Writing unwritten languages
— a guide to the process —

- Working Paper -

Drafted by
Clinton Robinson
with Karl Gadelii

October 2003
Languages and writing

1.1 Introduction

Language issues are at the heart of quality education for all. It is essential that children in school and adults in community have the opportunity to use their own languages to engage with the world around them – orally and in written form. Many minority and indigenous people groups do not have that opportunity – quite simply because their language is not written down.

It is a situation which is changing, though only slowly. The population of Papua New Guinea, for example, speaks around 800 different languages, and gradually more and more of them are being used in education. Children in primary school and adults in literacy groups have the chance to gain literacy skills in their own languages – and through that experience acquire literacy in other languages as well. That takes books and libraries, adequate teaching, literature and a dynamic literate environment – all in the local language. With the commitment of local communities all that is beginning to develop… but it starts with developing an adequate writing system.

These web pages explore what it takes to write unwritten languages – one way of offering new opportunities of expression and learning to the world’s linguistic minorities and indigenous peoples.

1.2 Languages – written and unwritten

There are around 6800 languages in the world today (Grimes 2000), but they differ widely in how many speakers each one has. Some languages are spoken only by those who are mother tongue speakers of it, while others are widely spoken as additional languages (second, third … languages). Eleven languages each have 100 million mother tongue speakers or more and they account for 51 percent of the world’s population: Mandarin Chinese, English, Spanish, Arabic (all varieties), Hindi, Portuguese, Bengali, Russian, Japanese, French and German (Skutnabb-Kangas 2000). These are, of course, all written and all have vast stores of literature. A further 200 or so languages account for another 44 percent of the world’s population. This means that over 6000 languages are spoken by groups of less than a million people each, with some languages of less than 1000 speakers. In other terms, 95 percent of the world’s languages are spoken by only 5 percent of its population. The world distribution of languages is as follows, according to the Ethnologue (Grimes 2000):

![Diagram of language distribution](image)

How many of these are written? It is extremely difficult to estimate how many written and unwritten languages there are in the world, and there is no established source of information. The difficulty in counting comes in part from a lack of information of what
is happening on the ground. The world currently has no systematic way to collect data on the number of communities which are developing their languages, what stage they have reached, whether existing writing systems are actually used, or whether attempts have been made to develop writing systems that are not in use. The *Ethnologue* notes sporadically whether a language 'has an orthography' or 'has an official orthography', but does not present information on writing systems for each language.

**1.2.1 Minorities and ‘small’ language groups**

Africa, Asia and the Pacific are the regions where there are many small language groups, although there is also a large number of small indigenous groups in Latin America. Most of these small languages are not written. The same is true of some larger groups in Africa and Asia, numbering in a few cases more than one million speakers. While some groups may not wish to see their language in written form, for many communities lack of a writing system is yet another factor of marginalisation, often compounding others such as:

- small population numbers
- minority facing a majority
- ‘remote’ location (from a metropolitan perspective)
- economically poor
- low resource base
- politically without voice
- socially marginalised or stigmatised
- little access to quality social services, such as education and healthcare

Developing a writing system will not by itself change these realities, but may interact with other factors to increase opportunities, for instance:

- a greater chance of literacy, and so education and opportunities for economic development;
- increased access to the learning of additional languages;
- opportunity for cultural expression and wider communication of cultural values and particularities;
- increased cultural and linguistic self-confidence and thus greater security in one’s own identity;
- appreciation by others of the unique richness of the language;
- the option to use the language in the electronic media.

Developing a writing system for an unwritten language is perfectly feasible – such work has gone on for centuries. Linguistic tools have improved and speeded up the process, and there is today a greater understanding of the influence of the social context, and of the cultural impact of writing a language down. UNESCO is convinced that it is the right of every language community to use their own language in written form, if they so desire.

**1.3 Why does it matter?**

For those who have mastered major international languages – and are therefore able to access web pages like this one – it may seem unnecessary or even irrelevant to spend the effort developing writing systems for unwritten languages. They are by definition small and peripheral, spoken by groups of people who will very likely need to learn other languages in addition to their own. Many arguments are advanced to challenge the concern for the development of these languages:
- There are too many languages, particularly in some highly multilingual countries
- People should simply learn languages that are already written
- Small languages are of limited use anyway
- It would be too expensive to develop writing systems for all languages
- People don’t really want to write their own languages, they want to learn English

These arguments look at languages from the point of view of managing a problem, rather than as developing an asset. While they raise important issues which must be considered in each context, they are dismissive of the rich diversity of human speech and communication. UNESCO has documented the many international declarations and agreements (UNESCO 2003) which call for attention to developing the world’s language resources, recognising that each language is a unique instrument of communication and identity for a particular group of people and a specific culture. Developing writing systems matters because of the place each language occupies in today’s world.

1.3.1 Globalisation

The increase of globalisation has placed focus on international communication and the dominance of English and a few other major languages as vehicles for it. The growth of the electronic media, the internet and e-mail have multiplied global interaction and speeded up the transfer of information. To access these opportunities, the user must know the language of the medium. By definition this is a written language and, when communication is international, one which is understood widely across the world. Does this mean that languages as yet unwritten have no place in a globalised world? It means rather that a globalised world is a multilingual world where we all know more than one language and where special attention is given to enabling smaller languages to find their place in the electronic media. This is the thrust of UNESCO’s programme Multilingualism in cyberspace.

1.3.2 Endangered languages

Nevertheless, the economic and social forces of globalisation, and the electronic media which are their vehicle, have resulted in posing an unprecedented threat to the world’s languages. It is estimated that at least half of the 6800 languages in the world may disappear by 2100. Many of these endangered languages have no writing system or a weak written tradition, that is with few documents, literature or ongoing written production. Developing an orthography is the first step in fostering the written use of the language, although this must be accompanied also by the training of writers, the development of publishing and distribution as well as by increases in literacy. At the very least the writing down of an endangered language will preserve it for future generations, enable others to learn or re-learn it, and give opportunity for studying its structure. When an undocumented language disappears not only does a community lose its culture-specific means of communication, but the unique structures of the language are deleted from the store of human knowledge.

1.3.3 Human rights

International concern for languages, their development and use is understood within a human rights perspective. Every language presupposes a community of speakers, small or large. The use of their language as a means of expression and as a vector of education is a human right, with the implication that whatever is necessary must be done to make that a reality. The right to use the mother tongue, to receive instruction in it and through it, to express cultural identity – these human rights,
encapsulated in a number of international declarations (UNESCO 2003), cannot be fully realised without developing a written form of the language.

1.3.4 Education

Using languages in education clearly requires a written form of the language, whether for use in schools with children, or in non-formal learning opportunities for adolescents and adults. Literacy – producing and using text – is a key learning tool. UNESCO and the international community have long recognised the pedagogical advantages of beginning schooling in the mother tongue of the child, and the greater ease of acquiring literacy first in the language the learner already knows. None of this is possible without a practical orthography in the relevant language. In many parts of the world, concern for adequate and high-quality basic education drives the process of writing system development forward in an increasing number of languages.

1.3.5 Multilingualism

In education as in wider society multilingualism is increasingly the norm across the world. Not only does the majority of the world’s people live in multilingual situations, but individual daily use of more than one language is common. One of those languages is the mother tongue – the language of family, home, village and local culture. Where multilingualism is a feature of society, both oral and written communication should be possible, in accordance with local patterns of language use and communication. Where small and minority mother tongues remain unwritten they will be compared unfavourably with languages of wider communication, both in terms of their communicative power and their prestige.

1.3.6 Identity

Language is also a key marker of identity. A language does not need to be written in order to fulfil this function – indeed the oral use of language is an effective marker of boundaries between in-groups and out-groups. However, the expression of cultural identity over time and distance will require that a language be written down, all the more so in the age of digital storage and transmission of knowledge.

1.4 Meaning, sound and symbol

The earliest writing systems of the world appeared in Egypt and Asia Minor around 3000 B.C., China (roughly at the same time), and Central America (AD 1000). As Dahl (2000) points out, these civilizations had in common a fairly high level of development which evidently necessitated a writing system of some kind. Successors of the writing systems of Egypt and Asia Minor are the ones most commonly found in the world today.

Early attempts to represent language through an abstract system of marks did so by reproducing an image of the concept in question – an iconic or pictorial form. Sumerian and Mayan hieroglyphic writing are examples of such a system. The picture represented the meaning of the concept/object, not its sound in spoken form. This worked well enough for material objects, but did not suffice for representing abstract notions. Thus an alternative system gradually developed which linked the representation of speech to the sound of words – now any word that was spoken could be represented in written form through phonetic representation. All writing systems, including those based on ideographic representation include some form of phonetic representation.

Syllabic and alphabetic scripts are based on this principle – in these systems writing is an abstract symbolisation of the sounds of speech, and the meaning is derived from the reader’s knowledge of the spoken form of the language. This is not to say that the correspondence between sound and symbol is necessarily direct or straightforward, as other sections show. Ideally, in alphabetic scripts, one symbol
represents one phoneme, and each phoneme is represented by a single symbol. Spanish and English provide an interesting example of how this is NOT the case: The phoneme /k/ is represented by more than one symbol in both languages:

<table>
<thead>
<tr>
<th>Spanish</th>
<th>English (British)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/komer/</td>
<td>/ka:/</td>
</tr>
<tr>
<td>&lt;comer&gt; – to eat’</td>
<td>&lt;car&gt;</td>
</tr>
<tr>
<td>/kerer/</td>
<td>/kil/</td>
</tr>
<tr>
<td>&lt;querer&gt; - to want, like</td>
<td>&lt;kill&gt;</td>
</tr>
<tr>
<td>/kwolitt/</td>
<td>&lt;quality&gt;</td>
</tr>
</tbody>
</table>

Similarly, the symbol <c> represents more than one sound, again in both languages:

<table>
<thead>
<tr>
<th>Spanish (Castillian)</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;cerveza&gt; ‘beer’</td>
<td>/θɛrveθə/</td>
</tr>
<tr>
<td>&lt;cara&gt; ‘face’</td>
<td>/kəɾə/</td>
</tr>
<tr>
<td>&lt;acid&gt; &lt;cello&gt;</td>
<td>/ɑsid/</td>
</tr>
<tr>
<td></td>
<td>/tʃelo/</td>
</tr>
</tbody>
</table>

These contrast, for example with Finnish where
- /k/ is always and only represented by <k>, and
- <k> always and only represents /k/

In syllabic scripts, the syllables of spoken language are represented by single symbols. As Garlén (1988:154) observes, this works quite well for languages with straightforward syllable structure, for example Japanese, which only exhibits syllables with a structure of consonant+vowel (CV). It would be impossible in languages such as many Germanic and Slavic languages which have a wide range of syllable patterns, including clusters of consonants.

The following diagram, adapted from Coulmas (1989), shows how the relationship of the written form evolved with regard to the meaning and sound of language:

From concept-based to sound-based writing (after Coulmas 1989)

1. unique relation between the meaning, sound and written form of the word
2. written sign represents the sound of the word
Different writing systems

Among different scripts, distinctions are usually made between logographic, syllabic and alphabetic. The latter two may be regarded as phonological systems, whereas logographic ones are meaning-based.

2.1 Logographic scripts

Logographic, ideographic or pictographic systems utilise separate symbols for different words or morphemes, without any obvious relation to the phonetic form of the lexical element. Logographic signs tend to have an iconic origin – they resemble the object or concept they refer to. There is some evidence to suggest that this is also true of alphabets, for example the letter <A> in its original inverted version represented an ox, but many signs in a logographic language such as present-day Chinese are not strikingly iconic. Furthermore, despite a commonly held view, Chinese is not entirely logographic, but also includes reduced signs related to sounds and not concepts. It is this more complex relationship between speech and writing which prompts Coulmas (1989) to prefer the term ‘morpheme-syllable writing system’, since ‘there is no one-to-one mapping between characters and morphemes’ (ibid: 107). The underlying principle is the representation by characters and combinations of characters of simultaneous cues to both the sound and the meaning of the morpheme. Some writing systems make use of combinations, such as Japanese (see below). As Garlén (1988:153) points out, alphabetic systems also display syllabic properties (InXS, IOU,...) and logographic ones (%, §, 2,...).

2.2 Syllabic scripts

Japanese is the best-known example of a language using syllabic script. However, in addition to the syllabic writing systems called Hiragana and Katakana, Japanese also makes use of logographic writing (Chinese-inspired Kanji) as well as an alphabetic one, for example in loan-words. The Ethiopic script is also syllabic and is used to write Amharic and other Ethiopian languages. However, after the political changes in the early 1990s some language groups chose to switch over to the Roman alphabet – a complete change of system from syllabic to alphabetic. Consonantal graphemes are modified according to their combination with seven different vowels, with an extra series for labialised velar combinations:

2.3 Alphabetic scripts

The spread of different scripts are to a great extent associated with different religions. Among the alphabetic scripts, the Latin or Roman one first spread in the Roman empire and later became associated with Western European Christianity (catholic and protestant). As Dahl (2000:217) points out, most new orthographies which are being created make use of the Roman script. Among languages outside the western world who use the Latin script one may mention Kurdish (also Arabic script), Swahili, Turkish, and Vietnamese, and there are also conventions for writing Japanese with Latin letters (Romaji) as well as Chinese (Pinyin). Most African languages use Roman script.

Roman, or Latin, script is read from left to right, top to bottom, whereas other writing systems may use other conventions, such as right to left + top to bottom, and vertically top to bottom, right to left as in Chinese. It is well known that computers are not very well equipped to deal with other sequentiality than left to right and top to bottom.

2.3.1 Greek and Cyrillic

The Greek and the Cyrillic scripts resemble each other. The Greek alphabet is nowadays only used for writing Modern Greek, but was earlier used also for certain
Albanic varieties. It contains 24 or 25 letters. The Cyrillic alphabet is used for writing, among others, Russian, Mongolian, Bulgarian and Serbocroatian (actually Serbian but not Croatian). Mongolian has the largest number of graphemes and Bulgarian the smallest. Dahl (ibid.) notes that during the era of the Soviet Union, this script was imposed on communities earlier using the Latin or Arabic alphabet.

Sample of Cyrillic script - Russian

Все люди рождаются свободными и равными в своем достоинстве и правах. Они наделены разумом и совестью и должны поступать в отношении друг друга в духе братства.

2.3.2 Arabic script

Arabic script is common in countries where Islam is widespread. It can also be used for writing Persian, Urdu and sometimes Kurdish. Arabic is written horizontally from right to left, using consonantal symbols. Thus it can be said to be a mixture between a syllabic and an alphabetic script. However, vowels are marked in certain genres of writing with sub- or superscript signs. Arabic is written continuously with each letter attached to the previous one. This creates a certain amount of variation in the way of writing a certain letter depending on the particular context – in initial, medial or final position; such variation is known as allography.

[sample of Arabic script]

عندما يرقد العالم أن يتكلم، فهو يتحدث بلغة يونيكود (Unicode Conference)، الذي سيعقد في 10-12 آذار 1997 بمدينة مانشستر، ألمانيا. وسيجمع المؤتمر بين خبراء من كافة قطاعات الصناعة على الشبكة العالمية إنترينيت ويونيكود، حيث ستتم، على السعيدين الدولي والمحلي على حد سواء مناقشة سبل استخدام يونيكود في النظم القائمة وفيما يخص التطبيقات الحاسوبية، الخطوط، تصميم النصوص والحوسبة متعددة اللغات.

Aramaic is written in one of three manners, called Nestorian, Jacobite or Estrangelo. As in Arabic, Aramaic writing is consonantal and goes horizontally from right to left. Vowels are indicated by Greek letters in the Jacobite variety and by dots in the Nestorian one.

2.3.3 Indian scripts

Any traveller in India is struck immediately by the many different scripts used in various parts of the country. They represent a complex web of linguistic, religious and cultural influences, but all derive originally from the Brahmi script, dating from about the fifth century BC, and itself based most probably on Semitic roots. Indeed the
influence of this script has been felt much farther afield in the alphabets of south-east Asia. There are over 200 scripts in India, and Pattanayak (1979) notes that they are still being created, as there are many unwritten languages on the sub-continent. The Devanagari script is used to write Hindi, the largest Indian language, although the Arabic script is used for Urdu, which is mutually intelligible with Hindi. Devanagari is an alphabetic script with 48 letters, 13 vowels and 35 consonants. It is also used to write Nepali and some north Indian and other Nepali languages and is frequently employed for the writing of Indian languages which are now developing a written standard.

[Sample of Devanagari]

Other scripts with Brahmi roots are those of Punjabi, Gurmukhi and Bengali. Bengali uses a system resembling the Arabic one, which can be regarded as a mixture of an alphabetic and a syllabic script. The letters in Bengali are consonantic and are pronounced using the "default" vowel /a/. If another vowel is required, this has to be indicated specifically in writing. Bengali also has numerous ligatures, which have fused to unsegmentable graphemes.

2.3.4 East and South-East Asia

Countries under the influence of China, confessing to Confucius, use systems emanating from the Chinese writing system. However, Korean uses an alphabetic writing system with 24 symbols, 14 representing consonants and 10 vowels. Some characters are digraphic, thus ejective /p’/ is achieved by doubling the letter used to write /b/. Korean is written horizontally from left to right, but individual syllables can be compressed vertically (Korean orthography).

Thai also makes use of an alphabetic system where the number of graphemes is considerably larger than that of phonemes, which means that certain sounds are written in more than one way. This is an undesirable situation for a learner and one which should be avoided when designing new orthographies. Garlén (1988:179) mentions as an example the aspirated stop /tʰ/, which can be written in no less than six different ways. Thai has five distinctive word tones which are indicated by diacritics in writing. Vietnamese is an exception among East and Southeast Asian languages in that it uses a Roman alphabetic writing system.
10

samples of Thai

บท นั้น ได้คลังภาษาจีนถูกตีว่า ดังนี้ ในสิ่งที่ใช้เป็น Unicode เราจึงขอ
เข้าขั้นลงไปของเว็บไซต์ International Unicode Conference ครั้งที่ 10 ซึ่งจะจัดให้ที่เมื่อนั้น เมื่อสิ่ง Mainz ประกาศยกมั่น ในระหว่างวันที่ 10-12 มีนาคม ค.ศ. 1997 เข้าแต่งบัดนี้ โดยในงานประชุมดังกล่าว ทำนั้นจะมี
โอกาสได้พบกับบรรดาภูมิ传送จากธุรกิจและ Unicode บริษัท Internationalization และ Localization จากพยากรณ์ถ้วนโลก ฟรังซ์รับ
ระบบการใช้ประโยชน์จาก Unicode รวมถึงระบบปฏิบัติการและโปรแกรม
ต่างๆ พบและรูปแบบข้อความ รวมทั้งวัตถุการทั้งหมดเพียงอยู่ในภาษา
ต่างๆ

Vietnamese (Tower of Babel text)

1. Và, cả thiên hạ đều có một giọng nói và một thứ tiếng.
2. Nhưng khi ở Đông phương đổi đời, người ta gặp một dòng bằng trong xứ Si-nê-
-a, rồi ở tại đó.
3. Người này nói với người kia rằng: Hãy chúng ta hãy làm gạch và hầm trong
lửa. Lúc đó, gạch thế cho đá, còn chai thế cho hồ.
4. Lại nói rằng: Nào! chúng ta hãy xây một cái thành và dựng lên một cái tháp,
chót cao đến tận trời; ta hãy lo làm cho rạng danh, e khi phải tản lạc khắp trên
mặt đất.
5. Đức Giê-hô-va bèn ngự xuống đặng xem cái thành và tháp của con cái loài
người xây nên.
6. Đức Giê-hô-va phán rồng: Nay, chỉ có một thứ dân, cùng dòng một thứ tiếng;
và kia kia công việc chúng nó dường khởi làm; bây giờ chẳng còn chi ngăn
chúng nó làm các điều dâ quyệt định được.

The following links give samples of various scripts from around the world:
http://www.omniglot.com/writing/atoz.htm
http://people.w3.org/rishida/scripts/samples/
3 Developing a writing system

3.1 A multi-faceted issue

For those who use writing every day it seems a simple matter to have some way of representing speech and words on paper or on the computer screen. We give little thought to how those marks were designed or decided upon. We are hardly aware that the system we use evolved over time and may have been the subject of much debate or even dispute. However, developing a writing system is not a simple or straightforward matter – it concerns the place that language has in human society and so has connections with many different aspects of life: social, political and cultural. Some of the processes are technical, concerned with the structure of particular languages and the procedures for analysing them scientifically in ways that reveal their uniqueness and the systematic way in which they are organised. Such an analysis must form part of the basis of deciding how to write a language which has hitherto been used only in oral form.

The technical, linguistic basis is only part of the story. Linguists may plead for a writing system to take the structure of the language fully into account; indeed, it may be that the linguistic reality and its analysis offer solutions which are elegant and relatively simple. However, a writing system is not only a linguistic reality, it is also a social convention, to be adopted and used by a community of speakers with their particular history, social relations, political context and cultural heritage. Thus all these factors must form part of decisions about how to write a language.

3.1.1 Social relations

If a writing system is to be accepted by the community of speakers of a language, decisions must take into account social relations within the community and with other communities:

- Within the community: what is the relationship between different dialect areas of the language? How are gender relations structured and how can women’s and men’s input be organised? Which institutions use or may use the language in written form and what is their influence on how the language should be written?
- With other communities: which of the neighbouring languages has a written form and when/how was it developed? What is the attitude of neighbouring groups to developing a writing system: will they use it, feel threatened, learn from the experience and develop their own, …?

3.1.2 Political context

Deciding how to write a language is not a politically neutral enterprise. Governments, the status of the language community and other interest groups have an influence on the decisions and how they are made:

- What degree of autonomy in decision-making does the language community have, and why?
- What other interests may have to be taken into account: linguists’ views, government policies, institutional power, …?
- How flexible or constraining is government policy with regard to writing systems for the country’s languages?
- How much influence may foreign governments have, for example, former colonial powers whose languages are spoken in the country in question?
3.1.3 Cultural heritage
Writing a language may change the relationship of the speakers to their culture, to the possibilities of access to their cultural heritage and transmission of it to future generations and to the wider world:

- What attempts have already been made to write the language? By whom and for what purposes?
- What kinds of oral literature exist and what will happen to them when the language is written?
- How has local history been recorded, and how will this change if the local language is written?
- Who are the creators and guardians of local knowledge, and what will be their role once the language is written? Will they become authors and writers? Will they be marginalised? How will local knowledge be better validated once the language is written?

3.1.4 Language structure
Languages have diverse and fascinating structures which a good writing system should seek to represent. However, analysis takes place in the context of the social, political and cultural factors, and these will carry more weight ultimately in determining how a language is written. Some of the broad questions which linguistic analysis must ask are the following:

- What are the distinctions of sound which must absolutely be represented in the language to avoid confusion?
- What should be the relationship of these essential sounds differences to the way they are spelled? Should one sign represent one sound? Should each distinctive sound unit be represented by only one sign?
- What is the grammatical structure of the language and how does this influence how the language is written, for example, word breaks, elision, parts of the verb?
- What light do the structures of related languages shed on the development of a writing system?

Section 4 (hyperlink) introduces the principles involved in analysing the sounds of a language.

The following diagram illustrates how all these factors – social, political, cultural, and technical/linguistic feed into the decision-making process.
3.2 Key Elements in Developing a Writing System

**Factors**

- **Social Relations**
  - within the group
  - with other groups

- **Political context**
  - government policies
  - competing interests

- **Cultural heritage**
  - history, oral literature
  - local knowledge

- **Language structure**
  - sound system
  - grammar

**Processes**

- dialogue and negotiate
- lobby, negotiate, compromise
- recognise and validate
- raise awareness
- analyse and compare

**Outcome**

- debate, consensus-building
- decision-making

Writing system

*as much a social convention as a linguistic reality*
4 Analysing sounds

When you listen to someone speaking a language you do not know, you hear a constant stream of sounds, with pauses from time to time. Except at those pauses, it is not possible to know where one word starts or stops, nor to know which sounds belong in the same word. Indeed it will probably be difficult to identify exactly which sounds you are hearing. In developing a writing system it is crucial to know which sounds a language has and how they are put together. Once the raw data have been collected, the next step is to find out which sounds must be represented in the writing system, subject to all the non-linguistic considerations mentioned elsewhere (hyperlink to section 3.1). In order to undertake this analysis, linguistics provides some useful tools.

4.1 Identifying the sounds

The first step of analysis towards writing an unwritten language is to capture the range of sounds of which it is made up. Phonetics and the phonetic alphabet provide ways of doing this. The International Phonetic Alphabet is an inventory of symbols and marks which enables linguists to represent the range of sounds produced in human speech. Although the IPA is a powerful tool, there may still be sounds in the languages of the world that it does not cover.

A phonetic transcription can capture minute variations in speech – differences between dialects of one language, for example, or even differences between speakers. However, for a writing system these small variations will not have great significance. It would overload writers and readers if phonetic variation were to be represented in writing – a writing system should give enough information for the user to make the meaning plain, neither providing too much so that reading/writing become unnecessarily complex, nor providing too little so that words cannot be distinguished with ease. Between these two extremes lies a huge amount of debate among linguists and others! Beyond scientific linguistic debate, many non-linguistic factors enter into consideration. In short, what matters is how sounds are used in a language to differentiate meaning. Thus the raw data of phonetics must be subject to analysis – this is called phonology: discovering the sound system of a language.

4.2 Understanding the sound system

Phonological analysis aims to establish the inventory of phonemes in a language. A phoneme is an abstract unit of analysis – based on a systematic distinction which a language makes. For example, [pit] and [bit] mean different things in English, and so it follows that /p/ and /b/ are different phonemes, and should normally be represented by different graphemes. Thus contrast is a key principle of phonology – since those two words differ only in their initial sound, that sound difference must be important in the language. Without it, a speaker could not communicate meaning in English.

Contrasting sounds

In order to establish the phoneme repertoire of a language, minimal pair tests are normally used. The following are examples minimal pairs in different languages, showing contrast between consonants, vowels and tone, in different positions in the word:
4.2.1 French

<table>
<thead>
<tr>
<th>French</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>/klā/</td>
<td>‘clan’</td>
</tr>
<tr>
<td>/glā/</td>
<td>‘acorn’</td>
</tr>
</tbody>
</table>

**Nugunu** (Cameroon)

<table>
<thead>
<tr>
<th>French</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>/áncê/</td>
<td>‘leaf’</td>
</tr>
<tr>
<td>/áncâ/</td>
<td>‘shout’</td>
</tr>
</tbody>
</table>

**Isirawa** (Papua New Guinea)

<table>
<thead>
<tr>
<th>French</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>/o’meta/</td>
<td>‘robber’</td>
</tr>
<tr>
<td>/o’mita/</td>
<td>‘star’</td>
</tr>
</tbody>
</table>

4.2.2 English

<table>
<thead>
<tr>
<th>English</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ko”lt/</td>
<td>‘colt’</td>
</tr>
<tr>
<td>/ko”ld/</td>
<td>‘cold’</td>
</tr>
</tbody>
</table>

**Cashinahua** (Peru and Brazil)

<table>
<thead>
<tr>
<th>English</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>swollen hand</td>
<td>/msu/</td>
</tr>
<tr>
<td>dark, black</td>
<td>/mʃu/</td>
</tr>
</tbody>
</table>

**Yoruba** (Nigeria)

<table>
<thead>
<tr>
<th>English</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘to meet’</td>
<td>bá high tone</td>
</tr>
<tr>
<td>‘to hide’</td>
<td>bã mid tone</td>
</tr>
<tr>
<td>‘to perch’</td>
<td>bà low tone</td>
</tr>
</tbody>
</table>

However, sometimes it is not possible to find perfect minimal pairs, and one then has recourse to so-called near minimal pairs, i.e. the systematic appearance of certain sounds in certain contexts, which can be taken as an indication that different phonemes are at stake. The languages of the world exhibit great variation when it comes to phoneme inventories, stretching from languages including only a dozen phonemes (Hawaiian) to languages with almost a hundred (certain African languages) (Garlén 1988). However, some phonemes tend to be present in all languages, such as /i/, /e/, /a/, /p/, /t/, /k/, with the exception for example of certain Polynesian languages.

Another key principle in phonological analysis is complementary distribution – the occurrence of fairly similar sounds in different contexts, such that it is predictable which sound will occur in which context:

In English /p/ in initial position before a vowel is aspirated, i.e. is followed by a little puff of air – [pʰm] ‘pin’ – but is unaspirated after /s/ – [spɛn] ‘spin’. The sounds [pʰ] and [p] never make a difference in meaning in English and are conditioned by the contexts in which they occur. They are therefore allophones of the same phoneme /p/ and are represented by the same symbol or grapheme <p>.

As a rule, purely phonetic, or allophonic contrasts should not to be represented in writing. This goes for both free variation, for example, between rhotic (back of the mouth) and apical (behind the teeth) /r/ in French or Swedish, and distributionally conditioned allophones, such as [d] -> [t] word-finally in German, for example:

<Hand> /hant/ in final position, but <Hände> /hɛnd̩/ before a vowel

Phonological analysis addresses all the different kinds of sounds of language, and the next sections give examples of them: consonants, vowels, and prosodic features.
4.3 Consonants

4.3.1 Stops

Stops are produced with a short total closure of the mouth, after whose release the sound in question "explodes" (cf. the name "plosives"), and a burst of air can be felt if you hold your hand close to your mouth.

The most common stops in the languages of the world are voiceless /p t k/, which often have voiced counterparts, /b d g/. The difference between the two can be felt on the pharynx, which vibrates when the voiced but not the voiceless sounds are produced. Finnish does not have the voiced series, nor does Samoan, which in addition does not seem to display /t/. On the other hand, it has the glottal stop /ʔ/. Some languages exhibit the full series with the exception of /p/, such as Arabic, and /g/ (Dutch).

The number of stops increases if a language displays phonemically distinctive aspiration: /pʰ tʰ kʰ/. There are also cases where stops are pronounced with so-called breathy or creaky voice, thus creating phoneme distinctions. The former is represented in IPA as a subscripted dieresis (e.g. /k'/), the latter as a tilde (/k/~).

Normally, the airflow originates in the lungs, with the airflow going outwards ("egressive" sounds). However, the source of the airflow can also be glottis, in which case the stop delivered is called "ejective". Ejective stops, which are always voiceless, have phoneme status in certain languages, such as Hausa in West Africa. They are rendered as /p' t' k'/ in the IPA.

The airflow can also go inwards, in which case we talk about ingressive sounds. These can either involve in the glottis, so called implosives, or velum ("clicks"). Implosive stops are characteristic of many Bantu languages, and are represented as /ɓ ɗ ɠ/ in IPA. It is a matter of controversy in Bantu orthography how these should be written. In some Southern Bantu languages <ɓ> represents the "ordinary" stop, whereas in others it may symbolize implosive /ɓ/, in which case "ordinary" /ɓ/ may be written <bh>. In general frequency of occurrence is the determining factor. That is, if either sound is quite unusual in the language in question, it should be rendered with a complex character, whereas the simpler symbol should be reserved for the more frequent sound. A third alternative is to write both sounds in the same manner – people will still pronounce them correctly when they appear in a particular context.

Click sounds can be found in Khoisan languages in southern Africa, and have also been borrowed into certain Bantu languages. They come in various guises and are e.g. written <q> /ǃ/, <gq> /gǃ/, n’q /nǃ/.

Many languages also display prenasalized stops, rendered as /m'b, n't, n'd/, etc. in IPA. These kinds of stops have phoneme status in various African languages, and there is a debate as how to distinguish them from so called syllabic nasals. In Bantu languages we find sequences of nasal + stop which are not prenasalized stops but where the nasal is syllabic and thus constitutes a syllable of its own.

As an illustration, consider the Cisena (Mozambique) word <nkuta> 'cuckoos', which is the plural n-kuta of the singular form mi-kuta. The n-sound is thus a plural-forming prefix, and not part of a pre-nasal /'k/. It has been suggested that this syllabic nasal be written with an apostrophe, thus <n'kuta>, but this convention has also been criticized for being overly ambitious for the ordinary language user.

Finally, we find examples of doubly articulated stops, existing for example in West African languages. The most common type is the labio-velar stop, which is produced simultaneously using the lips and the velum. Some examples are <kp> and <gb>, as in the language names Kpelle and Igbo, which are rendered in IPA as /k'p/ and /g'b/.
4.3.2 Affricates

In case a stop is released as a fricative, notably a sibilant (hyperlink to 4.3.4), we talk about affricates. These can be found, for example, in the English words *church* and *jungle*. Affricates are common in many languages, including Indo-European, and are notoriously difficult to represent orthographically. The previously mentioned words illustrate the most common ways of spelling English affricates, i.e. /tʃ/ is represented as <ch> and /dʒ/ as <j>, but the former is spelled differently in words like *nature, match, question*, etc., and the latter in *edge, soldier, age, exaggerate,...*. An affricate can also be released laterally – with /l/ – or nasally, e.g. with /n/. In e.g. English, such sounds mainly occur word-medially or -finally (<bottle> /bɒtl/, <mutton> /mʌtn/ and do not have phoneme status. The situation is different in other languages. In Polish, we find nasally released affricates word-initially in cases such as <dno> 'bottom', and many Bantu languages have a number of laterally released affricates, spelled <hl, tl, dl> etc., and represented in IPA as /ɭ/, /tl/ and /dl/, respectively. However, their status as separate phonemes has been debated, and some linguists regard them as sequences of separate phonemes, much as /t/ + /l/ in English <bottle>.

4.3.3 Nasals

Nasal consonants are continuous sounds produced with a closure in the mouth directing the flow of air through the nose (for nasal vowels, see 4.4).

All languages in the world have nasals, the most common one being /n/, followed by /m/. If languages have more than two nasals, these are normally either velar /ŋ/, palatal /ɲ/ or labiodental /n̥/. Some languages have phonemically distinct voiceless nasals, such as Icelandic and Burmese. Consider, for example, the voiced nasal in Icelandic <naust> [nɔ₁ʌst] 'boat house' versus the voiceless one in <hnaus> [hɔ₁ʌs] 'lump of soil'.

4.3.4 Fricatives and sibilants

Fricatives are a class of sounds pronounced with partial closure of the speech apparatus, thus causing friction and producing a "hissing" sound.

Examples of fricatives are /f v s z ʃ θ δ h/. There are languages which do not have any fricatives at all but this is unusual and seems to exist only within a few languages in the Austronesian family.

Sometimes the sounds comprising an /s/-property are singled out as a special group called sibilants. The /s/ sound itself is the most common sibilant among the languages of the world. If languages have two sibilants, the second one tends to be /ʃ/. As Garlén (1988:41) points out, a symmetric four-way pattern, consisting of /s z ʃ θ/ is not uncommon among the world's languages.

4.3.5 Liquids, laterals and vibrants

When pronouncing liquids, the air flows through the mouth over both sides of the tongue. Vibrants are so named because one part of the mouth (e.g. tongue or uvula) vibrates against another.

The most common lateral is /l/ and the vibrant /ɾ/. As to laterals, some languages distinguish between "ordinary" dental or alveolar /l/ and palatalized /ɭ/, rendered as /ɭ/ in IPA. This phoneme is spelled in a variety of ways, such as <figlia> 'daughter' in Italian, <fill> 'son' (Catalan), <moļho> 'pier' (Portuguese), <ļjudo> 'people' (Serbo-Croatian), and <llama> 'flame' (Spanish). Further, a number of Indian languages display a retroflex lateral, rendered as a /ɭ/ in IPA. Icelandic has a voiceless
dental/alveolar lateral /l/ written <hl>, and Welsh exhibits a dental lateral fricative /l/, which is written <ll>, as in <llyn> 'lake'.

The vibrant /r/ can be pronounced in a multitude of ways, in particular at the back of the mouth (uvular) and at the front (dental/alveolar). Again, Icelandic has a voiceless vibrant /r/ written <hr>, and Czech displays a fricative vibrant, phonemically distinct from the "ordinary" one and written <ř>. In Spanish we find intervocalic variation between ordinary [r] and a "tap" variety, /ɾ/. The contrast is illustrated in a pair like <carro> 'carriage' vs. <caro> 'dear'. Finally, we find a retroflex flap, /ɾ/, called "cacuminal l" in Nordic languages. This sound has phoneme status in several languages spoken in South Asia and Africa.

### 4.3.6 Semivowels

**Semivowels are produced in the same way as vowels, but function as consonants.**

The most common semi-vowels are /j/ and /w/, based on /i/ and /u/. Often, these are allophones of a corresponding vowel, but in some languages they have phoneme status. Some varieties of /r/ in British English have been called semi-vowels, as has the labio-dental /r/ in Dutch (spelled <w>, contrasting with /r/ (<r>) and /r/ (<v>)). This sound is also present in various Bantu languages. French exhibits an unusual contrast between the semi-vowels /ʁ/, /ʁ̃/ and /ʁ̃/, as illustrated in the minimal triplet <miette> – <muelette> – <mouette>. Rumanian displays no less than four semi-vowels. Apart from the more common /j/ and /w/, it also has semi-vocalic /e/ and /o/.

### 4.4 Vowels

**Vowels are sounds produced by the uninterrupted passage of air through the mouth, with the shape of the mouth giving different vowels their particular quality.**

Many languages have five vowels, a few have three and some have seven or more. The vowels in a given language generally contrast maximally with each other. Thus, if a language has only three vowels, these tend to be /i u a/, located in the extreme corners of the vowel space. This is the vowel system we find in various Semitic languages as well as in certain languages of the Americas. Some American languages also have a four vowel system, which in addition to the former three phonemes includes /e a/ or /i/. The most common system among the languages of the world comprises five vowels, which are /i e a o u/. This picture is found in many Indo-European and African languages and also in Japanese and Basque. These five vowels also participate in a six vowel system, which in addition has /i i o a/ or /al/, which in the last case tends to cause /al/ to be replaced by /aw/. Languages displaying six vowels include various Slavic languages, Persian and Singhalese.

As to languages with seven or more vowels, these can exhibit distinctive lip rounding, leading to /y/ and /w/, or make use of central vowels such as /i/ and /a/. In yet other languages such as Bengali, Burmese, certain Romance languages and Yoruba, the peripheries of the vowel space are further used, yielding /i e a o u/. Roughly the same phenomena can be observed for languages having eight vowels. Igbo, a Niger-Congo language related to Yoruba, uses the same strategy as the latter but adds /au:/ /i e a o u/. Nine vowels are found in Estonian, Norwegian, Swedish and Thai. Swedish has "outrounded" /y/ as well as "inrounded" /u/ with a bar, and Estonian makes an unusual distinction between rounded and unrounded mid vowels. Danish has nine vowel phonemes and French no less than eleven, namely /i y e o æ a ø o u/. In addition, languages can have nasal
vowel phonemes. French displays four (or three), and Portuguese five (nasal varieties of /i e a o u/).

4.5 Prosodic features

Prospodic features are variations of pitch, stress and length which may be found on a vowel or syllable, a consonant, a word or over a whole utterance. All these features may combine over a phrase or sentence to give patterns of intonation.

Should prosodic phenomena in languages be represented in writing and, if so, how? It often happens that western linguists neglect to represent tonal phenomena, for example in writing (see below), with the excuse that the meanings of words can be inferred from the context. This may to some extent be true, but many speakers of tonal languages find it difficult to read or pronounce written language where tones are not indicated. In fact, this is part of a larger issue within orthography where we can find two opposites: linguists who do not want to mark too many distinctions with the argument that context disambiguates everything, and people who advocate the overt marking of distinctive features.

Prosody is a complicated issue which necessitates defining the particular prosodic phenomena which are of interest, and then analysing them in the stream of speech, which is not always easy. In the following, we discuss stress, length (quantity), tone and intonation.

4.5.1 Stress

Stress is the force or loudness with which particular syllables or parts of words are spoken. Languages can have fixed or floating stress. If stress is fixed, it normally falls on the first, the last or the penultimate syllable – examples of these differences are:

- Stress on the first syllable: Bengali, Korean and Finnish.
- Stress on the last syllable: French and Kurdish.
- Stress on the penultimate syllable: Polish and Kiswahili.

None of these languages mark stress in their orthographies since it would be entirely redundant to do so.

Languages with floating stress can make use of this property for distinguishing meaning. Sometimes the different stresses are marked in writing, sometimes not.

In English one distinguishes two pronunciations of <permit>, the one with initial stress being a noun and the one with final stress a verb. This difference in pronunciation is not signalled in writing.

In contrast, languages like Spanish and Portuguese use diacritics to indicate cases where the stress falls on another syllable than the expected one (which in most cases is the penultimate one): Portuguese <falaras> with penultimate stress 'you had spoken' and <falarás> 'you will speak' with stress on the last syllable and marked in the orthography.

4.5.2 Length

Length or quantity is a distinctive phonological feature in some languages but not in others. In languages which use length as a distinctive feature, this may fall on vowels or consonants or both, for example:

- Consonant length can be illustrated by Italian <papa> 'pope' and <pappa> 'gruel'.
- Vowel length occurs in Czech '<pas> 'passport' and <pás> 'belt'.
- Finnish displays both consonant and vowel length:
  <palo> 'fire' and <pallo> 'ball'
As these examples show, length can be represented orthographically by a doubled vowel or consonant symbol, or by an acute accent.

### 4.5.3 Tone

Tone contrasts are achieved by varying the basic pitch of the sounds of a language. Pitch can vary at both the word level and the utterance level – at word level it is generally known as tone, whereas at the level of a whole utterance we speak of intonation (hyperlink to Intonation). Here we are concerned with tone phenomena at the word-level. In cases of word-level pitch differences, which should be carefully distinguished from stress, we can distinguish between word accent and word tone.

Word accent is conditioned morphophonologically (i.e. by the inter-relationship between sound and form), such as Swedish /änden/ 'the duck' from <and> 'duck', vs. /ànden/ 'the spirit' from <ande> 'spirit'. Note that both words have initial stress.

#### Tone contrasts

Word tone is found in a large number of languages in Africa, Asia, Oceania and the Americas, i.e. everywhere except in Europe.

In the Nigerian language Yoruba, the sequence <ba> can mean three different things depending on the tone used: 'meet' if pronounced with high tone, 'hide' with mid tone and 'perch' with low tone (hyperlink to 2.2.2). These distinctions in word tone should be manifested orthographically, e.g. as <bá>, <bà> with a macron, and <bà>.

Vietnamese has six distinctive tones, five of which are normally represented in writing. The absence of tonal marking indicates the presence of the sixth tone. Consider the sequence "ma", which has six different meanings, indicated as follows:

- `<má>` 'cheek'
- `<mà>` 'but'
- `<má>` (with half a circle above) 'grave'
- `<mà>` 'horse'
- `<mà>` (with a dot below) 'rice plant'
- `<mà>` 'devil' (unmarked)

Tone is also used to distinguish between grammatical categories, such as verb forms. In Nugunu, a Bantu language of southern Cameroon which developed a writing system in the late 1970s, the verb /éda/ 'to go' has the following forms distinguished only by tone:

- `/éda/ 'go!' (imperative) – high-low, written as `<éda>`
- `/éda/ 'he/she goes' (present) – rising-low, written as `<ééda>`
- `/éédá/ 'he/she went' (recent past) – rising-high, written as `<ééédá>`

Thus the Nugunu writing system has one diacritic – to mark high tone. Low tone is left unmarked, and combinations of tones are spread over two vowels (or three in one or two rare cases).

In addition, some languages have floating tones, which means that the different word tones vary depending on the linguistic context. These result in extremely complex
tone patterns, which are controlled and used automatically by native speakers, but which frequently baffle outsiders and present difficult problems of analysis. This means that the representation of tone in writing in such cases is often a compromise between at least three competing demands:

- the feeling of the native speaker as to what it is appropriate or right to represent
- the linguistic analysis, often incomplete or very complex
- the need to be able to reproduce the language relatively easily on paper.

Representing tone

It is a difficult question to know how far tone should be represented in the orthography. Many, if not most, of the languages not yet written in the world are tonal, where tone distinguishes between both lexical and grammatical meaning. While linguistic analysis will show the structure of tone in a language, as we saw earlier, other questions emerge when it comes to assessing how much tone it is useful to write for the reader.

Linguists often argue that tone must be fully represented, or at least sufficiently to differentiate between words that are otherwise similar. However, writers and educators sometimes take the view that tone does not need to be represented at all, since the reader will be able to disambiguate similar words by context. Experiments in Cameroon (Bird 1997) led researchers to conclude that different questions should be asked. Instead of asking: Should an alphabetic orthography for a tone language include tone marks? a rather more subtle approach should be taken, by asking:

Which combination of tone orthography and tone teaching method is best for a given language, taking the language’s tone system and sociolinguistic setting into account? (Bird 1997: 25)

The experiment showed that it may be a viable option not to mark tone at all, in terms of the ability of readers to read fluently and with understanding, although this only applies to the language – Dschang of western Cameroon – in which the experiment was carried out. This finding was based on an analysis of reading and reading errors and took into account the fact that the deep structure of tone in a language is frequently different from what is manifested on the surface. For tone as for other aspects of a newly developed orthography, it is clear that readability as well as community ownership will be crucial in making decisions, as implied in the question quoted above.

4.5.4 Intonation

Intonation describes the meaning conveyed not through the words we speak, but by the way we say them – quickly or slowly, with a rising, falling or ‘musical’ voice, loudly or softly. The meaning of intonation patterns varies from one language to another and is not represented in the consonants, vowels or other symbols representing speech. It is however represented in writing by punctuation marks. Consider the following short sentence in English:

She’s bought a car

Without any punctuation mark at the end of the sentence, we assume it is a straightforward statement of fact. However, once punctuation is added, the sentence can take on several different intonation patterns:

She’s bought a car. a statement of fact
She’s bought a car? a question: is it true?
She’s bought a car! an exclamation: I don’t believe it
She’s bought a car?! a surprised question: it can’t possibly be true, can it?
She’s bought a car, … an innuendo: … and that’s not the end of it
She’s bought a “car”.  an implication: it may not be a car in the normal sense
… and there may be others. Other languages use intonation in different ways, and
may use words or parts of words (morphemes) to indicate things like questions,
instead of any particular change in intonation. When a language is written for the first
time, decisions have to be made about how punctuation is used, and that depends
on how far intonation is used in the language as a distinctive device, and on the
range of intonational patterns. Frequently such decisions are not consciously or
deliberately made, but conventions of punctuation use may emerge as written
material is produced and more and more people write in the language.

4.6 Matching sound to symbol

In an alphabetic writing system the most economical and efficient way of
representing speech is when one phoneme is represented by a single symbol and
each symbol represents only one phoneme – the principle of sound-symbol
faithfulness. Very few alphabets achieve this, although Finnish is a good example of
an alphabet that seeks to respect these principles. Alphabets like English (hyperlink
to Counter-example 2.3.1) and French are far removed from this principle, since the
writing systems do not closely follow the spoken form. The reasons for this are
largely historical in nature – spelling and pronunciation have evolved over time, but in
some cases never were in harmony with each other or have diverged considerably
over the centuries.

Most European languages use the Roman alphabet, but it is noticeable that each one
is different. Some use fewer than the 26 letters associated with English while others
use other vowel symbols (e.g. Danish æ, Norwegian ø), modify vowel symbols (e.g.
German ö, ä, ü, Swedish å, French é, è, ê, ù, à), indicate stress (e.g. Spanish á, ó,
etc), use different or modified consonant symbols (e.g. German ß, French ç, Spanish
ñ). This illustrates two principles:

- alphabets should be adapted to the structures of the language they
  represent;
- scripts in use by other (even closely related) languages can rarely be
  adopted without modification.

Neglect of this latter principle led to attempts to write African languages using English
and French alphabets, sometimes the same language being written in two different
ways on either side of an international border. Since neither French nor English
respect sound-symbol faithfulness, it also resulted in the same sound being
represented in many different ways within and between African languages. This is
the sort of problem which language academies and general alphabets sought to
solve.

4.6.1 A counter-example: the spelling of English

As a way of illustrating the value and advantage of sound-symbol faithfulness in
designing a writing system, the counter-example of English is shown below to
illustrate the unpredictable and multiple ways in which sounds are represented. This
evidence makes it all the more amazing that English is not only so widely spoken, but
so widely written and read as well!

English uses the following alphabet: <a b c d e f g h i j k l m n o p q r s t u v w x y z>.
Starting with a basic case, the sequence <hit> vs. <bit> can be said to be a
graphematically minimal pair, which is paralleled by the phonemically minimal pair
/hɪt/ vs. /bɪt/. However, the English spelling system is quite far from the ideal sound-
symbol faithfulness principle. The following charts are supplied by courtesy of
Phoneme | Graphematic correlates in English
---|---
p | pen, happen, shepherd
k | key, cool, school, biscuit, lock, saccharine, tobacco, cheque, lough, walk, lacquer, khaki, queen (=/kw/), queue (=/kj/)
b | back, rubber, bhang
d | day, ladder, called, could
g | gay, bigger, ghost, vague, guard
t | tea, butter, Thomas, walked, yacht, ptarmigan, doubt, fright (AmE: phthisic)

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>Graphematic correlates in English</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>p</code></td>
<td>pen, happen, shepherd</td>
</tr>
<tr>
<td><code>t</code></td>
<td>tea, butter, Thomas, walked, yacht, ptarmigan, doubt, fright (AmE: phthisic)</td>
</tr>
<tr>
<td><code>k</code></td>
<td>key, cool, school, biscuit, lock, saccharine, tobacco, cheque, lough, walk, lacquer, khaki, queen (=/kw/), queue (=/kj/)</td>
</tr>
<tr>
<td><code>b</code></td>
<td>back, rubber, bhang</td>
</tr>
<tr>
<td><code>d</code></td>
<td>day, ladder, called, could</td>
</tr>
<tr>
<td><code>g</code></td>
<td>gay, bigger, ghost, vague, guard</td>
</tr>
<tr>
<td><code>ŋ</code></td>
<td>cheer, nature, match, question, cello, Czech, righteous</td>
</tr>
<tr>
<td><code>dʒ</code></td>
<td>jump, edge, soldier, age, exaggerate, gradual, adjust, sandwich</td>
</tr>
<tr>
<td><code>m</code></td>
<td>sum, bomb, hammer, autumn, calm, drachm, phlegm, government</td>
</tr>
<tr>
<td><code>ŋ</code></td>
<td>sung, sink, tongue, handkerchief</td>
</tr>
<tr>
<td><code>f</code></td>
<td>few, coffee, cough, physics, half, often</td>
</tr>
<tr>
<td><code>θ</code></td>
<td>thing</td>
</tr>
<tr>
<td><code>s</code></td>
<td>soon, city, nice, psychology, scene, mess, schism, fasten, sword</td>
</tr>
<tr>
<td><code>ʃ</code></td>
<td>fishing, ocean, sure, chivalry, station, tissue, fascism, fuchsia, conscious, passion, tension, politician, (BrE: schedule), (AmE: nauseous), luxury (=/kʃ/)</td>
</tr>
<tr>
<td><code>h</code></td>
<td>hot, whom</td>
</tr>
<tr>
<td><code>v</code></td>
<td>view, of, Stephen, navvy</td>
</tr>
<tr>
<td><code>ð</code></td>
<td>then, bathe</td>
</tr>
<tr>
<td><code>z</code></td>
<td>zero, was, scissors, xylophone, dazzle, example (=/gz/), discern</td>
</tr>
<tr>
<td><code>ʒ</code></td>
<td>pleasure, rouge, vision, seizure, usual, luxurious (=/gzh/), (AmE: glazier)</td>
</tr>
<tr>
<td><code>l</code></td>
<td>led, ball, battle, pedal, tunnel</td>
</tr>
<tr>
<td><code>r</code></td>
<td>red, marry, wriggle, rhubarb, diarrhoea</td>
</tr>
<tr>
<td><code>j</code></td>
<td>yet, onion, Europe, beauty, use, new, halleluja, strenuous, tortilla, queue (=/kj/)</td>
</tr>
<tr>
<td><code>w</code></td>
<td>wet, one, choir (=/kw/), queen (=/kw/), when, Don Juan, patois</td>
</tr>
<tr>
<td><code>ks</code></td>
<td>box, accident, except, sticks, forks</td>
</tr>
<tr>
<td>Sound</td>
<td>Examples</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>i:</td>
<td>sheep, field, police, team, key, people, scene, quay, amoeba, Caesar (AmE: busy)</td>
</tr>
<tr>
<td>i</td>
<td>ship, savage, women, carriage, valley, mountain, village, foreign, always, coffee, lynch, guilt, sieve, busy (BrE: busy, AmE: appearance)</td>
</tr>
<tr>
<td>e</td>
<td>bed, any, said, bread, says, guest, bury, leopard, leisure, friends (AmE: aesthetic)</td>
</tr>
<tr>
<td>æ</td>
<td>bad, plaid (AmE: laugh, calf)</td>
</tr>
<tr>
<td>o:</td>
<td>calm, father, heart, bazaar, sergeant, Shah (BrE: laugh, AmE: bother, honest)</td>
</tr>
<tr>
<td>o (BrE only)</td>
<td>pot, entree, bureaucracy, John, watch, cough, laurel, honest</td>
</tr>
<tr>
<td>o (:)</td>
<td>caught, ball, board, draw, haunt, four, floor, port, extraordinary, George (AmE: dog)</td>
</tr>
<tr>
<td>u</td>
<td>put, wood, wolf, could</td>
</tr>
<tr>
<td>u:</td>
<td>boot, move, shoe, group, flew, blue, too, fruit, rude, through, rheumatism, manoeuvre, new, cwm, leeward (AmE: lieutenant)</td>
</tr>
<tr>
<td>η</td>
<td>cut, some, does, blood, young</td>
</tr>
<tr>
<td>η:</td>
<td>bird, burn, fern, worm, journal, earn, myrtle, err, Guernsey, connoisseur, myrrh (AmE: chauffeur)</td>
</tr>
<tr>
<td>θ</td>
<td>about, fountain, clarity (AmE: physician), bureaucrat, parliament, purpose, luncheon, dangerous, tortoise (AmE: mullein), nation, restaurant, autumn, the, sergeant, cupboard, actor, theatre, bigger, surprise, furniture, beggar, soldier, colour, chauffeur, guerrilla (AmE: collegiate)</td>
</tr>
<tr>
<td>e1</td>
<td>make, pay, steak, vein, weigh, straight, prey, gauge, gaol, café, matinée, train, Gaelic, eh (AmE: melee)</td>
</tr>
<tr>
<td>e2o</td>
<td>note, sew, soap, soul, grow, toe, oh, brooch, beau, yeoman, mauve, owe, though, folk</td>
</tr>
<tr>
<td>ai</td>
<td>bite, eye, pie, buy, aye, try, dye, guide, sigh, height, aisle (AmE: coyote, geyser)</td>
</tr>
<tr>
<td>au</td>
<td>now, ounce, plough, sauerkraut</td>
</tr>
<tr>
<td>au</td>
<td>boy, poison, lawyer, buoy</td>
</tr>
<tr>
<td>ao</td>
<td>here, appear, idea, fierce, beer, souvenir, weir, atmosphere, theory</td>
</tr>
<tr>
<td>e3</td>
<td>there, hair, bare, where, bear, their, prayer, scarce, aeroplane, mayor, heir</td>
</tr>
<tr>
<td>u3</td>
<td>poor, insure, tour, cruel</td>
</tr>
<tr>
<td>e13</td>
<td>player, weigher</td>
</tr>
<tr>
<td>a1o</td>
<td>lower, sewer, boa</td>
</tr>
<tr>
<td>aυ3</td>
<td>tire, buyer, dyer, higher, quiet, lion, giant, fiery, tyrant (BrE: Isaiah)</td>
</tr>
<tr>
<td>aυυ</td>
<td>tower, our, Howard, sauerkraut</td>
</tr>
</tbody>
</table>
The *Longman Dictionary of Contemporary English* further notes that the following consonantal graphemes are silent in English:

<table>
<thead>
<tr>
<th>Letter</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>bomb, doubt</td>
</tr>
<tr>
<td>c</td>
<td>yacht, muscle, scene, victuals</td>
</tr>
<tr>
<td>g</td>
<td>sign, gnaw</td>
</tr>
<tr>
<td>gh</td>
<td>ought, light, through, high, Edinburgh</td>
</tr>
<tr>
<td>h</td>
<td>honest, hour, heir, exhausted, rhubarb, yacht, vehicle, which, John</td>
</tr>
<tr>
<td>k</td>
<td>know</td>
</tr>
<tr>
<td>l</td>
<td>half, walk, calm, could</td>
</tr>
<tr>
<td>m</td>
<td>mnemonic</td>
</tr>
<tr>
<td>n</td>
<td>autumn, government</td>
</tr>
<tr>
<td>nc</td>
<td>blancmange</td>
</tr>
<tr>
<td>p</td>
<td>psychology, ptarmigan, pneumonia, corps</td>
</tr>
<tr>
<td>r</td>
<td>BrE: car, card</td>
</tr>
<tr>
<td>s</td>
<td>aisle, island, corps, patois</td>
</tr>
<tr>
<td>t</td>
<td>fasten, often, ballet</td>
</tr>
<tr>
<td>th</td>
<td>asthma</td>
</tr>
<tr>
<td>w</td>
<td>wrong, sword, answer, dawn</td>
</tr>
<tr>
<td>x</td>
<td>faux pas</td>
</tr>
</tbody>
</table>

As the above tables show, English orthography abounds in violations of the sound-symbol faithfulness principle, but, as we saw earlier, this is not uncommon in languages whose writing systems were established many centuries ago.
5 Choosing a script

Analysing the sound system of a language is a necessary but not sufficient condition for developing a practical writing system. Once the analysis is done, there is still the question of which script to use, or which version of which script to use. Linguists tend to advocate alphabetic systems for a hitherto unwritten language. However, as some specialists point out, logographic systems may actually be preferential to phonetic ones. In the introduction of writing among certain South American Indian groups, phonetics-based scripts were disfavoured by the speakers concerned, since their language varied considerably in pronunciation. However, with the use of logograms, everyone could write and read the language, although the signs were pronounced somewhat differently. The same situation can be observed concerning the Chinese languages, which have a unified writing system but are different spoken languages, such as Mandarin, Hokkien, Cantonese and Hakka.

Also, Dahl (2000:218) notes the highly interesting fact that syllabic systems are preferred from the users’ point of view in some fairly recently designed orthographies, for example for the North American languages Inuktitut and Cree. If choosing a phonetic script, there may be a choice as to what kind of alphabet should be used. In this respect it is useful to look at how Arabic is used to write African languages not related to Arabic; it has also been used for Yiddish, the German-related spoken variety associated with the Jewish people.

The choice of a script raises many social and political issues – these are illustrated in what follows by examples from around the world. A number of basic questions must first be posed:

- Should the script of more widespread or more powerful languages be adopted, for example Roman script in many former British and French colonies?
- How far is the learning of a national or official language a factor? For instance, should a tribal language in India adopt (a modified form of) Devanagari script to facilitate learning Hindi subsequently?
- Is there a government policy or guideline regarding the choice of script and symbols? In Cameroon, the General Alphabet of Cameroon Languages was adopted in 1979 and is a guiding framework for the writing of unwritten Cameroon languages, using modified Roman script.
- How much weight should be given to history and tradition? Over the centuries Berber languages have used Arabic, Roman and Tifinagh scripts, with the latter being associated with traditional culture.

The examples given below from Ethiopia, Peru, Guatemala and Mongolia show how these and other factors play a role in decisions about scripts – from a historical and a contemporary perspective.

The very fact of writing a language down for the first time is not without political implications. In situations of high linguistic diversity, such as parts of Africa, Asia and the Pacific, the languages, or the dialectal varieties, that are written come to enjoy higher prestige than the unwritten languages around them, thus changing the balance of power and influence among language groups. Mühlhäusler (1996) refers to such situations as changes in the linguistic ecology. While recognising that writing a language down makes possible new opportunities of literacy, education and cultural expression, he sees the process as an intervention which is costly in terms of the changes it brings to the status of languages; some end up being promoted at the expense of others. Over time the predominance of one written language or variety could lead to the disappearance of others – a factor therefore in the reduction of
language diversity. Further research would be needed to determine how far this is the case in particular countries and regions and how the process is related to other factors of language endangerment such as language spread. While developing writing systems for unwritten languages can be a factor in their survival and development, Mühlhäusler reminds us that it can also lead to nefarious effects on the linguistic ecology in areas of high diversity.

5.1 Syllabary to alphabet in Ethiopia

In many cases choice of script is a historical process with many different kinds of influence brought to bear on the decision. Often these influences are external and they eclipse many of the functional arguments for choosing one script over another. Until recently this was the case for Oromo.

The Oromo language of Ethiopia is spoken by about 25 million people, making it the largest language group in the country. During the colonial era Oromo was written with a variety of scripts, including Roman and Arabic scripts and the Sabean or Ethiopic syllabary. The ‘Galla Spelling Book’ (Galla is an alternative name for Oromo) was published using the 250 characters of the Ethiopic syllabary in 1884.

In the 1950s an Oromo nobleman, Shaykh Bakri Sapolo devised a writing system for the language and taught it to his students. Although it did not come into general use, it was an attempt to adapt the Ethiopic syllabary to the particular structures of Oromo. This resulted in over 300 symbols derived from 28 basic consonantal forms. The following quotation highlights the originality – as well as the political implications – of the enterprise:

> It is not really clear why Shaykh Bakri returned to his home area to devote time to the alphabet, unless it was for the purpose of keeping the thing secret, for the authorities would certainly have been adamantly opposed to the idea of Oromo being written in any form, let alone in a script other than Ethiopic. Be that as it may, it does seem highly likely that Bakri was the first Oromo who saw clearly the problems inherent in attempting to write the Oromo language by means of orthographic systems which had been devised primarily for other languages. (Hayward and Hassan 1981)

Other languages in Ethiopia, which were developed in written form later than Oromo, also largely adopted the Ethiopic syllabary. In the 1970s both Ethiopic and Roman scripts were used for Oromo, but the writing of Oromo was officially banned until the Mengistu regime came to power in 1974. At that time the ban was lifted and there was liberty to write Oromo using the Ethiopic syllabary, although its use in education was proscribed.

About five months after the fall of the Mengistu regime in 1991, Oromo intellectuals held a meeting to discuss the question of the script. After six hours of debate there was a unanimous decision to adopt the Roman script, on the basis of three types of consideration: linguistic, pedagogical and practical.

The chief linguistic consideration was that the Ethiopic syllabary did not accommodate the structure of Oromo, for instance in the representation of lengthened vowels and consonants, as well as in the number of vowels. There was also a feeling that the syllabary was more adapted to writing Ethiopia’s Semitic languages, such as Amharic, Tigre and Tigrinya, rather than the Cushitic languages to which Oromo belongs. Pedagogical reasons focused on the reduced number of symbols to be learnt if an alphabet of about 30 symbols is used rather than a syllabary of 250. Under the heading of practical reasons, ‘global functional considerations’ were adduced, that is, the argument of easier transition to other languages, with English as a prime example.
While these reasons are entirely adequate in themselves, there may also have been some underlying political considerations, as suggested by Bakri’s clandestine project in the 1950s. The Ethiopic syllabary had come to be associated with political and cultural domination. The use of a distinctive script and its application to other languages was a powerful symbol of hegemony. Thus once the political landscape shifted, part of the new-found liberty consisted in challenging the linguistic symbols of domination – the adoption of a different script was thus an assertion of political freedom.

It is worth noting that other Cushitic language groups in Ethiopia, such as the Hadiyya and Kambaata continue to debate which script to adopt, Roman or Ethiopic. Amharic, used as a lingua franca in the country, continues to use the syllabary.

5.2 Vowels and politics in Peru

The controversy over vowels in Quechua in Peru illustrates how decisions about writing systems can become political. The question was whether Quechua should be written with three vowels or five. Hornberger (1995) documents the debate, pointing out the different interests of the players involved. They were:

- Peruvian linguists and bilingual educators, specialists in Quechua although few were speakers of it.
- Quechua academicians, belonging to the native Quechua academy in Cusco, founded in 1953. All are native speakers of Quechua.
- Foreign linguists, working with communities in linguistic research, literacy and Bible translation, often learning the local variety of Quechua.

These all had views on the number of vowels, depending respectively on whether linguistic arguments prevailed (three vowels), long-term written practice was foregrounded (five vowels), or native speaker/reader reaction was given prominence (mostly favouring five vowels). The problem arose because Quechua has three vowel phonemes /i/, /a/ and /u/, with [e] and [o] occurring as allophones of /i/ and /u/ in the proximity of the uvular consonant /q/. However, 400 years of writing Quechua represented [e] and [o] in writing, probably by analogy with Spanish.

However the problem is not purely linguistic. As Hornberger points out, other issues are stake, notably the language planning process, the basis of authority on the language, and defence of the language’s purity and autonomy – all politically charged issues. Thus any decision has implications for the relative power of the different groups involved and, beyond that, rests on a different view of who the important sections of the Quechua population are:

The choice for three vowels implies the rural monolingual Quechua speaker as a primary target group and an autonomous, cross-regional, and cross-national community of Quechua readers and writers as goal. whereas the choice for five vowels implies the urban, bilingual Quechua-Spanish speaker as primary target group and communities of Quechua readers and writers linked perhaps more directly to the Spanish-speaking Andean world than to each other as goal. (Hornberger 1995: 201-202)

What is the way forward in such a situation? The question was all the more urgent, as the educational use of Quechua depends on a stable writing system. While groups of specialists may take their respective positions, what is needed is the many voices of the Quechua-speaking population themselves.

Links

http://www.ullanta.com/quechua/
http://dolphin.upenn.edu/~scoronel/quechua.html
http://www.shef.ac.uk/q/quechua/
5.3 **Alphabets and identity in Guatemala**

Alphabets have been proposed for the Quiché (or Ki-che, or K'iche’) language of Guatemala since the 1940s, and the process has paralleled closely the assertion of ethnic identity, as Lewis (1993) describes. The first alphabet originated from the First Linguistic Congress of Mayan languages in 1949 and was designed to facilitate transition to learning Spanish – the alphabet was thus based on Spanish orthographic convention, including for the phoneme /k/ using <c> before <a, o, u> and <qu> before <e, i>. This alphabet was therefore not aimed at maintaining or affirming Mayan identity, but at integration into the national, ladino culture. In 1959 a Quiché speaker, Adrián Chávez, developed a set of symbols based on Mayan glyphs – a deliberate attempt to assert the distinctiveness of cultural and ethnic identity through the symbolism of an alphabet. While the alphabet was not generally adopted, his attempt was a milestone in the Mayan ethnic movement.

There followed a period when several competing alphabets came into use, prompting the holding of the Second Linguistic Congress in 1984. Orthography questions were top of the agenda, but no concrete solution emerged. This was left to a committee which soon formed an unofficial Mayan Language Academy and held a consultation on orthography in 1987. The outcome of this meeting was a unified set of symbols to be used in representing all Mayan languages. Thus, similar sounds in different Mayan languages should be represented by the same symbol, symbolising ethnic solidarity across different language groups. The inventory includes the use of <k> for the phoneme /k/, and the use of <q> to represent the post-velar stop /q/ – this usage makes a clear statement about the distinctive nature of Mayan languages vis-à-vis Spanish. Lewis (1993) concludes that orthographic decisions symbolised transitions in the consciousness and projection of ethnic identity, moving from passive assimilationist tendencies to the assertion of a wider Mayan nationalism.

Links

http://www.chichen.com.mx/chilambalan/
http://www.uady.mx/sitios/mayas/literatura_narrativa/cuentos/index.html

5.4 **Centuries of scripts in Mongolia**

Mongolian has had a writing system for about a thousand years. There is evidence in Chinese texts of the sixth century of a written form of the language of the nomadic ancestors of the Mongols, but the first scripts used by Mongols themselves date back to the tenth century. The Classical Mongolian script was developed from Uighur and Sogdian sources in the thirteenth century and was promoted by the great Mongolian emperor Chingis Khan. It is unique in that it is the only script written from top to bottom and from left to right across the page.

Other scripts such as the following appeared over the intervening centuries, but did not survive:

- **Square or Phagsba script:** designed by a Tibetan monk in the thirteenth century, this 'square writing' (which is what Phagsba means) combined features of Tibetan and Chinese and was written vertically.
- **Clear script:** this script, invented in 1648, sought to improve the classical Mongolian script by eliminating homographs (different words spelled the same way) and bringing written representation closer to the oral form.
- **Soyombo script:** a Mongolian monk, Zanabazar, invented this script in 1686 on the basis of ancient Indian writing. It served particularly to record
religious terms from Mongolian, Tibetan and Sanskrit, which all served at some time as literary languages for Mongolian scholars.

- Vaghintara script: also based on the classical Mongolian script, this variety was developed in 1905 and did away with homographs and positional allographs (that is, letters written different ways according to their position in the word).

It was the classical Mongolian script which was in common use until 1941 when, under Soviet influence, a new alphabet based on the Cyrillic alphabet was introduced with 35 letters (two more than Russian). Official documents stated two reasons for the change: first, the classical Mongolian script was too far removed from the oral form, and, second, it was not suitable for transcribing foreign words. The second reason was not in fact true, while the first reason applies to many scripts/languages, not least English.

Since the transition to democratic rule in 1990, there have been efforts to revive the classical Mongolian script, and it was taught in schools for some years. However, many of those born between the mid-1930s and the 1980s had never learnt to read it, and so it caused some friction. Also, literature was in Cyrillic script. Today, the Cyrillic script is taught and is in general use. With globalisation, some suggestions have been made to move to a Roman alphabet, but there seems little functional value in this, particularly given the ease with which transliteration can be effected between the Roman and Cyrillic alphabets.

Links
http://www.omniglot.com/writing/mongolian.htm
http://www.indiana.edu/~mongsoc/mong/language.htm
http://www.koreanhistoryproject.org/Jta/Mo/Lan/MoLAN1.htm

5.5 Writing minority languages in China

It comes as something of a surprise to find that some minority languages in China are written with Roman script. How did this come about?

In the 1950s the Chinese government carried out research on minority languages, including the development of scripts for those which did not yet have one. Chinese pinyin, a Romanised form of Chinese, was used at that time – and still is – as an initial introduction to reading for children in school. Thus there was a perception that a Roman script would be easier to learn. Also, fifty years ago, the idea was being floated of using this Romanised form for writing Chinese, rather than using ideographic characters – this suggestion, since abandoned, may at the time have contributed to the use of Roman script for minority languages.

These scripts were promoted in the early 1960s, then left on one side during the period of the Cultural Revolution. Interest returned in the 1980s, when there were a number of literacy efforts in minority languages, as well as the publication of research articles and dictionaries. For example three-month literacy classes for adults were held in the Dong language, and an experimental literacy programme took place in the Bai-speaking Jianchuan area between 1986 and 1992. Both Dong and Bai have Roman-based scripts.

Currently it is largely up to local communities to develop and use their minority languages, if they wish to do so. A limited bilingual education programme in the Dong-speaking area offers the chance to learn literacy skills in their own language.

Attitudes to Roman script reflect the long tradition of writing and literature in Mandarin Chinese – Chinese characters are felt to reflect a high view of culture and an accomplished skill of beautiful calligraphy. Roman scripts carry none of those connotations, but rather evoke the early years of children’s schooling, with all the
overtones of simple and childlike learning. These attitudes represent barriers to the use of these minority languages in written form, even though the Roman scripts have been in use, at least experimentally, for several decades.

5.6 Community ownership and use

Language is not only a means of communication, but also a symbol of identity. Writing an unwritten language is therefore a matter of great cultural significance – it changes the nature of the relationship of the community to its language. The language moves from one category to another – for some speakers it means that the language has come of age, or has acquired a new status, alongside the other written languages of the world, near and far. For others the development may not be seen in such a positive light, because writing opens the language up to others, and makes it in some sense less the property of the original community. The sense that a language belongs to the community of its speakers can be very strong, especially when the community is relatively small and language constitutes a major feature which marks that community out from surrounding ones.

When a language is in the process of being written for the first time, it evokes many different kinds of feelings. It is important that they are expressed and that the community of speakers is vitally involved in the process of developing a written standard. There are a number of important questions which a community will have to debate and decisions it will have to make:

- which dialect(s) of the language should be used for the written standard?
- what script should be used?
- which language(s), if any, should be used as a model for the writing system?
- should the writing system resemble that of related languages, or neighbouring languages, or national or official languages, or should it be quite distinct?
- how will words be spelled – joined up or in their separate grammatical or lexical parts?
- how will the written standard be promoted?
- how will it be taught, to adults, to children?
- who will produce literature in the language?
- is there a need for a linguistic standard-setting body (such as a language academy or institute), and if so how will it be organised?

These questions cannot be dealt with quickly or easily. Some may be settled by a deliberate process of consultation and decision-making, others may evolve slowly with a consensus gradually developing. Language issues may form part of a broader cultural or political agenda, such as indigenous people’s rights or government decentralisation, or they may be the object of focus in themselves.

5.6.1 Language committees in Cameroon

In 2003 there were 62 language committees in Cameroon, each one representing a distinct language group. It is estimated that there are upwards of 250 language groups in total in Cameroon, many of which have no written form or have a very recent written tradition. Each language committee is composed of speakers of the language and takes responsibility for the following functions:

- the development of the language in written form
- literacy promotion, including the production of literature and the organisation of literacy acquisition
• stimulating and overseeing bilingual programmes in the schools using the mother tongue and the official language
• cultural development and preservation
• developing and identifying resources for these functions
• links with other language committees

In setting up a language committee, the following criteria were taken into account in determining its composition (cf. Sadembouo 1988):
• representatives of all the dialects of a language
• men and women, young and old
• people from all religious groups
• people from different socio-economic groups
• educators and teachers
• literate and non-literate people

The Gulu Linguistic Committee (GULICO) of central Cameroon started life in 1978 as the initiative of local community leaders and a foreign linguist undertaking research into the language. It began to meet monthly and quickly addressed issues of how to write the language. Discussions focused particularly on the similarities and differences with two other languages with which the committee members were familiar in written form: Ewondo and French. Ewondo, the language of the Yaoundé region, had been used previously for religious purposes and was the only written African language that GULICO members were familiar with up to that point. In addition, the University of Yaoundé had recently adopted general principles for writing the languages of Cameroon, some of which conflicted with the practice in Ewondo.

For instance, the sound [t̥]; had been represented as <ts> in Ewondo, but was recommended to be written as <c> in the University’s principles. After much discussion, <c> was adopted. This was in large measure a result of the committee’s desire to adopt new general Cameroon practice, rather than maintaining a historical precedent whose scientific basis was not apparent and which, in any case, had been a colonial invention.

Once the alphabet had been adopted and spelling rules developed, the committee worked together on an initial booklet to promote the written form of the language. This was presented in visits by committee members to villages of the region, with the result that some villages requested literacy instruction. Again, the committee members were involved in setting these up and organising instruction.

The role of the committee expanded in the early 1980s through visits to neighbouring language communities, creating a regional dynamic for language development which, by 2003, resulted in joint working sessions and regular mutual exchange of information.

In the late 1980s a national federation of Cameroon language committees was formed, in order to organise training, facilitate publication of materials, undertake research, and coordinate the search for external funding support. The National Association of Cameroon Language Committees (NACALCO) today supports and promotes local responsibility for the development of Cameroon’s languages, with an emphasis on their role in education and development.

5.6.2 Building community ownership in Uganda

Lugbara is a language spoken in Uganda, southern Sudan and the Democratic Republic of Congo and has a number of dialects. In common with many African
languages, Lugbara crosses international borders which frequently divide a language area into two, or even three, as in this case. In 2000, representatives of the Lugbara-speaking community came together in Arua (northwest Uganda) for a conference on standardising the Lugbara writing system. The aim was to move forward as a whole community so that the resulting orthography would be owned and used by all sections of the Lugbara population— that is, recognised as belonging to all parts of the community, not merely one dialect area or sub-group. The conference set as its goals (LULA 2000: 6):

- to examine the challenges facing the existing Lugbara orthography
- to find out the most appropriate ways through which a standard orthography could be arrived at for wider communication in the Lugbara communities

The meeting strove therefore to develop ownership of the process of orthography standardisation rather than to decide directly on what orthography should be adopted. The reasons for this emphasis are clear: only when such ownership across the community is in place can the vigorous use of the language in written form be sustained. Earlier history showed how necessary this process was.

The first attempts at writing the language were undertaken by Catholic and Protestant missionaries from 1918, with differences between the denominations in the way the language was written. This religious use resulted in the Ayivu dialect being used as the standard for the written form, although some used the Aringa dialect. The orthography used in Congo differed again, under the influence of French. When mission schools were taken over by the government in 1965, the publication of books in Ugandan languages was abandoned. The Lugbara Literature Association – LULA – was founded in 1994 (registered as an NGO in 1999) with the aim of promoting research, publication, cultural expression and standardisation in Lugbara. It also sought to bring together all the dialect areas. Against the background of all these initiatives and divergent attempts at orthography development, LULA organised the 2000 conference as a way of establishing a commonly agreed approach.

It should be noted that Lugbara was one of the languages which the government of Uganda designated in 1992 as one of five languages of wider communication to be used in primary education. Thus the policy environment was supportive, as witnessed by the co-sponsoring of the conference by the Arua District authorities, where Lugbara is spoken. However, the use of Lugbara in schooling has never been implemented, nor is the language an object of study at university level, in contrast to the other four officially recognised languages.

What did the conference achieve? First, it brought together representatives of the different dialects, religious groups, government departments, educational NGOs, civil society groups and networks. Second, it issued a strong appeal for a standard Lugbara orthography and stated the reasons why. Third, it did not get into technical details, but appointed a committee to carry that process forward and set parameters for its work. Fourth, it highlighted the importance of research as the underpinning of language development, and, fifth, it emphasised the promotion of the use of the language as a vehicle of literacy for adults and primary schooling for children.

The conference served as the scaffolding for building the written use of the language: without such a process there can be no guarantee that a writing system, however elegant and scientifically based, will actually serve all the speakers of a language community.
5.7 Developing a writing system: stages in the process

Community dialogue about writing the language and about its written use, and about the decision-making process (eg setting up a language committee, …) → Choice of a reference variety of the language → Community consultations → Choice of script → Promotion of the language development process

Consider: govt policies, dialects and in-group relations, neighbouring language orthographies, alternative analyses, …

Orthography testing → Adjustments and refinements, further testing → Community acceptance → Official backing → Writer training → Publications → Literacy acquisition → Multilingual education → Promotion of a literate environment
6 Writing and technology

6.1 Typing new symbols

One of the issues in developing a writing system is whether it can easily be reproduced in print, by a computer or a typewriter. Typewriters are still in use in many of the remote or marginalised communities where languages are only now being written down. Traditional typewriter keyboards using Roman script do not cope well with anything beyond the European character sets, and have very limited possibilities of adding diacritics, such as tone marks. In the 1980s, typewriters in Cameroon could be modified in order to represent the symbols of the General Alphabet of Cameroon Languages for a particular language. No typewriter could accommodate the range of symbols necessary to type the symbols for all of Cameroon’s languages. Electric typewriters using the ‘golf ball’ system offered more possibilities by adapting different ‘golf balls’ to a different set of symbols. These solutions required special fonts and were not easily available to communities wishing to type their languages.

Computers changed the situation. Fonts are now part of the software and can be designed, adapted and printed almost at will. This is the case at least for Roman scripts and others, such as the International Phonetic Alphabet, which are adaptations of Roman script. The possibilities for non-Roman scripts are much more limited. Although there is no reason why other scripts should not be represented in electronic form, both on the computer screen and in print, much less investment in this has been made; software and other companies have concentrated on the major scripts where mass use ensures a viable market. Even less attention has been given to the adaptation of non-Roman scripts for minority languages. For instance, a south Asian language group may choose to use the Devanagari script for writing their language, however the script may need adaptation to the particular sound system of the language. This implies re-designing the software for a relatively small number of speakers – something which is economically unattractive for commercial firms. The Non-Roman Script Initiative addresses this issue.

Faced with the need to use a fairly standard keyboard, the most practical thing is to avoid diacritics, seeking to get by with the 26 letters from A to Z present on most keyboards. However, most languages have more than 26 phonemes, and different solutions will then have to be envisaged. Even if a language has less than this number of phonemes one may run into problems since the grapheme inventory does not fit the phoneme one. For example, it would be strange to represent affricates by means of <a> or <e>, if these graphemes happened to be available. Normally, the solution chosen in cases like this is to use two graphemes to represent one phoneme or to use diacritics. If one uses a phonetic font, this limitation to 26 characters disappears since the shift and alt keys can be used to produce an additional 26 + 26 characters, i.e. 78 in total, which in principle ought to be sufficient for writing down most languages in the world (actually one could also combine shift + alt, yielding 26 additional possibilities increasing the expressive power to 104 characters, and one could make use of the keys displaying the digits 0–9). The SIL Doulos Regular IPA font exploits this principle in a nice way when it lets the <r> key alone produce <r>, shift + <t> = <t> and alt + <t> = <t>. Other more sophisticated systems are also available, e.g. Lucida Sans Unicode. That is, the r-key allows the production of symbols related to /r/, which is practical from the users’ point of view. Now the problem is that we do not find this nice symmetry in phonological systems, rather, some phonemes have a lot of related varieties (/r/ is a case in point), whereas others have far less (for example the uvular stop /q/). This means that each key does not have exactly three related phonemes associated with it. Some keys do, whereas others have less, and some phonemes with variants need more than one key. Also, a
number of keys have to be reserved for diacritics and punctuation signs, as well as for pictograms like <&> and <§>.

A greater problem, however, is the fact that one would like to be able to use an ordinary text font provided with standard word processing systems. For one thing, downloading and installing extra fonts might be complicated for people not used to computers. Secondly and more importantly, who wants to write running text using something as complex as the IPA font? The disadvantage with an ordinary font like Times or Courier is that the shift key is used for producing capital letters. If we reserve in addition a number of key options for diacritics and other symbols as mentioned above, the possibilities to represent phonemes as graphemes in a simple way diminish considerably. However, the main question in this context is how to map phonemes whose representation in IPA does not have an obvious equivalent on a traditional Roman keyboard. That is, where on the keyboard should graphemes representing e.g. the affricate /dʒ/ or the fricative /θ/ be put? The former could be placed on the <d>, <j> or <z> key, and the latter under both the <t> and the <l> key, given that there is still free space in the shift or alt mode of the keys in question.

6.2 Harnessing the new technologies

With globalisation and the spread of electronic communication, there is much talk of the digital divide. This focuses often on access to computers, on production of Internet content, on the availability of electricity and phone lines, and on the development of expertise to harness the power of information technology. There is less talk of the linguistic aspect which means that for some groups there is currently no possibility for their language to be used in the electronic media, because their writing system cannot be represented on the computer screen. In spite of the huge capacity and great flexibility of electronic text processing, little has been done to enable all languages, in whatever script, to be represented on the computer screen and in word-processing. Only about 400 languages are currently included in the international standard identification system which is the starting point for their use in electronic form. To handle the needs of the wide variety of linguistic communities there is need for an extensible mechanism adequate to handle any orthographic phenomenon encountered in any writing system based on any modern script.

The Non-Roman Script Initiative (NRSI) of SIL International, in cooperation with UNESCO and software companies tackles the issue of new and distinctive scripts for minority languages based on systems other than the Roman alphabet. It links with UNESCO’s programme ‘Initiative B@bel’ which seeks to support linguistic and cultural diversity on the Internet and create space for all languages in the electronic media. The NRSI seeks to address the language aspect of the digital divide, making it possible for linguistic minorities and indigenous peoples, many of whom use complex scripts, to gain access to the electronic media. The processing of text by computer was designed on the basis of Roman script, with a number of assumptions, for example that one symbol requires just one key to be pressed, and that one letter always has the same form. Since these assumptions do not hold true for many of the world’s languages, new ways of processing text must be developed.

It is not merely a question of designing scripts for each language for use on the computer. The aim is to create an international standard so that any and every language is assured of its place in the digital world. NRSI works with the Unicode Consortium, founded to establish a universal standard for representing each character of all the world’s writing systems on the computer. The aim of these partnerships is to encourage a more inclusive and multilingual use of information technologies.
7 Who writes languages down?

The writing of unwritten languages has a tradition stretching back thousands of years and it is still unfinished work. Scholars, priests, administrators and many others have taken a hand in this enterprise, often on an individual basis. Today, various groups and organisations are involved, working increasingly in partnership together since the writing of an unwritten language is a community-wide concern with many social and cultural consequences. A number of groupings and organisations are today particularly active in developing writing systems, of which the following may be mentioned:

- **UNESCO** provides a framework and principles to guide policy formulation and implementation. It seeks to promote the use of the mother tongue in education, implying therefore that a written form of the language exists or needs to be developed. UNESCO’s role is to offer policy guidance and facilitate the exchange of experience and expertise.

- **Language communities** themselves are the principal actors in language development. Countless individual speakers of the world’s languages have taken the initiative and invested effort in developing writing systems. In addition community and civil society groups such as schools, churches and development associations have worked on projects to promote the use of the local language in written form. Some community efforts have sought technical help from external sources to assist in the linguistic aspects of their endeavour.

- In a more organised and deliberate fashion, **language committees and indigenous movements** have made language development part of their aims. With a particular concern to ensure that any writing system is indigenous and does not reflect past colonial patterns, the technical aspects of developing a writing system have been accompanied by vigorous and often effective efforts to promote the use of the language and the validation of the culture with which it is associated.

- Although governments set up the policy framework for developing the languages within the country’s borders, they rarely engage directly in writing system development. This is delegated to **national universities and research institutes**, working in conjunction with speakers of the languages concerned. These national institutions often seek to coordinate such work by setting up guidelines for the selection of symbols, or the representation of particular features such as tone.

- **Non-governmental and civil society organisations** active in socio-economic and cultural development are often faced with the need to include an educational component in their work. Where this involves literacy in the local language, they have often initiated or facilitated the writing of the language concerned.

- For many centuries **Christian missions** have been involved in designing alphabets and writing systems, based on a desire to present the Christian gospel and to launch educational initiatives. While some of these efforts were very much the work of linguistic amateurs, others produced viable and durable writing systems. Many of the writing systems of African languages were designed by this means.

- **SIL International**, a faith-based NGO, is the largest organisation specialising in the development of writing systems, having undertaken research in over 1500 languages. It also offers training in the techniques of linguistic analysis leading to orthography development. More recently, **SIL**
has cooperated with UNESCO and others in developing software for the analysis of language structures and the representation of scripts in electronic format.

- A number of organisations and funds devoted to endangered languages support or undertake research to preserve languages in danger of extinction, for example the Foundation for Endangered Languages, the Endangered Languages Fund or the Rausing Foundation. This work frequently includes phonological and grammatical descriptions which offer a basis for developing writing systems.

- An innovative project, the Basic Standardisation of All Unwritten African Languages – BASAL, is a unique initiative seeking over time to give viable writing systems to all of Africa’s languages in order to enable them to be used for educational purposes. Working with volunteer African graduates who receive initial training in the relevant techniques, the work is conducted in communities, working with local leadership and institutions. A useable alphabet, based on linguistic analysis, basic literacy materials and an initial dictionary of each language are planned as outcomes of the project.
8 Writing systems and education

Developing a writing system for a hitherto unwritten language often has an educational purpose – whether to record and transmit local history and knowledge to the next generation, to use the language in formal schooling, or to facilitate adult literacy acquisition. The educational use of a language depends, except in the most informal settings, on a written form which can be employed in learning. UNESCO’s three principles of education in a multilingual world underline this necessity: mother-tongue instruction, bilingual/multilingual education and intercultural education:

- UNESCO supports mother tongue instruction as a means of improving educational quality by building upon the knowledge and experience of the learners and teachers.
- UNESCO supports bilingual and/or multilingual education at all levels of education as a means of promoting both social and gender equality and as a key element of linguistically diverse societies.
- UNESCO supports language as an essential component of intercultural education in order to encourage understanding between different population groups and ensure respect for fundamental rights.

(UNESCO 2003: 30)

This section explores the issues which surface when a newly written language is used in educational settings.

8.1 Newly written languages and literacy

If a newly written language is to be used by its speakers, that implies a process of literacy acquisition. After all the processes of writing system development already described, there is a new set of processes around the promotion of the written form and the facilitation of learning. These processes must take account of the reactions which a newly written form of the language may evoke:

- ‘My language can’t be written’ – this is often a reaction to the fact that the language never has been written, with the assumed implication that the language is somehow inferior to others which can be written down
- ‘I don’t need to learn to read it, I know it already’ – people react like this when their only access to literacy has been through a foreign language. Since learning literacy was equivalent to learning another language, if you already know the language, why would you need to learn to read it?
- ‘The writing system is too complicated’ – this reaction comes from the expectation that the local language will be written using the same system as the national or official language, which the person may already know how to read. In particular, certain symbols and the marking of tone may be new features.
- ‘My language isn’t worth learning’ – this negative reaction comes from feelings of linguistic and often cultural inferiority inculcated by decades of hegemonic domination by other languages.
- ‘It isn’t my dialect’ – a reaction to the fact that the written standard may not reflect all the dialect forms of a language. Choosing a reference dialect for the standard is a delicate process which must be based on much consultation.

Taking account of these reactions and promoting the use of a newly written language for the purposes of literacy and learning imply the following processes: testing the orthography, developing written literature, and promoting a literate environment
8.1.1 Testing the orthography

This is part of developing a writing system. It is not enough to propose an alphabet based on linguistic analysis and socio-cultural factors, it must be field-tested to discover how adequate it is, how easy it is to read and to write and what kinds of problems may emerge. Testing an orthography will involve reading and writing test passages, possibly presenting different alternatives to different subjects in order to assess the relative ease of use of this or that symbol. It will involve an analysis of reading problems and errors, seeking to note where there are consistent errors across different subjects which may be due to difficulties in the orthography, rather than to individual performance or fluency. Hollman (1992) carried out such a study for the newly developed Bari orthography in southern Sudan. In testing an orthography it is important to do so with a range of subjects: men and women, young and old, schooled and unschooled, literate and non-literate, those who know the official/national language and those who do not. Testing may need to be repeated after adjustments are made and alternatives proposed.

When a new and simpler orthography of Ewondo (central Cameroon) was introduced experimentally into primary schools in the 1980s, there was hope that children would find it easier to learn than the previous system. The new orthography respected the linguistic structures of the language and used the General Alphabet for Cameroon Languages, specially designed to cope with the special symbols required. It is noteworthy that one of the educators associated with the project remarked, when asked if the children did not in fact find the new system easier: “They would have learnt to read anyway, even with the old orthography.” This is a reminder that it is not the simplicity or elegance of the writing system which makes the most difference in acquiring literacy, but rather the motivation of the learner. This should, however, not be an excuse for inadequate or inappropriate writing systems!

8.1.2 Developing written literature

The existence of a writing system does not in itself lead to the use of a language in literacy and learning. There must be something interesting and worthwhile to read and opportunities for people to express themselves in writing. This entails workshops to train authors and writers in the language and the encouragement of local poets, historians, story-tellers, philosophers and language experts to ply their craft in written form. It should be noted that not every guardian of local knowledge feels that their wisdom should be shared in written form. The relationship of literacy/literature to oral production and transmission is a delicate matter which can only be addressed in a specific cultural context.

Literature in a newly written language is often religious or functional – indeed the development of a writing system may have been undertaken with those purposes in mind. However, reading and writing serve whatever purposes people wish to use them for, as is the case of environments with a long and strong written tradition. People write and read not merely for functional purposes, however, but also for pleasure. This domain is often neglected when a language is newly written: newspapers and newsheets, magazine-type publications, comic books, leaflets and posters will form an important part of an emerging literate environment where writing and reading add pleasure and interest to life.

8.1.3 Promoting a literate environment

This involves all of the above activities, but implies also that literacy practices become a valued aspect of daily life, used for learning, for social and religious purposes, for personal communication, for pleasure and recreation. A sustainable literate environment, using both newly written languages and those with a longer written tradition, is one in which literacy::
is embedded in the life of the individual and community as a tool to accomplish their goals, and is not merely a technical skill that has
responds to and builds on the wide variety of motivations for acquiring and using literacy: improving livelihood, gaining political voice, affirming identity, validating culture, personal and intellectual been/to be acquired.
serves to improve the quality of life in all its aspects: in health, family income, political participation, food security, conflict resolution, use of all kinds of technology (from pumps to computer networks), cultural and religious expression.
is accessible to women and men, children, youth and adults
is a regularly used tool for learning and enables further learningdevelopment, religious participation, communication over distance and time, …
is embedded in the life of local and national institutions in both its use and acquisition; literacy provision or instruction may take place in many different institutional contexts, from women's crop cooperative to mosque, from church to water management group, from micro-credit group to cultural association, …
is locally self-reproducing as text materials are written, produced and used (books, magazines, newspapers, radio scripts, web pages, texts of artistic creation, text with functional purposes, literature designed to last, literature which will be discarded, text for learning, inspiration, amusement, expression, communication, action …)

8.2 Newly written languages in school
Papua New Guinea, India, Philippines, Uganda, Ethiopia, Ghana, Mali, South Africa, Zambia, Ecuador – these and many other multilingual countries have policies favourable to the use of local languages in school, in line with the three UNESCO principles of education in a multilingual world (UNESCO 2003). Such policies imply the use of a written form of the language. Papua New Guinea, a country with four million people and 800 languages has gone a long way to implement such a policy, with perhaps 200 languages already in use in primary schooling. Each of these languages has acquired written form, sometimes with the express purpose of its introduction into school, sometimes developed earlier by linguists or missionaries.
When a newly written language, which is by definition local and small, is introduced as the medium of instruction in primary education, a number of factors must be taken into account: a multilingual approach to learning and teaching, trained teachers who can write and teach in the language, appropriate second language teaching of major language, supportive policies, and written materials and local knowledge.

8.2.1 A multilingual approach to learning and teaching
While the best strategy for initial education, especially the acquisition of literacy and numeracy skills, is through the mother tongue of the learner, this will take place using an approach which introduces gradually other languages which a child needs to learn. These include neighbouring languages and the national or official language of the country. There are various models of bilingual and multilingual education – what is essential is a balanced curriculum which enables the full cognitive development of the child while offering wider learning opportunities.

8.2.2 Trained teachers who can write and teach in the language
Children cannot be taught to use newly written languages unless the teachers themselves master the writing system. This cannot be taken for granted and requires
in-service training courses where teachers not only become familiar for themselves with the orthography and its conventions, but acquire pedagogical skills for teaching through the mother tongue of the learners. This is a totally new and different approach in many situations where teaching has traditionally taken place in a language foreign to the children.

8.2.3 **Appropriate second language teaching of major language**

In order to achieve a balanced multilingual curriculum, schools must be able to teach the first foreign language, often the official language of the country, using second (or foreign, depending on the context) language teaching methods. Children now using their own mother tongue as a medium of instruction will learn the official language through their first language. No longer will they learn it simply by being taught in it. This is again a major change of approach for which teachers will need training.

8.2.4 **Supportive policies**

In addition to positive policies at national level, there must be supportive policies at the level of the education authorities, the local administration and at the school level. Introducing and using a newly written language in schools implies in-service training, appropriate posting of trained teachers, and an understanding of the purposes and methods involved. Community leaders must also be on board, as the use of the local language will connect the school more directly and in new ways with the local community. Ensuring parental support will be part of the role of parent-teacher associations, and indeed may lead to cooperative inter-generational literacy learning in families, involving children and their parents or older relatives. The Family Basic Education project of Literacy and Adult Basic Education (LABE) in Uganda supports this kind of innovation (Education Action International 2003).

8.2.5 **Written materials and local knowledge**

A newly written local language will enable the use of written materials based on local knowledge. As indicated with respect to a literate environment, such writing must be promoted. For use in school, this will be a collaborative venture between teachers, local knowledge-makers and the educational authorities. It will become possible to produce learning materials on history, geography, social science, flora and fauna, language skills and other topics based on written expression of local culture. These new opportunities not only validate local knowledge, but enable children to use their intellectual and cognitive skills directly in and on the local environment, breaking the pattern of the domination in many post-colonial contexts of knowledge which is disconnected from the child’s daily life.

Trudell (2003) describes the difference in classroom participation and interaction that the introduction of a recently written local language makes – the children are livelier, more vocal and engage more actively in the learning process, right from the start of schooling.
9 Links

These links provide background material on language and language development, and give information on languages in different parts of the world. Some sites offer links to further resources.

World languages

http://www.ethnologue.com
http://www.rosettaproject.org/live
http://www.yourdictionary.com/languages.html#table
http://www.inalco.fr/pub/

Scripts around the world

http://www.omniglot.com/writing/atoz.htm
http://people.w3.org/rishida/scripts/samples/

Glossaries of linguistic terms

http://www.sil.org/linguistics/GlossaryOfLinguisticTerms
http://www.sil.org/mexico/ling/glosario/E005ai-Glossary.htm#Alpha-A
http://qsilver.queensu.ca/french/Cours/215/chap3.html (French)
http://www.sil.org/linguistics/glossary_fe/?lang=en (English/French)
http://www.orbilat.com/General_References/Linguistic_Terms.html
http://www.cus.cam.ac.uk/~cjp16/spanish/linggloss.htm
http://www.englishbiz.co.uk/grammar/main_files/definitionsa-m.htm
http://www.englishbiz.co.uk/grammar/main_files/definitionsn-z.htm
http://sps.k12.mo.us/khs/linguistics/lingtrms.htm

Latin America

http://www.ini.gob.mx/monografias
http://www.sil.org/mexico/gob_mex/Himno_nal.html#Cora
http://www.sil.org/mexico/nahuatl/24e-OrtografiaNah.htm
http://www.indians.org/welker/nahuatl.htm
http://www.umar.mx/zapoteca/lengua.html
http://www.zapata.org/Tzotzil/
http://www.cimi.org.br/educacao/educ281102.htm
http://orbita.starmedia.com/~i.n.d.i.o.s/menu.html
http://www.unb.br/ii/lali/
http://educaterra.terra.com.br/almanaque/indios_5.htm
http://www.m-w.com/cgi-bin/dictionary?va=Tupi-Guarani
http://www.banrep.gov.co/blaavirtual/letra-l/lengua/clas02.htm#2
http://www.unisi.it/ricerca/centri/cisai/gnerre.htm#1
http://www.serindigena.cl/territorios/recursos/biblioteca/diccionarios/espa_mapu/fone.htm

Australasia
http://www.ntu.edu.au/yolngustudies/

North America
http://wings.buffalo.edu/linguistics/ssila/learning.stm
http://www.ipola.org/resources/directory/index.html
http://jan.ucc.nau.edu/~jar/RIL_8.html
http://www.yourdictionary.com/languages/north.html#na-dene

Africa
http://www.bisharat.net/A12N/
http://www.bisharat.net/Documents/Niamey78annex.htm
http://www.uio.no/english/about_uio/international/north-south/nufu/afrika/zimbabwesp.html
http://www.dokpro.uio.no/allex/gsd/fm/1-frontpage.htm
http://www.humnet.ucla.edu/humnet/aflang/Hausa/References/hrefs.html
http://african.lss.wisc.edu/yoruba/font/index.html
http://africanlanguages.com/
http://libtext.library.wisc.edu/cgi-bin/Africana/Africana-idx?id=Africana.Hair01
http://www.cis.yale.edu/swahili/
10 References


