DOCUMENT RESUME

ED 044 678 AL 002 652

AUTHOR Gisa, Charles F.
TITLE Language Classification and Indexing, With an Annotated Bibliography.
INSTITUTION Center for Applied Linguistics, Washington, D.C. Language Information Network and Clearinghouse System.
SPONS AGENCY National Science Foundation, Washington, D.C.
REPORT NO LINCS-F-70
PUB DATE Jun 70
NOTE 24p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.20
DESCRIPTORS *Annotated Bibliographies, *Indexing, *Information Retrieval, Information Sources, Information Storage, *Language Classification, Languages, Research Tools

ABSTRACT

This document discusses problems and methods of language classification, especially with regard to questions of information storage and retrieval in connection with an information network for the language sciences such as that envisioned by the LINCS (Language Information Network and Clearinghouse System) Project at the Center for Applied Linguistics. An introductory section discusses the need for a comprehensive list of languages and stresses the necessity for at least some degree of grouping if such a list is to be useful and manageable. Section 2 presents some approaches to language classification, while Section 3 illustrates the following methods: alphabetic listing, genetic classification, areal classification, sociolinguistic classification, and typological classification. Section 4 then considers the features necessary for a language classification scheme which can be employed as part of an indexing tool in an information retrieval system, and the final section briefly discusses the Language Names Component of the proposed LINCS indexing tool. The appended annotated bibliography contains 32 entries. (TNP)

CENTER FOR APPLIED LINGUISTICS

LANGUAGE INFORMATION NETWORK AND CLEARINGHOUSE SYSTEM (LINCS)

LANGUAGE CLASSIFICATION AND INDEXING

By Charles A. Zisa

With an Annotated Bibliography

LINCS PROJECT DOCUMENT SERIES / NATIONAL SCIENCE FOUNDATION GRANT

LINCS #5-70 June 1970 NSF GN-771

CENTER FOR APPLIED LINGUISTICS, 1717 MASSACHUSETTS AVENUE, N.W., WASHINGTON, D.C. 20036

LANGUAGE CLASSIFICATION AND INDEXING

By Charles A. Zisa

With an Annotated Bibliography

CONTENTS

1. Introduction

There are several thousand languages spoken in the world today with estimates of their number running from 3,000 to 6,000. If distinct speech forms above the level of idiolect are considered, the number probably stands well into the millions. Each of these distinct speech forms is potentially the subject of special study (e.g., the English of the children of a geographically and nationally defined non-English speaking group was described in one doctoral dissertation).

Along with contemporary speech forms, prior stages of existing languages and languages which are totally extinct and have left no descendants may also be the subject of special study. These fall into two categories: those languages which have left behind a body of literature, and those which have to be reconstructed. In the former case, the text material takes the place of the informant used in the analysis of a contemporary speech form (e.g., the syntax of Old English was described using one of the extant manuscripts in another doctoral dissertation). In the latter case, existing forms from related languages must be compared to produce hypothetical protoforms. Thus, for example, a doctoral candidate reconstructed certain features of proto-Colloquial Arabic.

Although the number of documented extinct speech forms is relatively small (no more than a few hundred), the number of potential reconstructions is huge. For every language and its nearest relative, a mutual proto-form can theoretically be reconstructed; similarly for these two and the next closest relative; likewise for two related groups of languages; and so on up the hierarchy to the originial proto-form of all the languages grouped under the highest rubric. One need only reflect on this to envision the enormousness of the number of potential reconstr ctions. A group of ten languages, for instance, related at the same level, could easily yield 45 reconstructed prototypes.

A complete list of all languages, dialects, and subdialects, both contemporary and extinct, would be of great value and interest. Obviously, however, such a list is an impossibility, as an unknown number of languages have developed, existed, and died without leaving any trace of their existence. Some languages are known only through their descendants or through the effect they have had on existing languages as, for example, in the case of place names which cannot be explained in terms of the existing speech of the area.

Three recent attempts have been made at compiling a complete list of languages which have been specifically identified. Only one of these has been published to date. If, however, such a list were solely an alphabetic listing of names, its value would be slight. The information provided by such a list is only that a particular speech form does (or did) in fact exist and that its name is spelled in a particular fashion. Some degree of grouping is needed to make the list useful and manageable.

2. Some Approaches to Language Classification

In his Language Typology, Horne (1966) lists four approaches to language classification: genetic, areal, sociolinguistic, and typological. He does not mention alphabetic listings, which would be a fifth approach.

The most commonly used of the four approaches mentioned by Horne is the genetic. According to this approach, languages are grouped in terms of common ancestors and by the closeness of their historical relationships as shown by the presence or absences of shared features. Thus, whereas Bulgarian, Macedonian, Russian, and Slovene are all decendants of a common ancestor, Bulgarian and Macedonian are grouped as a subgroup under South Slavic, as they have more features in common than either has with Russian or Slovene. Next the Bulgaro-Macedonian subgroup is put with Slovene because these three share more features than does any other of the three with Russian. Finally, all four are grouped under the general rubric of Slavic to show their common descent from proto-Slavic. The popularity of this approach is probably a result of the extensive comparative studies which were conducted especially during the 19th century.

Under the areal approach, languages are grouped by their location. This approach is commonly used where information concerning genetic relations is lacking. Indeed, most classifications which are basically genetic incorporate elements of the areal approach at various levels. Thus, among the classifications found in a basically genetic classification, African languages and American Indian languages are frequently included, as well as, at lower levels, New Hebrides languages. The use of these designations is not meant necessarily to imply the existence of a single proto-form. Unfortunately, areal classifications have a way of maintaining themselves even when genetic

relationships are found which cut across areal groupings. There is a reluctance by some linguists to admit even the possibility of an American Arctic-Paleo-Siberian language family (including Eskimo, Aleut, Koryak, Kamchadal, and Chukchee) or to depart from the traditional Indonesian-Melanesian-Micronesian-Polynesian division of the Austronesian family.

Groupings based upon common structural features are characteristic of the typological approach. Typological considerations are sometimes used to supplement areal classifications where genetic information is lacking and where the areal groupings are too large. Older language lists, for example, commonly grouped the languages spoken in Australia under the major areal rubric, Australian languages. Subgroups were based upon such criteria as whether the languages were prefixing or not, a typological consideration. Subsequent investigation has indicated that, with two or three exceptions, all Australian languages are probably genetically related, but that the structural similarities probably do not coincide with genetic nearness. Another example of the latter is that English shows, in many respects, greater structural similarities to Persian than to German, which is genetically closer. As research into universals of human language deepens, greater use of the typological approach may be anticipated. At the present time, however, there is too little precise knowledge about the structure of many languages to permit basing a major classification upon this approach.

The basis of the sociolinguistic approach to language classification is function. Languages are grouped in such a way as to reflect their use in the community. This approach has its greatest application in describing a particular language situation rather than languages in general. Some use is made of this approach within basically genetic classifications, when describing social dialects or when treating diglossic situations.

None of these approaches to language classification can, or even should, be labelled as the 'best' one. Each has to be considered in the light of the ultimate purpose to which the classification is being applied. If, for example, one is constructing an index or a 'finder' list, the alphabetic approach is the most satisfactory.

The approach which comes the closest to being an all-purpose approach is the genetic. Its shortcomings are felt only when

it is necessary to account for a series of interrelated speech forms used by the same community under different circumstances. The genetic approach is based upon the concept that languages are discrete units and has but a weak mechanism for describing complex situations or the historical stages of a language.

The areal approach has its greatest utility when the goal of the classification is other than linguistic. In an encyclopedia, for example, in articles discussing the various countries of the world, a listing of the languages spoken in the area may be given. The basic shortcoming of the areal approach is that it leads to non-unique classifications, with some languages under more than one rubric, as they are spoken in several areas.

The sociolinguistic approach is in many respects a refinement of the areal approach. It not only considers the location of the language but also adds the dimension of status in the community using the language. As stated above, this approach is most effective in handling specific situations such as the language situation in Haiti or the Arabic-speaking countries. Its basic shortcoming for use as the basis of a general classification is that the possible number of rubrics under which languages may be grouped is too small and would result in categories with too many members.

Both the sociolinguistic and the typological approaches give insight into the results of languages in contact. The major problem with the typological approach at present is that it has not received sufficient attention to have developed a fully defined technique.

3. Approaches to Language Classification Illustrated

The four approaches listed by Horne (1966), as well as the alphabetic approach, are exemplified below. The samples are based upon the languages spoken and used in the Balkans. The Balkans are defined here as that area of Europe which includes Albania, Bulgaria, Greece, Rumania, Turkey (European part only), and Yugoslavia.

The Balkans were chosen for several reasons. The area forms one cultural unit and is easily delimited. Several languages are spoken in the area, most of which are fairly well known. While they have certain sets of characteristics in common, they represent several different language families. Along with the languages which are specifically associated with the Balkans, there are several languages which are recently intrusive. The Balkans also represent interesting sociological situations.

The data have been restricted to the languages used in the Balkans and to their use in the Balkans. None of the classifications presented is intended to be exhaustive or to be the only possible classification using the particular approach.

A. Alphabetic listing

Albanian	Macedonian
Arabic	Old Slavic*
Armenian	Polish
Balkan Turkic*	Romany*
Bulgarian	Rumanian*
Czech	Russian
German	Serbocroatian
Greek*	Slovak
Hebrew	Slovene
Hungarian	Turkish
Italian	Ukrainian
Judaeo-Spanish	Yiddish
Latin	

* Denotes a cover term for several unspecified languages or dialects.

B. Genetic Classification

Indo-European
- Indic: Romany
- Armenian: Armenian
- Albanian: Albanian
- Hellenic: Greek
- Romance: Latin
  Italian, Rumanian
  Judaeo-Spanish
- Slavic: Old Slavic
  Russian, Ukrainian
  Czech, Polish, Slovak
  Bulgarian, Macedonian, Serbocroatian, Slovene
- Germanic: German, Yiddish

Afro-Asiatic
- Semitic: Arabic
  Hebrew

Uralic: Hungarian

Altaic: Balkan Turkic
Turkish

C. Areal Classification

1. Languages centered in the Balkans

Albanian	(Old Slavic)
Balkan Turkic	Romany*
Bulgarian	Rumanian
Greek	Serbocroatian
Judaeo-Spanish	Slovene
Macedonian	Turkish*

* Although these languages have large bodies of speakers outside the Balkan area, they are sufficiently identified with the Balkans to be included here.

2. Languages intrustive from the Middle East

Arabic	Hebrew
Armenian	Turkish

3. Languages intrusive from Western Europe

German
Italian
(Latin)

4. Languages intrusive from Central and Eastern Europe

Czech	Slovak
Hungarian	Ukrainian
Polish	Yiddish
Russian	

D. Sociolinguistic Classification

Three categories are used to illustrate this approach: (1) official (the language is recognized as an official language in some countries in which it is spoken); (2) vernacular (the language is used in every-day activities but is not recognized as an official governmental language); (3) religious (the language is used in the liturgy of a religious group). Old Slavic is used as a cover term for all liturgical Slavic. No consideration is made of the use of any language outside the Balkan area.

Language	Official	Vernacular	Religious
Albanian	x	x	x
Arabic			x
Armenian		x	x
Balkan Turkic		x	
Bulgarian	x	x	
Czech		x	
German		x	
Greek	x	x	x
Hebrew			x
Hungarian		x	
Italian		x	
Judaeo-Spanish		x	
Latin			x
Macedonian	x	x	
Old Slavic			x
Polish		x	
Romany		x	
Rumanian	x	x	x
Russian		x	
Serbocroatian	x	x	
Slovak		x	
Slovene	x	x	
Turkish	x	x	
Ukrainian		x	
Yiddish		x	

E. Typological Classification

This classification is based upon the position of a segmentable definite article. It does not take into account such features as the definite adjective declension of Serbocroatian or the definite objective case of Turkish. In the first instance, the definite marker is not segmentable from the case marker; in the second, the use of the definite marker is too restricted to be considered a definite article.

1. Preposed definite article

   a. Gender marked
      German
      Greek
      Italian
      Judaeo-Spanish
      Yiddish

   b. Gender unmarked
      Arabic
      Hebrew
      Hungarian

2. Postposed definite article

      Albanian
      Armenian
      Bulgarian
      Macedonian
      Rumanian

3. No definite article

Czech	Serbocroatian (standard)
Latin	Slovak
Old Slavic	Slovene
Polish	Turkish
Russian	Ukrainian

4. Ungrouped
      Romany
      Balkan Turkic

4. User-oriented Language Classification

In the construction of a language classification scheme to be used as part of an indexing tool in an information retrieval system, a prime consideration should be the manner in which the potential user of the system will view his subject matter. If a specialist in Rumanian were asked to provide a grouping of those languages which have the greatest relevance to the study and analysis of Rumanian, he might propose the following grouping:

1. Primary: Latin, Old Slavic, Bulgarian, Turkish, Greek, French.

2. Secondary: Italian, Russian, Serbocroatian.

If, however, Rumanian is considered as it occurs in the sample classifications of the preceding section, it is found in the following groups:

1. As a Romance Indo-European language, together with Latin, Italian, and Judaeo-Spanish.

2. As a Balkan-centered language, together with Albanian, Balkan Turkic, Bulgarian, Greek, Judaeo-Spanish, Macedonian, Old Slavic, Romany, Serbocroatian, Slovene, and Turkish.

3. As a language having official, religious, and vernacular status, with Albanian and Greek.

4. As a language showing a structural feature (the postposed definite article) in common with Armenian, Bulgarian, and Macedonian.

The two sets of groupings do not correspond to each other in whole or in part. The only feature most of the languages in the first set and in the second set have in common is that they have some use in the Balkans. The considerations used by the specialist incorporate many factors, some of which have been used in the second set of classifications. These include areal considerations for Bulgarian, Turkish, Greek, and Serbocroatian and genetic considerations for Latin and Italian. Other considerations have also been taken into account: political and historical (Turkish and Russian); religious (Old Slavic and Greek); sociological (French).

It is possible to draw up such user-oriented groupings for each known language. No one person, however, has the knowledge to construct complete groupings for all languages. It is, therefore, necessary to use other sources. Three types of sources which may be used are (1) university departments, (2) bibliographic references, and (3) biographic descriptions.

The structures of the various language sciences departments at institutions of higher education provide some insight into the way in which specialists group languages (Center for Applied Linguistics 1966; Rütimann 1969). The structures of these departments play a dual role: (1) they reflect the way specialists have structured the field, and (2) they influence the way future specialists will view the field.

Bibliographic references are of two types: classification systems and text references. The former is also divisible into two categories: external and internal. External systems are classifications devised to index printed materials, whereas internal systems refer to indexes or tables of contents which classify and are a part of specific materials. The most representative of the external systems are library classifications. Most of these systems suffer from being either too general or antiquated. Some are more concerned with languages of publication than languages being discussed and thus limit themselves to languages with extensive literatures. Internal classifications tend to be limited to the material being discussed in the work in which they appear and are frequently too personal to be used for a general classification. The most useful members of this category are the systems used in bibliographies. In spite of their shortcomings, classification systems are very significant in the construction of a user-oriented classification.

Text references are also significant in arriving at a picture of the direction which the interests of the specialist may take. In the particular context being discussed here, the languages to which an author makes reference while describing another language are important.

Biographic questionnaires which query the respondent about his special interests provide the most specific indication of how interests pattern. The major problem in using them is that they are directed toward the individual, not the subject matter.

A user-oriented classification of languages is not without problems. It would be impossible to base a classification totally upon patternings of interest of specialists, for the majority of the world's languages have not been studied or analyzed. These languages would not appear, therefore, in the sources listed above. Non-unique classifications would be common as interests do not form discrete units but overlap considerably.

5. The Language Names Component of the Proposed LINCS Indexing Tool

To construct an efficient and usable indexing tool, a classification system is needed which will reflect the patterns for the Language Information Network & Clearinghouse System (LINCS) of interest to the potential clientele but which will also

permit the addition of topics not yet covered by research. Its structure should be one which will be consistent but not static. That is to say, it should have the capacity to adapt to new developments and to accept the addition of new material without violent upheavals in its structure.

The classification should follow a basically genetic approach, the most generally applicable. The genetic hierarchy should be represented through the use of the broader and narrower term concepts in the indexing tool. For example, Germanic (Western) is a broader term with respect to German, and Swiss German, a narrower, in keeping with the genetic relationships within the Germanic language family. Alternate names should appear with a USE designation; for example, German is to be used for Hochdeutsch and German (High).

The two categories of related terms, reciprocal and non-reciprocal, should be used to represent non-hierarchical and non-genetic relationships; that is, to reflect the patternings of interest of the specialists in the fields. Yiddish, for example, has Hebrew, a genetically unrelated language, as a non-reciprocally related term. It would be anticipated that the specialist in Yiddish might have some interest in investigating Hebrew because of the strong influence the latter has had upon the former.

Thus, the proposed classification consists of two subsystems: genetic and user-oriented. The first subsystem, the genetic, satisfies the criteria set up by Greenburg in Essays in Linguistics (1957) for scientific classifications: it is non-arbitrary, exhaustive, and unique. The second, it is hoped, will satisfy the needs of the user community.

References

Center for Applied Linguistics University Resources in the United States for Linguistics and Teacher Training in English as a Foreign Language. Washington, D.C.: Center for Applied Linguistics, 1966.

Greenberg, Joseph H. Essays in Linguistics. Chicago: University of Chicago Press, 1957.

Horne, Kibbey M. Language Typology: 19th and 20th Century Views. Washington, D.C.: Georgetown University Press, 1966.

Rütimann, Hans. "Departmental and Language Information Available in the MLA List of Chairmen." PMLA 84:4. 685-687, 1969.

Appendix

An Annotated Bibliography of Sources of Languages Names and Language Classification

Included in this bibliography are materials which list language names and/or present schemes for language classification. Only materials which have been examined by the author have been included. The emphasis has been upon the more recently developed classifications.

Although some articles from journals have been included, coverage of this source of information is by no means complete. Articles having to do with language classification are to be found in practically all journals focusing upon linguistics. Examples of these journals are Language, International Journal of American Linguistics, and Anthropological Linguistics.

A second category of materials which have not been included are basic linguistics textbooks, most of which contain at least a chapter on the languages of the world and their genetic relationships.

The bibliography has been divided into three sections: generalized language lists, specialized language lists, and minor language lists.

1. Generalized Language Lists.

   The scope of the materials in this section is not limited by geography or language family, although within each item geographic or genetic groupings may be employed. These materials are particularly useful as sources of language names.

1.1 Educational Resources Information Center. Thesaurus of ERIC Descriptors. Washington, D.C.: U.S. Government Printing Office, 1968.

   Contains some language classifications, but is directed to the classification of the materials in the ERIC system rather than to the development of a language classification as such.

1.2 Encyclopaedia Britannica. Chicago: Encyclopaedia Britannica, 1963.

   Articles under "Language" and the names of specific language families give much useful information about classification and the membership of the groups. Different authorship of related articles, however, sometimes results in conflicting information.

1.3 Fraenkel, Gerd. Languages of the World. Boston: Ginn, 1967.

A description of the major languages and language families of the world directed to the non-expert.

1.4 Gołąb, Zbigniew, Adam Heinz, and Kazimierz Polański. Słownik terminologii językoznawczej. Warsaw, 1968.

In Polish. A dictionary of linguistic terms with some comments about specific languages and language groups.

1.5 Hamp, Eric P. "Selected Summary Bibliography of Language Classification." Studies in Linguistics 15:1-2.29-46, 1960.

A substantial bibliography of materials, primarily journal articles, having to do with language classifications.

1.6 Library of Congress Classification. Washington, D.C.: U.S. Government Printing Office, 1965.

A bibliographic classification which tends to be dated.

1.7 Meillet, A., and Marcel Cohen. Les langues du monde. Nouvelle édition. Paris: H. Champion, 1952.

In French. A classic in the field of language classification, although much of the information it contains is now dated.

1.8 Muller, Siegfried H. The World's Living Languages. New York: Frederick Ungar, 1964.

An annotated list of major languages of the world grouped by family.

1.9 Parlett, D.S. A Short Dictionary of Languages. London: English Universities Press, 1967.

Concentrates upon the languages of Europe and Indo-European and other significant languages in an alphabetic format. Gives classificatory, geographic, and other information.

1.10 Pei, Mario A. The World's Chief Languages (formerly Languages for World and Peace). New York: S.F. Vanni, 1960.

Contains grammatical sketches of several major languages with some discussion of language families.

1.11 Pei, Mario and Frank Gaynor. Dictionary of Linguistics. Totowa, N.J.: Littlefield, Adams, 1967.

A dictionary of linguistic terms with numerous languages included. Gives information concerning their relationships, numbers of speakers, and location. Some of the classifications are dated.

1.12 Research Center in Anthropology, Folklore, and Linguistics. Multilingual Thesaurus of the Languages of the World. [Incomplete]

This project was unfortunately never completed, and none of the information collected is available to the public. It is included here primarily to report upon its fate. It represents the most carefully controlled of the recent classifications.

1.13 Trager, George L. "A Bibliographic Classification System for Linguistics and Languages." Studies in Linguistics 3:3-4.54-108, 1945.

_____. "A Bibliographical Classification System for Linguistics and Languages (Alphabetical Indexes). SiL 4:1-2.1-50, 1946.

_____. "Revisions to A Bibliographical Classification System: 2." SiL 9:4.91-93, 1951.

Directed primarily to the classification of bibliographic materials. It is now somewhat dated.

1.14 Voegelin, C.F., and F.M. Voegelin, eds. "Languages of the World." Anthropological Linguistics 6:3-7, 1964; 7:2,3-7 (Part 1 of each), 8,9, 1965.

A series published as supplements to Anthropological Linguistics from 1964 to 1965. The most comprehensive classificatory list published to date. Some typographical errors and contradictory classifications. The last two issues published are an alphabetic index.

1.15 Winick, Charles. Dictionary of Anthropology. Totowa, N.J.: Littlefield, Adams, 1968.

Lists several languages of interest to the anthropologist with classifications, numbers of speakers, and location.

1.16 Zisa, Charles A. Directory of Language Names. [In preparation] An alphabetic list of language names, dialects, and alternate names with classifications.

2. Specialized Language Lists.

The limitations upon a specialized language list may be geographic (e.g., the languages of Africa); genetic (e.g., Indo-European languages); or other (e.g., languages with more than one million speakers). The particular value of these lists is that, in most cases, they have been compiled by experts in the area covered by the list. They are, therefore, especially valuable in low-level classification.

2.1 Amankwe, Nwozo. Classification of African Languages I. West Africa. Nsukka, East Nigeria: University of Nigeria, n.d.

A listing of the languages of West Africa and an attempt to develop a classification for bibliographic purposes. Contains gross classification of the languages and alternate names.

2.2 Baskokov, N.A. Tjurkskie jazyki. Moscow: Izdatel'stvo Vostočnoj Literatury, 1960.

In Russian. A discussion of the Turkic languages, their history, numbers of speakers, internal relationships, and structural characteristics.

2.3 Capell, A. A Linguistic Survey of the South-Western Pacific. New and revised edition. Nouméa, New Hebrides: South Pacific Commission, 1962.

Contains maps, classifications, and numbers of speakers, along with grammatical notes and finder lists of the languages of New Guinea, the Solomons, the New Hebrides, New Caledonia and the Loyalties, and Nauru.

2.4 Cense, A.A., and E.M. Uhlenbeck. Critical Survey of Studies on the Languages of Borneo. The Hague: Martinus Nijhoff, 1958.

Primarily a bibliography with comments about the languages of Borneo.

2.5 Collinder, Björn. Survey of the Uralic Languages. Stockholm: Almqvist and Weksell, 1957.

_____. Comparative Grammar of the Uralic Languages. Stockholm: Almqvist and Weksell, 1960.

_____. An Introduction to the Uralic Languages. Los Angeles: University of California Press, 1965.

All three contain detailed classifications of the Uralic languages.

2.6 Cust, Robert N. A Sketch of the Modern Languages of the East Indies. London: Trubner, 1878.

Discusses the languages of India, Southeast Asia, and Indonesia. Much valuable, if dated, information.

2.7 Dauzat, Albert. L'Europe linguistique. Nouvelle édition. Paris: Payot, 1953.

In French. A discussion of the historical, geographic, and sociological aspects of the languages of Europe.

2.8 De Bray, R.G.A. Guide to the Slavonic Languages. New York: E.P. Dutton, 1951.

Gives detailed information about the morphology of each Slavic language and includes notes about the dialects and history of the individual languages.

2.9 Entwistle, W.J., and W.A. Morison. Russian and the Slavonic Languages. London: Faber & Faber, 1949.

A philological discussion of the Slavic group with some information about internal relationships.

2.10 Geiger, B., Tibor Halasi-Kun, Aert H. Kuipers, and Karl H. Menges. Peoples and Languages of the Cacasus. The Hague: Mouton, 1959.

Gives ethnographic information and languages, together with the dialects, numbers of speakers, and status of each language.

2.11 Grace, George W. The Position of the Polynesian Languages within the Austronesian (Malayo-Polynesian) Language Family. Bloomington: Indiana University, 1959. (Published as IJAL Memoir 16.)

Primarily a discussion of methodology of language classification.

2.12 Greenberg, Joseph. The Languages of Africa. Bloomington: Indiana University, 1963. (Published as IJAL 29:1 [Part II].)

An expansion and revision of Studies in African Linguistic Classification which appeared in 1955. While not complete, it is the major current classification of African languages.

2.13 Grierson, G.A. Linguistic Survey of India. Delhi: Motilal Banarsidass, 1928.

An eleven-volume series containing descriptions of the languages of India.

2.14 Handbook of African Languages series. London: International African Institute.

A series of several volumes concerning the locations, classification, numbers of speakers, and salient structural features of African languages.

2.15 Hollyman, K.J. A Checklist of Oceanic Languages. Auckland: Linguistic Society of New Zealand, 1960.

An alphabetic list of language names of Melanesia, Micronesia, New Guinea, and Polynesia giving locations and broad classifications with bibliographic references.

2.16 Leenhardt, Maurice. Langages et dialectes de l'Austro-Mélanésie. Paris: Institut d'Ethnologie, 1946.

In French. Contains structural sketches of the languages of New Caledonia.

2.17 Linguistic Circle of Canberra Publications series. Canberra: Australian National University.

A series concerning the languages of New Guinea and Australia with much valuable information.

2.18 Linguistic Comparison in South East Asia and the Pacific. London: University of London, 1963.

A discussion of possible and demonstrated relationships among Southeast Asian and Pacific languages.

2.19 Matthews, W.K. Languages of the USSR. Cambridge: Cambridge University Press, 1951.

A listing with structural descriptions. Minor languages have been omitted.

2.20 Mayers, Marvin K., ed. Languages of Guatemala. The Hague: Mouton, 1965.

Contains ethnographic comments and structural notes together with texts of indigenous Guatemalan languages.

2.21 McQuown, Norman A. "Los Lenguajes Indigenas de América Latina." Revista Interamericana de Ciencias. 1:1.37-207, 1961.

In Spanish. A list of Latin-American Indian languages with variants, classification, and location.

2.22 Miller, Roy Andrew. The Japanese Language. Chicago: University of Chicago Press, 1967.

The chapter "Genetic Relationship" goes deeply into the relationships between Japanese, Korean, Okinawan. Gives a good description of comparative technique.

2.23 Sarkar, Amal. Handbook of Languages and Dialects of India. Calcutta: K.L. Mukhopadhajay, n.d.

A list of languages reported in various language surveys and censuses taken in India with locations, numbers of speakers, and classification. Includes a discussion of some questionable entries.

2.24 Shafer, Robert, ed. Bibliography of Sino-Tibetan Languages. Wiesbaden: Otto Harrassowitz, 1957.

"A bibliography of all known Sino-Tibetan languages" in alphabetic order. Lists variant names.

2.25 Thomas, Cyrus. Indian Languages of Mexico and Central America and their Geographical Distribution. Washington, D.C.: U.S. Government Printing Office, 1911.

2.26 Tovar, Antonio. Catálogo de las Lenguas de América del Sur. Buenos Aires: Editorial Sudamericana, 1961.

In Spanish. Probably the most complete list of South American languages and possible languages. Gives bibliographic references for each.

2.27 Trager, George L., and Felicia E. Harbin. North American Indian Languages: Classification and Maps. Buffalo: University of Buffalo, 1958.

2.28 Uhlenbeck, E.M. A Critical Survey of Studies of the Languages of Java and Madura. The Hague: Martinus Nijhoff, 1964.

Primarily a bibliography with comments about the languages of Java and Madura.

2.29 Waterman, John T. A History of the German Language. Seattle: University of Washington Press, 1966.

Contains a good if general description of the branches of Indo-European.

2.30 Watson, James B., ed. New Guinea: the Central Highlands. Menasha, Wisconsin: American Anthropological Association, 1964. (Published as a special publication of American Anthropologist 66:4, Part 2).

Contains much useful information about the relationships of the languages of New Guinea highlands with structural notes.

2.31 Welmers, William E. A Survey of the Major Languages of Africa. Washington, D.C.: Center for Applied Linguistics, 1959. (Published as a supplement to Linguistic Reporter 1:2).

A listing of the languages of Africa with 500,000 or more speakers indicating classification and location. Contains notes on the classifications of African languages in general.

2.32 Zograf, G.A. Jazyki Indii, Pakistana, Cejlona i Nepala. Moscow: Izdatel'stvo Vostočnoj Literatury, 1960.

In Russian. Gives classifications and structural information concerning the languages of India, Pakistan, Ceylon, and Nepal.

3. Minor Language Lists

Much valuable information concerning the genetic relationships and dialects of specific languages can be found in materials describing the individual language. A fairly typical example is the following from Teach Yourself Icelandic by P.J.T. Glendening (London: The English Universities Press, 1961)

> The surviving members of the Germanic branch of the Indo-European family of languages are: of the Western branch, German on the one hand and Dutch and English on the other; and of the Northern branch, Swedish, Danish, Norwegian, Icelandic and Faroese.. The development of certain differences which made it possible to divide this Germanic branch into Western and Northern (the Eastern being Gothic and related languages, all long since dead) occurred about 400 B.C. to 100 B.C., while the emergence of significant differences in the Northern branch became decisive about the year A.D. 800. At this time the Scandinavian dialects were but variations on an original theme, while English, or rather the Anglo-Saxon dialects, were not far removed from Norse in structure, sounds, or vocabulary.

Such information can be found in the introduction to grammar books and descriptive articles in linguistic journals. Although they are occasionally unreliable, they are collectively an important source of information. The number of individual items, however, is enormous, and no attempt has been made to list them here.