

THE GRAPHEMES ⟨y⟩ AND ⟨h⟩ IN ENGLISH AND CZECH

A contribution to contrastive graphemics

Josef Vachek

As is well known, the links connecting graphemes as the smallest linear items of written utterances with phonemes reflected by them as the basic items of corresponding spoken utterances are not of static character but are subject to changes taking place during the development of language within one and the same linguistic community. The impetus to such changes is regularly motivated by the alterations which take place within the phonological system and which, as a rule, are followed by analogous alterations affecting the corresponding graphemic system serving the needs of the same community. However, the latter alterations do not necessarily operate automatically (in view of persisting conservative habits of the writing and reading community members) and the results of their operations thus need not be reflected as closely parallel to the outcome of those alterations which have taken place on the phonemic level lying below the corresponding spoken utterances.

Among the most interesting cases of the absence of such close parallelism are those which arise when an item disappears from the phonological system of the language in view of its weakened position within that system, and in consequence of this the phoneme which originally reflected it was to become redundant in the system of the graphemic means used by the given community. In such cases, the simplest solution offering itself to the linguistic community would seem to be a parallel cancellation of the grapheme which no longer had a partner in the corresponding spoken norm to which it might refer. However, such a seemingly obvious solution of the given problem is resorted to only very rarely – as a rule the redundant grapheme keeps its place among the graphemic means of the given language. It only pays for its retention within that system by being functionally reevaluated, in other words, by becoming involved in some other link with items of the corresponding spoken norm.

One of the very instructive specimens of such functional reevaluation is provided by the case of the grapheme ⟨y⟩ in the course of the development both of English and of Czech. To discuss the case of English first, in the OE period the

grapheme ⟨y⟩ reflected a high mixed vocalic phoneme in two quantitative varieties, the short /y/ and the long /ȳ/. (As is commonly known, its historical origin goes back to a prehistoric /ǔ/-phoneme, which was changed to /ȳ/ by the well-known process of the ‘i-umlaut’ (as in *fyllan* < *fuljan, *mȳs* < *mūsiz). However, in Late OE the said vocalic pair /y/, /ȳ/ was to change into a front high vocalic pair /i/, /i:/, so that *fyllan* > *fillen*, *mȳs* > *mīs*. Still, the grapheme ⟨ȳ⟩ was not to be automatically changed in all instances but it was often preserved in the written norm of LOE, and even later, of Middle English, the two letters ⟨y⟩ and ⟨i⟩ being for a long time evaluated as optional allographs of one and the same grapheme ⟨i/y⟩. Still, in accordance with the universal tendency found in all languages to functionally differentiate two formally identical language means, the ME scribes were to use the two symbols for two different functions. The symbol ⟨i⟩ was to be classified as the unmarked member of the pair, while ⟨y⟩, the other member, clearly ranked as marked, being used in some specific situations, mainly before and after the graphemes denoting the nasal phonemes, as well as before and after the graphemes ⟨u⟩ and ⟨uu⟩ (corresponding to the ModE ‘double’ ⟨u⟩, i. e. ⟨w⟩). In all such cases, the use of the marked member of the pair was clearly prompted by functional motives, inasmuch as it guaranteed easier legibility of the given context. (Here one should recall that as early as 1909 A. Frinta, the pioneer of Modern Czech phonetic research, formulated the task of orthography – more exactly, of the rules underlying written utterances – as ‘speaking quickly and distinctly to the eyes, so that the pertinent idea can easily be mobilized’.) The same purpose of clear and quick functional differentiation was served by another, somewhat later ME scribal practice (very commonly attested in Chaucer’s writings) in which ⟨y⟩ was used for the long /i/-vowel while ⟨i⟩ was employed to refer to its short counterpart.

Besides, it should be noted that in some LOE manuscripts the grapheme ⟨ȳ⟩ could be used in another function which might be denoted as lexico-stylistic. Since, that is to say, at least some writers of the LOE period were aware of the grapheme ⟨f⟩ going back to an earlier ⟨ȳ⟩ in a relatively very large number of instances, they sometimes used the ‘ancient’ symbol in historically unjustified cases – they did this with the intention of imparting such words a more archaic graphemic outlook (on this point, see especially C. L. Wrenn 1943, pp. 19ff). And finally, lexico-stylistic motivation of the use of ⟨y⟩ can also be found, since the EModE period, in words which could be identified by the scribes or printers as loans from classical Greek, in which the ⟨y⟩-grapheme even now still reflects the ancient Greek *v* (see, e. g., lexical items like *crypt*, *hymn*, *myth*, etc., cf. J. Wright – E. M. Wright 1924, par. 26). Clearly, the use of ⟨y⟩ in such words was to underline – at least in the period of New Learning – their specific character of learned items of the English vocabulary.

In another category of instances the use of the redundant ⟨y⟩ was to be motivated by the morphemic structure of the written words, helping again the written utterances ‘to speak quickly and distinctly to the eye’ by preserving the grapheme ⟨y⟩ before inflectional endings (e. g. *boy* – pl. *boys*, *obey* – 3. sg. *obeys*, pret. *obeyed*, etc.), despite the otherwise very frequent tendency to use ⟨y⟩ at the end

of words while ⟨i⟩ was preferred in word-initial and word-medial positions (see, e. g., *it, idea; beauty – beautiful, very – verily*) and also before inflectional endings after the basic stems ending in a consonant grapheme (cf. *try – tried, rally – rallied*, and the like). Even such orthographic practice, of course, was in full conformity with the demand that written utterances should provide clear and quick information about the morphematic structure of the registered lexical items. (Cf. also J. Wright – E. M. Wright 1924, par. 24.)

One must not overlook, however, that the redundant grapheme ⟨y⟩ was also entrusted with another new function, basically different from the ones discussed here so far: it could also reflect since LME some consonantal phonemes. The most important part performed by it in this functional category was its replacement of ⟨ȝ⟩, the letter traditionally termed ‘yogh’ (which, as is well known, had been of fairly respectable age, going back to the OE grapheme ⟨ȝ⟩). The ME sound represented by ⟨ȝ⟩ was, of course, implemented either as a velar or as a palatal fricative, and it was only in the latter of the two implementations that the ‘yogh’ was replaced by ⟨y—⟩ (and sometimes also by a prevocalic ⟨i—⟩). This act of replacement was, naturally, motivated by very technical reasons – no letters of the yogh-type had existed in the inventory of continental symbols available to most of the post-Conquest scribes, following, as a rule, the Norman models (writers like Orm, building on the native graphemic bases, having constituted rare exceptions). The said technical problem was to be felt even more intensely by printers of the post-Caxtonian period, but already in the EME era the missing symbols started to be replaced by graphemes of more common shapes. Thus earlier *ȝaf* gave way to *iaf, yaf, ȝok* to *iok, yok* etc.

More or less as a footnote should be added here that in some very few words, particularly in some Scottish personal names, the printers were choosing another graphemic substitute for yogh, i. e. ⟨z⟩ – one can find it in names like *Dalziel, Menzies*, which originally included ⟨ȝ⟩ in their graphemic structures. The reason of this other replacement was again the purely graphical similarity of ⟨z⟩ and ‘yogh’. – Finally, a very marginal use of ⟨y⟩ reflecting a consonantal reference may be found in some very archaic spellings of the type *Ye Olde Cheshire Cheese* in which ⟨y⟩ was used in EModE for ⟨p⟩ which otherwise was replaced by the digraph ⟨th⟩: here again the unusual type of replacement had been motivated by the increasing graphical similarity of the symbols ⟨p⟩ and ⟨y⟩ in LME manuscripts (cf. K. Luick 1914 – 21. par. 56, 61).

To sum up, it has been found here that the redundant English grapheme ⟨y⟩ could be revaluated into a very useful item of the written norm of English inasmuch as it was to help make the written utterances of English more efficient in performing the basic demand imposed on the written utterances of English (and thus also on the written norm underlying them) viz. to speak quickly and distinctly to the eye. It should particularly be noted that the revaluated functioning of the English ⟨y⟩ referred not only to the basic, phonological level of English but also to its ‘higher levels’, particularly to the morphematic aspects of word-formation as well as to the facts of lexico-stylistic level.

After having discussed the ways of functional revaluation of the redundant grapheme ⟨y⟩ in English we may now turn our attention to the graphematic situation of Czech. In its development, too, the grapheme ⟨y⟩ became redundant as early as in the 14th century (the date is evidenced by orthoepic criticism of the pronunciation of the speakers of Czech at that time by John Huss). At that period of time, the old phonemic opposition /i/ : /ǝ/, inherited from Common Slavic, had been cancelled by the merger of the two phonemes into /i/. However, the redundant grapheme ⟨ǝ⟩ was again to be preserved in the written norm of Czech (just as it had been in English) because it was found useful to be entrusted by a relatively large number of new functional tasks, to be briefly commented in the following paragraphs.

As is well known, in Present Day Czech digraphs ⟨ty, dy, ny⟩ their second element ⟨y⟩ signals the non-palatal status of the consonant phoneme denoted by the first elements of the digraphs, where as the second item ⟨i⟩ in the parallel digraphs ⟨ti, di, ni⟩ signals that the status of the consonant phoneme referred to by the preceding elements of the digraphs is palatal (thus, /ti/, /di/, /ni/).¹ Just as in the written norm of English, in the written norm of Czech, too, the functional revaluation of ⟨y⟩ was not to be confined to signalling facts of the basic, phonological level alone but is also reflected in some data of the 'higher' language levels.

First of all, it is well known that the opposition of ⟨i⟩ : ⟨y⟩ has been utilized in Czech for the signalling of an important morphological differentiation within some important inflectional paradigms. This concerns the morphological category of masculine animate nouns of the type *kos* ('blackbird') whose Nom. Acc. Instr. plural are phonemically quite identical /kosi/ but whose graphemic shapes are differentiated, opposing the form of the Nom. pl. ⟨kosi⟩ to those of Acc. Instr. pl. ⟨kosy⟩.² Besides, there exists in Czech another case of underlining a morphological opposition by graphemic differentiation — it is found in the adjectival category of the paradigm *chudý* 'poor' where the opposition of the long graphemes ⟨i⟩ : ⟨y⟩ distinguishes the forms of the Nom. pl. against those of the Nom. sg., e. g. Nom. pl. ⟨slabí⟩ 'weak' as opposed to Nom. sg. ⟨slabý⟩ : similarly Nom. pl. ⟨malí⟩ 'small' : Nom. sg. ⟨malý⟩: the form of the plural, of course, are found in the adjectives only when they relate to animate nouns.

Finally, there exists in Czech a number of word-pairs the members of which are perfectly homophonous but which are lexico-semantically differentiated in the written norm by the very graphemic opposition of ⟨i⟩ and ⟨y⟩ (see, e. g., *bit* 'beat-

¹ The above formulation is, of course, valid only for the synchronically native words of the Czech wordstock; in its synchronically foreign lexical items the digraphs ⟨ti⟩, ⟨di⟩, ⟨ni⟩ refer to phonemic groups /ti/, /di/, /ni/, with non-palatal first members of the groups; see, e. g. *titul* 'title', *diktát* 'dictation', *nitrát* 'nitrate'. For the synchronic distinction between the native and foreign words see V. Mathesius 1935.

² The merit of asserting the morphematic distinction of ⟨i⟩ : ⟨y⟩ in this point must be attributed to V. Hanka (in traditional grammars of Czech the establishing of this morphological differentiation is usually referred to by the term "analogical reform of the orthography").

en': *byt* 'lodging': *vír* 'torrent' *výr* 'owl' etc.)³. Further, it should be recalled that the ModE grapheme ⟨y⟩ sometimes serves lexico-semantic purposes (documented by instances like *crypt*, *hymn*, *myth*) and that this fact can be regarded as due to stylistic motivation, since words of that category must have ranked as learned, highly specialized lexical items. And it is very interesting to find an analogous case of stylistic motivation also in the Czech vocabulary — as a matter of fact, all the English lexical items just quoted can also be found in Czech in a very remarkably allied orthographical structuration (*krypta*, *hymna*, *mýtus*). A rather long series of Czech words might be added to the list — see, e. g., terminological items built up on Graeco-Latin models, such as *fyzika* 'physics', *psychologie* 'psychology', *typ* 'type', *analýza* 'analysis', *hypotéza* 'hypothesis', and many others.

It has thus been ascertained that the Czech redundant grapheme ⟨y⟩ has been utilized in its linguistic community in just as many functional roles and areas as its opposite number found in the English graphemic system, discussed here above.

* * *

One may pass over now to the grapheme ⟨h⟩, referring in a number of European languages to the phoneme /h/. Both the phoneme and the grapheme are found both in English and in Czech, so that it might seem to be out-of-place to discuss them in a paper dealing with redundant graphemes. Nevertheless, a thorough analysis of the systemic situation of the ModE phoneme /h/ reveals that its position in the ModE phonological system very markedly differs from the systemic situation of its counterpart /h/ in the phonological system of Present-Day Czech. As the present writer demonstrated in detail (see J. Vachek 1964, pp. 9–21), the English /h/, unlike its Czech counterpart, constitutes a very markedly peripheral item of its phonological system. It can now occur only in one single position of words, i. e. before a following vowel (or a semivowel) at the beginning of word-stems. In the substandard varieties of English the phoneme /h/ is apt to disappear even in such positions, with the result that the sound [h] tends to become in English rather a sort of 'prosody' (in J. R. Firth's sense of the term) than as a phoneme. As a result of this weakened systemic position of ModE /h/ also the grapheme ⟨h⟩ corresponding to it must be diagnosed as very strongly tending to the status of a redundant item of the system. In consequence of this, one finds that the ModE ⟨h⟩ is free to function as a component of digraphs, mostly of the type ⟨cons + h⟩. Such digraphs, of course, are not confined to the graphemic system of ModE, and some of them are of fairly respectable age. Especially one of them, ⟨ch⟩, has played a very important part both in English and in Czech in the courses of their developments. It is for this reason that we shall devote to it our particular attention in the following paragraphs.

³ On such instances see also B. Havránek 1929.

First of all we want to discuss the part played by ⟨ch⟩ in Czech in whose written norm it reflects the phoneme /χ/ of the corresponding spoken norm. It is interesting that the same relation of ⟨ch⟩ and /χ/ is also found in Modern German, Slovak and Polish – in all the three written norms ⟨ch⟩ clearly represents an item inherited from the Latin cultural tradition. In Latin, of course, the said digraph had constituted, in its turn, an item inherited from the Greek tradition, together with two other digraphs of the kind, ⟨ph⟩ and ⟨th⟩. (In Latin, of course, the shape of the old Greek digraph ⟨kh⟩ was written as ⟨ch⟩, in conformity with the fact that the sound [k] was regularly written there as ⟨c⟩). In post-classic Greek the second element of the three digraphs functioned as a sort of diacritical symbol signalling the non-explosive implementation of the phoneme regularly denoted by the first element of the digraph (thus, ⟨ph⟩ stood for /f/, ⟨th⟩ for /θ/, and ⟨kh⟩ for /χ/⁴. The first and the third of the digraphs were to be accepted in the cultural area of the former Roman empire (at least in its Central European section) with the implementations /f/ and /χ/, respectively, while ⟨th⟩ was to be treated in that area in the same way as ⟨t⟩. In the German speaking part of that area, as well as in the West Slavic linguistic communities, this implementation of ⟨ch⟩ was to be preserved until the present day, whereas in the western and southern sections of that area ⟨ch⟩ was to follow other, specific ways (some of which will be mentioned here later on).

In discussing the Czech digraph ⟨ch⟩ one should realize that its phonemic partner /χ/ is phonetically implemented by a voiceless sound, and that the systemic partner of /χ/ is, rather unexpectedly, the voiced phoneme /h/, implemented phonetically by a voiced glottal fricative [h]. On the other hand it should be realized that in the Western European languages (and particularly in English) the phoneme /h/ is as a rule not phonetically implemented as voiced: its articulation is, as a matter of fact, identical with that of the vowel immediately following it, except for the very fact that it is not voiced. (Thus, e. g., ModE /hot/ might be phonetically transcribed as [çot], similarly /hit/ as [jit] etc.) This phonetic essence of the English /h/ can account for the fact that at least since the EME period the English grapheme ⟨g⟩ had been employed as a diacritical symbol signalling the voiceless implementation of the phoneme denoted by the first element of the digraph. A particularly clear specimen of this case can be instanced by the ModE digraph ⟨wh⟩, the phonetic implementation of which in ME was – and in some geographical varieties of English still is even now – the voiceless [w] – sound. In opposition to these ME and ModE phonic realities one should again recall the above-noted fact of the consistently voiced phonetic implementation of the Czech /h/-phoneme: it was, in fact, this voiced quality which was to prevent the use of the Czech grapheme ⟨h⟩ for the purpose of signalling a voiceless implementation of the preceding voiced phoneme such as has been instanced here above in the ME digraph ⟨wh⟩.

⁴ It should also be noted that in Old Russian the /θ/-phoneme found in loans from Greek was to be substituted by /f/, se, e. g. *Fëdor* from *Theodor*.

Some more remarks are due here for the Czech digraph ⟨ch⟩. The interesting point is that this digraph is psychologically always perceived by members of the Czech community as one single graphic entity, not as a sum of two elements, ⟨c⟩ + ⟨h⟩. As evidence of this fact may be adduced some data of everyday practice: when it is necessary, for practical reasons, to reduce written words to their first elements (e. g. for the purpose of entering them into a vocabulary, or for arranging the alphabetical lists of personal names in a telephone directory, etc.), the Czech words or names beginning in ⟨ch-⟩ are never entered into the lists of words (or names) beginning in ⟨c⟩ (or, of course, ⟨c⟩) but claim their own lists of words or names under the heading ⟨ch⟩ (or, of course, ⟨ch⟩). As is well known, the English practice is exactly the opposite: words or names beginning in ⟨ch⟩ (or ⟨Ch⟩) are listed together with those beginning in ⟨c⟩ or ⟨C⟩ without claiming their own specific lists. A similar practice can be found in references used in Czech contexts to personal names beginning in ⟨Ch-⟩: such names are never abbreviated into ⟨C⟩ but always into ⟨Ch⟩ (as is well known, the practice of English – and similarly of German – is again exactly the opposite). Another interesting piece of evidence for the specific character of the Czech ⟨ch⟩ is supplied by the fact that this digraph has never been objected to by the partisans of a radical orthographic reform of Czech, despite the very obvious antiphonemic nature of the digraph. And finally it should be noted that in the very exceptional cases of the digraph ⟨ch⟩ covering a real sequence of the graphemes ⟨c⟩ + ⟨h⟩ (referring to a real phonemic sequence of /c/ + /h/) it appears advisable to pinpoint the very exceptional character of such a case by some graphemic means (thus, the standard dictionary of Contemporary Czech by B. Havránek and coll. 1960–1971, provides the item *mlachuba* ‘tattler’ by inserting a hyphen between ⟨c⟩ and ⟨h⟩: *mlac-huba*).

To pass now over to the graphemic problems of English, one can establish in it a relatively large number of digraphs of the type ⟨cons + h⟩. In all of them, again, the second element ⟨h⟩ signals some qualitative adaptation of the phoneme usually denoted by the first element of the digraph. As two most firmly rooted digraphs of the kind may probably be pinpointed those of ⟨sh⟩ for /š/ and ⟨ch⟩ for /č/: the latter of the two was taken over from the scribal practice of Norman French (in ModFrench the /č/ originally denoted by it was to be later changed into /š/, cf. the name *Charles* differently implemented in the two languages).⁵ Very well known is, of course, also the ModE digraph ⟨th⟩, representing mostly the phoneme /θ/, but also /ð/ in a relatively small but very frequent group of formal grammatical words. Other digraphs of the category ⟨cons + h⟩ are functionally less clear than the two just registered – see, e. g., ⟨gh⟩, reflecting either the phoneme /f/ or phonemic zero (cf. *laugh: though*): sometimes it is, however, combined with a preceding ⟨i⟩ into a very clear and frequently occurring trigraph ⟨igh⟩, reflecting the phoneme /ai/.

⁵ In the synchronically foreign items of the wordstock, of course, the digraph ⟨ch⟩ regularly has other phonemic correspondents (see, e. g. /š/ in *machine*, /k/ in *mechanical* etc.).

It should be noted, however, that some digraphs of the ⟨cons + h⟩ type played an important part in the history of English since the EModE period. One of them was mentioned here above, i. e. ⟨wh⟩, reflecting the voiceless /w̥/-phoneme, in some geographical varieties of English surviving until the present day but in the Received Standard replaced by the voiced phoneme /w/. As was pointed out in another of our papers (Vachek 1964, pp. 29–46), the EME graphemic practice also knew three more digraphs parallel to ⟨wh⟩, viz. ⟨lh⟩, ⟨nh⟩, and ⟨rh⟩, the phonemic counterparts of which were the EME phonemes /l̥/, /n̥/, and /r̥/, respectively. Since the functional load of these EME phonemes had been very low, they were to become merged with the corresponding voiced phonemes /l/, /n/, and /r/, respectively. This merger took place very early, but in the parallel case of ⟨wh⟩, reflecting the phoneme /w̥/, an analogous merger was to be effected much later (and in some geographical varieties of the language not at all). The reason why the ⟨wh⟩-digraph was to preserve so long its very conservative phonemic counterpart may be looked for in the circumstance that the digraph ⟨wh⟩ had been found to be characteristic of an important grammatical group of interrogative as well as relative pronouns and adverbs (*who*, *what*, *which*, *where*, etc.): for the exceptional grapho-phonemic correspondence of ⟨wh⟩: /h/ in the pronominal form *who* see also Vachek 1964, pp. 40f).

It should only be added that the firm footing of the digraph ⟨sh⟩ in the graphemic system of Present Day English is also evidenced by the rise of a new digraph which was obviously modelled on the traditional ⟨sh⟩. The new item is that of ⟨zh⟩, very often employed in English in transliterations of Russian names containing the phoneme /ž/ — see instances like *Zhirmunsky*, *Zhivago*, etc. It can thus be seen that it was the deep-rooted status of ⟨sh⟩ that was to prepare the ground for the frictionless acceptance of the new item of the digraphemic category.

To sum up, our examination of the graphemic situation obtaining both in Modern English and in Modern Czech has revealed that the two redundant graphemes ⟨y⟩ a ⟨h⟩ have been preserved in the two graphemic systems because they were found capable of being revaluated in those systems for other functional purposes. It has also been found that such revaluations are not effected with exclusive regard for the needs of the basic, phonological level of the corresponding spoken norm used by the given linguistic community, but also with a view to the exigencies of the inter-level relations obtaining in the system of the spoken language considered as an integral whole (i. e., with particular consideration of the needs ascertainable on its morphematic as well as lexical levels). It can hardly be doubted that further contrastive research of the mutual relations of graphemic and phonemic structures used in other linguistic communities may be expected to discover other interesting results.

REFERENCES

- Firth, J. R. (1948). Sounds and prosodies, *Transactions of the Philological Society* 1948. 127–52 (Oxford).

- Frinta, A. (1909). *Novočeská výslovnost* (Prague).
- Havránek, B. (1929). Influence de la fonction de la langue littéraire sur la structure phonologique et grammaticale du tchèque littéraire, *Travaux du Cercle Linguistique de Prague* 1.106-20 (Prague).
- Havránek, B., ed. (1960–71). *Slovník spisovného jazyka českého I–IV* (Prague).
- Luick, K. (1914–21). *Historische Grammatik der englischen Sprache* I (Stuttgart). Reprinted in Oxford 1964.
- Mathesius, V. (1935). Zur synchronischen Analyse des fremden Sprachguts, *Englische Studien* 70. 21–35.
- Vachek, J. (1964). On peripheral phonemes of modern English, *Brno studies in English* 4. 7–109 (Prague).
- Wrenn, C. L. (1943). The value of spelling as evidence, *Transactions of the Philological Society* 1943.14–39 (Oxford).
- Wright, J. and Wright, E. M. (1923). *An elementary middle English grammar* (Oxford).
- Wright, J. and Wright, E. M. (1924). *An elementary historical new English grammar* (Oxford).

GRAFÉMY ⟨Y⟩ A ⟨H⟩ V ANGLIČTINĚ A ČEŠTINĚ

Moderní angličtina i moderní čeština zachovávají v svých grafémových systémech redundantní grafémy ⟨y⟩ a ⟨h⟩, a to proto, že je byly schopny přehodnotit pro nová funkční upotřebení. K takovým přehodnocením nedochází jenom vzhledem k potřebám základní, fonologické roviny dané mluvené normy, ale také vzhledem k potřebám mezirovinových vztahů existujících v systému mluveného jazyka nazíraného jako integrální celek (tj. zvláště vzhledem k potřebám zjistitelným na rovině morfematické a lexikální).

