

FOREIGN GRAPHEMES AND GRAPHEME-PHONEME CORRESPONDENCES IN MODERN ENGLISH

KLAUS HANSEN

Humboldt University, Berlin

'Written language', writes Professor Vachek,¹ 'must be based on a system of graphic oppositions capable of differentiating meanings in the given community... The units of this system may be called graphemes.' There exists thus a system of graphemes which underlies the written norm, similar to the system of phonemes which constitutes the lowest level of the spoken norm, the simplest—and, at first sight, ideal—correlation between them being a one-to-one correspondence of graphemes and phonemes. No language with some tradition of writing, however, seems to fully attain this ideal, since the development of the graphemic system does not directly reflect the development of the phonemic system, but almost necessarily lags behind, so that a more or less wide discrepancy between the two is quite typical. Besides, one may observe in the written norm of some languages—and of English in particular—'an increasing tendency to loosen the very close ties that were originally linking... phonemes and graphemes, and to supplement the correspondence on the lowest level by a relatively high percentage of instances which reveal correspondences based on higher levels of language,'² viz. at the level of morphemes and, especially, of words. Last but not least, the one-grapheme-to-one-phoneme correspondence has been severely disturbed in English under the influence of languages 'whose written norms had been built up on correspondences often differing from those found in English'³, particularly under the influence of French, which was most powerful in the Middle English period. As a result, some foreign graphemes have become 'naturalized' and are now used as integrated elements of the domestic system, both in loan-words and in native words. Thus the French (or Anglo-Norman) graphemes *o* (*money*), *ou* (*stout*), *v* (*move*), *ch* (*chief*), *gu* (preceding *e*, *i*, *y*: *guide*), and *qu* (*question*) have, partially or completely, superseded the English graphemes *u*, *f*, *c*, *g*, and the cluster *cw*, as regular graphic correspondences to the phonemes (or phoneme combinations) /u/ (ModE /ʌ,(u)/), /u:/ (ModE /au/), /v/, /tʃ/, /g/, and /kw/: *wonder*, *loud*, *love*, *child*, *guest*, *queen*, and it would be misleading to include them in a discussion of (synchronically) foreign graphemes in Modern English.

In addition to the above cases, however, there are foreign elements in present-day written English that are still, from a synchronic point of view, non-integrated: graphemes that have not yet been naturalized as such or occur in positions where they would not normally be used in the domestic vocabulary.

The former have been referred to by J. Nosek in his 'Systematic Analysis of Modern English Graphics'⁴. The examples he quotes (p. 64) are *cz* (*czar*, *Czech*; cf. also *czardas*), *kh* (*sovkhoz*; *khan*, *khaki*, *Sikh*, *Khedive*), *ts* (*tsar*, *intelligentsia*), *tz* (*waltz*, also *chintz*, *quartz*; *Fitzjames*), and the cluster *æ + u* in the word *manœuvre*, to which one might add *aa* (*Afrikaans*, *bazaar*, *kraal*, *laager*, *salaam*), *ae* (*aesthetic*, *anaemia*, *Caesar*,

*medi(a)eval*⁵, also used as a Latin plural ending: *formulae*; Gaelic, *maelstrom*), *eau* (bureau, plateau), and all vowel or consonant graphs with diacritics:⁶ *é* (*café, détour, fiancé(e), protégé(e), résumé*), *è* (*crèche, manège, suède*), *ê, â, ô, û* (*fête, débâcle; rôle; goût*), *ç* (*façade*), *ñ* (*cañon* [also *canyon*]); besides, *cch* (*Bacchanal, saccharine*), *cqu* (*laquey, racquet* [also *lackey, racket*]), *kk* (*pukka* [also *pucka*]), *trekker*, *ph* (*orphan, philosophy, phrase, triumph*),⁷ *rh* (*rhetoric, rhomb, rhythm*⁸); cf. also *catarrh, diarrhoea*), and *zh* (*muzhik* [also *moujik*]); mainly used to transcribe the sound /ʒ/ in Russian or Ukrainian names: *Voronezh, Zhitomir*). These graphemes do not belong to the English system, nor do they constitute a sub-system.

But J. Nosek also mentions another kind of foreignism. 'Graphical foreignisms', he writes (p. 64), 'may also be observed in the unusual distribution of a single letter I. This is an organic part of the native grapheme stock where it does not occur at word terminals (after consonants). If it does occur finally (after consonantals) it is indicative of foreign origin: *ski, khaki*.' Other examples of familiar graphemes or grapheme clusters in unfamiliar positions are interconsonantal *y* (*hymn*; cf. *my, kindly*, but *him*), initial or medial *oe* (*oecumenical, oedema; amoeba*; cf. *toe*); final *c* (*sac, zinc*; cf. *cell, clear, place*, but *sack, sink*), initial *k* (preceding *a, o, u, l, r*) and final *k* (following a vowel) (*kaleidoscope, kotow, kulak, kleptomania, kremlin; trek*; cf. *keep, king, knock, token*, but *call, cool, cut, clear, creep, stock*), initial *x* (*xylophone*; cf. *axe, six*), initial *ll* (*llama*; cf. *silly, tell*, but *lamb*), final *gn* (*sign*; cf. *gnaw*), initial *ps, pt* (*pseudonym; ptomaine*⁹), etc.

From these examples we may infer a certain parallelism (though not necessarily direct correspondence) between foreignisms at the graphemic level (of the written norm) and foreignisms at the phonemic level (of the spoken norm).¹⁰ There are unfamiliar, synchronically foreign graphemes (*aa, é, cz*, etc.) as there are unfamiliar, synchronically foreign phonemes (e.g. /ã/ in *fiancé(e)* or /x/ in *loch*); and there are familiar graphemes in unfamiliar, synchronically foreign positions (*-i, x-*, etc.) as well as familiar phonemes in unfamiliar, synchronically foreign positions (e.g. intervocalic /θ/ in monemes like *ether*; cf. /ð/ in *other*).

Another group of foreignisms, and they are the main group we are concerned with here, is based on foreign 'pronunciation' (and 'orthography' respectively).¹¹ The words in this group contain familiar graphemes in familiar positions, but there is an unfamiliar, 'irregular' correspondence between graphemes and phonemes, even though they are often familiar phonemes in familiar positions. That is to say, their foreign status sometimes shows only if 'the correlative relations undoubtedly existing between the written utterances and their spoken counterparts'¹² are taken into consideration. Again such an irregular correspondence may occur in many or all positions of the grapheme (cf. *ch* — /f/) or in a particular position only (cf. *e* — /i/; intervocalic *th* — /θ/). Both cases are illustrated by the following examples:¹³

a — /a:/ in stressed 'open'¹⁴ syllables instead of /ei/ (e.g. *navy, late*) or (preceding *r*) /eə/ (e.g. *vary*): *banana, drama* (AĒ rarely [-e:-]), *saga, soprano, tomato* (AĒ also [-e:-]); *promenade* (also [-ei-]), *vase* (also, esp. AĒ, [-ei-/e:-]); *aria* (AĒ also [¹erɪə, ¹æriə]); rather seldom also /a:/ in a stressed 'closed' syllable instead of /æ/ (e.g. *pan*); *Koran, Slav* (also [slæv]); cf. also /a:(ʒ)/ in usually unstressed *-age* instead of /i(dʒ)/ (e.g. *breakage*): *barrage, espionage* (also [-idʒ]), *fuselage* (also [-idʒ]), *garage* (also [-idʒ, -a:dʒ]), *sabotage* (also [-idʒ]). — Unstressed final *e* — /i/ instead of '∅', indicating the value of the preceding vowel graph (e.g. *smile*): *catastrophe, recipe, simile*; cf. the combination *-es* — /i:z/ instead of /ɪz/ (e.g. *buses*): *bases* (plural of *basis*), *crises*. — *i* — /i:/ in stressed 'open' syllables instead of /ai/ (e.g. *tiger, nine*): *kilo, mosquito, sonatina*;

critique, fatigue, machine, oblique (rarely [-ai-]), *police*; occasionally also /i:/ in a 'closed' syllable instead of /i:/ to *chagrin* (but more often [ʰægrɪn]), *chic* (also [ʰik]). — *au* — /ou/ instead of /ɔ:/ (e.g. *cause*): *chauffeur, mauve, vaudeville; noyau* (cf. usual -aw). — *ou* — /u:/ instead of /au/ (e.g. *loud*) or /ou/ (e.g. *soul*): *outré; acoustic* (old-fashioned [-au-]), *boudoir, debouch* ([di'baʊtʃ, di'bu:f]), *joust* ([dʒaʊst, dʒu:st]), *route* (army also [raʊt]); *bijou*; but sometimes also in native words: *wound* 'injury', *youth, you*, etc. — *ou(r)* — /uə(r)/ instead of /ɔ:(r)/ (e.g. *mourn, four*), /ə:/ (usually preceding a consonant) (e.g. *journey*) or, occasionally, /auə/ (e.g. *our, flour*): *bourgeois* (but [bɔ:'dʒɔɪs] for the printing type), *bourn(e)* 'limit', *tour(ist), tourbillon* (also [-ə:-]), *tourmalin* (also [-ə:-]), *tournament* (also [-ə:-]), *tourney* (also [-ə:-]); sometimes also in native words like *bourn* (but normally *burn*) 'small stream' or *dour* (Sc., poss. < Lat. *durus*). — *g* (preceding *e, i*) — /ʒ/ instead of /dʒ/ (e.g. *gin, age*) (see also the words in -age): *gendarme; bourgeois* (but [bɔ:'dʒɔɪs] for the printing type), *régime; prestige, rouge*. — *j* — /ʒ/ or rarely /j/ instead of /dʒ/ (e.g. *jump*): /ʒ/: *jabot, jupeon* (also [dʒ-]), *bijou*; /j/: *jodel, junker, majolica (maiolica)* (also [-dʒ-]). — *ch* — /k/ or /ʃ/ instead of /tʃ/ (e.g. *child, much*): /k/ (always preceding *l* and *r*: *chlorine, chrome*): *character, chasm* (AE sporadically [tʃ-]); *lichen* (also [ʰlitʃɪn]), *mechanical; conch* (also [-tʃ]), *stomach, technical*; cf. *sch* — /sk/ in *scheme, school*, etc. — /ʃ/ (especially preceding final silent -e): *chaise, chassis, chauffeur, chemise; machine, parachute; douche, moustache, niche* (more often [nitʃ]), *debouch* (more often [-tʃ]). Very seldom *ch* stands for /x/: *loch* (also [-k]). — *qu* — /k/ instead of the cluster /kw/ (e.g. *queen, equal*): *quadrille* (but usually [kw-]), *quay, questionnaire* (but more often [kw-]), *queue; bouquet, coquette, liquor, mannequin*. — Intervocalic *th* (in monemes) — /θ/ instead of /ð/ (e.g. *other*): *ether, method*; cf. intervocalic *s* — /s/ instead of /z/: *basin, crisis*.

Some typical examples of 'mute graphemes' in words of foreign origin ought to be mentioned here too, e.g.: Initial *h* in a stressed syllable (cf. *house*): *heir, honest, honour(able), hour*; AE also *herb, humble* (especially in the South), *homage* (beside [h-]). — *p* in *coup, corps, receipt*¹⁵. — Final or, sometimes, preconsonantal *s*: *bas-relief* (but more often [ʰbæs-]), *chassis, précis, demesne, puisne*. — Final *t* in unstressed syllables (cf. *thicket*): *ballet, beret* (also [ʰberɪt]), *buffet* (refreshment bar; sideboard: [ʰbʌfɪt]), *cabaret* (also [ʰkæbæret]), *depot, format* (more often [ʰfo:mæt]), *parquet* (also [ʰpa:kɪt]), *ricochet* (more often [ʰrɪkɔʃet]), *touriquet* (rarely [-et]), *valet* (more often [ʰvælit]), also *trait* (besides, esp. AE, [treɪt]).

Of course, the foreign character of many of the words quoted above does not only show in the particular irregular grapheme-phoneme correspondence mentioned. Quite often it is their graphic appearance as a whole that is the indication of their foreign status, all the more so as it is sometimes next to impossible to split them up into graphemic segments.

From a historical point of view we may claim for the examples listed above that there is a coincidence of native (or naturalized foreign) graphemes and foreign graphemes resulting in a kind of 'homography'. Synchronically, however, it would be as wrong to distinguish several graphemes (e.g. *au*¹ /ɔ:/ and *au*² /ou/ or *ch*¹ /tʃ/, *ch*² /k/ and *ch*³ /ʃ/) in such a case only because they correspond to different phonemes, as it would be to postulate separate phonemes (e.g. /ʌ¹/ in *sun* and /ʌ²/ in *son* or /s¹/ in *sell* and /s²/ in *cell*) only on account of the fact that they are represented in conventional orthography by different graphemes. This, as well as the basic irregularity in the grapheme-phoneme correlation of the words mentioned, is strikingly illustrated by the strong tendency towards spelling pronunciation to be observed in these words, i.e. by a tendency towards the predominant, 'regular' English grapheme-phoneme

correspondences based on orthography.¹⁶ Cf. above *promenade, vase, Slav, espionage, fuselage, garage, sabotage, oblique, chagrin, chic, acoustic, debouch, joust, route (army), bourgeois (printing type), tourbillon, tourmalin, tournament, tourney, jupon, majolica, lichen, conch, niche, quadrille, questionnaire, bas-relief, beret, buffet (sideboard), cabaret, format, parquet, ricochet, tourniquet, valet, trait*. To these examples, whose spelling pronunciation is still rivalled by their 'foreign' pronunciation in Modern English, one might add the great number of words for which spelling pronunciations have become the usual form (see, e.g., W. Horn—M. Lehnert, *Laut und Leben* [Berlin, 1954], index sub *Schriftbildausssprache*).

Finally, the phenomenon dealt with here has also some bearing on the problem of foreign words v. domestic words, and of non-naturalized, synchronically foreign words v. naturalized foreign words in particular. It may perhaps even lead to a re-evaluation of the status of some of the words quoted (see especially group 4 below), which has already been suggested by V. Mathesius, who wrote: 'Es scheint, daß das Verhältnis der Orthographie zu der Aussprache im Englischen besonders zur Unterscheidung der eingebürgerten und der noch als fremd gefühlten Entlehnungen dienen kann. Für eine lesende und schreibende Welt sind die orthographischen Abweichungen ein sehr wichtiges Unterscheidungsmerkmal, und meine englischen Freunde haben mir bestätigt, daß sie manchmal erst durch die ungewöhnliche Orthographie auf den fremden Charakter eines Wortes aufmerksam gemacht werden.'¹⁷

Regarding its graphemic and phonemic structures a word may thus, synchronically, appear foreign in four ways:

- (1) Both the phonemic and graphemic structures of the word are foreign, i.e., it contains foreign phonemes and graphemes and/or familiar phonemes and graphemes in unfamiliar, foreign positions or combinations, e.g. /f'i:ã:(n)sei/ *fiancé(e)*.
- (2) The phonemic structure of the word is foreign, but its graphemic structure corresponds to the rules of the domestic vocabulary, e.g. /i:θə/ *ether*.
- (3) The phonemic structure of the word corresponds to the rules of the domestic vocabulary, but its graphemic structure is foreign, e.g. /flem/ *phlegm*.
- (4) Both the phonemic and graphemic structures of the word — taken as such — correspond to the rules of the domestic vocabulary, but the correlation between them is foreign, e.g. /'aʊə/ *hour*.

The last case particularly implies that a linguistic analysis which considers only the phonemic system or the graphemic system as such, while neglecting the correlations existing between them, would be incomplete and would overlook a problem that the language user has permanently to face.

NOTES

¹ 'Some Remarks on Writing and Phonetic Transcription', *Acta Linguistica* 5.88 (Copenhagen, 1945-9).

² J. Vachek, 'Two Chapters on Written English', *Brno Studies in English* 1.18 (Prague, 1959).

³ *Ibid.* 17.

⁴ *Prague Studies in English* 9.53-67 (Prague, 1961).

⁵ There is a tendency to substitute *e* for *ae* and *oe* (see below): *medieval, ecumenical*.

⁶ As the words are integrated they tend to drop the diacritics: *detour, manege, role*, etc. This is quite typical of American English.

⁷ On the analogy of such words *ph* was also written in *nephew* (OF *neveu*), where the corresponding phoneme is, however, /v/.

- ⁸ On the analogy of *rhythm* also in *rhyme* (OF *rime*).
- ⁹ The spelling is pseudo-etymological in *piarmigan* (< Gaelic *tarmachan*).
- ¹⁰ Cf. e.g. B. Trnka, 'On Foreign Phonological Features in Present-day English', *In Honour of Daniel Jones 185-90* (London, 1964).
- ¹¹ 'Orthography is a kind of bridge leading from spoken to written utterances. More exactly, it is a set of precepts enabling the language user to transpose spoken utterances into written ones. (Conversely, what is popularly called "pronunciation", that means actual reading of printed texts, can be defined as a set of precepts enabling the language user to transpose written utterances into spoken ones.)' (J. Vachek, footnote 2, 16f. See also *A Prague School Reader in Linguistics* 442, 446 [Bloomington, 1964].)
- ¹² J. Vachek, footnote 2, 8.
- ¹³ For want of space we can give here (and elsewhere) only a small selection of examples.
- ¹⁴ 'Open' according to the spelling.
- ¹⁵ An 'etymological' respelling (*receipt* < ONF *receite*; cf. Lat. *recepta*). This is also true of *b* in *debt*, *doubt* and *subtle* (OF *dette*, *doute*, *sotil*; Lat. *debitum*, *dubitus*, *subtilis*), *c* in *indict*, *muscle* and *victuals* (OF *enditer*, *musle*, *vitailles*; Lat. **indictare*, *musculus*, *victualia*) or *s* in *isle* (OF *île*; Lat. *insula*). See also J. Vachek, footnote 2; 25f.
- ¹⁶ In German the reverse process prevails. Foreign words are assimilated orthographically: *Büro*, *Fabrik*, *Foto* (Photo), *Maschine*, *Schofför* (Chauffeur), *Telefon* (Telephon), etc.
- ¹⁷ 'Zur synchronischen Analyse fremden Sprachguts', *A Prague School Reader in Linguistics* 409 (Bloomington, 1964). See also J. Vachek, *The Linguistic School of Prague* 72 (Bloomington — London, 1966).

RESUMÉ

Cizí grafémy a cizí korespondence mezi grafémy a fonémy v moderní angličtině

Článek se zabývá synchronicky cizími grafémy (např. *aa* v *kraal*), běžnými grafémy v pozicích jevičích cizost (např. koncové *-c* v *zinc*) a zvláště pak cizími korespondencemi mezi grafémy a fonémy (např. *a* — /a:/ místo /ei/ v *vase*). Slovo tak může být synchronicky cizí, jestliže je cizí jeho fonémová i grafémová struktura (např. /fɪ'ā:(n)sei/ *fancé(e)*), jestliže je cizí pouze jedna z těchto struktur (např. /'i:θə/ *ether*; *phlegm* /flem/) nebo jestliže korespondence mezi nimi je cizí (např. /'aʊə/ — *hour*).

