

## WRITING SYSTEMS

*Tjeerd TICHELAAR*

### **Meaning, sound and looks**

Geographical names, as all names, can be viewed from three different aspects: semantic, phonetic and graphic.

Semantically, names carry a meaning at the time they are coined. Because of its meaning as ‘ford on the Roman road’, there could, at least to local people, be no misunderstanding about the location of the settlement named Stratford. The meaning may be lost in the course of time, either because the language the name springs from is no longer current, or because the name itself degenerates to such an extent that its meaning can no longer be recognized. More often, the original meaning ceases to be appropriate. The Greek settlement named Neapolis or ‘New Town’ by its founders, to set it apart from the ‘old town’ they had fled (neighbouring Kymai, that had been established six or seven generations earlier by colonists from the Ionian city state of Chalcis), retained its Greek name as it was incorporated in the Roman republic in the 4<sup>th</sup> century BC. As long as Greek remained the (local) official language in Roman Neapolis, which it did for about seven more centuries until the reign of Constantine the Great, the name of the city, its meaning still being understood, was fairly safe. After that, the state language Latin, in a regional variety spoken for long by the majority of the commoners in the city, replaced Greek as the language of administration. To the speakers of the regional variety of Latin, that consequently developed into the Romance Napolitan language (or dialect), the meaning of Neapolis was lost; to its citizens, the city actually appeared ancient instead of new. Kymai had completely been overshadowed even before the Romans came; what remained of it was, ironically, destroyed by the Napolitans in the early 13<sup>th</sup> century. The name Neapolis ultimately degenerated into Napoli in the Napolitan and Italian languages that succeeded Latin as the language of administration (the latter actually as late as the 19<sup>th</sup> century).

Unlike the Latin-speaking heirs of Napoli, new foreign-language users of existing names are occasionally known to make efforts to retain the meaning of a name after its language has been discarded. Translation of names was never a Roman policy – another famous ‘New Town’, the Phoenician colony of Qart Hadash, was also incorporated into the empire under its own native name, be it in this case latinized into ‘Carthago’ – but was very popular

during the manifold changes of sovereignty between the new national states of 19<sup>th</sup> and 20<sup>th</sup> century Eastern and South-Eastern Europe: examples are Austrian Karlsbad changing into Czech Karlovy Vary, Hungarian Újvidék becoming Croat Novi Sad. An example outside Europe, but equally associated with the emergence of national sentiments, is the translation of Spanish Casablanca (itself a translation of Portuguese Casa Branca – the Portuguese founded the place in 1515) into Arabic ad-Dar al-Bayda’. In other cases, name changes following transfer of ownership explicitly reflect a change in the meaning the object has to the new owner compared to the one attached to it by his predecessor. The seaport city of Reval was just a ‘little sand bank’ to its Danish founders in 1219. Although it successively passed into Teutonic (1346), Swedish (1561), and Russian hands (1710), it was allowed to retain its name; but upon the emergence of the independent Estonian nation-state in 1917, the town now chosen to be the nation’s capital reverted to a co-existing Estonian appellation reflecting what it had from the beginning on meant to the Estonians: the ‘Danish fortress’ (Tan-linn, through the centuries contracted to Tallinn).

Phonetically, names consist of a succession of vocal sounds that as a whole reminds the recipient of the objects they represent. The very fact that names do degenerate – only incidents like Roma, Cremona, Capua (Latin) and Argos (Greek) are known to have subsisted through larger spans of time without phonetic damage, and even these did so only after they became latinized c.q. hellenized – indicates that the sound aspect has for long been dominant above both the semantic and the graphic aspects in determining the evolution of names. This dominance can be explained by the heavy dependence on oral communication in ‘passing the word’ throughout history, and, in connection with this, maybe by the role played by the well-trained auditive memory in the process of learning/memorizing information conveyed through language.

It has been demonstrated earlier that the phonetic qualities of a name can lead to misinterpretation and thus influence the evolution of the name itself. The generic element ‘-wic’ (settlement) in the Anglo-Saxon name of York, Eoforwic, was misunderstood as ‘-vīk’ (bay) by the Danish invaders taking over the settlement in the 9<sup>th</sup> century, demonstrating that the mere suggestion of phonetic

recognition was enough for the Danes to set aside any semantic objections against this interpretation: the town they named after a 'bay' was then, as it is now, situated at a solid inland location. Likewise, the Anglo-Saxons before them had misunderstood the specific element 'ebur-' (yew-tree) of the preceding Brythonic-Roman name Eburacum as 'eofor' (wild boar), also resulting in semantic mutilation of the original name. The latter is an example of a very commonly occurring process called popular etymology, in which the historic background of a name is reconstructed without taking into account the rules of phonetic change as they are known by historical linguists, resulting in erroneous semantic interpretations and, in the worst case, wrongful (hyper)correction of the name.

As long as names were predominantly transmitted through oral communication, their graphic representation showed little stability. The need of (ortho)graphic standardization was rarely felt, and both between and within documents names referring to the same object could be spelled in different ways. The fact that this was considered acceptable suggests that written documents were essentially meant to support oral transaction: when reading it aloud, the exact orthography of the name, as of any word, was of little relevance.

It is a fairly recent development that writing takes absolute preponderance over the spoken word. More than in the past, public administration pervaded into all segments of social and economic life. A completely new situation was created by the introduction of computers in the administrative processes. Unprecedentedly, the computer made it possible to process graphic data, translated into digital code, without human intervention in the form of (phonetic) reading. As the pronouncing, understanding and correcting human interpreter was by-passed, this short-circuiting of the linguistic data interpretation process necessarily demanded a stricter than ever graphic standardization.

### **The graphic aspect of language: writing**

The need for graphic representation of the sounds of language became felt when the span of time and distance of man's utterances was requested to exceed the physical limitations of the speaker. An administration of greater volume and complexity than that what an individual was able to remember, required some kind of notation that could be separated from that individual and that specific point in time. The embryonic predecessors of writing

were most likely systems of notation of administrative data, consisting of numbers and... names. The data taken down were trade and cadastral data, that, as it was foreseen, needed to be remembered and consulted at a yet unknown point in the future, by a yet unknown individual. Once proven successful, these systems held a promise to allow local rulers to spread their word, and thus their law and authority, beyond the range of their voice and muscles. The first known scripts thus developed in the Sumerian city states of the late 4<sup>th</sup> millennium BC, and enabled these local centres of power to grow into territorial states exceeding the vision of the central *zigurrat* (tower). Shortly after this innovation was introduced in Sumeria, an elaborate script emerged on the banks of the Nile, where a powerful centralized Egyptian empire was about to see the light.

Whether the idea of writing spread throughout the world from one Sumerian origin or it emerged independently in different cultures, the development of writing systems seems to follow a universal sequence from purely pictorial representation (*pictograms*) to sets of abstracted sound-representing symbols (*phonograms*). Pictograms convey meaning without intervention of sound values; there may be a symbol meaning 'town', 'river' or 'mountain' irrespective of what the word for 'town', 'river' or 'mountain' sounds like, and thus regardless of any specific language. Such a symbol is named a *logogram*. The advantage of logograms is their universal applicability – because they are language-independent – but they have the obvious disadvantage that there must be a separate symbol for every word. A system consisting of pictograms only cannot be expected to be completely learned and memorized by anyone, as only a limited number of concrete substantives could possibly be represented by naturally drawn pictures.

All complete writing systems the world has ever known, from the highly pictorial Egyptian hieroglyphs to the modern alphabets, do effectively contain both logograms and phonograms. As purely pictorial 'proto-scripts' develop into 'scripts' or writing systems, naturally drawn pictograms are stylised and augmented with drawings for abstract phenomena (hence called *ideograms*), and will ultimately contain logograms for all basic words of a specific language. Phonograms are developed out of logograms through a process starting with the *rebus* principle: the sound values (in a specific language!) of monosyllabic words are attached to the logograms representing these words, thus creating a phonetic syllabary or syllabic script. A fully syllabic script would contain as many symbols as the language it is used for contains syllables. A syllabary can develop further into an alphabetic

system, in which single phonemes (units of sound) instead of syllables are represented by symbols – thus requiring even less symbols. Alphabets may contain both the consonants and the vowels used by a language, or be consonantal (containing consonants only). To the symbols (*letters*) of consonantal alphabets, the vowels following consonant sounds may, either optionally or obligatory, be added to the letters by diacritical marks (*vocalization*), as may certain phonetic modifications of the consonants (nasalization, aspiration etc.).

As said, even in alphabetic scripts some logograms persist: examples are the ciphers (0,1,2,..., in English: ‘one’, ‘two’, ‘three’...) and signs like + (in English: ‘plus’), – (in English: ‘minus’), & (in English: ‘and’), and, recently added, @ (in English: ‘at’).

### Scripts of the world

The cradle of most of the modern phonographic scripts is the Middle East. The oldest known Sumerian and Egyptian pictographic inscriptions considered to be scripts date from the 4<sup>th</sup> millennium BC. Egyptian hieroglyphic ideograms were progressively phoneticized, receiving the phonetic value of the consonantal content of the words they represented. Likewise, the Sumerian logograms were phoneticized as early as in the 3<sup>rd</sup> millennium BC, giving rise to an essentially syllabic writing system known as *cuneiform* (after its wedge-shaped characters).

It was the unique endurance of the introvert Egyptian culture (Egypt succeeded to remain Egyptian and ‘egyptophone’ for thousands of years), that left the contemporary Sumerians with the greater immediate posterity, as far as writing is concerned. In the 24<sup>th</sup> century BC the Sumerian states were subdued by the Semitic Akkadians, whose language supplanted Sumerian as the language of administration in the Mesopotamian area. From the Sumerians, whose language eventually disappeared, the Akkadians adopted the syllabic cuneiform script. As their language was totally unrelated to Sumerian, the script now definitely became phonographic in essence.

After its invention by the Sumerians, the cuneiform script survived for more than 3000 years. Cuneiform scripts were used to represent the words and sounds of about 15 different languages – among which Semitic languages, Elamitic (a presently extinct language isolate) and Indo-European languages (Hittite, Old-Persian). Most, but not all of the scripts (Old-Persian) made use of the symbols introduced by the Sumerians.

The ancient world saw a number of other scripts appear and disappear, among which two scripts used by the Minoan language of pre-Mycenaean Crete: a hieroglyphic script and a (probably) syllabic script called Linear A; an adaptation of Cretan Linear A by the Mycenaean (Greek) invaders of the island to accommodate for the syllables of their own Indo-European language, that was consequently exported to mainland Greece and Cyprus (Linear B); the as yet undeciphered script of the Indus civilization, the language of which is also unknown; and a Canaanite cuneiform introduced in the Levantine city state of Ugarit (in present-day Syria) as a, for the Semitic Canaanite language, preferable alternative to the Sumerian-based Akkadian script.

The Ugarit cuneiform is thought to deserve the honour of having been the first real alphabet: it contained only about 30 letters.

A not too certain genealogy links the Semitic scripts emerging after the ‘sea people’ migrations around 1200 BC to Akkadian cuneiform. The alphabetic pioneer of Ugarit was, together with the city itself, wiped out during this major turmoil that affected the whole Mediterranean area. South of Ugarit, a new Canaanitic nation emerged out of the ashes, that from a girdle of well-defended sea cities was about to build a network of trading colonies of an unprecedented territorial extent: it reached from the Levantine homeland all the way to the Atlantic Ocean. This nation, called Phoenicia, obviously needed an effective writing system to hold itself together; that it was focussed on effective communication was demonstrated by its impressive beaconing system, that made it possible to send a message from Tyre (now Lebanon) along the southern coast of the Mediterranean all the way to Gades (now Cadiz, south-western Spain) within hours. The Phoenicians invented an alphabet of 22 consonants, that was later borrowed and adapted by their major competitors in the Mediterranean: the Greeks.

To accommodate the Semitic script to the phonology of their own Indo-European language, the Greeks applied some unneeded characters for the representation of vowels, and augmented the alphabet with some newly designed letters. Initially there were two separate Greek alphabets, a western (Chalcidian) and an eastern one (Ionian); the latter, more specifically the Miletan variant, became the standard script as Athens acquired political and economic supremacy among the Greek city states. It spread into Asia with the conquests of Alexander the Great, and remained influential during the hey-day of the Byzantine Empire.

In the Middle Ages, the Greek script served as the basis for the development of two new alphabets for the liturgies of the Slavic-Orthodox churches: Glagolitic and Cyrillic. The Glagolitic script was in use for some time in Croatia and Dalmatia (in Croatia as long as the 17<sup>th</sup> century), while the Cyrillic alphabet subsisted in Serbia, Bosnia and Hercegovina, Macedonia, Montenegro, Bulgaria and the Soviet Union. In Russia it was revised during the reign of Peter the Great (1708–1710): some letters were discarded, others were reshaped so that they became more similar in appearance to the letters of the Latin alphabet. Through the expansion of tsarist Russia, the alphabet reached the shores of the Pacific Ocean (and temporarily across the Bering Strait into Alaska), its position being consolidated in the 20<sup>th</sup> century by the political equilibrium between the Soviet Union and the Latin-writing ‘Western world’. Although for the phonetic transcription of some minority languages Soviet linguists preferred to use the Latin alphabet, most of the languages of the Soviet Union (with as most important exceptions Estonian, the Baltic languages, Georgian and Armenian) used, up to the collapse of the Union in 1992, Cyrillic scripts. The former Socialist Soviet Republics of Moldavia, Azerbaijan and Turkmenistan reverted to Roman scripts after gaining independence.

In Italy, where the Greeks of Magna Graecia (and especially Campania) closely interacted with the Etruscans, the Chalcidian Greek alphabet was taken over by the latter, who passed the script (but not their non-Indo-European language) on to the Romans eventually subduing them. The success of the Roman (Latin) alphabet, that spread all around the world – initially together with the advances of the Roman Empire and the Roman Catholic faith, later as a consequence of Western colonization – is evident. The languages of the non-Orthodox nations of Europe are now almost all written in alphabets the letters of which are practically identical to those of the classical Latin script. Only in Ireland, with its early tradition of writing, a Gaelic alphabet coexists with the modern Latin script. The Gaelic alphabet was derived from the Latin script as early as the 5<sup>th</sup> century AD.

A score of other Italic languages also developed scripts out of the Chalcidian, either by direct adoption from the Greeks in Southern Italy, or through mediation of the Etruscan alphabet.

Even before the Roman conquests (that didn’t reach them anyway) and christianisation, the peoples of Northern Europe already adopted some scripts that were most probably derived from, or at least influenced by, the Italic alphabets: the Ogham script of the British Celts and

Picts, and the runes of the Germanic peoples. A few runic symbols still survive as augmentations to the alphabets of Icelandic and Faeroese.

In Egypt, the Ionian Greek alphabet was modified and augmented with some letters from Demotic (a stylised descendant of the ancient hieroglyphs) to create the Coptic script, that to date is still in use for the liturgy of the Coptic Church.

The Phoenician alphabet is classified as a Canaanite version of the North-Semitic script. North-Semitic further includes an Aramaic version, that competes with Phoenician in successfulness. Influential descendants of this Aramaic script are: 1) the Hebrew alphabet; 2) the Nabataean script, out of which the Arabic alphabets developed; 3) the Pahlavi script; 4) the Sogdian script. The Hebrew and Arabic alphabets conveyed the Holy Scriptures of the Jewish and Islamic religions, which was of great consequence to their dissemination. Where the use of the Hebrew alphabet stayed confined to the Hebrew and Yiddish languages as well as the languages of some numerically less important Jewish sects like the Turkic Karaim in Ukraine and Lithuania, the Arabic alphabet spread, from the 7<sup>th</sup> century on, together with the Islamic faith, from the Atlantic coast of Northern Africa all the way to the Malay Archipelago. The Arabic script is currently used mainly by Afro-Asiatic (Arabic) and Indo-European (Persian, Tadjik, Urdu) languages; in the past it served Altaic (Turkish, Azeri) and Austronesian languages (Malay, Javanese, Madurese) as well. Likewise, the Pahlavi alphabet of Sassanide Persia spread with the Zoroastrian religion over large parts of Central Asia. As Pahlavi was an impoverished version of the Aramaic consonant script, in spite of a large number of complex ligatures poorly equipped to accommodate the phonology of the Indo-European Persian language, it was in the Zoroastrian parishes eventually substituted with the Avestan script, that was better suited to represent the Persian phonemes. The supremacy of the Islamic religion prevented this script, however, to play any role of importance in Persia: modern Persian is written in the Arabic script, augmented with a couple of characters for phonemes not occurring in Arabic. The Pahlavi script did possibly influence a few alphabets that are currently still in use: Georgian Mkhedruli, and the Armenian script dating from the 5<sup>th</sup> century AD.

A Sogdian adaptation of the Aramaic script was adopted by the Turkic-speaking Uyghur nation. In spite of many specifically Uyghur modifications, the script remained defective for the representation of the Turkic consonantal phonemes: especially the meaningful discrimination

between voiced and unvoiced consonants, foreign to the Iranian (Indo-European) Sogdian language, could not be represented. With the introduction of Islam, the Uyghurs adopted the (Perso-)Arabic script. The Sogdian alphabet revived in Mongolia, as the script of the 13<sup>th</sup> century empire of Kubilai Khan. Until 1941 it was in use in Outer Mongolia, after which it was, because of the political dominance of the Soviet Union over the People's Republic of Mongolia, supplanted with a modified Cyrillic alphabet. In the People's Republic of China, the Sogdian-Mongolian alphabet is reported to be still in use amongst speakers of the Mongolian and Manchu minority languages.

More recent adaptations of the Aramaic alphabet are the so-called Neo-Aramaic alphabets, to date still in liturgical use by the Maronite Church (Estrangelo), the Syrian-Jacobite Church (Serto) and the Nestorian Church (Nestorian).

Southern Semitic alphabets, it is thought, may have been the source of both the Ethiopian alphabet and the Indian Brahmi script, although the latter is disputed. The original Ethiopian script survived, with addition of vocal diacritics, in the alphabet used to date for Amharic and other languages of Ethiopia and Eritrea (Tigre, Tigrinya).

The Brahmi alphabet was, in the course of the 1<sup>st</sup> millennium BC, developed to convey the rich Indo-Aryan literature, thereby replacing the lost script of the pre-Indo-European Indus culture. Initially there was some competition from the Aramaic-based Kharosthi script, but from the 5<sup>th</sup> century BC onwards Brahmi-based alphabets emerged throughout the Indian sub-continent to provide a vehicle for the diversity of Indo-European and Dravidian languages used in the area. Their dissemination followed the spread of the Hindu religion and literature. A northern variety of Brahmi, the Gupta script, gave in the 7<sup>th</sup> century AD birth to Nagari, which is still the alphabet used by Sanskrit, Hindi and a number of other Indo-Aryan languages (e.g. Marathi, Nepali, Bihari), as well as some Dravidian and Munda languages.

Also derived from the Nagari script were: 1) the Bengal alphabet, currently in use in different varieties for Bengal (spoken in Bangladesh and West Bengal), Assamese and the Munda-language Santali; 2) the Gujarati script, still very close to Nagari but lacking the horizontal lines above the letters; 3) Gurmukhi, the alphabet used by Punjabi, attached since the 11<sup>th</sup> century to the Sikh religion; 4) the Tibetan Fagspa script.

Southern varieties of Brahmi (Grantha, Pallawa) differentiated into: 1) a number of highly divergent alphabets

for the Dravidian languages Tamil, Telugu, Kannada, Malayalam and Tulu, spoken in the southern part of India; 2) the Singhalese script of Sri Lanka; 3) the alphabet of Oriya spoken in Orissa; 4) the Indonesian Kawi-scripts used by Javanese, Balinese, Sundanese and some other native languages. Since the arrival of Islam in Java, the Kawi scripts were progressively superseded by the Arabic alphabet and, in the 20<sup>th</sup> century, also by Latin script, but they subsisted in Bali and around the Javanese *kratons* (royal palaces). At present they are taught again in elementary and secondary schools.

Other representatives of the Indian alphabets are the *Âksâr* script, from which the modern Khmer and Thai alphabets were derived, and the Mon script, that laid at the base of the Sino-Tibetan languages Lao and Burmese and Austronesian Cham, spoken in Vietnam and Cambodia.

The major modern systems of writing that cannot be traced back to a Semitic origin have developed from the largely (but not exclusively) logographic script of China. The Chinese script, dating back at least until the 2<sup>nd</sup> millennium BC, contains some 50,000 different characters. The (unrelated) Japanese language adopted Chinese logograms starting from the 3<sup>rd</sup> century AD. The Chinese-based Kanji script is currently in use together (and intermixed) with the two times 50 syllabic Kana characters (Hiragana for general use, and Katakana specifically for the transcription of foreign words and names).

The Korean script was not derived from Chinese, but devised in the 15<sup>th</sup> century on an alphabetic basis. The alphabet, consisting of 25 letters (vowels as well as consonants), is sometimes used in combination with Chinese logograms.

Alphabets of unknown origin are, or were in a not too distant past, in use for the Berber language Tuareg (Tafinagh, a consonantal script possibly derived from the ancient Numidian alphabet), and some regional languages in Indonesia: Batak (in North Sumatra) and Buginese (in Southwest Celebes). The latter is supposed to have had its origins in Java, and thus be related to the scripts of Southern India.

In order to allow representation of the phonemes of as many known languages as possible, the International Phonetic Association (IPA) developed an artificial script on the base of the Latin alphabet. This script is occasionally used to put as yet unwritten languages to paper, but because of its complexity its usage is effectively confined to circles of linguistic specialists.