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 20th 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 21st 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.

This spoken surface of the linguasphere, supplemented by written and other derivative forms of communication, is composed of thousands of interactive sub-systems or *languages*, all in a state of gradual change and interchange among succeeding generations and within the minds of plurilingual speakers. The complexity and fluidity of relationships among languages requires the establishment of a finely meshed system of reference. It also requires a shift in the emphasis of linguistic classification away from the reconstruction of distant *levels of prehistorical relationship*, often hypothetical, towards the observation and cataloguing of close *layers of contemporary relationship* among modern languages and dialects. Small variations of language are often very significant in marking the identity of neighbouring communities, and it has proved useful to replace the traditional dichotomy of language and dialect with a sequence of three layers of immediate relationship: *outer language*, *inner language* and *dialect*. Although these three layers form the base of a relatively complex system of reference (see Table One), it must be emphasised that this system has been designed as a practical device to assist the observation and 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*, which together cover the entire linguasphere. These ten sectors comprise five *phylosectors*, corresponding to the five well documented linguistic "families" which together cover the majority of the world's population, 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 Table Two. 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* (see Volume One of the *Linguasphere Register*).

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 Table One.

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 20th 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 15th to the end of the 19th 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 Table One.

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.

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 bilingual communities
➢	notes on languages	#	notes on nomenclature
🔄	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 Two).

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 🏴. Use in two or more states is marked by the icon of crossed flags 🎌.

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 20th century, while the icon ● marks idioms known or believed to be extinct before the end of the 19th.

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

Arterial languages are defined in Volume One as all those outer languages (or networks of largely interintelligible 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 speech within the linguasphere.

PLEASE CONSULT VOLUME ONE OF THE *LINGUASPHERE REGISTER* FOR A MORE DETAILED INTRODUCTION TO THE TERMINOLOGY AND CONVENTIONS USED.

Table One

LINGUASPHERE REGISTER: 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 693 linguistic *sets*

1. Numerical framework of worldwide reference

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

linguasphere key = a fixed two-digit **numerical code** (99 as an example)	marking ➢ two **layers of worldwide reference**	for an inventory of sectors and zones *see Table Two*	TOTALS
(uncoded)	(LINGUASPHERE)	= totality of the world's languages	1
9=	**SECTOR**	= **phylosector** (odd digit 1, 3, 5, 7, 9) *or* **geosector** (even digit 0, 2, 4, 6, 8)	10
99=	**ZONE**	= **phylozone** *or* **geozone**	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-)

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

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

Table Two

THE LINGUASPHERE: REFERENTIAL SECTORS & ZONES

The *Linguasphere Register 1999/2000* enumerates & classifies 4,994 outer & 13,840 inner languages within 694 sets. Each set is assigned to one of 100 referential zones, within one of 10 geosectors or phylosectors.

5 GEOSECTORS = **22** phylozones + **28** geozones	*total* **487** *sets*	**5 PHYLOSECTORS** = **50** phylozones	*total* **207** *sets*
0=AFRICA *geosector*	44 *sets*	**1=AFRO-ASIAN** *phylosector*	43 *sets*
00=MANDIC	**4**	**10=TAMAZIC**	**1**
01=SONGHAIC	**1**	**11=COPTIC**	**1**
02=SAHARIC	**3**	**12=SEMITIC**	**1**
03=SUDANIC	**2**	**13=BEJIC**	**1**
04=NILOTIC	**3**	**14=CUSHITIC**	**7**
05=EAST-SAHEL *geozone*	**16**	**15=EYASIC**	**2**
06=KORDOFANIC	**4**	**16=OMOTIC**	**6**
07=RIFT-VALLEY *geozone*	**4**	**17=CHARIC**	**7**
08=KHOISANIC	**2**	**18=MANDARIC**	**9**
09=KALAHARI *geozone*	**5**	**19=BAUCHIC**	**8**
2=AUSTRALASIA *geosector*	223 *sets*	**3=AUSTRONESIAN** *phylosector*	72 *sets*
20=ARAFURA *geozone*	**26**	**30=TAIWANIC**	**11**
21=MAMBERAMO *geozone*	**22**	**31=HESPERONESIC**	**18**
22=MADANGIC	**23**	**32=MESONESIC**	**5**
23=OWALAMIC	**11**	**33=HALMAYAPENIC**	**1**
24=TRANSIRIANIC	**22**	**34=NEOGUINEIC**	**7**
25=CENDRAWASIH *geozone*	**25**	**35=MANUSIC**	**9**
26=SEPIK-VALLEY *geozone*	**22**	**36=SOLOMONIC**	**6**
27=BISMARCK-SEA *geozone*	**26**	**37=KANAKIC**	**4**
28=NORTH-AUSTRALIA *geozone*	**21**	**38=WEST-PACIFIC**	**8**
29=TRANSAUSTRALIA *geozone*	**25**	**39=TRANSPACIFIC**	**3**
4=EURASIA *geosector*	35 *sets*	**5=INDO-EUROPEAN** *phylosector*	10 *sets*
40=EUSKARIC	**1**	**50=CELTIC**	**1**
41=URALIC	**3**	**51=ROMANIC**	**1**
42=CAUCASUS *geozone*	**3**	**52=GERMANIC**	**1**
43=SIBERIA *geozone*	**4**	**53=SLAVIC**	**1**
44=TRANSASIA *geozone*	**3**	**54=BALTIC**	**1**
45=EAST-ASIA *geozone*	**3**	**55=ALBANIC**	**1**
46=SOUTH-ASIA *geozone*	**11**	**56=HELLENIC**	**1**
47=DAIC	**1**	**57=ARMENIC**	**1**
48=MIENIC	**1**	**58=IRANIC**	**1**
49=DRAVIDIC	**5**	**59=INDIC**	**1**
6=NORTH-AMERICA *geosector*	63 *sets*	**7=SINO-INDIAN** *phylosector*	22 *sets*
60=ARCTIC	**1**	**70=TIBETIC**	**1**
61=NADENIC	**3**	**71=HIMALAYIC**	**3**
62=ALGIC	**3**	**72=GARIC**	**2**
63=SAINT-LAWRENCE *geozone*	**2**	**73=KUKIC**	**4**
64=MISSISSIPPI *geozone*	**3**	**74=MIRIC**	**1**
65=AZTECIC	**1**	**75=KACHINIC**	**2**
66=FARWEST *geozone*	**26**	**76=RUNGIC**	**4**
67=DESERT *geozone*	**5**	**77=IRRAWADDIC**	**3**
68=GULF *geozone*	**8**	**78=KARENIC**	**1**
69=MESO-AMERICA *geozone*	**11**	**79=SINITIC**	**1**
8=SOUTH-AMERICA *geosector*	121 *sets*	**9=TRANSAFRICAN** *phylosector*	60 *sets*
80=CARIBIC	**1**	**90=ATLANTIC**	**16**
81=INTER-OCEAN *geozone*	**16**	**91=VOLTAIC**	**9**
82=ARAWAKIC	**2**	**92=ADAMAWIC**	**3**
83=PRE-ANDES *geozone*	**20**	**93=UBANGIC**	**2**
84=ANDES *geozone*	**13**	**94=MELIC**	**2**
85=CHACO-CONE *geozone*	**10**	**95=KRUIC**	**1**
86=MATO-GROSSO *geozone*	**16**	**96=AFRAMIC**	**13**
87=AMAZON *geozone*	**22**	**97=DELTIC**	**2**
88=TUPIC	**10**	**98=BENUIC**	**11**
89=BAHIA *geozone*	**11**	**99=BANTUIC**	**1**