Demographics, ministries, and scriptures via 13,500 language profiles

1. A WORLD LANGUAGE CLASSIFICATION

A large part of the Christian mission in the world centers on understanding and utilizing the vast world of languages. From Apostolic days the church pioneered translation and the uses of mother tongues, vernaculars, and lingua francas in the proclamation and spread of the gospel.

Many observers of the language scene have realized for a long time that the church today has not been fully aware of this vast world. For this reason the first section of this Part 9 develops a global survey and taxonomy of languages, their demography, and their relation to race, ethnicity, cultures, religions, and statistical enumeration.

Classifying the world of language and its users

The next few pages summarize the schema worked out over 25 years by linguist David Dalby in his 2000 publication *The Linguasphere: register of the world's languages and speech communities* (Linguasphere Press, Wales: Contributing Editors David B. Barrett, Michael Mann). Readers interested in the sources used, assumptions made, and methodology employed should study this seminal publication. Its application here to the world of Christian language ministries is based on the 1997 abbreviated version which uses the same codes except in a handful of cases. Full-length versions with all dialects and alternate names are available on electronic media.

This Part 9 thus sets out a new and original classification of all the world's living languages. It does this by assigning to each language a unique 7-character code, and to each of its dialects an 8-character code, such as 01-AAAA-aa. By means of this code the reader can then search or navigate through large areas of data describing geography, linguistics, demography, Christian resources, translations of the Scriptures, agencies at work, ministries, and the like. By comparing the codes of any 2 languages the reader can immediately discover how close or how distant they are to or from each other.

Although the basic system is relatively easy to grasp, there are many useful implications to assist readers working with 2 or more languages or wanting to locate, explore, or investigate particular groupings. For this reason the next few pages develop a variety of tables that explain the classification and its codes from different starting-points.

Table 9-1. Shorthand terms for comprehension levels.

	Within... 1	*All share a minimum of...* 2	*Which means intercomprehension is...* 3
a	MACROZONE	0%	zero
	GLOSSOZONE	5%	negligible
	GLOSSOSET	30%	acquirable
	GLOSSOCHAIN	50%	potential
	GLOSSONET	70%	partial
	GLOSSOCLUSTER	80%	general
	language	85%	adequate
	dialect	90%	mutual

Eight groupings of speechforms

The 7 characters of each language's code, or the 8 for each of its dialects, represent distinct varieties or levels of groupings of their related speech forms. These are as set out in Table 9-1.

Getting the idea through single adjectives (Table 9-1)

A way of understanding this schema therefore is to draw up the simplified table which we can verbalize as follows: 'Within 1 (a particular level or category, in the first column), all speech forms share a minimum of 2 (second column, %), which means that the poorest intercomprehension between any 2 of them is a minimum of 3 (third column).'

Starting near the bottom of the table, the term **language** is here defined as a speech form in which all its related component or subsidiary speech forms such as dialects share with the language 85% or more of basic vocabulary of human experience. This thus gives their speakers adequate intercomprehension—they can all understand each other at least adequately. Moving down in the table, within a **dialect** all speech forms then share 90% or more, providing mutual intercomprehension.

Moving up in the table, a language and any other related languages which share 80% or more of basic vocabulary are here defined as forming a GLOSSOCLUSTER. They share general intercomprehension. If several of these exist sharing 70% or more, this forms a GLOSSONET, sharing partial intercomprehension. Several related glossonets may then make up a GLOSSOCHAIN, sharing 50% or more and having potential intercomprehension. Several of these may make up a GLOSSOSET (30%). And lastly to complete this schema there are 2 top levels, GLOSSOZONE and MACROZONE, with virtually no intercomprehension but essential to complete this worldwide classification of all languages and speechforms.

Identifying 8 levels by codes

The 8 levels of speech forms can now be coded by assigning a character to each level. The result is both a unique classification, and also an organic or independent one. This is set out in Table 9-2.

Forming a speech form's glossocode

The 8 levels described above result in 8 characters which together form what we here term a glossocode, as with 01-AAAA-aa as mentioned above. This code can be seen to form a proximity scale. It will now be divided into 3 smaller scales. The first scale is composed of the first 2 characters, always one digit each; they describe the first 2 lines of Table 9-1's and 9-2's list. The second scale consists of 4 capital letters, describing the next 4 lines of that table. The third scale closes with 2 lowercase letters. The makeup of these scales needs now to be explained in detail.

LANGUAGE PROXIMITY SCALES

The main entity listed in this classification is a language, being defined as the mother tongue of a distinct, uniform speech community with its own identity. Each language is characterized here by a 7-character computer code or language code or language proximity scale (scale of linguistic proximity), which locates the language in its relationships with other living languages. Likewise, each dialect is described by an 8-character code. This proximity scale is divided into 3 indexes or scales: first a wider scale, which is a 2-digit reference grid situating the language within the wider world; second a closeness scale, which is a 4-letter lexical similarity index, grouping each language by its closeness to other related languages; and third a comprehension scale, which is a 2-letter intelligibility index, classifying speech forms into languages (defined as speech forms each of which needs, requires, or already has its own separate literature and broadcasting) and dialects (defined as speech forms each of which is sufficiently interintelligible with its parent language and with sister dialects to not need or require separate literature or broadcasting).

Table 9-2. Recognizing partial or abbreviated codes in the World Language Classification.

Partial codes in this classification all start from the left end. They have the following meanings:

a 1-character code by itself	=	a **MACROZONE**
a 2-character code by itself	=	a **GLOSSOZONE**
a 3-character code by itself	=	a **GLOSSOSET**
a 4-character code by itself	=	a **GLOSSOCHAIN**
a 5-character code by itself	=	a **GLOSSONET**
a 6-character code by itself	=	a **GLOSSOCLUSTER**
a 7-character code by itself	=	a **language**
a 8-character code by itself	=	a dialect
name in capitals, any code	=	anglicized cover-name
name in lowercase	=	reference-name (autoglossonym) with affix (prefix or suffix) where used, added in medium type
in medium type with capital	=	anglicized name where different from foregoing, or geographic or personal name
name in lowercase medium (only in full *LinguaMetrics)*	=	dialect or alternate name/names

The first 2 of these scales present cover names in anglicized form, where such exist. The third scale presents each language's or dialect's own autoglossonym—what it calls itself.

The 3 indexes or scales can be elaborated on as follows, in words and in tables.

1. WIDER SCALE
(a 2-digit reference grid)

This first scale or grid situates the language within the wider world. It is a schematic grid dividing the world's languages for arbitrary convenience into 100 linguistic or geolinguistic zones.

Table 9-3. The world of languages divided into 10 macrozones.

This digit provides the first character of the 8-character proximity scale.

Code	*Name*	*Code*	*Name*
0	AFRICAN	1	AFRO-ASIAN
2	AUSTRALASIAN	3	AUSTRONESIAN
4	EURASIAN	5	INDO-EUROPEAN
6	NORTH AMERICAN	7	SINO-TIBETAN
8	SOUTH AMERICAN	9	TRANSAFRICAN

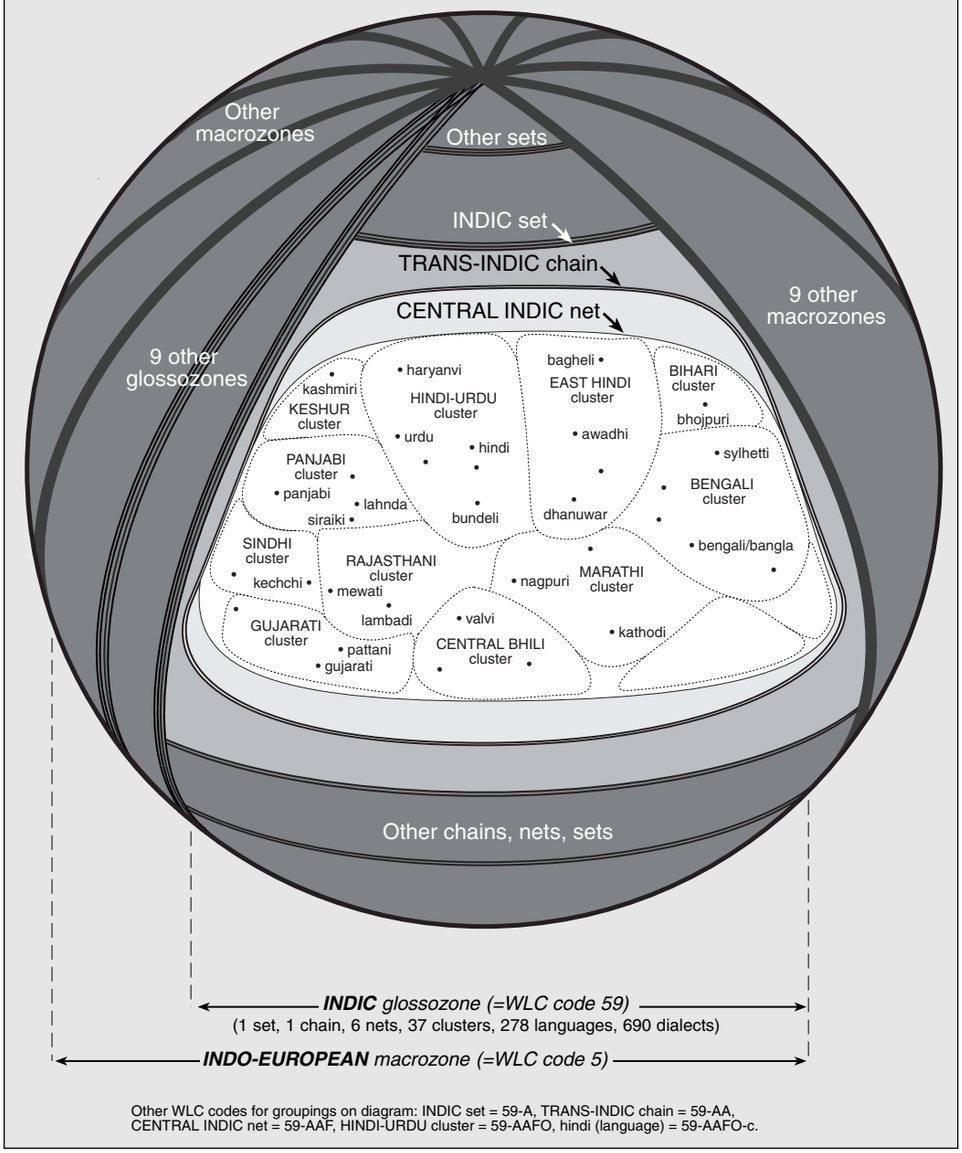
First character (Table 9-3)

The initial digit divides the entire world into 10 primary reference zones that we here call **macrozones**, which are major areas of linguistic affinity or continental reference. The digit thus assigns the language to one of the 10 macrozones shown below. Further, we recognize 2 types of macrozone (though we leave them uncoded, excluded from the coding system): of either geographic or linguistic character, and for which we coin respectively the terms geozone and phylozone. Firstly, a macrozone can be a geozone, defined here as one of the world's 5 geographical continents (continental land-masses), regarded as geolinguistic regions or continental language regions covering the 2,000 or so languages in the world which do not belong to the 5 recognized major language phyla or families. These geozones are shown in the lefthand column of Table 9-3 coded by an even digit 0, 2, 4, 6, 8. Alternatively, a macrozone can be a phylozone (one of the world's 5 recognized major language phyla or families, which contain over 70% of the world's distinct languages). These phylozones are shown in the righthand column of Table 9-3 coded by an odd digit 1, 3, 5, 7, or 9. These even and odd digits have been allocated in sequences which are

Graphic 9-1. Closeness or distance in relationships between any 2 or more languages on Earth, AD 2000.

The diagram sets out a schema illustrating the World Language Classification. Every language has a unique 7-character code, enabling immediate estimates to be made of its proximity to any other languages. Thus if 2 languages share the first six characters of the code, they belong to the same cluster or outer language. This means they share over 80% basic vocabulary of common human experience, differing only in under 20% of vocabulary.

The diagram illustrates this by zeroing in on one small but highly significant part of the world of languages—the Central Indic network in northern India.



Other WLC codes for groupings on diagram: INDIC set = 59-A, TRANS-INDIC chain = 59-AA, CENTRAL INDIC net = 59-AAF, HINDI-URDU cluster = 59-AAFO, hindi (language) = 59-AAFO-c.

both alphabetic and logical at the same time.

To recapitulate: The first character of each language code (or proximity scale) indicates either the geographic position of the language within one of 5 continents or geozones = initial even digit; or its linguistic position within one of the 5 major language phyla or phylozones = initial odd digit. The titles we use are as follows.

(a) The 5 geozones are arranged to be both alphabetical and also logical, i.e. geographically anticlockwise they cover the world. (b) Phylozones 1, 3, and 5 each represent an intercontinental phylum. (c) The 5 phylozones are named using the current universally-recognized family names except for Trans-African, which stands for the old widely-used Niger-Congo family, minus the Mande languages of the Niger basin and the Kordofanian languages.

Second character (Table 9-4)

The second digit divides each macrozone into 10 subordinate reference zones called **glossozones**, 10 to each macrozone. These are defined and constructed in order to fit, as closely as possible, the linguistic and geolinguistic realities of each macrozone. Each glossozone can be either geographic or linguistic in character. Again, we coin 2 more non-coded terms and call each glossozone respectively either a topozone (a local geographic or geolinguistic area, having only geographic significance, covering 2 or more geographically adjacent sets of languages), or a glossozone (a discrete linguistic grouping, having linguistic significance, covering one or more related sets of languages).

In conjunction with the first digit, the second digit therefore assigns the language to one of 100 glossozones, coded 00 to 99, each covering one or more sets of languages. In practice, virtually all glossozones have linguistic meaning, being sub-families of related languages; in general, it is found that each glossozone's constituent languages share something like 5% of basic vocabulary in common. The rest of the glossozones are topozones (or geomicrozones); each's constituent languages may share no vocabulary at all, or some may share over 10%.

The names of the 100 glossozones are set out in full in Table 9-4. Each set of 10 is there arranged not in alphabetical order but in geolinguistic order, that is in the order which best corresponds to linguistic reality, keeping glossozones of related sets together and observing, as far as convenient, the sequences of west to east and north to south.

Defining the 10 levels used in this classification (Table 9-5)

In addition to the above 2 levels of macrozone and glossozone, the classification utilizes 9 other levels. Before describing these in detail, it would be useful at this point to briefly outline the whole 11 levels and how they relate to each other. This is set out in Table 9-5. This table is given as a general guide to understanding a complex situation. It attempts to present the world situation and then to present and define 10 levels or categories or subdivisions or groupings of relationships among the world's languages. Of these, the 8 shown in boldface type are the major levels or categories used and coded in this classification, and a minor subcategory is added at the end to assist in understanding the classification. The percentages shown should be interpreted only as indicators of

Table 9-4. Ten macrozones and 100 glossozones covering the whole world.

This table sets out the first 2 digits of the 8-character proximity scale.

Macrozones. The 10 macrozones are set out below in large bold capitals. The left-hand pair of columns below, codes plus names, list the 5 macrozones which are geographic (also here termed and explained as geozones). The right-hand pair of columns, codes plus names, list the 5 macrozones which are linguistic (also here termed and explained as phylozones).

Glossozones. All the 2-digit codes represent 100 glossozones. The 9 followed by an asterisk (*) may be called 'open glossozones' in which their unity may not be external. Sets included within such a glossozone will normally, but not necessarily exclusively, be more closely related among themselves than any one of them will be with a set in another glossozone.

0 AFRICAN
00 Mandic
01 Songhaic
02 Saharic
03 Sudanic
04 Nilotic
05 Nilo-Sahelic*
06 Kordofanic
07 Riftic*
08 Nama-Tshuic
09 Kalaharic*

1 AFRO-ASIAN
10 Tamazic
11 Coptic
12 Semitic
13 Bejic
14 Mid-Cushitic
15 Para-Cushitic
16 Omotic
17 East Chadic
18 Biumandaric
19 West Chadic

2 AUSTRALASIAN
20 West Irianic
21 North Irianic
22 Madangic
23 South Irianic
24 Transirianic
25 West Papuasic
26 Sepic
27 East Papuic
28 Darwinic*
29 Pama-Nyungic

3 AUSTRONESIAN
30 Formosic
31 Hesperonesic
32 Mesonesic
33 Halyamapenic
34 Neoguineic
35 Neobritannic
36 Solomonic
37 Neocaledonic
38 West Pacific
39 Transpacific

4 EURASIAN
40 Euskaric
41 Uralic
42 Caucasic
43 Siberic
44 Transasiatic
45 East Asiatic
46 South Asiatic
47 Daic
48 Mienic
49 Dravidic

5 INDO-EUROPEAN
50 Celtic
51 Romanic
52 Germanic
53 Slavonic
54 Baltic
55 Albanic
56 Hellenic
57 Armenic
58 Iranic*
59 Indic

6 NORTH AMERICAN
60 Arctic
61 Athabaskic
62 Algonkic
63 North Pacific
64 Iroquo-Dakotic
65 Circumgolfic
66 Aztecotanic
67 Oto-Mangic
68 Mayanic
69 Mesomeric

7 SINO-TIBETAN
70 Bodic
71 Himalayic
72 Garic
73 Kukic
74 Miric
75 Kachinic
76 Rungic*
77 Lolo-Burmic
78 Karenic
79 Sinitic

8 SOUTH AMERICAN
80 Caribic
81 Arawakic
82 Tupic
83 Interoceanic
84 Pre-Andinic
85 Andinic
86 Chaconic
87 Matogrossic
88 Amazonic
89 Bahianic

9 TRANSAFRICAN
90 Atlantic*
91 Voltaic*
92 Adamawic
93 Ubangic
94 Melic
95 Kru-Grebic
96 West Akanic
97 Deltic
98 Benuic*
99 Bantuic

Table 9-5. Meaning of proximity scale by 10 levels or groupings of languages, 8 being coded.

Column 1: Examples of typical values of the coded components of the proximity scale (from 1-8 characters horizontally, representing 8 coded types of groupings vertically).
Column 2: Our 8 basic levels or categories of classification used in the scale, all coded, shown in boldface type, with 4 other subsidiary levels or categories which are uncoded or only partially coded, shown in medium type. The 6 levels in capital letters are the classification's anglicized cover names for groupings and families of languages.
Column 3: A statement about how much common basic vocabulary (%) is shared by any 2 of each level's major components; the percentage shows the minimum threshold.
Column 4: Validity or value of lexical relationships.
Column 5: A single adjective expressing degree of lexical similarity shared by all components within each level, or minimum shared by any 2 components.
Column 6: A single adjective expressing degree of intercomprehension (interintelligibility) shared by all components within each level, or minimum shared by any 2 components.

Code	*Level or category*	*Within this level-- (minimum threshold for % vocabulary shared)*	*Lexical relationships shared across this level*	*Lexical similarity shared by all within this level*	*Intercomprehension shared by all within this level*
column					
1	*2*	*3*	*4*	*5*	*6*
-	World	Any 2 macrozones share 0%	None measurable	zero	zero
0	**MACROZONE**	Any 2 glossozones share under 5%	Little quantifiable	nil	zero
01	**GLOSSOZONE**	Any 2 glossosets share 5% or more	Somewhat quantifiable	minimal	negligible
01-A	**GLOSSOSET**	Any 2 glossochains share 30% or more	Apparent to native speakers	occasional	acquirable
01-AA	**GLOSSOCHAIN**	Any 2 glossonets share 50% or more	Facilitate learning	partial	potential
01-AAA	**GLOSSONET**	Any 2 glossoclusters share 70% or more	Obvious to all	general	partial
01-AAAA	**GLOSSOCLUSTER**	Any 2 languages share 80% or more	Facilitate communication	sequential	general
01-AAAA-a	**language**	Any 2 dialects share 85% or more	Functional understanding	similar	adequate
01-AAAA-aa	**dialect**	Any 2 voices share 90% or more	Close understanding	close	mutual
-	variety	Any 2 idioms share 95% or more	Very similar	full	high
-	voice	One speaker, 100%	Identical speech forms	identical	complete

general order of magnitude, claiming accuracy perhaps only to plus or minus 10%.

The description of scales can now continue with an examination of the second of the 3 scales.

2. CLOSENESS SCALE
(a 4-capital lexical similarity index)

This second scale or index describes relationships and groups the language by its closeness to other languages. Comprehension and intelligibility between languages cannot be measured solely by lexical closeness, grammar, pronunciation, and discourse markers; but they can be measured by tested or reported intelligibility and systematic and careful linguistic comparison, including direct testing as well as lexicostatistics (comparison of word lists) and other pointers to the language's closest known lexical relationships).

The 4 levels represented by the next 4 capital letters of the code are shown in detail in Table 9-5 and also on the Quick-Reference Schema in Table 9-11. The full names for these 4 levels throughout the classification form its structure, consisting of 9,843 **cover names**, which are always given there as anglicized and capitalized names followed in every case by one word describing its level—glossoset, glossochain, glossonet, glossocluster—always appearing there in lowercase (noncapitalized) letters.

Third character (first capital letter)
This first letter assigns the language to a **glossoset,** defined as a grouping of languages each of which shares at least around one third (30%) of their basic vocabulary of common human experience, as measured by the use of phonologically related forms with the same meanings, using wherever possible Swadesh's 200-item comparative wordlist. Glossosets can be identified in the classification by (a) the abbreviated term 'set' after each's capitalized cover-name and (b) each's 3-character code.

Fourth character (second capital letter)
In cases where a glossoset is very complex, it may be divided among 2 or more subdivisions, each called a **glossochain**, a linked grouping or groupings within a glossoset. A glossochain shares within its component languages at least 50% of their basic vocabulary of common human experience. Glossochains can be identified in the classification by (a) the abbreviated term 'chain' after each's capitalized cover-name and (b) each's 4-character code.

Three varieties of glossochain are shown in the classification, all coded by a single capital as fourth character:

1. a chain (or, chain proper, being a multinet chain linking 2 or more nets);
2. a minimal chain, and
3. a monochain (absence of chain).

Fifth character (third capital letter)
This third letter assigns the language to a **glossonet** (implying a network of words and meanings). A glossonet is defined as a grouping of languages each of which shares at least around two-thirds (70%) of their basic vocabulary of common human experience. Glossonets can be identified in the classification by (a) the abbreviated term 'net' after each's capitalized cover-name and (b) each's 5-character code.

Sixth character (fourth capital letter)
In cases where a glossonet is very complex, this letter divides it into 2 or more subdivisions, each called a **glossocluster** (or outer language, or wider language, broad language, or tongue) which is a grouping of languages each of which shares at least around 80% or more of their basic vocabulary of common human experience. Glossoclusters can be identified in the classification by (a) the abbreviated term 'cluster' after each's capitalized cover-name and (b) each's 6-character code.

Three varieties of glossocluster are shown in the classification, all coded by a single capital as the sixth character:

1. a cluster (or cluster proper, being a multilingual cluster linking 2 or more languages);
2. a monolanguage cluster (an internal cluster of idioms or dialects within a single language); and
3. a minimal monolanguage cluster (a language or grouping with a marked absence of clusters of any sort).

An additional clarifying definition is that it may be helpful to regard our level 'glossocluster' as 'broad language', and our level 'language' as 'narrow language'. A typical glossocluster would consist of several narrow languages including any literary languages and any colloquial or popular or vehicular (or even ecclesiastical) languages.

3. COMPREHENSION SCALE
(a 2-miniscule intelligibility index)

The third scale or index deals with the internal relationships between a language and its dialects. This final part of the proximity code has 2 miniscules (lowercase letters, a-z). It describes which speech forms are languages eligible for their own literature and broadcasting, and which speech forms are dialects interintelligible enough not to be eligible for their own literature and broadcasting. Obviously other considerations, sociocultural and not purely linguistic, come into the evolution of such eligibility, but it forms a useful and usable approximation to reality.

Again, these 2 levels are shown in detail in Tables 9-5 and 9-11.

Seventh character (first lowercase letter)
This letter, a miniscule (lowercase), identifies the individual **language** (recognized and named as such by its speakers), which usually does not share more than 85% or more of basic vocabulary (of common human experience) with other languages. It is defined as needing, requiring, or having its own separate literature and broadcasting.

Languages can be identified in the classification's listing by (a) being in lowercase type and (b) each's 7-character code.

Eighth character (second lowercase letter)
This final letter, a miniscule, identifies a **dialect** sufficiently close to its parent language and sister dialects to not need or justify separate literature or broadcasting. A dialect is defined here as a speech form whose component varieties and idioms (if any) share 90% or more of basic vocabulary. Dialects can be identified in the classification's listing by each's 8-character code (these being listed here only on CD).

The ordering and listing here of glossosets within glossozones, of glossochains within glossosets, of glossonets within glossochains, of glossoclusters within glossonets, of languages within glossoclusters, and of dialects within languages, are not alphabetical but geolinguistic (approximating to linguistic reality). Codes are then applied at the left of the listing in strict alphabetical sequence.

Example:
LANGUAGE CODE = 2 digits + 4 capitals + 2 small letters
Illustration: 52-AAAD-ra

macrozone = Indo-European phylozone
glossozone = Germanic zone
glossoset = Germanic set
glossochain = Nordic chain
glossonet = Nordic net
glossocluster = Nordic East cluster
language = svea-svensk (Swedish)
dialect = halsinglandsk

Lexicostatistics as one of several guides
As mentioned above, to establish degrees of intelligibility, and lack of intelligibility, between languages needs empirical testing followed by systematic and careful linguistic comparison. Lexical similarity or lexicostatistics normally only measures a small sample of the total vocabulary, and for this reason should be treated with caution. Nevertheless, they are valuable as an indicator in the absence of more detailed testing, and are certainly a better guide than nothing at all. Table 9-5 sets out their meaning in the present classification.

In dealing with relationships between languages, it is often possible to measure how much basic vocabulary 2 languages share in common. Lexicostatistics takes a scientific approach by using a standard word list (Swadesh's 200-word list, or a similar one). The result for 2 languages is expressed as a percentage of words thus shared. In the present analysis, these percentages as we define them have a broader meaning than merely sharing basic vocabulary. They cover vocabulary, but also to some degree phonology (accent, pronunciation), grammar (morphology and/or syntax), discourse structure. They measure: closeness, intercomprehension, interintelligibility, similarity. For ease of reference, however, we usually abbreviate the meaning of the percentages to shared vocabulary, although we fully recognize the limitations of lexicostatistics as a measure of closeness of languages.

To sum up, these 8 levels measure degrees of proximity or closeness between speech forms. The percentage ranges of each level represent approximate areas of magnitude and symbolize the cumulative effects of similarities which are not only lexical but also

Table 9-6. Implications of percentages of closeness.

90%-95%	Two speech forms sharing this much vocabulary are sometimes called idioms; they understand each other mutually and more than adequately; they differ one from another principally in terms of pronunciation or "accent".
85%-90%	Functional or adequate intercomprehension exists between the 2 speech forms for communication, conversation, or use of literature.
80%-85%	Relationship between 2 languages sharing these amounts facilitate communication and provide general intercomprehension.
70%-80%	Partial or general interintelligibility or intercomprehension, often functional, exists between the 2 languages or idioms, or can be readily acquired, and this is obvious to all.
50%-70%	Languages are close enough to facilitate acquisition of them as additional languages.
30%-50%	Relationships between 2 languages or other speech forms are apparent to their speakers, and indicate that knowledge of the other is acquirable as a second language.
5%-30%	Lexical relationships, though still somewhat quantifiable, become less obvious to speakers and less useful for acquisition.
1%-5%	Lexical calculations become less quantifiable and less reliable, and grammatical relationships become relatively more important for classification.

grammatical, morphological, syntactic, and phonological.

It needs to be understood that all such resulting percentages should be treated as approximate, giving only the general order of magnitude of the relationship rather than exact figures fully valid for comparative purposes. Moreover, for many languages, no such tests or detailed calculations have yet been conducted. In these cases, estimates have been utilized here.

Understanding 8 percentage levels (Table 9-6)
Before presenting the classification in detail, some observations can be made about different levels of this numerical relationship, as follows. Table 9-6 shows on the left a percentage range within this concept of 'closeness', as measured by common basic vocabulary shared by 2 speech forms; on the right, some explanatory comments.

Using proximity codes to estimate intercomprehension (Table 9-5, -6, -7)
Yet another way of understanding this schema, or of presenting and verbalizing our classification with its proximity scales, is as follows. Any 2 or more languages or speech forms or aggregates thereof which

Table 9-7. Moving from codes shared to intercomprehension.

If languages share these code characters	*they also share this vocabulary*	*and they all become assigned to this level..*
No first character	0%	The World
First character only	0-5%	a MACROZONE
First 2 characters only	5-30%	a GLOSSOZONE
First 3 characters only	30-50%	a GLOSSOSET
First 4 characters only	50-70%	a GLOSSOCHAIN
First 5 characters only	70-80%	a GLOSSONET
First 6 characters only	80-85%	a GLOSSOCLUSTER
First 7 characters only	85-90%	a **language**
All 8 characters	90-100%	a dialect

share proximity codes or parts of codes (as shown in first column of Table 9-7), also share basic vocabulary within a clear percentage range (given in the second column), as a result of which they are classified as being within the same level or category (in the third column).

Comparing 2 speech forms
The justification for Table 9-7 needs interpretation and explanation. This is done by greatly expanding words into detailed statements, and this is set out in Table 9-8.

Summarizing this whole schema of coding (Table 9-8)
The briefest way, and probably the easiest way, to comprehend the classification and codes is to arrange the schema in 8 vertical blocks corresponding to the 8 characters of the proximity code, with closeness ranged vertically from 0% at the bottom to 100% at the top. This is set out in Table 9-8. Thus, for example, a glossoset consists of languages which share at least 30% basic vocabulary; some glossosets consist each of a single glossonet only, and some of a single language within that glossonet, whereupon obviously intercomprehension in the glossoset is 100%.

Spacing and rules (Tables 9-9 and 9-11)
The presence of thin rules (lines) and/or linespaces across the page at any point throughout the classification indicates a break or end in intercomprehension between adjacent nets, or sets, or microzones, or macrozones, above and below the rule or linespace. All blocs of languages listed together without rules or linespaces can therefore be regarded as generally similar lexically, and as partially or generally interintelligible or intercomprehensible, i.e. as each a single language cluster whose languages share at least 80% basic vocabulary.

A full linespace without a rule normally signals the beginning of a new cluster whose constituent languages share at least 80% basic vocabulary. Two languages separated by one such linespace share only from 80% or more basic vocabulary.

Rules plus linespaces signal the ends of occasional lexical similarity and of any interintelligible blocks of languages, and the beginnings of unrelated new parts of the classification.

The whole combination of rules and linespaces therefore conveys the following meanings.

Setting out the various levels on paper (Table 9-9)
These levels of classification enable the user to divide up lengthy listings of thousands of languages into understandable and manageable blocs, and to display them for immediate understanding as follows:

a. Languages within a glossocluster (which share over 80% vocabulary) are shown listed on adjacent lines with no blank lines or linespaces between them; their 7-character codes are given on the left, their names on the right. This type of listing is the major one in the whole classification, and the reader should familiarize himself or herself with its 'solid' appearance and its position on the page so that it can be identified at once. Around this listing, to its left or (occasionally) within it, are various groupings of languages using cover-names in capital letters.
b. GLOSSOCLUSTERS (groupings of related languages) are likewise shown on adjacent lines but separated from each other by one linespace.
c. GLOSSONETS (larger groupings of geoclusters) are shown as separate blocs separated from each other by one line.
d. GLOSSOCHAINS (chained or linked or related glossonets) are shown separated from each other by 2 rules.
e. GLOSSOSETS (larger groupings of glossochains) are shown separated from each other by 3 rules close together.
f. GLOSSOZONES (groupings of glossosets) are shown separated from each other by 4 rules close together.
g. MACROZONES (groupings of glossozones) are shown separated from each other by 5 rules close together.

This layout is shown in Tables 9-9 and 9-11.

FEATURES OF THE CLASSIFICATION

The following are a series of notes, explanations, and technical comments on the features of the actual classification itself in the book version *Linguasphere* with its long lists of names of cover-names, languages, and other speechforms.

Cover-names of 6 kinds
Cover-names are anglicized names in use for our major levels or categories. They are always shown throughout the classification in capital letters. Cover-names are always shown preceded by a code which classifies the 6 basic levels or categories or groups of languages: *macrozones, glossozones, glossosets, glossochains, glossonets, glossoclusters*. These 6 basic categories are globally applicable and so are found regularly and consistently throughout the classification in all parts of the world. They are basic to the whole classification with its system for totalling or subtotalling the various statistical categories of macrozones, glossozones, glossosets, glossochains, glossonets, glossoclusters, languages, and dialects. They follow regular rules of coding and each's code describes it and its place in the classification.

Types of language names
After each 7-character code on the left, the relevant language names on its line on the right are presented (in the *Linguasphere* book) in the following standardized order, using parentheses, square brackets, periods, colons, commas, boldface, italics, or medium type. Note that all names of languages and dialects are shown with lowercase initial letter, this being the universal practice by the speakers themselves (except for English, and all anglicized names). The dozen or so various distinct elements of each entry are as follows:

> reference name, being one or both of 2 elements: (1) anglicized name with initial capital(s); and/or (2) autoglossonym, consisting of root-name in medium type, plus affixes if any in medium type (then in parentheses phonetic representation in italics, anglicized name(s) if any, lowercase names as used in French, German, Spanish, Russian or other major literature; other alternate names or close names or versions): after a colon, dialects, varieties and regional variants (and their anglicized and alternate names in parentheses); [Notes on bridges, continua, extended units, or relations with other codes]; Location of language in particular countries, provinces, regions, islands, etc.

Table 9-8. Comparing the closeness or proximity of 2 or more languages or other speech forms.

To find out the comparative closeness of (or distance between) any 2 speech forms, write the 2 codes one under the other, then begin at the left and see how far to the right the characters are the same, then count them (how many of the 8 code characters in sequence from the left are the same for both). The following 9 conclusions can then be drawn, arranged firstly in ascending order of closeness, then the same 9 in descending order.

In ascending order of intercomprehension or closeness:
- 2 speech forms which have different first characters share no vocabulary (and share zero intercomprehension).
- 2 speech forms which share only the first character share under 5% vocabulary (zero intercomprehension).
- 2 speech forms which share only the first 2 characters share 5-30% vocabulary (negligible intercomprehension).
- 2 speech forms which share only the first 3 characters share 30-50% vocabulary (and acquirable intercomprehension).
- 2 speech forms which share only the first 4 characters share 50-70% vocabulary (potential intercomprehension).
- 2 speech forms which share only the first 5 characters share 70-80% vocabulary (partial intercomprehension).
- 2 speech forms which share only the first 6 characters share 80-85% vocabulary (general intercomprehension).
- 2 speech forms which share only the first 7 characters share over 85% vocabulary (adequate intercomprehension).
- Any 2 speech forms which share the same 8-character code share over 90% vocabulary (mutual intercomprehension).

In descending order of intercomprehension or closeness:
- Any 2 speech forms which share the same 8-character code share over 90% basic vocabulary (mutual intercomprehension).
- 2 speech forms which share only the first 7 characters share over 85% vocabulary (adequate intercomprehension).
- 2 speech forms which share only the first 6 characters share 80-85% vocabulary (general intercomprehension).
- 2 speech forms which share only the first 5 characters share 70-80% vocabulary (partial intercomprehension).
- 2 speech forms which share only the first 4 characters share 50-70% vocabulary (potential intercomprehension).
- 2 speech forms which share only the first 3 characters share 30-50% vocabulary (acquirable intercomprehension).
- 2 speech forms which share only the first 2 characters share 5-30% vocabulary (negligible intercomprehension).
- Two speech forms which share only the first character share under 5% vocabulary (and zero intercomprehension).
- Two speech forms which have different first characters share no vocabulary apart from the occasional loanword (and share zero intercomprehension).

Table 9-9. Intercomprehension between languages set out by rules and linespaces.

The degree of interintelligibility between nearby languages is set out systematically by the following sequence of rules and line-spaces:

Example:
German: **deutsch** (alemão, allemand, nemetski, tedesco, tudesco): hochdeutsch (Standard German, High (Upper) German) [for Regional German see 52-ABCF]; Location Germany, Austria.

Note that alternate names are usually given within parentheses (round brackets). Any lists of such alternate names within parentheses, which are all equivalent, are shown separated by commas. By contrast, names of subdialects are given (on the CD) outside of any parentheses. Any lists of such subdialects, shown outside parentheses, describe different dialectal variants which are dissimilar or nonequivalent or nonidentical even though also shown separated by commas.

Reference-names
These need additional explanation. In this classification every language is given its own reference-name, specifically for use in large-scale comparative contexts. This is the first name on the line after each idiom's code, where unhyphenated. Usually, this is also the root-name—the shortest form of the language's own name for itself. Reference-names are shown in 2 forms: (1) if it is an anglicized name, it is shown in medium type with initial capital letter(s), or (2) if a reference-name is also the autoglossonym, or part of an autoglossonym, it is also shown in medium type. A complication is that small numbers of unrelated languages in different glossozones use identical root-names, and so, to avoid confusion in such cases we add a numeral (-1, -2, -3, etc.). This becomes part of this usage of the reference-name.

Autoglossonyms
In many cases, such as with the great majority of Bantu language and dialects, affixes (prefixes or suffixes) meaning 'the language of' are widely used attached to root reference-names. Each such longer name, shown hyphenated in this classification, is the speech form's autoglossonym—the name used by speakers themselves to identify their own language. This is given for every language and dialect except where unknown.

Some reference-names as just explained have numerals appended (-1, -2, -3, etc.). To get any particular language's actual autoglossonym as used by its speakers, one should of course remove this number.

Occurrence of homonyms
Homonyms or duplicate language-names are of 2 sorts: (1) those which result from the splitting of a language into 2 or more languages or dialects, or from the splitting of an ethnic group between 2 or more languages, both of which cases refer to different applications of the same name, related but distinct; and (2) those which result from the coincidental use of a similar name in different parts of the world. Homonyms of type (1) are found in the same set or in the same country and are shown in this classification followed by an arabic numeral, thus: malinke-1, malinke-2, etc. Homonyms of type (2) are relatively rare and are not differentiated in situ, but the reader is alerted to their existence in the indexes at the end of the *Linguasphere* volume.

Order of names
Language names are listed and set out vertically, not in alphabetical order, but in the manner which best represents linguistic reality, including chains of interintelligible languages where they exist, and/or geographic sequences. Cover-names of macrozones, glossozones, glossosets, glossochains, glossonets and glossoclusters are all given in capital letters. Those for macrozones happen to be coded in alphabetical order and so are listed throughout in this alphabetical order. Other cover-names, however, are listed not in alphabetical order but in a geolinguistic (lines of relationship) order or geographic order (west to east, north to south) corresponding to linguistic reality.

Chains and extended nets
Linguistic reality often consists of long chains of related language nets or languages in which adjacent nets or languages are lexically much closer to each other than those at the beginning of the chain are to those at the end. These realities are shown by the order of listing in the right-hand column, amplified by adjacent textual remarks in square brackets, but the requirements of coding mean that such chains are then subdivided into nets and languages/idioms as defined and presented here. These chains do not always or necessarily run in any single direction such as north to south. Sometimes, as with the Great Bantu Chain, chains may start off in one direction only to double back on themselves to near their point of origin. Sometimes, therefore, a more useful analogy than 'chain' is 'chain-link' or a 'chain-link fence', since the links run in all directions, weaker in some directions and stronger in others.

Transitional nets and linked nets
These 2 technical usages should be noted: (1) a 'transitional net' is a net which is seen to be located between 2 sets or subsets (such as 2 streams of Bantu languages) with relationships to both; and (2) a net is said to be 'linked with' one or more adjacent or nearby nets, as listed, when it shares with them a high degree of intercomprehension placed at over 65% of basic vocabulary in common.

Use of hyphens
Many names of languages in this classification contain one, 2 or even very occasionally 3 hyphens within the letters of the name itself (excluding at this point consideration of hyphen plus number, which is dealt with in the paragraph 'Occurrence of homonyms' above). A single hyphen usually separates the root-name from an affix (prefix or suffix) meaning 'the language of..'. Two hyphens usually delineate 2 levels of prefix (i.e. prefix and preprefix), with essentially the same meaning. Usual practice among language speakers, linguists, governments, churches, Bible societies et alia is to write these names with an initial capital and without hyphens. Usual practice also is to include affixes in some circles, but to omit them in others. All this makes it difficult or impossible for persons not familiar with the languages to identify them, to use the literature concerning them, or to find them in indexes. Our final recommendation to all users therefore after spending many years on this problem is (1) to follow existing practice and omit all hyphens when writing names for limited local contexts, but (2) to retain all hyphens when using numerous names in large geographical contexts (countries, continents, the world). For large-scale comparative use of language names, our recommendation is (3) to omit both hyphens and affixes, and to use root-names as reference-names, but at the same time to provide nearby at least a single listing of equivalents (full hyphenated names with root-name of each, and vice versa).

Dialects
Dialects of languages, including regional variants, are presented in *Linguasphere* and in the CD version in 2 distinct and different ways. (1) If they are sufficiently distinct from their parent language to warrant being regarded as separate languages/idioms (because of sharing less than 90% vocabulary), they are given a separate line each and a separate 8-character code each. In cases where their status is dialectal, this is shown by their names in medium type, since their associated autoglossonym is usually the parent language's autoglossonym, already given above either on its own line or often on that of the language net's or cluster's name. (2) If not sufficiently distinct from their parent language (because of sharing over 95%), they are simply listed after the same code as the parent language and after a colon (:) following it. Note that lists of dialects are never placed within parentheses and are shown separated by commas but all refer to different nonidentical dialectal variants. (Alternate names, by contrast, are shown within parentheses, separated by commas). Note also again that all carryover lines are indented 2 or more spaces to indicate continuation from the previous line.

Country names and geographical names
The countries and localities in which nets and languages are mainly or primarily located are given (in the electronic database version) after the italicized word Location; if several countries are listed, this is the order in which languages/idioms below it are shown. If component idioms are spoken in a different country or countries, this fact is noted on their own lines after their words Location ('also in... '). By this means, the dominant location or countries of every idiom can be seen by the reader.

This geographical locating is not done for every idiom or net, in the printed version of the classification: it is only done where necessary for clarification, or advisable to avoid confusion. However, in the computerized version of the classification, every idiom has a variable in which location is stated exactly (country, province, island, etc.) both (a) in words and through online maps.

All other country names or place names or geographical names used in the classification, but which do not follow a word Location, are adjectives describing differing varieties of languages/idioms or dialects/varieties and do not refer simply or primarily to places where a particular idiom or dialect is spoken.

Standard or literary languages
Many languages have a recognized standard form or version, often referred to as the literary language or the literary or written version of the language. In this classification, a clear distinction is made between 2 types of standard language:

(a) those standard languages which have been created as a compromise (e.g. Standard Italian, Standard Shona), which are listed as idioms in their own right, and which (at least initially) are spoken by no one as a natural mother tongue; in all these, the word 'Standard' is always shown preceding the idiom's name; and

(b) those already existing idioms which have been chosen as standards for a unit, or part of a unit, in which case they are listed with 'standard' following the idiom's name (e.g. Acoli Standard, West Dinka Standard).

Comments in situ
Additional short explanatory comments are given [in square brackets, thus] at a number of points in the listings. These pinpoint bridge languages, which provide a bridge between 2 adjacent nets; chains or continua or linked languages or nets or sets which although not linked by sufficient basic vocabulary in common (30% minimum) to be shown related on these definitions are nevertheless somewhat related (e.g. with 25% in common); the presence of extended nets; relations to other codes; and lower-limits nets or sets in each of which basic shared vocabulary is at its lowest limit on the definition and which is therefore a candidate for further subdivision into 2 nets or sets.

Further conventions employed
1. Quotes are placed around a language name in use or recorded in the literature but to be discouraged because of racist or other pejorative or unscientific connotations. Quotes around numbers, e.g. 'A25', refer to names in other existing clas-

Table 9-10. World totals of languages and cover-names.

10	language MACROZONES,
100	language GLOSSOZONES,
684	language GLOSSOSETS,
1,403	language GLOSSOCHAINS,
2,684	language GLOSSONETS,
4,962	language GLOSSOCLUSTERS,
13,511	**languages** with some 10,000 distinct and different autoglossonyms,
30,000	dialects, and
50,000	speech-form names of all kinds, including alternates.

Table 9-11. Quick-reference schema explaining the World Language Classification.

CODES (PROXIMITY SCALES AND NAMES)

Levels and codes
The 8 pairs of lines on the left below stand for the 8 main or usual or standard worldwide categories or levels or meanings or codes. The resulting proximity scale is then composed of 3 shorter more focused scales (named below in bold italic capitals at left margin) which are shown separated by 2 hyphens for ease of use (as with telephone numbers). Together they make up an 8-character code or proximity scale for a language or for any dialects.

Meaning
Note that the one, 2, 3, 4 or 5 rules (or one linespace) across the page shown below introduce new categories as explained on previous pages and represent differing barriers defining varying levels of lexical closeness or distance (implying varying degrees of intelligibility or intercomprehension or the absence of them). The %s represent *minimum threshold* figures for closeness (shared vocabulary, etc).

WIDER SCALE or reference grid (2 digits):		
First character (0-9) *Example of code:* 0	=	**LANGUAGE MACROZONE** (geozone or phylozone; each is a grouping sharing little or no basic vocabulary apart from handfuls of loan words, i.e. with zero, nil, or nonexistent intercomprehension)
Second character (0-9) *Example of code: 01*	=	**LANGUAGE GLOSSOZONE** (topozone or glossozone; abbreviated to 'zone'; each is a grouping sharing limited basic vocabulary (a minimum of 5%), i.e. with negligible, minimal, marginal, scant, slight, or limited intercomprehension).
CLOSENESS SCALE or similarity index (next 4 capital letters):		
Third character (A-Z, uppercase) *Example of code: 01-A*	=	**GLOSSOSET** (abbreviated to 'set'; a grouping sharing over 30% basic vocabulary, i.e. acquirable intercomprehension); note that a glossoset may sometimes have within it a chain of languages, or an extended unit, i.e. chain of units where there is closer relationship between adjacent members than there is between beginning and end.
Fourth character (A-Z, uppercase) *Example of code:* 01-AA	=	**GLOSSOCHAIN** (or chain or chains, or group or groups of linked units: it signifies a subdivision of a large or very extensive glossoset; it is a grouping sharing over 50% basic vocabulary, i.e.potential moderate, or consecutive intercomprehension).
Fifth character (A-Z, uppercase) *Example of code: 01-AAA*	=	**GLOSSONET** (abbreviated to 'net'; a grouping sharing over 70% basic vocabulary, i.e. partial intercomprehension); usually divisible into a number of glossoclusters, which are listed next, and which can be identified by the single linespace before each's name (and the term 'cluster' after its name).
Sixth character (A-Z, uppercase) *Example of code:* 01-AAAA	=	**GLOSSOCLUSTER** (or outer language, or broad language, or wider language, or tongue, or a cluster of languages; abbreviated to 'cluster'; a grouping of related languages sharing over 80% basic vocabulary, i.e. sharing general intercomprehension); 3 varieties are listed (but not separately coded): *cluster* (of languages; multilingual, i.e. with 2 or more languages), *monolanguage cluster* (internal cluster of idioms, within a single language), *minimal monolanguage cluster* (marked absence of clustering).
COMPREHENSION SCALE, or intelligibility index (last 2 letters):		
Seventh character (a-z,lowercase) *Example of code:* 01-AAAA-a	=	**language** (or inner language, or narrow language; a speech form or grouping of speech forms widely recognized by people and observers as a "language", based on political reality, ethnic or social affiliation, or literary history, or availability of literature or scriptures, et alia; usually a grouping of interintelligible speech forms here termed dialects and/or idioms sharing over 85% basic vocabulary and which are mutually intelligible to each other, i.e. with adequate intercomprehension. This reference name for a language is always shown in lowercase type, and is given in 2 forms: in most cases (a) as the **autoglossonym**, with in medium type separated by hyphen any affix (prefix or suffix meaning "the language of") where used; or, in cases where an autoglossonym does not appear to be in use, or where it is not known, (b) as an anglicized name given in boldface lowercase type but with initial capital(s). Any alternate or variant names then follow in parentheses (given on CD version only).
Eighth character (a-z, lowercase) *Example of code:* 01-AAAA-aa	=	dialect (a sequence or grouping of varieties or subdialects or idioms, sharing over 90% basic vocabulary, i.e. sharing mutual intercomprehension); shown only on CD version.
Uncoded	=	variety, subdialect, or idiom (a speech form identified and recognized as distinct by speakers, sharing (variety) more than 95% or (idiom) more than 99% or (voice) 100% vocabulary with other adjacent varieties or idioms).

sifications in widespread use, in this case Guthrie's widely-quoted numbers for Bantu languages.

2. The recommended reference-name for each language is the root-name shown in medium type which is often part of the autoglossonym (people's own name for their language), and is the first name cited and printed in lowercase type, but with affixes (prefixes or suffixes) if any (usually translating 'the language of..') in medium type. The recommended reference-name is thus the autoglossonym minus any affix. For comparative purposes, the root-name alone by itself would serve better than the autoglossonym. Even then, affixes should be retained with reference-names in cases where they serve to distinguish 2 otherwise identical language names. Where no autoglossonym is given first, the first name shown (which is in medium type) can be regarded as the recommended reference-name, and its autoglossonym is that shown for the net or cluster as a whole.
3. Although many languages use different prefixes to distinguish between the name of a language and the name of its speakers (e.g. the Baganda (people) of Uganda speak the language luganda), for other languages the practice is not so clearcut but is mixed. Since normally-used nomenclature is being recorded here, a number of people-names are included because, although not correctly language-names, they are used by adjacent peoples as language-names (this is frequently the case with Bantu names). Many other languages make no distinction but use the same name for both (e.g. the English speak English).
4. Initial capitals are used only to distinguish anglicized names (which are always given in medium type) or geographic (including directional) names, e.g. North Kono (always given in medium type). If no anglicized name is given, then normal anglicized usage can be assumed to be the same as the autoglossonym but with first letter capitalized and hyphens removed.
5. Where an autoglossonym covers a range of divergent dialects, sufficiently distinct to justify separate codes, it is often cited for the first of these only and can be assumed to cover any or all items following on succeeding lines as anglicized names.

Language totals
The total number of distinct languages in any net, or set, or microzone, or macrozone, or in the whole world, is defined here as equal to the total number of different 7-character proximity codes shown. Thus to find the total of languages in the Indo-European macrozone, simply total its distinct or different 7-character codes (which each terminate in a lowercase letter): this particular total comes to 300.

World totals
The grand total of languages for the whole world, defined in this way, comes to 13,511. This can be expanded. Our classification yields the global totals as shown in Table 9-10.

LAYOUT OF THE CLASSIFICATION

Languages and their groupings are listed in the classification after their respective codes. The following listing summarizes the various usages and features of this listing.

Typefaces
Throughout this classification, type styles and formats have the following standardized meanings:

- The first name on each line, immediately after its code on the left, is this classification's standard, definitive reference-name. This may be the anglicized form (with initial capital letter) in cases where such a form is widely known and used. If no anglicized form is in use, the reference-name is the autoglossonym, always shown here in lowercase type.
- All names in CAPITALS are anglicized cover-names for the first 6 levels: MACROZONE, ZONE, SET, CHAIN, NET, CLUSTER. In English text usage outside the classification, they should be written with an initial capital followed by lowercase letters ('Afro-Asiatic', 'Indo-European', etc).
- The great majority of names in the classification are the standard reference-names, often anglicized, for **languages**, and dialects. Languages (or 'narrow languages') may be instantly recognized by (a) each always having a 7-character code, (b) always being shown in lowercase letters, and (c) always forming the main vertically-aligned listing of names, the great majority of which are autoglossonyms shown in lowercase type.
- All names in lowercase are **autoglossonyms** (each being a people's own name for their own language), except dialects. These are always the first version of the reference name; they may be followed by any anglicized reference name for the language.

2. COMPILING THE LINGUAMETRIC DATABASE

Utilizing this classification, data were now added covering all of the church's worldwide ministries that focus on languages. These cover: Scripture translation and distribution, Christian publishing, Christian literature, books and periodicals, broadcasting and telecasting, audiovisual approaches, ministries to the blind, the deaf, the handicapped, with special reference to children of all ages and also to nonliterates. Of particular interest is Table 9-12, Names for God in 900 languages.

This extensive database is available on CDs related to David Dalby's *Linguasphere: register of the world's languages and speech communities,* including the forthcoming CD, *World Christian database.* It is partially reproduced here as Table 9-13 which lists all 13,511 languages of the world (but not dialects) together with many of the ministries listed above.

Of many new discoveries that flow from this material, one of the most significant is the relationship between languages with direct ministries (e.g. the Zulu Bible, or the 'Jesus' Film in Hindi) and their thousands of closely related languages. This can be stated in single-sentence form: Every language (also termed 'inner language' or 'narrow language') listed here benefits directly from language ministries in any other language shown as within the same language cluster (also termed 'outer language' or 'broad language'). Thus at the end of Table 9-13 it can be seen that a number of languages around Zulu, and within its cluster, in practice have access to the Scriptures. Though termed here 'indirect access', it is nevertheless adequate access. This analysis terms this further here by stating that a language has access to, or understands or uses, a *near-Bible*. This role of near-scriptures—near-Bible, near-New Testament, near-gospel, near-selection, near-'Jesus' Film, near-audio scripture, near-Braille scripture, near-signed scripture, near-broadcast, et alia—clearly revolutionizes the extent to which Gutenberg's original vision in inventing printing with movable type (to see the Holy Scriptures disseminated and available to all the peoples of the world) is being realized today.

CODEBOOK FOR LINGUAMETRICS TABLE 9-13

The 280 pages that follow set out the 13,500 distinct and separate languages spoken during the 20th century. Data for each language occupies one single line across one page only. Note that the unit 'a language' is a single entity independent of any country or countries it is spoken in. By contrast, the unit 'a people' refers to one ethnocultural ethnolinguistic people residing in one particular country; spread over 10 countries, it would count as 10 peoples.

Extinct languages
Note that some 1,000 languages spoken in the 20th century are now extinct. This is demonstrated by the firm '0', zero, in the population columns 5 and 6. Some 400 others are either nearly extinct (dying, with under 10 speakers), or endangered (under 100), or moribund (under 1,000).

Little-known languages
Note also that numerous languages have a blank space in those 2 columns, meaning that no population figure is assigned to them. At this stage in the evolution of this complex database, their populations (mostly unknown or relatively unknown in the literature because as yet unstudied by linguists or anthropologists) are combined with other better-studied and better-known languages within their language cluster. In many cases, also, their situation is one of duplication—their 'speakers' can also be said to be at the same time speakers of other closely-related near-languages.

For full understanding of the origins, compilation from 1975-1999, and rationale for this Linguasphere/World Language Classification's categories and codes, consult the definitive publication by David Dalby, *Linguasphere*. The version employed here is a slightly earlier published version differing only in the codes assigned to a few languages.

Column and codes
The following brief listing will enable the reader to use the lengthy compilation of data for 13,511 languages that follows. For more explanation, consult Part 3 "Codebook".

Note that most languages have several alternate names or spellings; these, together with several thousand dialects, are not given here but are published in Dalby's *Linguasphere*, and on related CDs.

Note also that although almost all reference names here are written out in full, a number as part of their name end with a capital letter, or occasionally a lowercase one, and period, with the following meanings:

A = proper	P = peripheral
C = Central	N = North
E = East	S = South
F = formalized, revived	T = traditional
G = generalized, standard	U = urban
H = historical	V = vehicular
M = Middle	W = West

Names for Scripture languages
Column 12 in this table records each language's biblioglossonym, if any exists. This is the official or formal name given as the language's name in connection with its translation of the Christian Scriptures. Usually, it is an anglicized name ('French', 'German', 'Zulu', etc) but in many cases it is named by the speakers themselves who term it in their own language ('français', 'deutsch', 'isiZulu', etc).

The 2 major Scripture translation agencies—United Bible Societies (UBS), and Wycliffe Bible Translators/Summer Institute of Linguistics (WBT/SIL)—often have confusingly different language names for the same autoglossonym (own language). Sometimes it is simply a small difference in spelling, but often it is a completely different name: WBT utilizes anglicized names while UBS uses many vernacular names. In the majority of cases, this database shows the UBS name.

The disadvantage of this difference is that the 2 biblioglossonyms in such cases are incompatible from the standpoint of each other's computers, databases, and hence search capability. Even if the difference is a single letter, ordinary programs will not note that these refer to the identical language.

This distinction is recorded in column 12 by means of an asterisk (*) attached to a biblioglossonym. *Its presence* means: either (a) in addition to this biblioglossonym, there is at least one other biblioglossonym (usually that used by WBT/SIL) not given on this printout; or (b) 2 or more biblioglossonyms are or have been used by one of the 2 agencies; or (c) in addition to the main biblioglossonym shown, one or more of this language's dialects have their own translations (not recorded here) and thus their own distinct biblioglossonyms.

Likewise, the *absence* of an asterisk means either (a) the 2 agencies use an identical biblioglossonym for the language under consideration, or (b) only one agency knows of or uses a biblioglossonym at this point.

Meanings of columns in Table 9-13
Column

1. Language code

REFERENCE NAME
2. Cover-name (in capitals)
   Autoglossonym (own name for language)
3. Countries where significantly spoken or used
4. Peoples using this language as mother tongue

MOTHER-TONGUE (NATIVE) SPEAKERS
5. In AD 2000
6. In AD 2025 (assuming current trends)

MEDIA
7. Countries broadcasting Christian programs in this language:

Code	*Meaning*
0	No broadcasts
1	Local only or in same-cluster language
2	National, within this country
3	External broadcasts from this country
4	International, from one foreign country
5	Plurinational, from 2–4 countries
6	Multinational, from 5–9 countries
7	Multicontinental, from 10-20 countries
8	Global broadcasts from 20 or more countries

CHURCH among language's native speakers
8. Affiliated Christians (AC), % of population
9. Evangelization, E (% of population evangelized)
10. Worlds A/B/C: location of most speakers

SCRIPTURES
11. Scripture Translation Status (a scale 0-92): see details at end of Part 1.
12. Biblioglossonym (official name of Scripture translation, if published or under way); sometimes the anglicized name is preferred by speakers, sometimes the autoglossonym)
    Note meaning of any asterisk after a biblioglossonym (see detailed explanation above): an asterisk * means: one or more additional biblioglossonyms for this autoglossonym (language reference name) exist
    No asterisk means: biblioglossonym is the only one in use for this autoglossonym

PRINT SCRIPTURES
13-15. Scriptures in print (...=none, P..=gospel only, PN.=New Testament, PNB=whole Bible, pnb=near-Bible)
16. Portion/gospel activity (year of first publication and year of latest, if any)
17. New Testament activity (year of first publication and year of latest, if any)
18. Bible activity (year of first publication and year of latest, if any)

AVAILABILITY OF AUDIOVISUALS
19. 'Jesus' Film year first published
20. 'Jesus' Film availability, viewership:

Code	*Meaning*
0	not available in mother tongue or its cluster of languages
1	Available in mother tongue (if under 10% of all speakers) or in its cluster
2	Available, viewers 10–50%
3	Available, viewers 51–100%
4	Vast impact in mother tongue (viewers>100%)

Next 4 lines 21–24: a dot in any of these 4 columns means: Nothing available

21. Audio scriptures available:

Code	*Item*	*Value*	*Meaning*
•	nothing	0	No audio scriptures available
c	materials	1	Audio materials available only in same-cluster language
s	selection	1	Selections/teaching/music purchasable on cassette
r	radio	2	Radio audio selections hearable
a	portion	3	Audio gospels purchasable
A	Testament	4	Audio NT purchasable
B	Bible	5	Audio whole Bible purchasable

22. New Reader Scriptures available = y
23. Braille scriptures available = u
24. Signed scriptures available = h

DIALECTS
25. Reference number: indicating a language's total of *dialects* (not listed here but named only on CD); subtract any language's reference number from the next reference number shown, minus 1 (e.g. 00-AAAA-a mandinka-kango has minus 7 plus 12 minus 1 = 4 dialects).

Table 9-13. **The globe's 13,500 distinct and different languages, with speakers, Christians, scriptures, audiovisual ministries.**

Code	REFERENCE NAME Autoglossonym	Coun	Peo	Mother-tongue speakers in 2000	in 2025	Media radio	CHURCH AC%	E%	Wld	Tr	Biblioglossonym	SCRIPTURES Print	P–activity	N–activity	B–activity	J-year	Jayuh	Ref
1	2	3	4	5	6	7	8	9	10	11	12	13-15	16	17	18	19	20-24	25
0	**AFRICAN macrozone**	37	557	69,440,731	120,922,018		31.68	63	B	68		PNB					4Asu.	1
00	**MANDIC zone**	18	136	22,935,471	40,401,572		6.03	50	B	68		PNB					4As..	2
00-A	**NORTHWEST MANDE set**	15	82	17,416,271	30,479,615		3.96	51	B	68		PNB					4As..	3
00-AA	MANDING chain	15	72	14,477,858	25,196,933		1.55	49	A	61		PNB					4As..	4
00-AAA	**MANDEKAN net**	15	59	13,131,504	22,926,335		1.69	50	A	61		PNB					4As..	5
00-AAAA	WEST MANDEKAN cluster	13	25	5,548,959	9,305,278		0.49	44	A	42		PN.					4a...	6
00-AAAA-a	mandinka-kango	7	8	1,746,145	3,002,540	4	0.48	44	A	42	Mandinka	PN.	1837-1966	1989		1992	4a...	7
00-AAAA-b	sijanka-kango	1	1	59,080	61,662	4	2.00	48	A	42		pn.					1c...	12
00-AAAA-c	maninka-xanwo	1	1	92,081	154,870	1	0.50	22	A	42		pn.					1c...	13
00-AAAA-d	kalanke-kango	1	1	2,611	4,302	1	0.00	22	A	42		pn.					1c...	14
00-AAAA-e	jakhanka-kango	1	1	27,564	48,674	1	0.00	18	A	42		pn.					1c...	15
00-AAAA-f	xasonka-xango	3	3	150,665	284,305	1	2.56	37	A	42	Kassonke	pn.					1c...	16
00-AAAA-g	kakolo-qango	1	1	25,405	48,160	4	1.00	38	A	42		pn.					1c...	17
00-AAAA-h	maninka-kan	7	8	3,440,117	5,690,735	5	0.38	45	A	42	Maninka*	PN.	1931-1964	1932-1966		1989	3a...	18
00-AAAB	EAST MANDEKAN cluster	12	21	6,536,498	11,753,300		1.99	56	B	61		PNB					4As..	22
00-AAAB-a	bamanan-kan	10	10	4,366,464	7,990,122	4	2.54	63	B	61	Bambara	PNB	1923-1942	1933-1995	1961-1987	1983	4As..	23
00-AAAB-b	manenka-kan					1				61		pnb					1cs..	32
00-AAAB-c	mikifore-kan					1				61		pnb					1cs..	37
00-AAAB-d	manya-kan	1	1	53,851	112,988	5	0.03	39	A	61		pnb					1cs..	38
00-AAAB-e	wasulunka-kan	1	1	740,441	1,403,622	4	2.00	45	A	61		pnb					1cs..	39
00-AAAB-f	konyanka-kan	2	2	147,933	235,100	4	0.10	33	A	61		pnb					1cs..	40
00-AAAB-g	tenenga-kan					1				61		pnb					1cs..	41
00-AAAB-h	mauka-kan	1	1	187,780	296,483	1	1.00	41	A	61		pnb					1cs..	42
00-AAAB-i	koroka-kan					1				61		pnb					1cs..	43
00-AAAB-j	baralaka-finanga					1				61		pnb					1cs..	44
00-AAAB-k	sienkoka-kan					1				61		pnb					1cs..	47
00-AAAB-l	wojeneka-kan					1				61		pnb					1cs..	48
00-AAAB-m	gbelebanka-foloka					1				61		pnb					1cs..	49
00-AAAB-n	boduguka-kan					1				61		pnb					1cs..	52
00-AAAB-o	tuduguka-kan					1				61		pnb					1cs..	53
00-AAAB-p	vanduguka-kan					1				61		pnb					1cs..	54
00-AAAB-q	nowoloka-kan					1				61		pnb					1cs..	55
00-AAAB-r	karanjanka-kan					1				61		pnb					1cs..	56
00-AAAB-s	woroduguka-kan					1				61		pnb					1cs..	57
00-AAAB-t	kanika-kan					1				61		pnb					1cs..	61
00-AAAB-u	nigbi-kan					1				61		pnb					1cs..	62
00-AAAB-v	sagaka-kan					1				61		pnb					1cs..	63
00-AAAB-w	koro-kan					1				61		pnb					1cs..	64
00-AAAB-x	koyaga-kan					1				61		pnb					1cs..	67
00-AAAB-y	siaka-kan					1				61		pnb					1cs..	68
00-AAAB-z	jula-kan	4	6	1,040,029	1,714,985	4	0.22	42	A	61	Jula	PNb	1992	1993-1994			1cs..	69
00-AAAC	MARKA cluster	2	3	272,781	532,515		11.53	44	A	20		...					0....	77
00-AAAC-a	bolon-kan	1	1	14,230	27,801	1	3.00	27	A	20		...					0....	78
00-AAAC-b	da-fin-kan					1				20		...					0....	81
00-AAAC-c	maraka-jalan-kan	2	2	258,551	504,714	2	12.00	45	A	20	Marka	...					0....	84
00-AAAC-d	meeka-kan					1				20		...					0....	85
00-AAAD	KURANKO cluster	2	3	415,721	694,815		1.53	32	A	51		PN.					0....	86
00-AAAD-a	falanko-kuranko	1	1	77,634	130,572	0	1.00	25	A	51		pn.					0....	87
00-AAAD-b	muso-kuranko					0				51		pn.					0....	88
00-AAAD-c	wasamandu-kuranko	2	2	338,087	564,243	0	1.65	33	A	51	Kuranko	PN.	1899-1911	1972			0....	98
00-AAAE	VAI-KONO cluster	5	6	356,601	638,582		7.52	45	A	35		P..					4....	99
00-AAAE-a	Central kono	4	4	232,770	386,410	4	11.20	53	B	35	Kono	P..	1919-1993				1....	100
00-AAAE-b	North kono					1				35		p..					1....	105
00-AAAE-c	kono-P.					1				35		p..					1....	109
00-AAAE-d	dama			0	0	1	0.00	0		35		p..					1....	116
00-AAAE-e	vai	2	2	123,831	252,172	1	0.61	31	A	35	Vai	P..	1995			1993	4....	117
00-AAAF	JELKUNA cluster	1	1	944	1,845		2.97	24	A	6		...					0....	118
00-AAAF-a	jelkuna	1	1	944	1,845	0	2.97	24	A	6		...					0....	119
00-AAB	**LIGBI-NUMU net**	2	2	19,051	33,592		0.28	23	A	4		...					0....	120
00-AABA	LIGBI-NUMU cluster	2	2	19,051	33,592		0.28	23	A	4		...					0....	121
00-AABA-a	ligbi	2	2	19,051	33,592	0	0.28	23	A	4		...					0....	122
00-AABA-b	hwela					0				4		...					0....	128
00-AABA-c	numu					0				4		...					0....	129
00-AAC	**SOSO-YALUNKA net**	6	11	1,327,303	2,237,006		0.19	37	A	41		PN.					4....	130
00-AACA	SOSO-YALUNKA cluster	6	11	1,327,303	2,237,006		0.19	37	A	41		PN.					4....	131
00-AACA-a	soso	5	6	1,065,843	1,792,660	4	0.19	41	A	41	Soso*	PN.	1869-1963	1884-1988		1994	4....	132
00-AACA-b	yalunka	4	5	261,460	444,346	1	0.18	21	A	41	Yalunka	PN.	1907	1976			1....	133
00-AB	SOUTHWEST MANDE chain	3	10	2,938,413	5,282,682		15.85	65	B	68		PNB					3....	137
00-ABA	**SOUTHWEST MANDE net**	3	10	2,938,413	5,282,682		15.85	65	B	68		PNB					3....	138
00-ABAA	LOKO-MENDE cluster	3	6	1,606,671	2,724,116		8.53	63	B	68		PNB					3....	139
00-ABAA-a	bandi	2	2	178,744	335,583	1	6.18	44	A	68	Bandi	Pnb	1954-1995				1....	140
00-ABAA-b	loko	2	2	142,420	237,287	4	3.65	49	A	68	Loko: Sierra Leone	PNb		1983			1....	143
00-ABAA-c	mende	2	2	1,285,507	2,151,246	4	9.40	67	B	68	Mende	PNB	1867-1954	1956	1959	1985	3....	154
00-ABAB	LOMA-TOMA cluster	2	2	348,350	655,894		10.60	54	B	42		PN.					0....	160
00-ABAB-a	loma	1	1	168,198	352,901	3	15.00	61	B	42	Loma	PN.	1949-1967	1971			0....	161
00-ABAB-b	toma	1	1	180,152	302,993	1	6.50	47	A	42	Toma	PN.	1961	1981			0....	167
00-ABAC	KPELLE cluster	2	2	983,392	1,902,672		29.67	72	B	42		PN.					2....	168
00-ABAC-a	kpele	1	1	597,530	1,253,700	1	25.00	69	B	42	Kpelle*	PN.	1922-1964	1967			1....	169
00-ABAC-b	kpelese	1	1	385,862	648,972	1	36.90	77	B	42	Kpelee*	Pn.	1945-1969			1997	2....	170
00-B	**SONINKE-BOZO set**	9	17	1,601,693	2,934,483		0.03	19	A	32		...					3....	171
00-BA	SONINKE chain	9	12	1,460,214	2,666,148		0.03	19	A	32		...					3....	172
00-BAA	**SONINKE net**	9	12	1,460,214	2,666,148		0.03	19	A	32		...					3....	173
00-BAAA	SONINKE cluster	9	12	1,460,214	2,666,148		0.03	19	A	32		...					3....	174
00-BAAA-a	Proper soninke	9	9	1,435,752	2,625,567	0	0.02	19	A	32	Soninke	...				1988	3....	175
00-BAAA-b	azayr	3	3	24,462	40,581	0	0.26	19	A	32		...					1....	176
00-BAAA-c	girganke			0	0	0	0.00	0		32		...					1....	177
00-BAAA-d	kinbakka					0				32		...					1....	178
00-BAAA-e	xenqenna					0				32		...					1....	179
00-BB	BOZO chain	2	5	141,479	268,335		0.02	15	A	20		...					0....	180
00-BBA	**TIEMA net**	1	1	2,954	5,601		0.10	14	A	4		...					0....	181

Continued opposite