

# FIELD COLLECTION OF NAMES IN MULTI-LINGUAL AND MULTI-CULTURAL AREAS

Tjeerd TICHELAAR

In the 1980s, the Indonesian national mapping agency Bakosurtanal set up an extensive base mapping project, to provide new topographic maps at scales of 1/50,000 and 1/25,000 needed to sustain the country's accelerating socio-economic development. The methodology followed in Indonesia, that was analysed and fine-tuned during a 1989 workshop held in the West-Javanese mountain resort Cipanas under the auspices of the UNGEGN, may serve as an example for field collection endeavours in other multi-lingual, multi-cultural areas.

To record geographical names in the field, Bakosurtanal prepared a standard form specifying exactly which names data had to be collected. The form consisted of three different parts. The first contained previously determined object identification data: a map sheet number, an object code (a unique code with which the object was marked on the map), a feature code specifying the category the object belonged to ('settlement', 'river', 'mountain top', etc.), grid reference, and, where already available, a recommended spelling of the name following document study preceding the field survey. The latter was for instance included in cases where the object possessed a *de facto* standardized name because it was, uniformly spelled, part of important national data files.

The second part of the form specified the information that had to be collected by the toponymist in the field. The third part, then, had to be completed in the base camp after finishing the day's survey: it included, for instance, locational object specifications like coordinates, dimensions, elevation and administrative affiliation. Today, most of these data could – using GPS and modern surveying instruments – more efficiently be added right on the spot in the field.

In the field, the toponymists had to record information on the name 'in locally official use', as well as any possible alternative and formerly used names. *In official use* meant: correct according to the local authorities. A name would only become an *official name* after it had been endorsed as the official standard by the national names authority, that was to be constituted following the conclusion of the 1989 workshop.

One of the tasks of the toponymist was to investigate whether a locally official name was in actual accordance with the name the object was known by among the local

population. Secondly, he had to check whether there was an unambiguous literal relationship between the meaning of the name and the current condition of the object. Further, he had to specify as exactly as possible the geographic range and delimitation of the named object.

Of the name itself, the toponymist had to record (1) the *writing* (the graphic quality) (2) the *pronunciation* (phonetic quality), (3) if traceable, the *meaning* (semantic quality), and (4) the *language* it was considered to belong to.

## Object classification

The feature code of an object – what generic class does the named object belong to – is sometimes less obvious and clearcut than one might think, and may have to be adjusted after the field check. Although the toponymist should be open-minded enough not to ignore any named phenomenon he encounters in the field, regardless of whether the occurrence of its generic class has been foreseen, a list of object classes and the feature codes assigned to them should be prepared before setting off into the field. As this is strongly region-dependent, a geographic expert on the region under survey may assist in composing this list.

A study of generic terms occurring in known geographic names also helps in shedding light on which generic distinctions are made in a certain region. A list prepared for the field-work in the Sundanese area of West-Java, for instance, included among many others the following generic terms:

**angkron** excavation in a river bank

**babakan** new settlement

**bantar** 1. shallow part of a river with rapids; 2. occasionally flooded river bank covered with pebbles or boulders

**baru** newly planted cropland – at the time of the *Cultuurstelsel* in the Netherlands Indies, an obligatory coffee planting

**beber** part of a river with little or no current

**bojong** peninsula in a curve (*meander*) of a river

**cadas** stony sediment or solid rock in a river bed

**cukang** small bridge made of bamboo

**dungus** dense forest

**karees** stony shallow in a river  
**lengkong** creek or inlet in a river or lake  
**leuwi** deep part of a river  
**nambo** former river bed  
**pabuaran** temporary shelter, for instance for forest or construction workers  
**panenjoan** viewpoint  
**penggalengan** complex of little dikes around wet rice fields (*sawahs*)  
**pangragajian** sawmill  
**pangumbahan** wash-place  
**parapatan** crossroads  
**parigi** 1. deep ditch; 2. dead end tributary near a river mouth  
**parung** rapids with steep fall

As this list was compiled through an analysis of names of settlements (villages, hamlets) on existing topographic map sheets, it is not to say that the above object type distinctions are actually in current use. What it does imply, is that at the time these names were assigned, distinctions like those between the different parts of a river's course – *angkron*, *beber*, *karees*, *leuwi*, *parung* etc. – were by the settlers or local population considered meaningful, and not just irrelevant detail.

To make it easier on the toponymist to assign any named object to a distinguishable object class, a hierarchical generic system might be devised. In the Indonesian base mapping project, the system contained the following classes:

...

On the base of an analysis of generic terms like the Sundanese example above, this classification might of course be elaborated to greater detail:

...

Not to overshoot the practical target of the toponymist, each hierarchical layer should be acceptable as an object class itself as well, and should be endowed with its own feature code.

Some examples of language-, country- or region-dependent generic distinctions in other parts of the world are:

- in French: *fleuve* vs. *rivière* (a river flowing into a sea or into another river)
- in Finnish: *vaara* vs. *tunturi* (a wood-covered mountain vs. a bald mountain)
- in North African Arabic: *erg* vs. *sarir* (a sandy desert vs. a rocky desert)

- in English: *marsh* vs. *swamp* (open wet ground vs. wet ground with trees)
- in Dutch: *sas* vs. *zijk* (lift-lock vs. discharging-sluice)
- in German: *kugel* or *kuppe* vs. *spitze* (a dome-shaped vs. a pointed mountain top)
- in English: *convent* vs. *monastery* (an abbey for nuns or monks)

Obviously the above list could be endless, and even to make it more or less exhaustive within one language area would require quite some study. As a general rule, the most dominant elements of both the physical and the cultural landscape show in any region the largest conceptual and onomastic differentiation. The Greenlanders distinguish many different types of snow, while toponyms in the Netherlands are full of swampy generics (*broek*, *hol*, *moer*, *veen*), mountainous France knows numerous models of mountain tops (*aiguille*, *ballon*, *butte*, *colline*, *crête*, *dent*, *dôme*, *mont*, *roche*), Norway all kinds of bays (*anger*, *bogen*, *fjord*, *våg*, *vik*), the Arabic and Berber world a great variety of deserts (*edeyen*, *erg*, *hammad*, *ouar*, *sahra*, *sarir*, *ténéré*, *tanezrouft*) and, toponymically dominant because of their scarcity, sources of water (*aïn*, *bir*, *hafar*, *hammad*, *hassi*, *umm*).

A mere listing of the generics encountered does not suffice for the toponymist faced with the task to classify the named objects he records. Before becoming applicable to this end, the listing will have to be studied by a specialist, or team of specialists, to set apart purely dialectal variations (exact equivalents or synonyms) from the conceptually (semantically) different variants.

It may not prove possible to establish all generic distinctions in a discrete, objectively manner. Whether a stream should be considered a 'river' or a 'brook' – depending on its size, which is both a matter of judgement and may not be stable in time or in space – is in many cases unclear. In flat, low-lying areas like the Netherlands, where most of the streams show little fall and current, and have typically been canalized along large stretches of their course, the distinction between a 'stream' and a 'canal' is never that clear. The characteristics of linear features, by their name defined as a single object, may considerably differ along their course: a torrent may turn into a river. Dimensional differences with generic consequences – hills vs. mountains, hamlets and villages (where no official status definitions are involved), gulfs and bays – cannot be unequivocally established, as they are based on comparison. A mountain (*berg*) in the Netherlands may be too low even to be considered a hill in an Alpine country.

To relieve the toponymist from the burden of this kind of subjective decisions, the generic classification should actually not include such distinctions. Both hills and mountains should be one and the same class that might be titled for instance ‘named elevation points’, just as rivers, brooks and torrents should better be combined in a common category like ‘streams’, or something like ‘linear hydrographic features of natural origin’ (as opposed to canals).

### Writing and pronunciation

As far as the writing is concerned, all different varieties encountered – for instance in documents, at landmarks, on mile-posts and signboards, and in written statements of respondents – have to be recorded. After analyzing these and the other information collected the field, an eventual choice may be made which writing should in the end be recommended as the (primary) standard.

Something of a phonetic notation will be required to lay down the pronunciation of the name in all cases where the language does not make use of a purely phonetic writing system – i.e., where there is no one-to-one relationship between letters (or combinations of letters) and phonemes. Languages like Finnish and Italian are highly phonetic; among the least phonetic languages we should count the English language. In English, even a four-letter combination like ‘ough’ can be pronounced in many different ways: cf. through, although, thought and rough. In many languages, inelarity of pronunciation is limited to a couple of letters. In Dutch, the letter e may represent three different phonemes; all of them occur in the place-name Enschede. To Dutch people, the pronunciation of the first E is regular, in that it follows the phonological rules of the Dutch language; the second and the third e we simply need to know in order to pronounce them correctly. To note the pronunciation (but not in the actual spelling), it would suffice to use diacritical marks to identify the different sounds (é, è, e).

To record the pronunciation for toponymic purposes, it suffices to note down those sounds that might lead to confusion. As this is obviously language-dependent, the toponymist needs to be well aware of the phonological means of expression used by the language(s) involved. In other words, he needs to know which variations in sound are (potentially) semantically discriminatory, or *meaningful*. This requires, in the preparatory stage, assistance of expert linguists. In English, for instance, the pronunciation of the words *thong* (‘belt’) and *tongue* differs only

in the articulation of the initial dental sound, the first being fricative, the second occlusive. In another language, such differences in articulation might be meaningless personal variations in speech. In Finnish, the difference in length of the first vowel in *vara* (‘reserve’) and *vaara* (‘wooded hill’) is meaningful, while it would not even be noticed by a speaker of a language as, say, Indonesian, that does not make use of vowel length to discriminate between words. The Finnish listener, on the other hand, might not perceive the difference in vocal sound between Dutch *ei* (‘egg’) and an *ui* (‘onion’). Likewise, the four different tonal accents with which the Chinese word *ma* may be pronounced, defining whether it means ‘hemp’, ‘horse’, ‘mother’ or ‘scold’, would definitely be too subtle not to elude the observation of any listener not being trained in the use of a tonal language like Chinese. The phonological variables employed may differ between otherwise closely related languages; in the Indonesian case, for instance, the Javanese language knows two semantically discriminating d-sounds (represented in the old Javanese alphabet by two different letters), whereas neighbouring Sundanese does not. To discriminate between these sounds, the toponymist might make use of either a conventional (a dot beneath the d) or a self-invented notation.

Variation of the sound volume, called *accentuation*, is sometimes meaningful in languages making use of a so-called free centralized accent. In Dutch, the word *voorkomen* means ‘to occur’ if the accent falls on the first syllable, but ‘to prevent’ if it falls on the second one. Sometimes accentuation combines with a morphological change to create a different meaning: in modern Greek, *o ánthropos* with the accent (stress) on *á*, means ‘man’ (in the meaning of ‘mankind’ – English omits the definite article here), while *tou anthrópou*, with the accent on *ó*, means ‘of man’.

A free accent may to a certain extent be subjective, as illustrated by the usual contextless accentuation of Dutch place-name *Amsterdam* on the third syllable (*dam*), against an intentional accentuation of the first syllable of both place-names in the phrase *Amsterdam en Rotterdam* (‘Amsterdam and Rotterdam’) to put emphasis on the distinction between the two.

Where a language employs a fixed accent, the change in sound volume is not used for semantic discrimination, but typically for the sake of intelligibility of speech. In Finnish, Czech, and Bengali, the accent always falls on the first syllable of each word, thus making it immediately clear where one word ends and another starts (the very reason why even foreign graduates majoring in the

French language still tend to find it difficult to follow a spoken French conversation).

Accentuation, even where it is semantically discriminatory, is rarely graphically represented in written language. Where it is, accents are usually shown by diacritical marks (also called *accents*). In Spanish and Italian orthography, for instance, acute resp. grave accents are used where the word accent is in contravention of the general rules, although in current Italian practice this is only done where it concerns the final syllable (the grave accent on e and o then simultaneously functions to indicate an 'open' vocal sound instead of the usual 'closed' one). Especially in Italian toponyms it is, under the influence of the regional languages/dialects the names spring from, not uncommon: for educational purposes school atlases sometimes even add acute accents to all irregularly accentuated Italian toponyms. Anyway, where accentuation is considered to be an essential aspect of pronunciation, it would be a waste not to take note of it at the event of a systematic toponymic field-work.

### Meaning

Semantic information obtained from interviews should always be verified with linguistic expertise before being accepted as true. *Popular etymology*, explanation of the origins of words or names without taking in consideration or even being aware of the phonological rules governing sound shifting, is a favourite pastime of all times and places. As demonstrated earlier by the example of the place-name York, it may even affect the orthography of a name. The following example from Indonesia shows that it is a pitfall that may trap the educated as well as the less educated.

In the Indonesian province of West-Java, the most important native language is Sundanese. Almost all Sundanese river names and more than a quarter of all settlement names – the settlements then typically developed on the banks of homonymous streams – follow a fixed pattern: they consist of a prefix *Ci-* followed either by an adjective or the name of a tree, a plant, or an animal. On the north coast of West Java lies the capital of an old sultanate, that in colonial days used to be called 'Cheribon'. As the town might be dating from the late Middle Ages, and has for long been known as a typical 'melting pot', a trading port where Javanese, Sundanese, Chinese and Europeans mingled, the name is thought to have derived from a Sanskrit word carrying exactly this connotation. In analogy with the well-known Sundanese pattern, the

local population however became used to refer to the town as 'Cirebon' (*rebon* is a kind of shrimp), even though it seemed not to be located on a stream carrying this name. After decolonisation, this version replaced 'colonial' Cheribon as the official name of the town.

### Language

The language(s) spoken in the area to be surveyed should be known, ideally also mapped, before going to the field. The toponymist should either master the local language, or be accompanied by a fluent interpreter. When conducting an interview or taking note of names information in another way, the peculiarities of the languages in question should be carefully taken into account. As every language has its own phonetic, morphological and grammatical characteristics, a common checklist cannot be presented. Some important aspects of language that one might be confronted with are mentioned below.

### Gender

In many languages, words like nouns, adjectives, articles, and also names may have a gender. In Romance languages for instance, names may be masculine or feminine. It is important to know the gender, because this governs the form adjectives and articles take in case the name is used syntactically. In French, for instance, the country name *France* is feminine (taking the article *la*; 'of France' would thus be translated 'de la France'); *Brésil* (Brazil) on the other hand is feminine; it is therefore *Plateau du Brésil* instead of *de la Brésil*. The river *Rhône* is masculine, so it is *Côtes du Rhône*, but the *Seine* is feminine, so it was the *Département de la Seine*. As *Guyane* and *Hollande* are feminine, it is *Guyane française* and *Hollande septentrionale* (with *e*); but masculine *Brabant* makes *Brabant septentrional*.

In Welsh, the first letter of female nouns (including names) are subject to consonant mutation when preceded by the definite article *Yr...*

### Cases

In some languages, names and other nouns are subject to morphological changes because of declination (Indo-European languages like, for instance, Russian) or the addition of affixes and infixes (synthesizing languages like Uralic Finnish) depending on their *case*. The basic form of the name would then be the nominative case. In

Finnish, for instance, the complete name of one of the provinces is *Uudenmaan Lääni*. *Uudenmaan*, however, is in the genitive case ('the Province of ...'); the name of the province (nominative case) is actually *Uusimaa*. *Uusimaa* itself consists of a generic element *maa* ('land') and a specific element, in this case the adjective *uusi* ('new'). To make it genitive, both elements receive an affix *-(e)n*, which in the case of *uusi* causes a consonantal mutation producing *uuden*. Someone ignorant of this aspect of the Finnish language would be in risk of drawing a wrong conclusion, and, knowing that *Lääni* means Province, suggest that the name of the province is *Uudenmaan*.

Likewise, the Russian *ostrov Karla-Aleksandra* is not an island (*ostrov*) named *Karla-Aleksandra*: the island is named after (male) Karl-Aleksandr.

### *Sociolinguistic peculiarities*

Sometimes the sounds of a language, or even the words themselves, are dependent on who is speaking or who is being spoken to.

The Paleo-Asiatic Chukchi language, spoken in the easternmost peninsula of the Russian Federation, contains a phoneme that is pronounced *k* by men and *ts* by women, and another phoneme pronounced *rk* by men and *tsts* by women. The word for 'walrus' is thus pronounced *kyrky* by men, and *tsytstsy* by women. Needless to say that this is a very useful piece of information for the toponymic interviewer operating in Chukchi territory.

In some Indonesian languages, like Javanese, Balinese, and Sundanese, completely different language versions are used depending on the mutual (social-hierarchical) relationship between the parties in discussion. In Javanese, there are five such language 'levels' called (from high to low) *kråmå inggil*, *kråmå biasa*, *madya*,

*ngoko madya*, and *ngoko biasa*; in Balinese there are even seven. The respect enjoyed by the visiting topographers, as representants of the Netherlands Indies colonial government, from the side of the Javanese villagers, made these villagers address their interviewers in the higher *kråmå* version. They would thus mention place names like *Mambeng*, *Pelujengan* and *Pengangsalan*. According to guidelines published by the Topographic Service in 1906, however, geographic names should be put on the map in their *ngoko* forms, which are for the above mentioned places *Malang*, *Besuki* and *Pekalongan*.

An anecdote in David Crystal's Cambridge Encyclopedia of Language points at another situation that might puzzle the toponymist. It tells about an ethnic group native to Columbia, the Vaupés Indians. Even though the group is numerically small – totalling about 5,000 in the early 1960s – they are subdivided into more than 20 tribes, each having their own language. Although they are ethnically quite homogeneous and do interact a lot – using a commonly mastered *lingua franca* called Tukano – they carefully keep their tribal languages apart, and even take care to remain mutually unintelligible. This they do for reasons of identity. The fact that their separate language actually *is* their identity – you are Bará if you speak Bará – guarantees the maintenance of the language as long as the tribes survive: as the Vaupés Indians may only marry outside their own tribe (read: 'family'), i.e. with people not sharing their own language, it is of vital importance that they will never adopt each other's language. In these circumstances, all Vaupés reportedly learn at least three languages: the language of their own tribe, the *lingua franca* Tukana, and the language of whoever they will be married to. For toponymy, it means that some places have different names in up to 20 languages, and that the toponymist has to be very certain of not only the ethnic background, but also the tribal affiliation of his informant.