

# THE LINGUASPHERE IN TEN PAGES (pp. 29=1-30=0)

## the world's major languages in 19=99=-20=00= (provisional)

The following provisional table lists all *languages* estimated to total a minimum of approximately 10= million *voices* (speakers/hearers) each, at the turn of the millennium. These *macrolanguages* include a category of 12= *megalanguages* (from [79=] Putonghua to [51=] Français in the list below) which are each spoken by an estimated 10=0 million voices or more. The totals include estimates of *primary* and *translingual* voices, plus *alternate* voices wherever relevant, rounded to the nearest million (or 10= million for languages above 12=5m).

**Please see commentary on Statistics in Chapter 2.6 above, and the Linguasphere Lexicon for italicised terms.**

Whereas macrolanguages and megalanguages are measured in absolute terms, an important intermediate category is that of *arterial languages*, comprising all those spoken by at least 1% each of the current world population (i.e. 60= million or more voices in 19=99=/20=00=). This category totals at least 28= languages.

\*The majority of all *macrolanguages* (54= out of 76=) belong to only 8 *phylozones* (of which 4 are within *phylosector* 5=Indo-European): [31=] Hesperonesic, [49=] Dravidic, [79=] Sinitic, [51=] Romanic, [52=] Germanic, [53=] Slavic, [59=] Indic, [99=] Bantuic.

\*Among these, 12= out of 28= *arterial languages* belong to only 3 *nets* of closely related languages: 51=-AAA Romance-West, 59=-AAF Hindi+Bangla (Indic-Central) and 79=-AAA Han-yu ("Wider" Chinese).

τ The world's most spoken language alternates daily. When the sun is over the western Pacific, the national language of China is the most in use. When the sun is over the Atlantic and China sleeps, English takes the lead. The world's second most spoken language also alternates daily, between Hindi+Urdu and Spanish.

Language-names linked in the following list may be treated as a single continuum of spoken communication (often linking two or more differently standardised idioms, e.g. [59=] Hindi+Urdu or [47=] Thai+Lao). In contrast, individual language-names sometimes cover wide degrees of internal spoken variation (often bridged by a single standardised and/or common vehicular idiom, e.g. [12=] 'Arabiyya or [52=] Deutsch)

### list of arterial languages each spoken by at least 1% of humankind

1. MACROLANGUAGES each estimated to total above 300m voices	NETS ( <b>bold</b> =3 most dominant)	(primary/translingual + alternate voices ) TOTAL in millions	principal countries where spoken
[79=] <b>Putonghua</b> (Chinese-Mandarin)	<b>79-AAA</b>	(800+200) 1000m	China; Taiwan; Thailand; Malaysia; Singapore; Vietnam
[52=] <b>English</b>	52-ABA	(400+600) 1000m	UK; Ireland; Canada; USA; Australia; New Zealand; South Africa; Botswana; Lesotho; Malawi; Namibia; Swaziland; Zambia; Zimbabwe; Kenya; Tanzania; Uganda; Rwanda; Ethiopia; Somalia; Sudan; Cameroon; Gambia; Ghana; Liberia; Nigeria; Sierra Leone; Gibraltar; Malta; Cyprus; Bahrein; Egypt; Israel & Palestine; Jordan; Kuwait; UAE; Mauritius; Seychelles; Bangladesh; Burma; China (Hong Kong); India; Malaysia; Pakistan; Singapore; Sri Lanka; Fiji; Kiribati; Micronesia; Nauru; Papua New Guinea; Philippines; Samoa; Solomon Islands; Tonga; Tuvalu; Vanuatu; Bahamas; Belize; Bermuda; Barbados; Dominica; Grenada; Guyana; Jamaica; Puerto Rico; St. Lucia; St. Kitts and Nevis; St. Vincent; Trinidad & Tobago; Virgin Islands
[59=] <b>Hindi+ Urdu</b>	<b>59-AAF</b>	(550+350) 900m	India; Pakistan; Nepal; Fiji; Mauritius; South Africa; UK
[51=] <b>Español</b> (Spanish)	<b>51-AAA</b>	(400+50) 450m	Spain; Mexico; Venezuela; Colombia; Ecuador; Peru; Bolivia; Chile; Argentina; Paraguay; Uruguay; Panama; Costa Rica; Nicaragua; Honduras; El Salvador; Guatemala; Cuba; Dominican Republic; Puerto Rico; Canary Islands; Morocco; Equatorial Guinea; USA
[53=] <b>Russkiy</b> (Russian)	53-AAA	(170+150) 320m	Russia; Ukraine; Belarus; Latvia; Estonia; Lithuania; Georgia; Azerbaijan; Armenia; Kazakhstan; Kirghizstan; Turkmenistan; Uzbekistan; Tajikistan; Moldova; Poland; Slovakia; <i>other states of central Europe</i> ; USA; Canada; Brazil

2. MACROLANGUAGES each estimated to total between 60m and 300m voices	NETS ( <b>bold=3</b> most dominant)	(primary/translingual + alternate voices ) TOTAL in millions	principal countries where spoken
[12=] <b>'Arabiyya</b> (Maghribi Arabic+ Mashriqi Arabic)	12-AAC	(200+50) 250m	Morocco; Algeria; Tunisia; Malta; Mauritania; Senegal; Mali; Niger; Cameroon; Chad; Libya; Sudan; Egypt; Israel & Palestine; Lebanon; Syria; Saudi Arabia; Yemen; United Arab Emirates; Oman; Kuwait; Bahrain; Qatar; Iraq; Iran; Djibouti; Somalia; Comoros; France
[59=] <b>Bangla</b> (Bengali)	<b>59-AAF</b>	(190+60) 250m	Bangladesh; India; UK
[51=] <b>Português</b> (Portuguese)	<b>51-AAA</b>	(180+20m) 200m	Portugal; Brazil; Angola; Cape Verde; Guinea-Bissau; Macau; Mozambique; São Tomé & Príncipe
[31=] <b>Malayu</b> (Malay+Indonesian)	31-MFA	(50+110) 160m	Indonesia; Malaysia; Singapore
[45=] <b>Nihongo</b> (Japanese)	45-CAA	(120+10) 130m	Japan; USA; Brazil; Peru
[52=] <b>Deutsch</b> (German)	52-ACB	(100+25) 125m	Germany; Austria; Switzerland; France; Italy; Belgium; Luxembourg; Liechtenstein; USA; Canada; Namibia; Kazakhstan; Romania; Russia; Brazil; Argentina
[51=] <b>Français</b> (French)	<b>51-AAA</b>	(90+35) 125m	France; Canada; Belgium; Luxembourg; Switzerland; Monaco; St-Pierre and Miquelon; Canada; Haiti; Guadeloupe; Martinique; French Guiana; French Polynesia; New Caledonia; Vanuatu; Laos; Cambodia; Vietnam; Réunion; Mauritius; Seychelles; Lebanon; Syria; Tunisia; Algeria; Morocco; Mali; Mauritania; Benin; Burkina Faso; Burundi; Cameroon; Central African Republic; Chad; Comoros; Congo/Zaire; Congo; Côte d'Ivoire; Djibouti; Gabon; Guinea; Madagascar; Niger; Rwanda; Senegal; Togo
[47=] <b>Thai+ Lao</b> (incl. Isan, Huang+ Buyi)	47-AAA	90m	Thailand; Laos; Vietnam; China
[59=] <b>Panjabi</b>	<b>59-AAF</b>	85m	Pakistan; India; Malaysia; Fiji; UK
[79=] <b>Wu</b> (Chinese-Wu)	<b>79-AAA</b>	85m	China
[31=] <b>Jawa</b> (Javanese)	31-MFM	80m	Indonesia; Malaysia; Singapore; New Caledonia / Kanaky; French Guiana; Surinam
[59=] <b>Marathi</b>	<b>59-AAF</b>	80m	India
[45=] <b>Hankukmal</b> (Korean)	45-AAA	75m	North Korea; South Korea; China; Japan; USA
[46=] <b>Viêt</b> (Vietnamese)	46-EBA	75m	Vietnam; Cambodia; USA
[51=] <b>Italiano</b> (Italian)	<b>51-AAA</b>	70m	Italy; Switzerland; USA; Canada; Argentina; Brazil
[49=] <b>Tamil</b>	49-EBE	70m	India; Sri Lanka; Malaysia; Singapore; Fiji. Mauritius; Trinidad & Tobago
[49=] <b>Telugu</b>	49-DBA	70m	India; Malaysia
[79=] <b>Yue</b> (Chinese-Cantonese)	<b>79-AAA</b>	70m	China; Malaysia; Vietnam; Singapore; Indonesia; USA
[59=] <b>Bhojpuri+ Maithili</b>	<b>59-AAF</b>	60m	India; Nepal; Bangladesh; Mauritius; South Africa; Trinidad; Guyana; Surinam; UK
[44=] <b>Türkçe+Azeri</b> (Turkish+Azeri, <i>incl.</i> Turkmen)	44-AAB	60m	Turkey; Bulgaria; Greece; Cyprus; Turkmenistan; Kazakhstan; Uzbekistan; Iran; Azerbaijan; Russia; Germany
[53=] <b>Ukrainska+ Belarusskaya</b> (Ukrainian+ Belarussian)	53-AAA	60m	Ukraine; Belarus; Russia; Moldova; Kazakhstan; Kirghizstan; Turkmenistan; Uzbekistan; Azerbaijan; Georgia; Armenia; Latvia; Lithuania; Poland; Slovakia; Hungary; Israel; USA; Canada; Brazil
[99=] <b>Swahili</b>	99-AUS	(7m+53m) 60m	Tanzania; Kenya; Uganda; Rwanda; Burundi; Congo/Zaire; Somalia; Comoros
[31=] <b>Tagalog</b> ( <i>incl.</i> Filipino)	31-CKA	(25m+35m) 60m	Philippines; USA; Canada

## list of other macrolanguages between 10= and 59= million voices each

3. MACROLANGUAGES each estimated to total between 10m and 59m voices	NETS ( <b>bold</b> =3 most dominant)	TOTAL in millions (incl. alternate voices)	principal countries where spoken
[79=] <b>Min-nan</b>	<b>79-AAA</b>	55m	China; Taiwan; Malaysia; Singapore
[79=] <b>Xiang</b> (Chinese-Xiang)	<b>79-AAA</b>	48m	China
[53=] <b>Polski</b> (Polish)	53-AAA	47m	Poland; USA; Lithuania; Ukraine; Canada; Brazil
[59=] <b>Gujarati</b>	<b>59-AAF</b>	45m	India; Uganda; UK
[19=] <b>Hausa</b>	19-HAA	44m	Nigeria; Niger; Cameroon; Chad; Benin; Ghana
[49=] <b>Kannada</b>	49-EBA	42m	India
[52=] <b>Wes-kos+ Caribbean Anglo-Creole</b> (Atlantic Anglo- Creoles)	52-ABB	40m	Cameroon; Nigeria; Ghana; Liberia; Sierra Leone; Gambia; Bermuda; Barbados; Grenada; Trinidad & Tobago; Belize; St. Lucia; St. Kitts and Nevis; St. Vincent; Virgin Is.; Dominica; Bahamas; Jamaica; Cayman Is.; Panama; Colombia; Costa Rica; Nicaragua; Honduras; Belize; Guyana; USA; UK
[58=] <b>Farsi</b> (Persian)	58-AAC	40m	Iran; Tajikistan; Afghanistan; Pakistan; Iraq
[49=] <b>Malayalam</b>	49-EBE	39m	India
[79=] <b>Hakka</b> (Chinese-Hakka)	<b>79-AAA</b>	35m	China; Taiwan; Malaysia; Singapore; Indonesia
[77=] <b>Bama</b> (Burmese)	77-AAA	33m	Burma; Bangladesh
[59=] <b>Oriya</b>	<b>59-AAF</b>	33m	India
[31=] <b>Sunda</b> (Sundanese)	31-MFN	30m	Indonesia
[52=] <b>Nederlands+ Vlaams+ Afrikaans</b> (Dutch+ Flemish+ Afrikaans)	52-ACB	30m	Netherlands; Belgium; Surinam; Netherlands West Indies; South Africa; Namibia
[51=] <b>Româneasca</b> (Romanian+ Moldavian)	51-AAD	27m	Romania; Moldova
[98=] <b>Yoruba</b>	98-AAA	26m	Nigeria; Benin; Togo
[12=] <b>Amarinya</b> (Amharic)	12-ACB	25m	Ethiopia; Israel
[58=] <b>Pashto</b>	58-ABD	25m	Afghanistan; Pakistan
[79=] <b>Gan</b> (Chinese-Gan)	<b>79-AAA</b>	20m	China
[44=] <b>Kazakhşa+ Kirghizca</b> (Kazakh+ Kirghiz)	44-AAB	20m	Kazakhstan; Kirghizstan; Tajikistan; Uzbekistan; Russian Fed.; China
[52=] <b>Svenska+ Dansk+ Norsk</b> (Swedish+ Danish+ Norwegian)	52-AAA	20m	Sweden; Denmark; Norway; Finland; Iceland; Greenland
[98=] <b>Igbo</b>	98-GAA	19m	Nigeria
[53=] <b>Srbski+ Hrvatski</b> (Serbian+ Croatian <i>or</i> Serbo- Croat)	53-AAA	19m	Serbia; Croatia; Bosnia-Herzegovina
[59=] <b>Sindhi</b>	<b>59-AAF</b>	18m	Pakistan; India
[44=] <b>Özbekça+ Uyghur</b> (Uzbek+ Uyghur)	44-AAB	18m	Uzbekistan; Tajikistan; Afghanistan; Kirghizstan; Kazakhstan; Turkmenistan

MACROLANGUAGES totalling between 10m & 60m voices	NETS ( <b>bold=3</b> most dominant)	TOTAL in millions (incl. alternate voices)	principal countries where spoken
[31=] <b>Cebuan</b> (Cebuano)	31-CKG	17m	Philippines; USA
[59=] <b>Nepali</b>	<b>59-AAF</b>	17m	Nepal; India; Bhutan
[53=] <b>Čestina+ Slovenčina</b> (Czech + Slovak)	53-AAA	16m	Czech Republic; Slovakia; Ukraine
[90=] <b>Fula / Peul</b>	90-BAA	16m	Nigeria; Guinea; Guinea-Bissau; Senegal; Gambia; Mauritania; Mali; Burkina Faso; Togo; Benin; Niger; Cameroon; Central Afr.Rep.; Sudan
[41=] <b>Magyar</b> (Hungarian)	41-BAA	15m	Hungary; Romania; Slovakia; Ukraine; Croatia; Yugoslavia
[58=] <b>Kurmanji+ Kurdi</b> (North+ South Kurdish)	58-AAA	15m	Turkey; Iraq; Iran; Syria; Lebanon; Germany
[99=] <b>Lingala</b>	99-AUI	14m	Congo; Congo/Zaire
[14=] <b>Borena+ Tulema</b> (Oromo)	14-FBA	14m	Ethiopia; Kenya
[99=] <b>Rwanda + Rundi</b>	99-AUS	14m	Rwanda; Burundi; Congo/Zaire; Tanzania
[59=] <b>Sinhala</b> (Sinhalese)	59-ABB	14m	Sri Lanka
[31=] <b>Madura</b> (Madurese)	31-MFL	13m	Indonesia
[31=] <b>Merina+ Tandroy</b> (Malagasy)	31-LDA	13m	Madagascar; Comoros
[00=] <b>Mandinka+ Bamanan+ Jula</b> (Manding)	00-AAA	13m	Gambia; Guinea-Bissau; Guinea; Mali; Côte d'Ivoire; Senegal; Burkina Faso; Sierra Leone; Liberia
[56=] <b>Helleniki</b> (Greek)	56-AAA	12m	Greece; Cyprus; USA
[79=] <b>Min-bei</b> (Chinese-Min-bei)	<b>79-AAA</b>	12m	China
[99=] <b>Sotho + Tswana</b>	99-AUT	12m	South Africa; Lesotho; Botswana; Namibia; Zimbabwe
[10=] <b>Tamasheq+ Qabaylith</b> (Tamazic or Berber)	10-AAA	12m	Morocco; Algeria; Tunisia; Mali; Burkina Faso; Niger; Libya; Egypt; France
[59=] <b>Axamiya</b> (Assamese)	<b>59-AAF</b>	11m	India; Bhutan; Bangladesh
[51=] <b>Català</b> (Catalan)	<b>51-AAA</b>	11m	Spain; France; Andorra; Italy (Sardinia)
[51=] <b>Créole</b> (Gallo-Creole)	51-AAC	11m	Haiti; Martinique; Guadeloupe; France; Mauritius; Seychelles
[99=] <b>Xhosa+Zulu+ Swati+Ndebele</b> (Nguni)	99-AUT	11m	Zimbabwe; South Africa; Swaziland; Tanzania
[53=] <b>Bulgarski + Makedonski</b> (Bulgarian+ Macedonian)	53-AAA	10m	Bulgaria; Macedonia; Yugoslavia; Albania; Greece; Turkey; Moldova
[14=] <b>Soomaali</b> (Somali)	14-GAG	10m	Somalia; Ethiopia; Kenya

end of table

## principles of the linguasphere register

The 1999/2000 framework edition of the Linguasphere Register has the following aims:

- to provide the first planetary outline of humankind's linguistic environment or *linguasphere*, at the beginning of a new era of global communication;
- to compile a first transnational classification or "roll-call" of human speech communities, at the end of the 20<sup>th</sup> century;
- to establish a stable framework of worldwide reference for the documentation and mapping of the world's languages and speech communities, from the beginning of the 21<sup>st</sup> century;
- to establish a flexible scale of linguistic proximity for the expansion and adjustment of a worldwide corpus of data on individual languages, independently of their demographic importance;
- and to begin the enumeration and transnational monitoring of all human communities, however small, isolated or socially disadvantaged, within a common framework of global relationship.

**The starting-point of the Register is the recognition of the linguasphere as a global environment.**

The linguasphere is the environment of spoken communication created around the globe by successive generations of humankind. It comprises a growing continuum of millions of words, organised within thousands of alternative patterns of speech-sounds and grammatical rules, and manipulated by billions of continuously renewed human voices. The weaving of this planetary mantle of speech is humankind's most important and most collective creation.

Since small variations of language are often significant in marking the identity of neighbouring communities, the traditional dichotomy of language and dialect has been replaced with a sequence of three layers of immediate relationship: *outer language*, *inner language* and *dialect*. Although these layers form the base of a relatively complex system of reference (see table p. 297), it must be emphasised that this system is designed as a practical device to assist the cataloguing of the contemporary linguasphere, and not as an end in itself.

The modern linguasphere has thus been classified in the Register in the form of over 13,800 *inner languages* (plus their internal dialects), organised within almost 5,000 *outer languages* and approximately 700 sets of related languages. These sets are in turn classified within 100 reference zones within 10 major sectors, together covering the entire linguasphere. These ten sectors comprise five *phylosectors*, corresponding to five well documented linguistic "families", and five *geosectors*, corresponding to five well defined continental areas within which all remaining languages are classified.

**Numerical codes provide a stable framework, and alphabetical codes a flexible scale of relationships.**

The 100 reference zones are coded, and their components identified, by pairs of digits [00= to 99=] within the ten sectors [0= to 10=]. This provides a stable framework of worldwide reference for the location of languages, as set out on page 300. Reference names are provided in the Register for all known languages, which are always unique within the relevant zone. Any reference name preceded by a code in the form of a bracketed pair of digits, e.g. [51=] *Italiano*, may therefore be readily located within the relevant zone and also within the *Index to the Register* above.



Within each zone, all component languages are classified according to a flexible scale of alphabetical codes, which may be modified between successive editions of the Register to reflect new or improved data and research. Sets within each zone are coded by means of a single alphabetical letter (A to a maximum of Z) suffixed to the numerical code of that zone, and further letters are added for narrowing layers of linguistic relationship within each set, through to the component outer and inner languages (and where necessary, also dialects). The full system of coding is set out on p.297.

**Information on each layer of linguistic relationship is presented in a series of tabulated columns.**

The Register provides an overview of the modern linguasphere in a continuous table of five columns, covering all languages known to have been spoken during at least part of the 20<sup>th</sup> century. Also included are written languages inherited from earlier centuries, but which are still read as literary or liturgical languages (and which thus still form part of the modern linguasphere). Entries have likewise been included for certain languages known to have become extinct during the previous four centuries (from the late 15<sup>th</sup> to the end of the 19<sup>th</sup> century), since these are directly relevant to any consideration of the modern impact of European languages on the state of the linguasphere. A raised star \* is suffixed to items of data which are unreliable or which require corroboration.

The five columns of the Register (compressed to three in the Synopsis) are organised as follows:

**Column 1** presents a coded classification of the world's *language-groups* (sets, chains and nets) and *idioms* (outer languages, inner languages and dialects). This classification is constructed around the numerical and alphabetical codes presented in the following table, on p.29=7.


**Column 2** presents a list of selected reference-names for all language-groups and idioms, their classificational hierarchy being visually apparent from sequences of typography (ranging from bold capitals to normal lower-case). The reference-names of idioms represent wherever possible speakers' own-names or *autoglossonyms* for their primary forms of speech. Most reference-names of language-groups in the Register are constructed from a combination of the names of two of their component elements, rather than utilising existing, often artificial or foreign "cover-names". Names of languages which are today read rather than spoken are prefixed by the icon of a book , whereas spoken languages modelled at least partially on the written word are preceded by the icon of a writing hand . For this first framework edition of the Register, names have been recorded only in the Latin script.

A series of capitalised suffixes is employed for outer languages, inner languages and dialects, to distinguish reference names which would otherwise be identical in the same zone (for example, by the use of contrasting points of the compass). The suffixes are as follows:


(directional suffixes, used also for place-names in column 3) -N north(ern); -E east(ern);  
-S south(ern); -W west(ern); -C central; plus combinations, eg -NW northwest;  
-CW west central, etc.



(other suffixes) -A "proper" (name); -F formal or standard; -G generalised; -L liturgical or pre-modern literary; -M middle; -U urban; -V vehicular

**Column 3** presents the alternative names recorded for many language-groups and idioms, including alternative reference-names in bold type. Other names applied to languages and communities are distinguished by the use of lower case initials, as opposed to initial capitals for geographical names. (This typographical convention does not apply to textual notes, printed in italics.) Notes are categorised by a series of icons:

- |   |   |   |                                    |
|---|---|---|------------------------------------|
| ⊕ | notes on locations or epicentres                    |  | notes on scripts or written models |
| ¶ | notes on speech communities                         |   |                                    |
| ➤ | notes on languages                                  | #   | notes on nomenclature              |
| X | notes on contacts and relationships among languages |   |                                    |

Cross-references to other languages are preceded by a zonal reference in square brackets, e.g. [49=] *Telugu* (i.e. classified in zone 49=, see table of sectors and zones on p.30=0).

**Column 4** lists the nation-state or states in which an idiom is spoken (with provinces in brackets), with any official status indicated by the icon of a flag .

**Column 5** presents a single-digit *scale of voices* (i.e. speakers) for individual languages, and for the combined languages of each zone. This digit records the order of magnitude of the number of primary and alternate speakers of every outer language in the Register (and of some inner languages), as known or estimated at the end of the 20th century. This estimate is expressed on a scale from 0 (extinct since 1900) through 1 (less than 100), 2 (100+), 3 (1000+), 4 (10,000+), 5 (100,000+), 6 (1,000,000+), 7 (10,000,000+), 8 (100,000,000+) to 9 (over one billion). The icon  marks complete nets, chains or sets of idioms which were extinct before the end of the 20<sup>th</sup> century, while the icon  marks idioms known or believed to be extinct before the end of the 19<sup>th</sup>.

**The Register and its synopsis make reference to the world's arterial languages, totalling at least 28=.**

Arterial languages are defined as all those outer languages (or nets of largely inter-intelligible outer languages) which are each understood by at least 1% of the world's total population and which therefore play a major role in the circulation of the world's *speechways*.

**PLEASE SEE CHAPTER TWO: GUIDE TO THE REGISTER, AND THE LINGUASPHERE LEXICON,  
FOR A FULL INTRODUCTION TO TERMINOLOGY AND CONVENTIONS.**

## linguaspHERE layers of classification

Between the planetary *linguaspHERE* and the *voice* of each person, the LinguaspHERE Register 1999/2000 identifies, classifies and codes 13,840 *inner languages* (plus 8,881 constituent *dialects*) within 4,994 *outer languages* (4,557 spoken in 1999) and within 694 linguistic *sets*.

### 1. Numerical framework of worldwide reference

Each *set* of languages is classified and coded within one of 100 referential *zones* within one of 10 referential *sectors* (one of 5 *phylosectors* or 5 *geosectors*).

<i>linguaspHERE</i> key = a fixed two-digit <b>numerical code</b> (99= as an example)	marking ➤ two <b>layers of worldwide reference</b>	for an inventory of sectors and zones, see table on p.30=0	TOTALS
(uncoded)	(LINGUASPHERE)	= totality of the world's languages	1
<b>9=</b>	<b>SECTOR</b>	= <b>phylosector</b> (odd digit 1, 3, 5, 7, 9) or <b>geosector</b> (even digit 0, 2, 4, 6, 8)	10
<b>99=</b>	<b>ZONE</b>	= <b>phylozone</b> or <b>geozone</b>	100

### 2. Alphabetical scale of linguistic proximity

Each *set* comprises two successive layers of close relationship:

*chain* (within each *set*) and *net* (within each *chain*) = upper-case alphabetical code (-AAA-)

<b>+ an alphabetical code</b> comprising three <b>upper-case (majuscule) letters</b>	marking ➤ three <b>layers of close relationship</b>	<i>ideally</i> , the following minimum of basic vocabulary may be shared by languages in the same <i>set</i> , <i>chain</i> or <i>net</i>	TOTALS
99-A	<b>SET</b>	<b>substantial minority</b> (say 25-30%+)	694
99-AA	CHAIN	<b>intermediate proportion</b>	1,410
99-AAA	<b>NET</b>	<b>substantial majority</b> (say 65-70%+)	2,694

Each *net* comprises two or three successive layers of immediate relationship:

*outer language*, *inner language* and (optionally) *dialect* lower-case alphabetical code (-aaa)

<b>+ two or three lower-case (miniscule) letters</b>	marking ➤ two or three <b>layers of immediate relationship</b>	up to <i>three</i> layers of relative proximity composed of largely inter-intelligible spoken (and/or written) <i>idioms</i>	TOTALS
99-AAA-a	<b>Outer language</b>	<b>basic demographic unit</b>	4,994
99-AAA-aa	<b>inner language</b>	<b>basic unit of classification</b>	13,840
99-AAA-aaa	dialect ( <i>as required</i> )	<b>local, social or written variety</b>	(< 8,881)
(uncoded)	(voice)	<b>the total linguistic repertoire and competance of each person in any language or languages</b>	6,000,000,000

## the linguasphere at the turn of the millennium

The Map of the Linguasphere, inside the back cover of this volume, shows the approximate distribution of the referential *zones* [from 00= to 99=] under which the modern languages and dialects of the world are classified in the accompanying Volume Two. This first attempt at a detailed classification of humankind, based on language rather than nation, was completed in Wales on 31<sup>st</sup> December 1999, in the bilingual village of Hebron. Work on its preparation had been undertaken in places as far apart as Normandy and Provence, Quebec and Virginia, London and Dakar, Gujarat and Maharashtra.

The work ahead is even greater, and the material now presented in the Linguasphere Register provides the framework for a global research project, in which individuals and organisations, from universities to schools and private companies, are already responding to the request for more and improved information on the languages and present location and size of the world's speech communities. The compiler of the Register is alone responsible for its shortcomings, but its future accuracy and comprehensiveness will depend on information received from participants and observers of speech communities throughout the world, including those which now form part of the multilingual fabric of so many great cities in every continent. All readers who wish to propose refinements or additions to the Register are invited to e-mail the Observatoire Linguistique at <[register@linguasphere.info](mailto:register@linguasphere.info)>. Sources of new information will be acknowledged in future editions.

### **The Observatoire Linguistique / Linguasphere Observatory**

The Observatoire Linguistique, founded in Quebec and Normandy as a transnational organisation in the early 1980s, is a virtual institute independent of all governmental, political, religious and commercial interests. With a current budget no larger than that of the smallest administrative office in an international organisation, the Observatoire serves as a scientific viewing platform in cyberspace, through which up-to-date information is collected and disseminated. It has no formal membership, apart from the active support of all who contribute information, who share in the evaluation and editing of incoming data, or who otherwise participate in the Observatoire's website <[www.linguasphere.info](http://www.linguasphere.info)>. Extracts from the Register are available for viewing and downloading from that site, which will focus increasingly on debate about the world's languages, from the global role of English to the needs of speech communities threatened with extinction.

The research and information services of the Observatoire are supported by the Linguasphere Press (Gwasg y Byd Iait in Welsh), which is responsible for the sale and distribution of the Register, and for on-line access by subscription. The website and e-mail address of the Press have the suffix "net" and orders for the Register should be addressed to <[register@linguasphere.info](mailto:register@linguasphere.info)>.

### **Classification of the linguasphere**

During the 19<sup>th</sup> and 20<sup>th</sup> centuries, the world's linguistic complexity became increasingly apparent, as data was gathered on thousands of languages, and on their variety of diverging vocabularies and structures. It was then more conventional to speculate about the origins of languages than to plan their collective future, and the only means available for partially classifying the world's languages was to assign them to a variety of 'family-trees', often hypothetical. The unsuitability of this historical classification was that its piecemeal structure depended on the most widespread relationships proposed, which were the most distant and least certain. There was no great concern with enumerating closely related modern idioms, and there was a resulting gap in the classification of the environment. Humankind had analysed and classified many aspects of its home planet, from subatomic particles to plate tectonics, and had begun to explore the distant structures of Mars. Yet one of the most remarkable elements on Earth – its multilingual cloak of human communities – had never been classified.

The Linguasphere Register now bridges this gap by providing the first full classification of the world's languages. For its principles of classification see pp. 295-296, and for fuller discussion see Chapter Two.

### **Evolution of the linguasphere**

Until modern times, the *linguasphere*, or global linguistic environment, consisted of a continuum of localised *speech communities* in regular or intermittent contact. Before the advent of agriculture, the typical speech community was a small band of hunter-gatherers, so that there were inevitably more languages spoken in the paleolithic world than now – as illustrated by around 900 *outer languages* still spoken in the relatively undisturbed environment of New Guinea: see zones 20= to 27= and 33= & 34= on the map. Working from this concrete case, one may deduce that the linguasphere may have comprised as many as 20,000 or more distinct languages before the spread of agriculture (in contrast to around 4,500 today). The Americas – and North America in particular – may have contained a far larger number of speech communities and languages in pre-Columbian times than hitherto believed.

During the last 500 years, the linguasphere has been dislocated by the spread of European languages, including the ethnic clearance of two continents led by speakers of [52=] English. This expansion has also



involved an extension of the power and use of the printed word, the brutal enslavement of millions of speakers of African languages, and the creation and use of new creolised languages among their descendants. The steady loss of hunter-gatherers' habitat, communities and languages still continues.

The 20<sup>th</sup> century saw the end of colonial empires and the proliferation of nation-states in their wake. The governments of both had an interest in controlling or influencing the news media to which their subjects had access, and in strengthening the administrative and educational use of national languages. Across the boundaries of nation-states, there was also continued rivalry among major religions, most of them associated with a written tradition of sacred texts in one or more languages.

The turn of the millennium marks the crossing of a new threshold in the development of human communications, the spoken word having escaped from its constraints of space and time and its dependency on writing. Although the majority of the world's population has been illiterate until modern times, and in spite of the failure of many noble resolutions to end illiteracy by the end of the 20<sup>th</sup> century, it is significant that most of the world's major spoken languages (see pp.291-294 above) are closely related to or heavily influenced by a few classical written languages, notably [12=] Classical Arabic, [51=] Latin, [53=] Church Slavonic, [59=] Sanskrit and Pali, and [79=] Classical Chinese.

The world is currently dominated by communication in a relatively small number of *arterial languages*, defined as those spoken - or understood - by at least 1% of humankind (i.e. by 60 million or more each, in 1999/2000). Of 28 such languages (see p. 291 and p. 300), all but 3 are classified within the 10 *zones* indicated in orange on the accompanying map: these are zones 12=, 31=, 45=, 49=, 51= to 53=, 59=, 79= and 99=.

### Threatened languages

At the opposite extreme are languages totalling a few thousand voices or less, most of which are classified within the 22 "microzones" indicated on the map exclusively by small digits in orange: see 07=, 09=, 11=, 20=, 21=, 22=, 27=, 28=, 29=, 36=, 37=, 43=, 63=, 64=, 65=, 66=, 67=, 68=, 80=, 83=, 86=, 89=. Although many of these zones comprise a large number of languages (see p.300), the total number of voices within each microzone is less than 0.1% of the total for any of the 10 "megazones" indicated on the map by large digits in orange: see 12=; 31=; 45=; 49=; 51=; 52=; 53=; 59=; 79=; 99=. Microzones are associated above all with areas where vestiges of hunter-gatherer economies were still found in the 20<sup>th</sup> century, in Australia, New Guinea, the Arctic, and parts of the Americas and southwestern and eastern Africa.

Since the late 20<sup>th</sup> century, much concern has been expressed about the threat posed to many languages by the process of globalisation. The now frequently cited claim that more than half the world's languages may become extinct during the 21<sup>st</sup> century appears exaggerated, however, and is not supported by the global view of the linguasphere provided by the Register. Notwithstanding the continuing extinction of hunter-gatherer communities and their languages, it seems unlikely that this process will extend rapidly to the majority of the world's other languages. The English language may appear *all-powerful*, but the attachment of other communities to their own languages should not be underestimated, as is well illustrated by the survival of the language of Wales, within a couple of hundred miles of London.

### Priorities for the 21<sup>st</sup> century

In assisting communities whose languages are under pressure, the greatest needs are to encourage the creation and recording of vernacular literature and audio-visual materials, and to provide bilingual education for their children. Most parents in privileged positions in the world ensure that their children have access to English, and there is no reason why other children should be denied the chance of participating in an increasingly global society. Transnational languages should be developed in the service of a multilingual world and as channels for the free exchange of ideas and information. They should not be controlled by the agents of monolingual cultures or of privileged economic power.

The future of humankind in the 21<sup>st</sup> century depends on communication, and on the ability to profit socially from the revolution in telecommunications. The technical means are becoming available to allow the establishment of a global community which respects and protects local communities everywhere. Different languages will have different roles to play, and a central function of education will be to teach the young to think and converse in alternate languages, to benefit mentally from the skills of bilingual expression, and to escape the cultural prisons of the past. No profession will have a greater role to play in the development of such an open global society than the media, including press, television, radio and the internet. Monitoring the freedom and coverage of the media in all languages will become an important aspect of the future observation of the linguasphere.

# linguasphere sectors and zones

The Register classifies 4,994 outer languages (incl. 13,840 inner languages and 8,881 dialects) within 694 sets. This table shows: sets (as a gauge of relative complexity); outer languages (minus ☞ extinct languages); total outer languages (spoken in 1999); arterial languages (each spoken by at least 1% of world population); ☰ = scale of voices: 3 1000> speakers; 4 10,000>; 5 100,000>; 6 1million>; 7 10m>; 8 100m>; 9 1billion>

## geosectors

0=AFRICA	sets	outer	-☞ total	arterial	☰
00=MANDIC	4	35			7
01=SONGHAIC	1	5			6
02=SAHARIC	3	5	-1		6
03=SUDANIC	2	37			6
04=NILOTIC	3	45			7
05=EAST-SAHEL geozone	16	57	-5		6
06=KORDOFANIC	4	18			5
07=RIFT-VALLEY geozone	4	7	-1		4
08=KHOISANIC	2	14	-2		5
09=FKALAHARI geozone	5	14	-7		4
<b>geosector totals</b>	<b>44</b>	<b>237</b>	<b>-16</b>	<b>=221</b>	<b>0</b>

2=AUSTRALASIA	sets	outer	-☞ total	arterial	☰
20=ARAFURA geozone	26	104	-2		4
21=MAMBERAMO geozone	22	78			4
22=MADANGIC	23	101			4
23=OWALAMIC	11	57	-1		5
24=TRANSIRIANIC	22	251	-1		5
25=CENDRAWASIH geozone	25	76			5
26=SEPIK-VALLEY geozone	22	96			5
27=BISMARCK-SEA geozone	26	101	-2		4
28=NORTH-AUSTRALIA geoz.	21	85	-8		3
29=TRANSAUSTRALIA geoz.	25	213	-134		4
<b>geosector totals</b>	<b>223</b>	<b>1162</b>	<b>-148</b>	<b>=1014</b>	<b>0</b>

4=EURASIA	sets	outer	-☞ total	arterial	☰
40=EUSKARIC	1	1			5
41=URALIC	3	16			7
42=CAUCASUS geozone	3	20	-1		6
43=SIBERIA geozone	4	11	-3		4
44=TRANSASIA geozone	3	16		Turkish	7
45=EAST-ASIA geozone	3	11	-2	Japanese; Korean	8
46=SOUTH-ASIA geozone	11	85	-10	Vietnamese	7
47=DAIC	1	26		Thai+Lao	7
48=MIENIC	1	9			6
49=DRAVIDIC	5	24		Tamil;Telugu	8
<b>geosector totals</b>	<b>35</b>	<b>219</b>	<b>-16</b>	<b>=203</b>	<b>7</b>

6=NORTH-AMERICA	sets	outer	-☞ total	arterial	☰
60=ARCTIC	1	14			5
61=NADENIC	3	23	-3		5
62=ALGIC	3	22	-5		5
63=SAINT-LAWRENCE geoz.	2	7	-3		4
64=MISSISSIPPI geozone	3	21	-5		4
65=AZTECIC	1	44	-21		4
66=FARWEST geozone	26	80	-33		3
67=DESERT geozone	5	15	-4		4
68=GULF geozone	8	23	-18		4
69=MESO-AMERICA geozone	11	87	-8		6
<b>geosector totals</b>	<b>63</b>	<b>336</b>	<b>-100</b>	<b>=236</b>	<b>0</b>

8=SOUTH-AMERICA	sets	outer	-☞ total	arterial	☰
80=CARIBIC	1	37	-15		4
81=INTER-OCEAN geozone	16	47	-15		5
82=ARAWAKIC	2	50	-14		5
83=PRE-ANDES geozone	20	73	-18		4
84=ANDES geozone	13	28	-10		6
85=CHACO-CONE geozone	10	25	-10		5
86=MATO-GROSSO geozone	16	36	-5		4
87=AMAZON geozone	23	55	-4		5
88=TUPIC	10	50	-9		6
89=BAHIA geozone	11	14	-11		3
<b>geosector totals</b>	<b>122</b>	<b>415</b>	<b>-111</b>	<b>=304</b>	<b>0</b>

<b>geosector global totals</b>	<b>487</b>	<b>2369</b>	<b>-391</b>	<b>=1978</b>	<b>7</b>
--------------------------------	------------	-------------	-------------	--------------	----------

## phylosectors

1=AFRO-ASIAN	sets	outer	-☞ total	arterial	☰
10=TAMAZIC	1	4			7
11=COPTIC	1	1			-
12=SEMITIC	1	24	-1	Arabic	8
13=BEJIC	1	1			6
14=CUSHITIC	7	25			7
15=EYASIC	2	6	-2		5
16=OMOTIC	6	30			6
17=CHARIC	7	41			5
18=MANDARIC	9	62			6
19=BAUCHIC	8	82	-4		7
<b>phylosector totals</b>	<b>43</b>	<b>276</b>	<b>-7</b>	<b>=269</b>	<b>1</b>

3=AUSTRONESIAN	sets	outer	-☞ total	arterial	☰
30=TAWANIC	11	24	-7		5
31=HESPERONESIC	18	462	-6	Malay-Indonesian; Javanese; Tagalog	8
32=MESONESIC	5	136	-5		6
33=HALMAYAPENIC	1	42	-1		5
34=NEOGUINEIC	7	173	-3		5
35=MANUSIC	9	82	-3		5
36=SOLOMONIC	6	45	-2		4
37=KANAKIC	4	39	-1		4
38=WEST-PACIFIC	8	119	-1		5
39=TRANSPACIFIC	3	57			5
<b>phylosector totals</b>	<b>72</b>	<b>1179</b>	<b>-29</b>	<b>=1150</b>	<b>3</b>

5=INDO-EUROPEAN	sets	outer	-☞ total	arterial	☰
50=CELTIC	1	5			6
51=ROMANIC	1	30	-2	Spanish; Portuguese; French; Italian	8
52=GERMANIC	1	25	-2	English; German	9
53=SLAVIC	1	8		Russian; Ukrainian+Belarussian	8
54=BALTIC	1	3	-1		6
55=ALBANIC	1	1			6
56=HELLENIC	1	2			7
57=ARMENIC	1	1			6
58=IRANIC	1	26			7
59=INDIC	1	49		Hindi-Urdu; Panjabi; Bhojpuri; Bangali; Marathi	9
<b>phylosector totals</b>	<b>10</b>	<b>150</b>	<b>-5</b>	<b>=145</b>	<b>13</b>

7=SINO-INDIAN	sets	outer	-☞ total	arterial	☰
70=TIBETIC	1	25	-1		6
71=HIMALAYIC	3	48			6
72=GARIC	2	23			6
73=KUKIC	4	49			6
74=MIRIC	1	6			6
75=KACHINIC	2	3			5
76=RUNGIC	4	9			5
77=IRRAWADDIC	3	27			7
78=KARENIC	1	12			6
79=SINITIC	1	16		Putonghua; Wu; Yue	9
<b>phylosector totals</b>	<b>22</b>	<b>218</b>	<b>-1</b>	<b>=217</b>	<b>3</b>

9=TRANSFRICAN	sets	outer	-☞ total	arterial	☰
90=ATLANTIC	16	53			7
91=VOLTAIC	9	76			6
92=ADAMAWIC	3	59	-1		6
93=UBANGIC	2	37			6
94=MELIC	2	7			6
95=KRUIIC	1	28			6
96=AFRAMIC	13	66	-1		7
97=DELTIC	2	8			6
98=BENUIC	11	209	-2		7
99=BANTUIC	1	259		Swahili	8
<b>phylosector totals</b>	<b>60</b>	<b>802</b>	<b>-4</b>	<b>=798</b>	<b>1</b>

<b>phylosector global totals</b>	<b>207</b>	<b>2625</b>	<b>-46</b>	<b>=2579</b>	<b>21</b>
<b>LINGUASPHERE TOTALS</b>	<b>694</b>	<b>4994</b>	<b>-437</b>	<b>=4557</b>	<b>28</b>

◀ SEE pp. 291-300 for *Linguasphere in Ten Pages*

