathing. *Transactions and Proceedings of the American Philological Association* 68, 1937, 112–119.
- Sylvester 2006: SYLVESTER, I.: *Grammatica Hungarolatina* [1539]. Bartók, I. [ed.]. Budapest 2006.
- Teuber 2020: TEUBER, K.: *Mein Weg nach Catan*. LangenMüller, Stuttgart 2020.

- Thümmel 2002: THÜMMEL, W.: Diakritische Zeichen. In: Bußmann, H. [ed.]: *Lexikon der Sprachwissenschaft*. Kröner, Stuttgart 2002, 161–162.
- Thurneysen 1909: THURNEYSEN, R.: *Handbuch des Alt-Irischen*. Vol. I. Winter, Heidelberg 1909.
- TLIO: Avere (1). In: *Tesoro della Lingua Italiana delle Origini* [on-line]. <<http://tlio.ovi.cnr.it/TLIOm/MobiVox.php?TLIOid=037936>> [cit. 02-09-2022].
- Tomelleri 2020: TOMELLERI, V. S.: Intorno alla nuova edizione dell' *Orthographia Bohemica*. *Studi Slavistici* 17, 2020, 2, 203–214.
- Vidmanová 1969: VIDMANOVÁ, A.: K výslovnosti a grafice středověké latiny v Čechách. *Listy filologické* 92, 1969, 4, 294–300.
- Vidmanová 1982: VIDMANOVÁ, A.: Ke spisku *Orthographia Bohemica*. *Listy filologické* 105, 1982, 75–89.
- Voleková 2018: VOLEKOVÁ, K.: Znaménko rozlišovací a zdůrazňovací ve staročeských rukopisech. *Slavia Occidentalis* 75, 2018, 1, 133–141.
- Voleková 2019: VOLEKOVÁ, K. [ed.]: *Orthographia Bohemica*. Akropolis, Praha 2019.
- Vykypěl 2020: VYKYPĚL, B.: Poznámky ke spisu *Orthographia Bohemica* a jeho novému vydání. *Linguistica Brunensia* 68, 2020, 1, 95–102.
- Zaborowski 1983: ZABOROWSKI, S.: *Orthographia seu modus recte scribendi et legendi Polonicum idioma quam utilissimus* [1514, 1518]. In: Urbańczyk, S. [ed.]: *Die altpolnischen Orthographien des 16. Jahrhunderts*. Stanisław Zaborowski – Jan Seklucjan – Stanisław Murzynowski – Jan Januszowski. Böhlau, Köln 1983, 51–115.

Slavisches Institut, Universität zu Köln
Cologne

Daniel Bunčič
daniel.buncic@uni-koeln.de
ORCID: 0000-0002-1090-8907