

The origin of the peculiarities of the Vietnamese alphabet^{*}

Translated by Alexis MICHAUD,
LACITO-CNRS, France

André-Georges HAUDRICOURT
LACITO-CNRS, France

Abstract

The missionaries who adapted the Latin alphabet to Vietnamese were Portuguese, Italian and French. The resulting spelling inherited some peculiarities from the spelling systems of Romance languages.

Aspirated consonants H, PH, TH, KH [IPA: /h/, /p^h/, /t^h/, /k^h/] are not found in Romance languages; however, in these languages, the combinations of letters PH, TH and KH are present in words of Greek origin, as equivalents for the Greek initials *phi*, *theta*, *khi* (φ, θ, χ), which were aspirated consonants in Greek; and thus these combinations were used to transcribe Vietnamese aspirated stop consonants.

Dorsal stops C, G are only used before the vowels /a/, /o/ and /u/. This is because, in Romance languages, it is the only position where these consonants preserved the obstruent pronunciation they had in Latin; GHE, GHI are used with the phonetic value they have in Italian; the combinations KE and KI resort to the letter K, used in Greek (*kappa*, κ) and in Germanic languages.

Labiovelar stops QU and GU are taken from Italian and Latin spellings.

Among prepalatals, the unvoiced stop CH is taken from the Portuguese and the Spanish, which themselves borrowed this notation from Old French, where it had been created to transcribe a new sound, not found in Latin.

The voiced stop D is used as an approximative notation for a sound not found in Europe, where D is the voiced counterpart of T. In Vietnam, a new letter, Đ, was coined [for a preglottalised alveolar stop: /d̪/], its horizontal bar pointing to a similarity with the letter T.

The voiced spirant¹ was written as GI, as in Portuguese and French (at the time, J was not yet in use in Europe).

The unvoiced spirant X is borrowed from Portuguese and the Northern dialects of Spanish: in these languages, the S is pronounced at the back of the mouth [IPA: retroflex /s̺/], as in Vietnamese, whereas elsewhere in Europe S is an anterior sibilant as in French [IPA: alveolar /s/].

NH [for the palatal nasal /ɲ/] is borrowed from the Portuguese; TR is an approximative notation for a sound which is not found in European languages [IPA: /t̪/].

^{*}Originally published as: L'origine des particularités de l'alphabet vietnamien, *Dân Việt Nam* 3:61-68, 1949.

Ô, Ê are taken from the Portuguese, which is why the pair Ê, E does not have the same phonetic value as in French. [In French, Ê stands for /ɛ/, E stands for /e/, whereas the opposite is found in Vietnamese *quốc ngữ*, namely Ê for /e/, E for /ɛ/.]

Y is used in Vietnamese *quốc ngữ* in a way similar to what is found in Spanish, where it replaces I in-between vowels or at the end of words. The letter Y comes from the Greek alphabet (*upsilon*, υ).

Ô and U were created to represent Vietnamese vowels that do not have equivalents in Western Europe [IPA /ɤ/, /ʉ/].

Translator's foreword

André-Georges Haudricourt's contribution to Southeast Asian studies is internationally acknowledged, witness the Haudricourt *Festschrift* (Suriya, Thomas and Suwilai 1985). However, many of Haudricourt's works are not yet available to the English-reading public. A volume of the most important papers by André-Georges Haudricourt, translated by an international team of specialists, is currently in preparation. Its aim is to share with the English-speaking academic community Haudricourt's seminal publications, many of which address issues in Southeast Asian languages, linguistics and social anthropology.

The article "The origin of the peculiarities of the Vietnamese alphabet" is not one of Haudricourt's most famous articles, and therefore it will not be included in the projected volume of collected papers. However, to this day, it remains an insightful and vivid account of the origin of the modern Vietnamese script. It traces the peculiarities of this spelling system back to the spelling habits of the Romance languages that were familiar to the authors of this system. The article illustrates Haudricourt's passion for reconstructing the historical origin of complex phenomena, and his skill for adducing evidence from an impressive range of sources.

The article was clearly intended for a broader audience than most of Haudricourt's other publications. Its style is colloquial; technical terms are avoided. This work appeared in the third and last issue of the journal *Dân Việt Nam* (The Vietnamese People) published by the *Ecole Française d'Extrême-Orient* in Vietnam in 1948 and 1949. The original is now difficult to find; moreover, many of its potential present-day readers may not have a command of French, whereas back in 1949, Haudricourt could confidently expect that the general public could read French. The present translation aims to make this document available to anyone interested in Vietnamese, and in writing systems in general.

The Vietnamese publisher obviously had a difficult time composing this article, which uses a wealth of symbols. The present translation corrects some typos: in the original edition, C had been substituted for G on page 64; the diacritic in p', t', k' had been rendered approximately, as pc, tc, kc; etc. Transcriptions in the present-day version of the International Phonetic Alphabet were systematically added at translation. Translator's comments are placed in square brackets or in footnotes.

Many thanks to Michel Ferlus for pointing out this article, to Boyd Michailovsky and Martine Mazaudon for corrections to the translation, and to Jean-Michel Roynard for his help with the figures.

Introduction

Alphabetic writing aims to represent the pronunciation of words. In theory, each letter stands for one sound [one phoneme], a given letter always stands for one and the same sound, and a given sound is always represented by one and the same letter. This ideal situation is not found in the case of the alphabetic writing of languages that have a long history and a literary tradition, because as time goes by, their writing systems remain the same whilst their pronunciation changes. Written texts are passed on from one generation of speakers to the next without changes (or only with minor changes, which lag behind the evolution of pronunciation). Thus the correspondence between written signs and spoken sounds changes as the spoken language evolves. What I refer to as the peculiarities of a given alphabet consists of the lack of one-to-one correspondence between letters and sounds: one and the same letter standing for different sounds, or the same sound being represented by different letters. These peculiarities originate in the historical evolution of pronunciation in the language at issue.

When a language is put into writing for the first time, the writing system devised for this language is generally based on the writing system of another language, and the new writing system inherits peculiarities from the writing system which served as a model. Thus the Vietnamese *quốc ngữ* alphabet was based on the orthography of several Romance languages, and it inherited peculiarities which can be explained by the way the pronunciation of Latin evolved in Europe. Recall that Latin, which in the 3rd century BCE was only spoken in the town of Rome, spread all over Europe at the time of the Roman empire (up until the 5th century CE), only leaving a few non-Latin-speaking areas, e.g. a small region of the Western Pyrénées where the Basque language was preserved. Later, the varieties of Latin spoken in different areas evolved in different ways: the Latin of the region of Florence became the Italian language, the Latin of Paris became the French language, that of the region of Burgos became Spanish. However, throughout the Romance area, Classical Latin continued to be used as the language of religion and science, although its accurate original pronunciation was not known anymore. (To this day, Classical Latin remains the liturgical language of the Roman Catholic church, and scholars doing research in natural history refer to plants and animals by Latin names.) The present-day French, Italian, Spanish and Portuguese languages are called “Latin languages”, “Neo-Latin languages”, or “Romance languages”. The Latin alphabet was adapted to other languages of Europe (Celtic, Germanic and Slavic) by scholars who knew the Latin transcriptions of Greek words, Greek being the language of culture of Antiquity.

In this article, I set out the peculiarities that the Vietnamese alphabet [today known as *quốc ngữ*] inherited from the Romance languages; why it differs from the French alphabet despite the fact that French is a Romance language; and how the Vietnamese alphabet and the French alphabet both differ from the scientific alphabets used by the scientists who study languages.

The order in which the letters of the alphabet are arranged goes back a very long way. It is known that the alphabet was invented in Phoenicia (North-West of Palestine) and spread in Greek-speaking countries; in Eastern Greece it gave birth to the Greek alphabet as we know it; in Western Greece it took on a somewhat different shape and spread into Italy, where it gave birth to the Latin (Roman) alphabet. The present-day order of letters in the Roman alphabet—which we use in dictionaries, for instance—is the same as that of the Greek alphabet and of the Hebrew alphabet (that of the inhabitants of ancient Palestine). It is a purely traditional order, which we are not able to explain or justify. This is quite different from the alphabets of Hindu origin, which are ordered in a scientific way, separating consonants and vowels, and classifying consonants according to their place of articulation, starting out at the back of the mouth and progressing forward up to the lips.

[Consonants are presented first (section 1 of Part I), then vowels (section 2); Part II consists of a short note on tones.]

Part I: Consonants and vowels

The Latin and Greek alphabets were very well adapted to the pronunciation of the Greek and Latin languages. Each letter stood for one given sound.

1. Consonants

Latin had two series of stop consonants (i.e. consonants produced with a full closure at some point in the mouth): a voiced series (i.e. with vibration at the larynx), /b/, /d/, /g/, written B, D, G, and an unvoiced series (without vibration at the larynx): /p/, /t/, /k/, written P, T, C, Q. This system of two series is still found in French. Ancient Greek, on the other hand, had three series of stop consonants: a voiced series (/b/, /d/, /g/, written Β, Δ, Γ), an unvoiced unaspirated series (/p/, /t/, /k/, written Π, Τ, Κ), and an unvoiced aspirated series (/p^h/, /t^h/, /k^h/, written Φ, Θ, Χ). Latin literati reproduced this last series by adding an H after an unvoiced consonant, yielding the combinations PH, TH, KH. But as early as the end of Antiquity, the pronunciation of Greek evolved: all the stops became spirants [fricatives], i.e. instead of a full obstruction of the vocal tract during the consonant, there was only a narrowing at one point in the tract. Thus, when the Roman alphabet was employed to transcribe Germanic languages at the beginning of the Middle Ages, the combinations PH, TH, KH were still in use but the words in which they appeared now had spirant consonants and not obstruents, and as a consequence PH, TH, KH were used by scholars with their new phonetic value, which was different from the one they had in Classical Latin: they were

used to transcribe spirants. This is why, to this day, TH is employed for spirants in English, and CH for spirants in German. In Latin, PH (/p^h/) got to be pronounced like F (as /f/), and TH got to be pronounced like T (/t/), hence the French pronunciation: /t/ for TH, /f/ for PH. In Vietnam, the situation is analogous to that found in Greek: TH remains an aspirated stop like that of Ancient Greek; PH is a spirant, as in Modern Greek. In scientific transcription, spirants are to be carefully distinguished from aspirated stops. Spirants are usually represented by means of the Greek letters ϕ, θ, χ for the unvoiced set (bilabial: /ϕ/, dental: /θ/, velar: /χ/) and β, δ, γ for the voiced set (bilabial: /β/, dental: /δ/, velar: /γ/). The aspirated stops are ph, th, kh [IPA: /p^h/, /t^h/, /k^h/]; some authors transcribe them as p', t', k', using a small apostrophe with its concave part turned to the right: the *spiritus asper* of Greek writing, created by the Greeks of Alexandria to note aspiration, which existed in their variety of Greek but not in Athens Greek, where the letter H was used for a vowel [written H as a capital letter, η as a small letter]. Indianists use the notations *ph*, *th*, *bh*, *dh*, whereas Sinologists use p', t', b', d' instead, to note the same sounds.

Latin had four types of stops: labials B, P (/b/, /p/), formed by joining the lips; apicals (dentals) T, D (/t/, /d/), formed by joining the tongue tip (by its Latin name: the *apex*) with the base of the teeth; dorsals (also called palatals) C, G (/c/, /ɟ/), formed by joining the back of the tongue with the palate. The letter Q represented a velar stop, formed by joining the back of the tongue with the velum; in Latin, this sound was always labialised [IPA /k^w/], i.e. accompanied by lip rounding (as found in the vowel U [IPA: /u/] of Vietnamese spelling). This is why, in the Latin writing system, Q is followed by U. QU [IPA /k^w/] is the unvoiced stop corresponding to the voiced stop GU [IPA: /g^w/]. This opposition is preserved in Rumanian, where Latin AQUA “water” became *apa* whereas Latin LINGUA “tongue” became *limba* [i.e. the earlier opposition of /k^w/ and /g^w/ in Latin remains a voicing opposition in Rumanian, between /p/ and /b/]. In Latin, QUI is a single syllable [IPA: /k^wi/], whereas CUI represents two syllables [IPA: /ku.i/].

The old (Latin) pronunciation of the groups of letters QU and GU is preserved only in Italian, and it is from Italian that this reading was borrowed for use in the Vietnamese script [where QU stands for /k^w/]. In French and in Spanish, labiovelars became simple dorsals (except before A). Thus, the Latin word QUAMQUAM was pronounced CANCAN [IPA: /kãkã/, with initial /k/, and not /k^w/] in the schools of Paris in the 16th century.

C, G

The history of C is very complex. To begin with, by comparing the Latin alphabet with the Greek alphabet one observes that the third letter of the Greek alphabet is a voiced dorsal, Γ (small letter: γ), not an unvoiced dorsal as in Latin (C). This is because the Latin alphabet was influenced by a neighbouring people, the Etruscans, who distinguished between aspirated and unaspirated stops [e.g., /t^h/ and /t/] but not between voiced and unvoiced stops [e.g., /d/ and /t/]. (The same holds true of the Chinese of Beijing and Guangdong.) The Etruscans thus used the third letter of the alphabet to note an

unvoiced dorsal. The Latins, on the other hand, did have a distinction between voiced and unvoiced stops; using the same letter for both was not satisfactory. They therefore added a horizontal bar to distinguish a G from a C. C remained in third position in the alphabet, whereas G went into the slot formerly occupied by the Z [small letter: ξ] of the Greek alphabet, which at that time was not employed in Latin. (The two Greek letters Y and Z are found right at the end of the Latin alphabet, because, as we shall see below, they were not borrowed until much later.)

In Latin, C had the same pronunciation in the syllable CE as in the syllable CO. This is still how things stand in the Celtic languages of the British isles (Irish and Welsh), which were the first non-Romance languages to adopt the Latin alphabet. But in Latin, from the 4th century onwards, when the dorsal consonant was followed by a vowel E, I (/e/, /i/), which requires a fronted position of the tongue, it became a prepalatal consonant, during whose articulation the tip and the back of the tongue make simultaneous contact with the front part of the palate. Such consonants have been part of the phonemic inventory of languages of Indochina for a very long time, and they remain stable in these languages; in Europe, on the contrary, these new sounds were difficult to perceive; their perceptual salience was enhanced by shaping the tongue in a channel-like form: if the channel is on the back of the tongue, this yields a non-anterior sibilant stop [in present-day IPA transcription, /tʃ/, a postalveolar affricate], like Italian CI, CE; if the channel is towards the tip of the tongue, this yields an anterior sibilant stop [IPA: /ts/, alveolar affricate]: this is how CI, CE were pronounced in Old French and Old Spanish. This last sound is currently found in Mandarin Chinese and Cantonese Chinese, but not in Indochina.

Meanwhile, the consonants of the Latin syllables TIA, TIO had become anterior sibilant stops [i.e. the affricate /ts/] in Italian as well as in the other Romance languages. The Greek letter Z was borrowed to represent the voiced anterior sibilant stop [i.e. /dʒ/]. In Italian, Z stands for both the voiced and unvoiced sounds [i.e. /dʒ/ and /ts/].

K, KH, GH

When the Latin alphabet was adapted to the Germanic languages (in the 7th century), then to the Slavic languages that had these stops, Z and C were employed. One had to borrow the Greek letter K (which in the Latin alphabet had remained at its original place, the one it had in the Greek alphabet) to note the dorsal stop.

This accounts for the Vietnamese spelling, in which one finds CA, CÔ, CU [for /ka/, /ko/, /ku/] but KÊ, KI [for /ke/, /ki/]: in Vietnamese, QU could not be used in the way it is used in French and Spanish [i.e. for /k/] because QU was used to transcribe another Vietnamese sound [IPA /kʷ/]; Vietnamese /k/ could not be transcribed as CH either (the way it is used in Italian), because CH was used for a Vietnamese prepalatal unvoiced stop, as we shall see below. This also precluded the use of CH as a notation for the aspirated stop [IPA: /kʰ/]. This led to the use of KH for the aspirated stop [IPA:

/k^h/, before all vowels. The voiced dorsal raised a similar issue: GA, GÔ, GU [for Vietnamese /ɣa/, /ɣo/, /ɣu/], but GHÊ and GHI [for /ge/ and /gi/, present-day /ye/ and /yi/] and not GÊ and GI; one could not use the combinations GUÊ and GUI, which were used in the case of a labiovelar initial, whereas the Italian notation could be used, hence GHÊ, GHI.

The scientific notation consists in using k for the unvoiced dorsal stop, and g for the sonorant dorsal stop.

CH

In Western Romance languages, a new prepalatal stop emerged and later turned into a non-anterior sibilant stop [IPA: postalveolar affricate, /tʃ/]. In French, it originates in the Latin syllable CA: for instance, *calidum* “hot” became French *chaud*, *caballum* “horse” became French *cheval*. The non-anterior sibilant stop [IPA: postalveolar affricate, /tʃ/] was written as CH, where the function of the H is only to indicate that the C is pronounced neither as in CO, nor as in CE. At the time of Charlemagne (8th century), it was still a prepalatal stop; in Old French (9th century), it had become a non-anterior sibilant stop [IPA: /tʃ/] — a value which is preserved in English. In Portuguese, it originates in initial clusters PL, CL; in Spanish, in the word-medial cluster CT, e.g. Latin *noctem* “night” first changes to *noite*, as preserved in Portuguese, then to *noche*. In Basque the prepalatal stop still exists, e.g. in *éche* “house”.

Given this situation in the languages they were familiar with, it was quite natural for the creators of the Vietnamese Roman alphabet to note the Vietnamese prepalatal stop as CH.

The scientific notation of the palatal stop, a rare sound in Europe, is not well established yet [present-day IPA standard: /c/: it is sometimes noted as k' or t', where the apostrophe indicates palatalisation, a phenomenon referred to in French as *mouillure*, literally ‘wetting’, because the horizontal position of the tongue dorsum *wets* the palate, as it were. This notation is borrowed from a Slavic language, Czech; it is familiar to the European public through the famous shoemaker’s brand *Bat’a*. English and Chinese phoneticians, objecting to the use of diacritics placed on the side of the letter, add a curl to the bottom of the letter. Indianists note it by C; actually, the prepalatal stop of Sanskrit was only preserved in Indochina, in Sanskrit borrowings into Khmer and Lao: in India, it became a non-anterior sibilant stop [IPA: postalveolar affricate, /tʃ/]. For sibilant stops, the Czech notations are commonly used: C for the anterior sibilant stop [i.e. /ts/], Č for the non-anterior sibilant stop [i.e. /tʃ/] (thus, “Czech” is written *česky*). The sign Č is an abbreviation for CZ, a combination still in use in Polish. Czech does not have a voiced anterior sibilant stop, and Z is used in Czech to note a voiced spirant [i.e. the fricative /z/], as in French.

Old French had such stops [IPA: affricates]; in the 13th century, they weakened into spirants. The anterior sibilant stop became an anterior sibilant spirant, and since then CE, CI have been pronounced in the same way as SE, SI. Meanwhile, the corresponding voiced sound became the voiced

correspondent to /s/, i.e. /z/. The non-anterior sibilant stop [i.e. the affricate /tʃ/] became a non-anterior spirant: in French, CH is not pronounced as the English CH [i.e. /tʃ/] anymore, but as the English SH [i.e. /ʃ/]. A Latin word-initial I followed by a vowel had by that stage become a voiced palatal stop (like G in GE, GI /ge/, /gi/), then a voiced non-anterior sibilant stop [IPA /ɟʒ/] (word-initial J, originating in Latin I, still has this value in English), and finally, in 13th-century France, a voiced, non-anterior spirant [IPA: /ʒ/].

GI, D, Ð

The notation of the spirant mentioned above (/ʒ/) as J in French only dates back to the 17th century. Earlier on, J was simply the form that the letter I took when at the beginning of a word; thus one wrote *jnde* (for present-day French *Inde*, “India”), *jure* (for present-day *ivre*, “drunk”), *jean* (for the proper noun still written as *Jean* today; English equivalent: “John”), and so on. In Italian, the sequence GI was used in all cases where the sound at issue was a consonant, e.g. *Giovannis* “John”. Thus, in the 17th century, when a romanised script was devised for Vietnamese, J did not exist as a distinct letter contrasting with I; one can easily understand why GI was then adopted as a notation for the prepalatal voiced spirant (/ʒ/). When it came to transcribing the voiced prepalatal stop², which to the European ear sounded like a palatalised /d/, a simple D was used, without any added diacritic. In Vietnamese there was also a voiced [and preglottalised] alveolar stop [IPA: voiced implosive /ɗ/]; to transcribe it, the creators of the Vietnamese alphabet coined a new letter, Ð, by adding to the letter D a horizontal bar by analogy with the one in the letter T, which is the unvoiced counterpart to Ð.

S, TR

The anterior sibilant stop of Old Spanish, written C and Z, weakened, but did not become confused with S as it did in French. It became a dental spirant, like the English TH [IPA: /θ/]. It is thus more fronted than the former S. In some areas (especially in Soule Basque), Z corresponds to the French S: it is a dental spirant [IPA: /s/], whereas the S moved further back in the mouth and became a cacuminal spirant [IPA: /ʃ/]. ‘Cacuminal’ means that it is pronounced at the top—Latin *cacumen*—of the palate; it can also be called ‘retroflex’, meaning that the tongue is curled back. Vietnamese has a cacuminal spirant, which was written as a simple S, by analogy with the sound

²Translator’s note: since the time when Haudricourt’s article was published (1949), it has been shown, on the basis of reconstruction and of Vietnamese borrowings into neighbouring languages, that the phoneme written as D in the Vietnamese orthography was not a voiced prepalatal stop (closest IPA equivalent: /ɟ/) at the time when the orthography was devised, but a dental spirant, written as *ð* by Ferlus 1982 (see also Ferlus 2001). The closest equivalent of this sound in the International Phonetic Alphabet is /ð/: however, Ferlus makes a difference between spirants and fricatives, following Martinet 1956:24-25, who points out that spirants and fricatives are two different classes of sounds: “It is advisable to distinguish between relaxed articulations which tend towards a vowel-like articulation, for which we reserve the term *spirant*, and firmly articulated consonants, clearly characterised by friction as the air passes through the place of constriction: the latter are *fricatives* properly speaking” [our translation]; see also Martinet 1981, 1985 and Thomas, Bouquiaux and Cloarec-Heiss 1976:29-31. However, Haudricourt does not make a distinction between the two in the present paper: he uses “spirant” throughout.

just mentioned, whereas the corresponding stop [IPA: /t/] was approximated through the combination TR. These consonants are rare in Europe; they are often confused by Europeans with non-anterior sibilant consonants. For instance, Chinese has a consonant transcribed as TCH in the EFEO (Ecole Française d'Extrême-Orient) system of romanisation, and as CH in the Wade-Giles system; these notations appear to suggest that it is a non-anterior sibilant stop consonant [IPA: /tʃ/], pronounced with the back of the tongue and with simultaneous protrusion of the lips, whereas in fact this is a cacuminal sound [IPA: /tʃ̺/], pronounced with the tip of the tongue curled upwards against the palate. Such consonants [IPA: retroflex consonants] are especially frequent in India; Indianists transcribe them with an added dot: ʃ̺ ʈ̺. English and Chinese phoneticians add a forward-pointing hook at the bottom of the letter: ʃ̺̹ ʈ̺̹ [this has become standard in the IPA]. These consonants are rare in Indochina: in the Tonkin delta, they turned into prepalatals.

X

In Latin, the letter X stood for two sounds: it was a graphic abbreviation for CS. In Romance languages, this group of consonants underwent the same evolution as the group CT: it changed to IS (e.g. Latin *coxa*, French *cuisse* [IPA: /kɥis/] “thigh”). But in the Middle Ages, there developed a habit of linking the combinations US and IS into a single sign, which looked somewhat like the Latin letter X; as a consequence, an IS in French words often corresponded to a Latin X; as a consequence, X came to be used instead of S after an U or an I. This is why the French pronounce *soixante* ‘sixty’ as *soissante* [IPA: /swasât/], *Auxerre* (a place name) as *Ausserre* [IPA: /osɛʁ/], and *Bruxelles* (Brussels) as *Brusselles* [IPA: /bʁysɛl/]; this also explains why the plural of *cheval* ‘horse’ is *chevaux*, that of *genou* ‘knee’ is *genoux*. It is simply a matter of writing habits. But when words were borrowed from Greek into French in the 16th century, the letter X was again pronounced as KS, GZ [IPA: /ks/, /gz/].

We have seen that in Spanish, the group IT had evolved into a prepalatal stop; the same happened to the group IS, written as X. In Old Spanish, X thus stood for a prepalatal spirant; it was heard by the French as a non-anterior sibilant sound: *Don Quixote* and *Ximena* were written by the French as *Don Quichotte* and *Chimène* [both pronounced with a /ʃ/]. But the name of the famous missionary *Xavier* (a Spanish Basque name, equivalent to French Basque *echeverry*, ‘new-house’) was treated like a learned word and pronounced *Gzavié* [IPA: /gzavje/].

These considerations explain why the Vietnamese prepalatal spirant was noted X.

The notation of this spirant in the scientific alphabet is ç (the letter c with a small curl below it), but if it is more like a sibilant (anterior, as in Vietnamese, or non-anterior), the diacritic for palatalisation is added to the corresponding sibilant consonant (as used in Polish: s', c'). The anterior

sibilant spirants are always noted S, Z [IPA: /s/, /z/]. (...) ³ According to English and Chinese phoneticians, sibilant stops should be written with two signs, decomposing them into a stop+spirant sequence: /ts/, /dz/ for anterior sibilants, /tʃ/, /dʒ/ for non-anterior sibilants. (...) The cacuminal stops of Vietnam and China are written tʃ, dʒ or tʃ̣, dʒ̣. The stops written by phoneticians with two letters are said to be affricated.

N, NH, NG

The nasals (voiced stops with a lowering of the velum, such that air comes out at the nose) N, M have the same value in all languages. But in French, when they come at the end of a syllable, they are not pronounced as such anymore: instead, the vowel is nasalised (the velum is lowered during the pronunciation of the vowel). In Portuguese, the same phenomenon took place, but the N, written on top of the vowel, became \tilde . Linguists took up this sign for nasal vowels, e.g. /ã/. In Spanish, a double N (a sequence of two Ns) was likewise written with a superscript \tilde , yielding ñ; later, this double N became a prepalatal nasal; Indianists adopted the sign ñ for this sound. In French and in Italian, the prepalatal nasal originates in the Latin consonant cluster GN; this orthography was preserved in French and Italian, whereas in Portuguese, in Provençal and in Gascon this sound is written NH by analogy with the corresponding stop, CH. The latter spelling was adopted for the Vietnamese alphabet. English and Chinese phoneticians draw out the first stroke of the N into a curl: /ɲ/.

Lastly, in Romance languages the dorsal nasal is only occasionally found, when N is followed by C or G; this sound exists word-finally in English and German, where it is written as NG. It was therefore natural to write it as NG in Vietnamese. English and Chinese phoneticians combine both letters by dragging out the extremity of the n as in a g, thus: ŋ. Indianists write it with a dot on top of the n, thus: ñ.

V

Let us recall the origin of V. In Latin, V and U were not distinguished: V was the notation found in inscriptions for the labiovelar vowel, and U was the cursive form for the same sound. When U was followed by a vowel, it became a labiovelar consonant [IPA: /w/], then a labiodental consonant (the voiced counterpart to F [IPA: /v/, as opposed to /f/]); this change took place around the 6th century CE, whereas no change took place in the writing until the 17th century, so that two different sounds were written the same way. The letter V was used word-initially, when it referred to a vowel (as in *vn* “one” [written as *un* in present-day French]) as well as when it referred to a consonant (as in *vent* “wind” [still written as *vent* in present-day French]). Word-medially, U was used, thus *chanure* “hemp” [present-day French: *chanvre*] and *mur* [present-day French: *mur*]. But in other languages, a new notation had been introduced much earlier than the 17th century: since the

³Translator’s note: Two sentences were deleted at translation; they are indicated as (...).

labiovelar consonant had already become labiodental in Romance in the 6th century, the Romance-speaking scholars who devised spelling systems for Germanic languages had to create a new notation for the labiovelar consonant that they encountered in these languages: a double U (still called by that name in English), which led to present-day W (called *double V* in French). English and Northern French language varieties (Picard, Wallon) have retained a labiovelar, hence the pronunciation of the word *Wallon* [/walɔ̃/, and not /valɔ̃/], whereas in German, Dutch and Polish, W transcribes a labiodental (hence *Wagram*, *Württemberg* are pronounced with an initial labiodental [/v/]). Linguists distinguish labiodental V and labiovelar W.

H

As a last note concerning consonants, let us recall that French has a consonant which is written but not pronounced: the so-called “aspirated H” prevents the elision of the article [thus *la hache*, *le hibou* “the axe”, “the owl” and not *l’hache*, *l’hibou*] but is not pronounced anymore, except in a few areas (such as Lorraine and Normandie); conversely, Vietnamese has a consonant which is pronounced but not written: the glottal stop (sudden opening of the larynx) found at the beginning of vowel-initial words. In scientific writing, it can be noted by means of the *spiritus lenis* of the Greek writing system, a small apostrophe with its concave part oriented backward: ‘

English and Chinese phoneticians use the interrogation point instead: ? [Present-day IPA notation: ʔ]

2. Vowels

A, E, I, O, U

Latin had five vowels:

(i) A, an open vowel (i.e. with maximum distance between the tongue and the palate);

(ii) and (iii): two maximally closed vowels. I, a prepalatal vowel, better described as an *anterior* vowel in terms of tongue position; I is not rounded, i.e. the lips are spread. U, a labiovelar vowel, better described as a posterior vowel (in terms of tongue position) which is rounded: the lips are protruded and rounded;

(iv) and (v): two intermediate vowels: E in-between I and A; O in-between U and A.

The U changed its phonetic value in Old French around the 10th century: it became an anterior, rounded vowel [IPA: /y/]. The scientific notation of this French sound is ü (and ö is used for the vowel in-between ü and A); this notation comes from the German, where the trema, *Umlaut* in German, stands for an abridged E: ü stands for ue, ö for oe. English phoneticians favour the Scandinavian notation Y.

Y

In Ancient Greek, the rounded posterior closed vowel [IPA: /u/], which was written Y, became an anterior rounded consonant. Latin scholars borrowed the Greek letter Y to transcribe this vowel, which was not present in Latin. In Greek, this vowel later got confused with I, and the Latin too came to pronounce it as I, hence the name “Greek I” which the letter Y retains to this day in French. Concerning its role in the Vietnamese alphabet: in French, it is commonly used to note the voiced prepalatal spirant, which stands in the same relationship to I as W does to U; this is the notation used by Indianists [and Africanists]. But the English phoneticians prefer to note this sound as j. [Notation as /j/ has become standard in IPA.]

Latin had a distinction between long and short vowels. The notations were: a for the long vowel, ă for the short one; the latter sign is used in Vietnamese. Indianists transcribe the long vowel, whereas they omit the short vowel altogether. English and Chinese phoneticians indicate length by means of a colon following the vowel, or by doubling the vowel: a: or aa. [IPA: /a:/]

Ê, Ô, Â

In Romance languages, long vowels appeared when two vowels came together; this was abbreviated by writing the vowel only once, and adding a circumflex accent ^, e.g. in French *aage* was rewritten as *âge*, *meur* became *mûr*. (In French, ^ often stands for a former s, e.g. *fête* “celebration” comes from an earlier *feste*, *pâte* “paste” from *paste*.) The same happened in Portuguese, where *oo* became *ô*, *ee* became *ê*. In Portuguese, the new vowels had a more closed pronunciation: *ô* was in an intermediate position in-between *o* and *u*, and *ê* was intermediate between *e* and *i*. This is where the notation for Vietnamese vowels was taken up from. Neither Spanish, nor Italian, nor Provençal offered appropriate means for noting two different *e* vowels and two different *o* vowels.

In French, *ô* [IPA: /o/] is more closed than *o* [IPA: /ɔ/], and *ê* [IPA: /e/] is more closed than *e* [IPA: /ɛ/]; in view of this fact, some linguists use the grave and acute accents: *ó*, *é* for the more closed vowels, vs. *ò*, *è* for the more open vowels. Another possibility is to use a subscripted dot for the more closed vowels and a subscripted hook for the more open vowels. Lastly, English phoneticians use *e*, *o* for the more closed vowels, and *ɔ*, *ɛ* (an inverted C, and Greek *epsilon*) for the more open ones. The sign *æ* is used for a vowel in-between a and *ɛ*, like the vowel of the English word *cat*. An italicised *a* (i.e., *ɑ*) is used for the vowel of the French word *pâte*.

Vietnamese has unrounded back vowels [IPA: /u/, /ɯ/]. They were noted as *ư* and *ơ*, perhaps drawing inspiration from the notation *ü*, *ö* mentioned above, though in fact *ư* is quite the opposite of *ü* in terms of lip position and tongue position: *ư* [IPA: /u/] is back and unrounded, *ü* [IPA: /y/] is front and rounded. These vowels are not found in Western Romance languages; however, they are found in a Romance language of Eastern Europe:

in Romanian, the word “dog” could actually be written in Vietnamese spelling as *củn*. These vowels are sometimes written with a trema: *ï*, *ë*, or with a subscript circle. English phoneticians write them as /ʉ/, /ə/ [IPA /ə/ would now be reserved for a truly central vowel, and /ɤ/ be used for a back, unrounded, close-mid vowel], and /ʌ/, i.e. an inverted m, e and v, respectively. The latter (/ʌ/) corresponds to Vietnamese *â*.

Second part: Tones

As a last note, let us indicate that the notation of tones was borrowed from Ancient Greek. Ancient Greek had two tones, written as ´ and ~, and a third accent was used to note the tone of unaccented words: ` . These accents were supplemented by punctuation signs. Indeed, in Romance languages, tone cannot be used to distinguish between words, but it can distinguish between two sentences: for instance, in French, interrogative *C'est vrai?* “Is that true?” and affirmative *C'est vrai.* “That is true.” Thus, a simple point (analogous to a full stop) placed under the vowel, and an interrogation mark placed on top of it, supplemented the notation of tones [here are, as an illustration, the six tonal categories for the vowel A: a à á â ã ä].

English and Chinese phoneticians use a vertical bar placed to the right of the word, against which a small horizontal bar is added, indicating by its shape and position the height and modulation of the tone [e.g. ˧ for a high tone, ˨ for a mid tone, etc].

Conclusion

The Vietnamese alphabet is thus the product of well-identified historical facts.

The first Europeans who reached the Far East by sea came from the Iberic peninsula: they were Portuguese, Spanish, and Basque. Among them, the scholars who adapted the Latin alphabet to the Vietnamese language belonged to the Catholic clergy, and therefore knew Latin, Italian and Greek. All the peculiarities of the Vietnamese alphabet can be explained in light of these facts.

Looking at the Roman-based scientific alphabets, which were briefly presented in the course of the discussion, it appears that they fall into two groups: (i) those employed for Oriental languages that already had their own alphabets, in which case the Latin transcription is essentially a transliteration, replacing a letter of the indigenous alphabet by a letter of the Latin alphabet with a view to facilitate printing by European printers; (ii) those employed by phoneticians, who aim to transcribe all the nuances of pronunciation. The latter kind of alphabet is used in the countries where there is no alphabetic writing (e.g. China) or where spelling only has a distant link with pronunciation (e.g. England: it is no easy matter to guess the actual vowel sound in an English

word, despite the fact that English uses the Latin alphabet). But the phonetic alphabet is designed for handwriting, and difficult to print.

Hebrew	Greek		Latin	
א ALEF	Α α	ALPHA	A	
ב BET	Β β	BETA	B	
ג GIMEL	Γ γ	GAMMA	C	
ד DALET	Δ δ	DELTA	D	
ה HE	Ε ε	E-ppsilon	E	
ו VAV	Ζ ζ	ZETA	F	
ז ZAYIN	Η η	ETA	G	
ח HET	Θ θ	THETA	H	
ט TET	Ι ι	IOTA	I	
י YOD	Κ κ	KAPPA	J	
כ KAF	Λ λ	LAMBDA	K	
ל LAMED	Μ μ	MU	L	
מ MEM	Ν ν	NU	M	
נ NUN	Ξ ξ	XI	N	
ס SAMEKH	Ο ο	O-micron	O	
ע AYIN	Π π	PI	P	
פ PE	Ρ ρ	RHO	Q	
צ TSADI	Σ σ ς	SIGMA	R	
ק KUF Q	Τ τ	TAU	S	
ר RESH	Υ υ	U-pilon	T	
ש SHIN	Φ φ	PHI	U	
ת TAV	Χ χ	CHI	V	
	Ψ ψ	PSI	X	
	Ω ω	O-mega	Y	
			Z	

Figure 1. From the Hebrew and Greek alphabets to the Latin alphabet.

		Scientific notations :	
Vietnamese	French	indianists and ethnologists	phoneticians and sinologists
th		th	t'
ph			φ
	f	f	f
	ch	ś	s
	tch	č or tš	ts
ch		c or t'	t or ṭ
d		j or d'	d or ḍ
x		ç or ś or s' or š'	ç or ś
gi	z' or ž'		ʒ
s		ś	s
tr		tʃ	ts
nh	gn	ñ	ɲ
ng		ṅ	ŋ
vocalic initial		'	ʔ
	eu	ö œ	ø
	u	ü	y
ư			
u	ou	u	u
ơ		ə	ə
	ill	y	j
ê	é	ɛ	e
e	è	e	ɛ
ô	o	o	o
	o	ɔ	ɔ

Figure 2. Equivalences across alphabets: Vietnamese, French, and scientific notations.

REFERENCES

Ferlus, Michel. 1982. "Spirantisation des obstruantes médiales et formation du système consonantique du vietnamien." *CLAO* 11(1):83-106.

Ferlus, Michel. 2001. "Les hypercorrections dans le thổ de Làng Lỗ (Nghệ An, Vietnam) ou les pièges du comparatisme." *Paper Presented at the Quinzièmes Journées de Linguistique de l'Asie Orientale*, Ecole des Hautes Etudes en Sciences Sociales, Paris.

Martinet, André. 1956. *La Description Phonologique Avec Application au Parler Franco-Provençal d'Hauteville (Savoie)*. Genève: Droz.

Martinet, André. 1981. "Fricatives and spirants." In B. P. Mallik, ed., *Suniti Kumar Chatterji Commemorative Volume* 145-151. Burdwan, West Bengal, India: Burdwan University Press.

Martinet, André. 1985. "Two proposals." *The Study of Sounds (Onsei no Kenkyuu)*, *Commemorative Volume for the 50th Anniversary of the Phonetic Society of Japan* XXI:67-72.

- Suriya, Ratanakul, David Thomas and Premsrirat Suwilai, Eds. (1985). *Southeast Asian Linguistic Studies Presented to André-G. Haudricourt*. Bangkok: Mahidol University.
- Thomas, Jacqueline, Luc Bouquiaux and France Cloarec-Heiss. 1976. *Initiation à la Phonétique: Phonétique Articulatoire et Phonétique Distinctive*. Paris: Presses Universitaires de France.

Received: 24 January 2008

7 rue Guy Môquet, 94800 Villejuif,
FRANCE
<alexis.michaud@vjf.cnrs.fr>