UNITED NATIONS GROUP OF EXPERTS ON GEOGRAPHICAL NAMES

Working Paper No. 45/15

Twenty-ninth session
Bangkok, 25 – 29 April 2016

Item 15 of the provisional agenda

Activities relating to the Working Group on Pronunciation

UNGEGN involvement in the dealing with pronunciation aspects of geographical names

Submitted by Working Group on Pronunciation*

* Prepared by Tjeerd Tichelaar (Netherlands), Convenor, Working Group on Pronunciation

United Nations Group of Experts on Geographical Names
29th session, Bangkok, 25–29 April 2016
Agenda item 15

UNGEGN involvement in the dealing with pronunciation aspects of geographical names[i]

Summary

Starting from a descriptive survey of the issues involved in the relationship between writing and pronunciation of geographical names and their historical background, exemplified by the Roman script, this paper examines the support UNGEGN and its Working Groups can provide to efforts to cover pronunciation aspects in connection with the general objectives of national and international geographical names management and standardization. Following this analysis, a proposal is made for the future direction as well as the immediate goals to be set for the Working Group on Pronunciation.

Pronunciation and names standardization

A little over half a century ago, the United Nations Group of Experts on Geographical Names (UNGEGN) was instituted in order to promote the standardization of geographical names, both on national levels and – making it a self-evident UN concern – in international communication. Although this was not explicitly expressed, the standardization effort was practically directed at the *written* form of names, as it was a problem of written communication the expert group was meant to address.

Because names are linguistically non-generic nouns and thus, as all words, essentially sequences of sounds rather than graphic characters, the standardization pursued by UNGEGN cannot intend to 'freeze' them for eternity: both speech and writing conventions are dynamic, they evolve and change over time. Shakespeare's 16th century English we now find hard to understand, and likewise the names in the maps of his contemporaries Ortelius and Blaeu are no longer the names we put in our maps today. The standardization endeavoured by UNGEGN, in other words, is not a one-time job but a process that cannot be expected ever to end. As long as our nations unite, the relevance of the expert group's mandate will not decrease.

Although it is the graphical expression of geographical names which is for practical reasons the central concern of UNGEGN, pronunciation aspects both underlying and interacting with the written forms may not be ignored when dealing with names. Hence the creation of the Working Group on Pronunciation in response to resolution no. 11 of the Eighth United Nations Conference on the Standardization of Geographical Names (Berlin 2002). As standardization of pronunciation is rarely attempted by national governments worldwide and, admittedly, way over the head of any international expert group whether or not it operates under the umbrella of UN – its very desirability being debatable, in this respect – the proper allocation of the subject of pronunciation within the focal range of UNGEGN deserves some discussion yet. The following is meant to support the debate.

Written vs. spoken language

Of the current 192 UN member states, the nationwide official languages of 190 exclusively use so-called *phonographic* writing systems. The written form of these languages essentially represents their pronunciation, as opposed to languages using *logographic* writing systems, in which the writing directly refers to the meaning of what is written. The latter are nowadays still in use for Chinese and (partly) Japanese.

Although phonographically written words and names are thus supposed to reflect their pronunciation, the relationship between writing and pronunciation is language-specific and in many languages far from straightforward. Similar letters and combinations of letters or letters combined with diacritical marks are pronounced differently in different languages. Even within a single language they may be pronounced in more than one way, depending on their position within a word or the sounds surrounding them, or, especially in the case of names, by historical influences the writing itself does not disclose. The letter combination *ough* in English, for instance, may be pronounced in multiple ways: cf. r*ough*, thr*ough*, th*ough*, th*ough*t, or, to list some English geographical names, Br*ough* (IPA notation: ʌf), *Ough*tibridge (uː) and Scarbor*ough* (ə) in Yorkshire (but Scarbor*ough* in Ontario: oʊ), Sl*ough* in Berkshire and L*ough*ton in Essex (aʊ), Br*ough*ton in Buckinghamshire (ɔː), St*ough*ton in Massachusetts (oʊ), L*ough*or in Wales (ʌx), Cl*ough* (ɒk), Cloughey (ɔːx) and Kill*ough* (ɒx) in Northern Ireland, Y*ough*al in Ireland (ɒh), G*ough* Island in the South Atlantic Ocean (ɒf). Sometimes this 4-letter combination is even pronounced differently within one name, as in L*ough*bor*ough* in Leicestershire (ʌ and ə). Likewise, the letter combination *augh* is pronounced differently in the names of three (written) homonymous places Cl*augh*ton, all located within miles from each other: Cl*augh*ton in the City of Lancaster (æf) and Cl*augh*ton in the Borough of Wyre (aɪ), both in Lancashire, and Cl*augh*ton in neighbouring Merseyside (ɔː).

Fig. 1. The British Isles: many ways to say ough (lines connect similar pronunciation)

The twelve different pronunciations of *ough* in English geographical names are obviously an extreme example. Although many languages do maintain a more systematic correspondence between writing and pronunciation than English, it is quite common for languages to either use the same character for several different sounds, or apply letters, diacritical marks and combinations thereof in a language-specific way to accommodate at least the meaningful sounds the language discerns. The reason for this is, that the writing systems applied for languages were more often adapted than specially created for the language employing them. Writing systems typically spread in the same way most technological innovations do: borrowed at first from foreign creators, then gradually adapted to the specific requirements of the borrowers – in this case the borrowing languages. In the case of the writing systems called alphabets, officially applied now by 158 UN member states, a complicating factor is that the ancient Phoenician script all these systems ultimately trace back to was a so-called *abjad* rather than an alphabet itself: a script representing consonants only. This must have sufficed for the purposes this script was originally devised for, which may have involved the administrative identification of a limited number of generally known objects and geographical names. The widely travelling Phoenician merchants undoubtedly needed to write down names that were foreign to them, and thus lacked the meaning allowing them to be written down in the logographic script of the time. Nevertheless the letters they devised represented the consonants of their own Canaanite language, to which foreign sounds were equated in accordance with what the Phoenicians believed to hear.

At this point, it is instructive to realize that of the numerous sound distinctions human beings are physically able to make, communities sharing a language typically use a limited number only to communicate. The sounds they set apart by such (to them) meaningful distinctions are called *phonemes*. They are defined by inherited consensus within the community of speakers of the language. Every language thus possesses its own specific set of phonemes. The members of a language community develop sensitivity towards their own phonemic sound distinctions (the sound distinctions meaningful to them), and are simultaneously trained to ignore any other distinctions that might be heard. People speaking different languages don't just fail to understand each other's words: they neither recognize each other's phonemes, to a level that they may believe they don't hear the difference between all of each other's sounds. This mechanism is nicely demonstrated through the word by which ancient Greeks generalized all non-Greek speakers: these people, according to their judgement, did not really speak a language but produced 'bar-bar-bar'-sounds instead (i.e.: sounds that to Greek ears all sounded the same). This habit reduced them to 'barbarians', a brand of people occupying a lower step of civilization. Similar references were made by foreigners in later times to indigenous people of northern ('berbers') and southern ('hottentots') Africa. Ethically speaking, most of us will currently agree that such appellations expose an intolerable degree of ignorance and indifference on the side of the name-givers, but actually it is an important quality to be insensitive to the sounds of others in order to be able to understand one's own.

Roman alphabets

The Roman alphabets applied internally by 75% of the UN member states today and for transliteration purposes by many others were all derived from the alphabet the Romans once optimized to represent the phonemes of their own Latin language. The Roman script was an adaptation of an Etruscan predecessor, which had been adapted again from the alphabet of the Greek city state of Chalcis, an early adaptor of the phonograms of its Phoenician trade partners in the first half of the first millennium BCE. By adding vowel signs, the Chalcidians (and other Greeks) had turned the Phoenician abjad into a proper alphabet. The fact that the Greek phonetic system

differed to a considerable degree from the Phoenician, the Etruscan from the Greek and the Latin from the Etruscan meant that at every adoption some letters were considered superfluous or redundant, while for some phonemic sounds the source script lacked the letters required. This problem was incidentally resolved by inserting characters taken from a different script (for instance the runic Þ in Gothic, Old English, Old Norse and Icelandic), but more often by making digraphs and ligatures (Ll in Welsh, IJ in Dutch, Æ in Danish, Œ in French, and ß in German – the latter initially used as a typographic ligature of 'long s' ſ and 'short s' z or 'round s' s) and/or combining letters with diacritical marks (Đ in Old Irish and again old Germanic alphabets and Icelandic, Ł in Polish, letters with acute, grave and circumflex accents etc.). In some languages, like German, French and also English (where this is the exclusive resolve), combinations of two to four letters were customarily applied to represent phonemes non-existent in the Roman Latin source of their alphabet. As a result of such local solutions to similar problems invented all over the world over a period of roughly a century and a half, the world now knows dozens of different Roman alphabets, between which the sound values of many letters and diacritic signs differ in a sometimes unpredictable way.

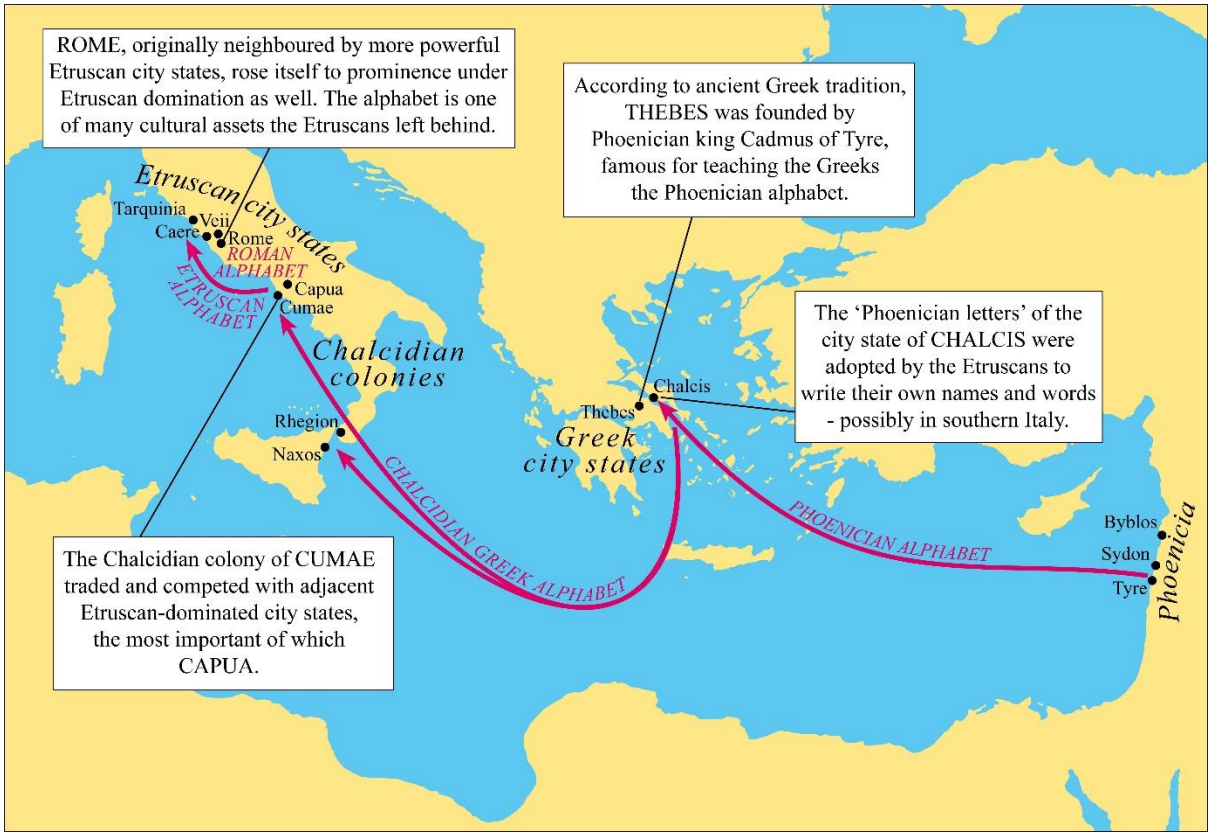

Fig. 2. Spread of the innovation of phonetic script: the origin of the Roman alphabet.

Synchronization of writing and pronunciation: spelling reforms

To maximize the accessibility of written language, national regulatory bodies regularly enact spelling reforms to ensure that the correspondence between speech and writing remains as uncomplicated as possible. Because of political and other reasons – such authorities do not always exist, and their mandates are not everywhere as far-reaching – languages quite spectacularly differ in the degree to which this policy is effective. Orthographies accurately reflecting current pronunciation are labelled *phonetic* – for instance in Finnish, as well as in Indonesian and other languages relatively recently put to (Roman) script – as opposed to *archaic* spelling where pronunciation and writing grew wide

apart. To the detriment of those concerned with geographical names, in quite a few cases (for instance in the UK and in the Netherlands), names tend to be completely or partially exempted from general pronunciation-based spelling reforms. Even when new spelling rules are meant to include toponyms as well, the mandate of language authorities does not always extend to legally registered names of administrative entities. In other cases, however, geographical names neatly follow suit: this happened, for instance, in the Swedish reform of 1906, the Indonesian reforms of 1947 and 1972, the Romanian replacement of î with â in 1993, and the simplification of the Bulgarian Cyrillic alphabet in 1945.

Treatment of pronunciation issues by UNGEGN

Although the standardization objectives of UNGEGN cannot be realistically extended to standardization of names pronunciation, the issue of pronunciation is both essential and important enough from a user's point of view to justify incorporation in the expert group's supportive activities. The involvement of the working groups in the subject of pronunciation could (and partly already does) focus on the following aspects:

1. **Toponymic Data Files and Gazetteers**
   Although the entity Name would be logically defined as a string of phonemes (a sound sequence) that may be expressed by one or more strings of graphic characters, the fact that written name forms are the main focus of names standardization as promoted by UNGEGN makes it an obvious choice to include pronunciation among the attributes of the written name in a toponymic database. In this structure, it should be noted that there may be a '1 to n' relationship between written name and pronunciation. In many cases, it might be worthwhile to facilitate the inclusion of both a (preferred) standard pronunciation and a local pronunciation: for instance New Orleans (/nuː ɔːrˈliːnz/, locally /nuː ˈɔːrlɪnz/), Toulouse (/tu.luz/, locally /tuˈluzə/), Toronto (/təˈrɒntoʊ/, locally also /ˈtrɒnoʊ/). It should be reminded that it may occasionally be debatable and/or politically sensitive to make an implicit statement about what should be considered standard and preferred. In any case, pronunciation is never merely dependent upon writing, but always at least upon a combination of writing and language: both in English and in Dutch, the writing of Amsterdam is the same but its pronunciation differs (/ˈæmstərdæm/ in English, /ˈɑmstərˈdɑm/ in Dutch), while Berlin is pronounced /ˈbərlɪn/ in English but /bɛɐ̯ˈliːn/ in German. As local pronunciation may actually belong to another language, confusion in this respect is imminent: Barcelona is pronounced /barθeˈlona/ in Spanish, but locally and in Catalan /bərsəˈlonə/. Likewise, in spite of its very obvious Spanish origin, the name of America's second largest city Los Angeles is pronounced /lɑˈsændʒələs/ by English-speakers but /losˈaŋxeles/ by its 1.5 million Spanish speaking inhabitants; similar situations occur all over the State of California and other parts of the southwestern United States, abounding as they are both in geographical names of Spanish origin and in Spanish-speaking inhabitants. As neither the United States federally nor the State of California constitutionally defined an official language, the location of the objects these names refer to does not automatically imply the language they belong to. Moreover, in cases where the connection between writing and pronunciation of names has essentially been lost, like in the English of the UK, the need to apply another writing to the dialectal form of a name will not be perceived, making it impossible to assign a language to a written name: the language of Newcastle (upon Tyne) is 'English English' (Received Pronunciation) when pronounced /ˌnjuːkɑːsəl/, but Northern (Geordie) dialect when pronounced /njuːˌkæsəl/. On the other hand, the dialectal

form of the name of the southern Dutch city of Maastricht (Dutch pronunciation: /maːˈstrɪxt/) is written Mestreech, to approximate (according, although debatably so, to Dutch writing standards) the local pronunciation /məˈstʁeːç/.

In a digital database, pronunciation may be included as a character string (IPA or any language-specific notation) and/or as a recorded or automatically generated sound file. The advantage of IPA notation is that, unlike sound, it does not depend on the trained ear: as argued above, the sounds one recognizes are to a certain degree dependent on the phonemes one is familiar with. Disadvantages are the unfamiliarity of the general public with the sound values corresponding with the IPA characters, and the fact that there is no assured unity or consensus (yet) concerning the proper way to represent every single pronounced name in IPA. Automatically generated sound files depend on the character sequences used as input to the sound generation application, so they share this latter disadvantage with IPA. An additional disadvantage of recorded sound files is the influence of the voice and the native language of the owner of the recorded voice. Language-specific complications may occur: in Chukchi, an indigenous language of Eastern Siberia written in a Cyrillic script, the proper pronunciation for the character combination кр is reportedly /rk/ when pronounced by men, but /ʦʦ/ when the speaker is a woman[ii].
Pronunciation details provide in either way a useful addition to serve the needs of those needing to communicate the names in speech (school teachers, news readers and reporters etc.). When including pronunciation details in a toponymic database considered to have official status, they should however be accompanied by a disclaimer explaining the exact status of the pronunciation details offered – unless these are official as well.

2. Toponymic Terminology

The Glossary of Terms for the Standardization of Geographical Names published in 2002 (ST/ESA/STAT/SER.M/85) does include terms required to discuss the relationship between writing and pronunciation.

3. Romanization Systems

Although, as demonstrated above, even in originally Roman-written names the relationship between writing and pronunciation is never self-evident, it was the observed disadvantage of transliteration as opposed to direct transcription of sounds to graphics – the obvious difficulty to interpret sound values of transliteration alphabet letters to those unfamiliar with the language involved – that made the 8th UNCSGN issue its resolution recommending the institution of the Working Group on Pronunciation in 2002. In the Romanization tables listed in the Technical reference manual for the standardization of geographical names (ST/ESA/STAT/SER.M/87), details of pronunciation were thus far nevertheless explicitly left out because of the difficulty to find complete and authoritative (official) sources for all the languages presented. When such sources become available, the sound values of single characters might be explained using IPA in print or pdf, or hyperlinks to exemplary sound files offered on the Romanization website. As the tables are exclusively meant for transliteration from non-Roman to Roman script and not for direct transcription of names from unwritten languages, however, such additional functionality in a way overshoots the target of the tables. Also, as was made clear above, in many a case there is no simple or unambiguous correspondence between a single written character and the pronunciation of a phoneme.

4. **Training Courses**

   Instructions on the recording of pronunciation details when collecting geographical names in the field are essential to the toponymic component of the base mapping process. Language and situation-specific directives should be drafted up and practiced using the expertise of local language specialists. Properly recorded pronunciation details are required during the office treatment of the names, when recommendations on the correct or most appropriate writing are being prepared for official approval.

5. **Exonyms**

   If pronunciation is to be included in the discussion about the discrimination between exonyms and endonyms, the definition of an exonym should be broadened from the written form of the name to its pronounced original. This would not be without consequences for the prevailing lists of exonyms: in spite of its coincidentally identical writing, English /ˈæmstərdæm/ (Amsterdam) would then for instance be considered an exonym of Dutch /ˈɑmstərˈdɑm/, and English /ˈbərlɪn/ (Berlin) an exonym of German /bɛɐ̯ˈliːn/. Although theoretically defendable, this would not serve the interest of (written) names standardization, and might thus not be advisable.

6. **Geographical Names as Cultural Heritage**

   For the cultural heritage content of a name, its pronunciation might be considered to weigh as heavily as its written form. As there is a tendency for the pronunciation of names to change faster than their writing, especially where the latter does not follow spelling reforms but keeps reflecting a pronunciation that no longer exists, written names sometimes expose a part of an object's heritage that its pronounced equivalent no longer reveals. On the other hand, there are many historical examples of name writings that were at one point of time adapted to a reinterpretation of the name's meaning following language change (a process called folk etymology). The mutual influence of writing and pronunciation is a central theme when it comes to the study of a name's etymology in order to reconstruct its cultural background.

7. **Task Team for Africa**

   In order to support the collection and standardization of names belonging to thus far unwritten languages of Africa – as well as other multilingual parts of the world that fell under foreign domination until recent times – it is important to map the phonemes of the languages concerned and agree on a way to systematically transcribe them. If no written name is available yet, pronunciation is obviously all we can fall back on. The experience of the Working Group on Training Courses may prove useful here.

8. **Toponymic Guidelines**

   Especially for languages where the correspondence between writing and pronunciation essentially remained systematic, toponymic guidelines can provide valuable assistance to those required to pronounce a name. As in many situations there are at least some archaic remains among the written names, or names of which the pronunciation does not follow the contemporary rules for reasons unknown, listings of the general rules for all languages in official use within the territory covered by the guidelines should be augmented by a list of the most commonly encountered exceptions, i.e. irregularly pronounced names of the most prominent geographical objects. This will only be practicable for languages where a fixed set of rules is generally followed: as demonstrated before, English is not one of them.

A quick survey of Toponymic Guidelines recently published online discloses the following:

Country	Year of publication/ latest update	General information on pronunciation	Standard pronunciation	Standard accentuation	Exceptions	Minority language/ dialect pronunciation
Austria	2012	x	x	-	-	x
Belgium	2009	-	-	-	-	-
Chile	2007	-	-	-	-	-
Czech Republic	2007	x	x	-	-	-
Denmark	2012	x	x	-	-	-
Estonia	2012	x	x	x	-	x
Germany	2010	-	-	x	x	-
Netherlands	2012	x	x	x	-	x
Ukraine	2012	x	x	(none)	-	-
Republic of Korea	2012	x	x	-	-	(none)
Hungary	2012	x	x	-	-	-
Poland	2010	x	x	x	-	x
United Kingdom	2009	x	(none)	(none)	(not listable)	(references)
Norway	2014	x	x	x	-	x
Finland	2014	x	x	x	-	x

Whereas some of the Guidelines provide quite extensive information on standard pronunciation details of the major official language, regional minority languages are typically treated more summarily, and pronunciation details are mostly omitted. Notable exceptions to general pronunciation rules were only incidentally listed; it would be recommendable to mention at least whether they occur, which in some cases might not be the case.

Amongst the pronunciation details, accentuation deserves special mention. Although in some languages it does not play a very prominent role, in many it is phonemic. Where standard accentuation rules apply, it is helpful when these are explained. They might be reflected in writing, either mandatory or optional, and either universally or just in case of exceptions to the general rule (e.g. Spanish): this too is helpful to know.

9. Pronunciation

The Working Group on Pronunciation, originally primarily instituted to produce pronunciation guides for sets of geographical names, facilitates the sharing of insights and experiences regarding the handling of all thinkable pronunciation aspects of geographical names between the UN member states. As language situations and the relationship between pronunciation and writing conventions vary widely between nations, the complexity and the exact nature of the problems faced as well as the direction into which solutions are to be sought and found differ accordingly. Parallels nevertheless occur worldwide, and where they do the wheel doesn't need to be invented twice: the opportunity to learn from each other should never be left unused.

Apart from being an obvious platform to share expertise in the recording, storing and disseminating of pronunciation details on a national scale, the Working Group is also mandated to provide support for the collection of audio files for the international UNGEGN Geographical Names Database.

Future direction and short-term goals for the Working Group – a proposal

In many cases the subject of pronunciation may be considered slippery or even politically sensitive, but in the context of standardization of the writing of names it is, as recognized by the UN Conference on the Standardization of Geographical Names – cf. resolution 7 of the 3rd (1977) and resolution 11 of the 8th UNCSGN (2002), certainly not irrelevant. By gathering accounts of practical experience and bringing together relevant expertise from the greatest possible variety of national situations concerning written vs. pronounced geographical names, as has been done before on a limited scale, the UNGEGN Working Group on Pronunciation can live up to a wide interpretation of its mandate to offer guidance in the treatment of pronunciation aspects of names. As the technical aspects of the storage and dissemination of digital names information, to which for instance audio files representing pronunciation may be reckoned too, are covered by the WG on Toponymic Data Files and Gazetteers, a logical focus of the Working Group on Pronunciation might be the theoretical underpinning of all pronunciation-related activities as well as their coordination on the input and output side, and the production of guidelines and directives to serve the needs of geographical names authorities worldwide. At short notice the Working Group should yield a document stimulating national names authorities to accept the challenge to capture, store and disseminate any kind of names pronunciation detail, by providing widely applicable best practice advice. This Best Practice document should ideally be presentable at the 11th United Nations Conference on the Standardization of Geographical Names scheduled for 2017.

[i] Prepared by Tjeerd Tichelaar (Netherlands), Convenor, Working Group on Pronunciation

[ii] The languages of the world. K. Katzner, London 1975