United Nations
Educational, Scientific and
Cultural Organization

UNESCO INSTITUTE FOR STATISTICS

Communication and Information Sector

SUSTAINABLE DEVELOPMENT GOALS

P.O.Box 6128
Succursale Centre-Ville
Montreal, QC H3C 3J7
www.uis.unesco.org

August 2018

Instruction Manual
Survey of World Languages

Table of contents

Introduction **3**

1. Overview of the questionnaire **4**

a. Format 4

b. Completion 4

c. Data 4

d. Structure 4

e. Questions 5

2. Questionnaire structure and definitions **6**

Section 1. General information 6

Section 2. Specific language information 7

Part I. Language status/status and affiliation 7

Part II. State of the language 9

Part III. Users of the language 10

Part IV. Use of the language 13

Introduction

The UNESCO Survey of World Languages seeks to collect data to develop the UNESCO World Atlas of Languages web platform, an interactive and dynamic online tool for the documentation, safeguarding and promotion of linguistic diversity around the world. The purpose of the survey is establish a database of languages data used around the world.

The first target of this questionnaire is to gather data on official languages (if available). These languages are statutory and used in government and the administration of states. An official language is not necessarily used throughout the entire country and may serve only a region or minority group within that country. A large number of languages are used in many countries but are not considered official. These languages, which can be regional, local or community languages, are the second focus of this questionnaire.

It is important to report information on both official and non-official languages to produce an accurate database.

Questionnaire submission

The electronic questionnaires are available in English, French, Russian and Spanish on the UIS website: http://uis.unesco.org/en/methodology

Completed questionnaires should be sent by email attachment to uis.ccsurvey@unesco.org

If you experience problems accessing the site or for any questions related to the data collection, please contact Arthur Nogacz at uis.ccsurvey@unesco.org or by telephone at (1) 514 343 6880.

Overview of the questionnaire

a. Format

There are two questionnaires to be completed: one for sign languages and one for spoken languages.

The nature of the data collected by both questionnaires is identical. Differences between the questionnaires refer to definitions that change according to whether a language is signed or spoken. In addition, some questions asked in the spoken languages questionnaire do no translate into sign languages (e.g. regarding the literacy of users) and vice versa.

b. Completion

The questionnaire is to be completed by:

- Official government organization of a UNESCO Member State (Ministry of Culture, language institution, national statistical office);
- Non-governmental organizations of language users on behalf of a UNESCO Member State; or
- Expert or group of experts on behalf of a UNESCO Member State.

Each questionnaire workbook can capture information for five different languages. If more than five languages are used in your country, more than one questionnaire can be used. All completed questionnaires are to be returned together by email attachment.

c. Data

The questionnaire should be completed with the most recent/complete data available (or best combination thereof). Additionally, the UIS encourages all countries to make their own estimations of missing or incomplete data.

d. Structure

The questionnaire is divided into two sections:

> **General information:** information on the person/institution completing the questionnaire* and the listing of languages (and their characteristics) which will be entered in the questionnaire. Only one general information sheet is to be completed per questionnaire.
>
> *The person identified as being responsible for completing the questionnaire will act as liaison should further clarification on responses be required.

Specific language information: specific information on the status of the language, demographics, users and general use. This section is divided into four parts:

i. Language status/status and affiliation
ii. State of the language
iii. Users of the language
iv. Use of the language.

Up to five specific language information sheets can be completed per questionnaire (one per language).

e. Questions

The questionnaire consists of four kinds of questions:

i. Single-choice selections, where only one option can be chosen. The options in single-choice selection questions are marked with a round tick box ○.

ii. Multiple-choice selections, where more than one option can be chosen at a time. "Please select all that apply" will appear in these questions. The options in multiple-choice selection questions are marked with a square tick box ☐.

iii. Open-format responses, where the question is to be answered in written form (either verbal or numeric, according to each individual question). These questions are answered within a textbox [ ].

iv. Scroll-down questions, where a limited number of options are provided and the respondent may choose one. These responses are selected in a scroll box [▼].

Questionnaire structure and definitions

The instructions below follow the same structure as the questionnaire (2 sections, with further subdivisions).

Definitions will not be provided on a question-by-question basis, as (1) questions may differ slightly between the two questionnaire formats (spoken and sign languages); (2) certain questions do not have specific definitions attached to them; and (3) certain definitions are used more than once during the questionnaire. Definitions are provided in the order that they appear in the questionnaires, but their numbering does not represent the equivalent question number (e.g. Definition 7 does not necessarily represent information for Question 7). The questions for which the definitions are used are listed at the top right side of each definition box.

If definitions apply to only one format of the questionnaire, this will be indicated. Definitions for terms used in the questionnaire are provided in the boxes, while an introduction to certain themes of the questionnaire is presented outside of these boxes. Not every definition box contains a theme-introduction.

Section 1. General information

This section identifies the person responsible for completing the questionnaire. This person will act as liaison should there be a need for further clarification on responses.

A list of languages used in your country is to be provided. Languages listed should be in the same order as the worksheets where specific language information is entered (e.g. a language entered at the “Language 1” slot should have its specific information entered on the “Language 1” sheet).

”Code” refers to the International Organization for Standardization’s codes for comprehensive coverage of languages (ISO 639-3). Each code consists of three letters, which represent a specific language.

Definition 1. General language information

Name: Common name of the language used in your country.

Status: Official or legal status as defined by the national administration and/or constitution.

Other names: Other names of the language (if any).

Code: Associated ISO 639-3 code (if available)[1].

[1] A complete list of ISO 639-3 codes can be found at: https://iso639-3.sil.org/code_tables/639/data

Section 2. Specific language information

Reminder: One "Specific Language Information" sheet should be completed for each language listed in Section 1.

Part I. Language status/status and affiliation

This part of the questionnaire focuses on the official, government-designated status of languages, as well as language's genetic affiliation (spoken languages only).

The most frequent status labels for languages are: i) primarily functional pragmatic official language; and ii) more political, ideological national language. Whereas the concept of a national language is always connected – *de facto* or *de jure* – with people and a territory, the term official language refers to the statutory language used by the government and in the administration of a sovereign state.

Definition 2. Language status

Applicable to Question 1 (both questionnaires)

Official language: A statutory language of the government and the administration of a sovereign state. The official functionality is either constitutionally anchored or results from the conventionalised practice in a country.

Regional language: (Co-)official language in an administrative area (federal state or province or district) of a sovereign state.

Minority language: Language used by members of a minority group or community within a dominant language context with the legal status of a (co-)official language in public formal domains, mostly within a defined territory of a sovereign state. If speakers of a minority language are dispersed across the territory of a state, their language is characterised as non-territorial.[2]

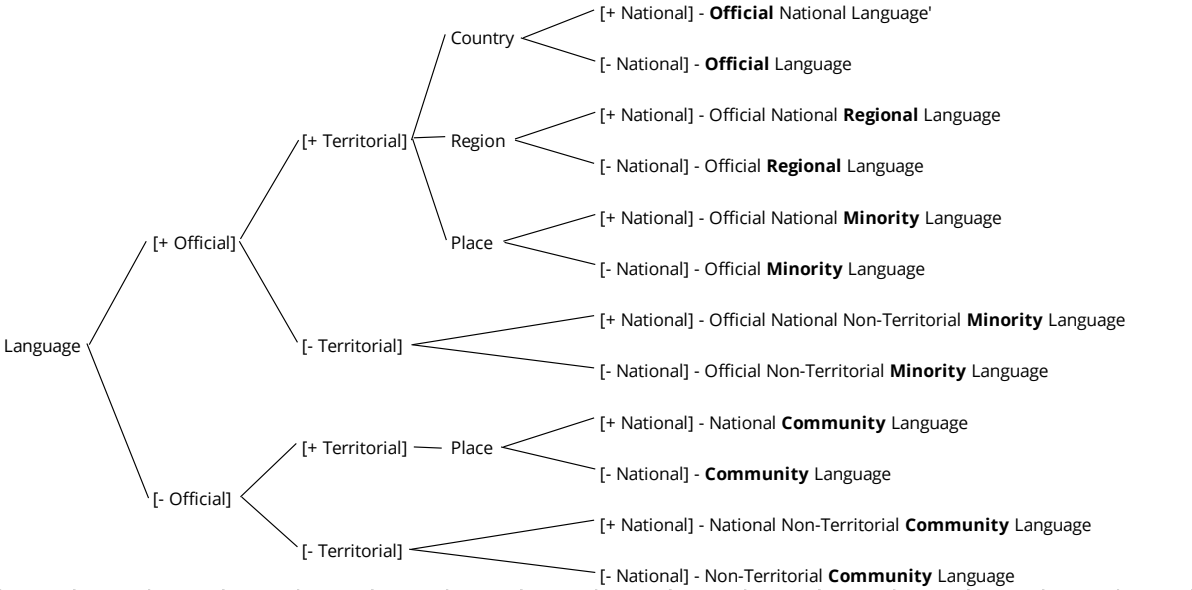
Community language: Language used by members of a minority group or community within a dominant language context.[3] Depending on whether users live concentrated in one area or are dispersed across the territory of a state, their language is additionally defined as territorial or non-territorial. Furthermore, a community language may be defined – *de facto* or *de jure* – as a national language. However, it generally does not function as an official language.

[2] According to the European Charter for Regional or Minority Languages, the term **non-territorial minority language** means the official language used by nationals of the state, which differs from the language(s) used by the rest of the state's population but which, although traditionally used within the territory of the state, cannot be identified with a particular area thereof.

[3] **Minority group** or **community** refers to a cultural, ethnic, religious or social sub-division of a society, subordinate to a dominant group in its political, financial or social power, regardless of the size of either group.

Territorial application: Federal state, province, region, district, town or village where the status of the language applies. If speakers are dispersed across the territory of a state, no entry is needed.

The underlying dichotomies of **official**, **territorial** and **national** designations result in 12 types of languages:



The genetic affiliation or genetic grouping of a language is based on two classification criteria: whether a language is related to other languages and whether the language is a mix of two or more languages.

Definition 3. Language's genetic information (spoken languages only)

Applicable to Question 2 (spoken languages)

Languages are differentiated as:

Language isolate: Language that is not part of a group (family) of languages.
Language: Language that is part of a group (family) of languages.
Bilingual mixed language: Natural language that consists of elements of two languages.
Creole language: Natural language that historically developed from a Pidgin.
Pidgin: Auxiliary language with reduced functionality.

Language affiliation:

Language: Family + branch + sub-branch + etc.
Bilingual mixed language: Two languages whose elements are used in the language.
Creole language and Pidgin: Language from which most of the vocabulary comes.
Natural language: Human spoken/signed language.

Part II. State of the language

This part of the questionnaire focuses on the state of a language in terms of the availability of materials, graphisation (spoken languages only), asynchronicity (sign languages only) and language standardisation.

Definition 4. Materials in the language

Applicable to Question 3 (spoken Languages) and Question 2 (sign languages)

Corpus: Collection of materials/text, assembled for the purpose of linguistic research. May be in written, audio and video formats.

Material: Archived and accessible example of language use. May be in written, audio and video formats.

Text: Original spoken/signed production that forms a coherent stretch of language that may be used for critical analysis.

Written use of natural human spoken languages ranges from limited and infrequent spontaneous writing to the existence of a standardised writing system used by (almost) all users of the language. The availability of script encoding (e.g. Unicode) is linked to the digital usability of a language in its written form.

For sign languages, video technology can perform many of the same functions that reading and writing does for spoken languages. Paramount in this regard is the extent to which sign language recordings are used in ways that parallel the use of spoken languages. These functions are summarised under the label asynchronicity.

Definition 5. Asynchronicity (sign languages)

Applicable to Question 4 (sign languages)

Asynchronicity: Refers to the design and use of a signed video for long-term reference, repeated (re)viewing, searching/scrolling, heightened meta-linguistic attention and explicit structuring for content other than casual conversation. Asynchronicity may or may not be linked to educational contexts. Examples of signed texts intended for asynchronous use include signed academic publications, lectures, translations of factual content into signing, bilingual libraries, children's stories and creative performances in sign language.

Standardisation of a language complies with the language's level of codification in conjunction with its legal status. It correlates with the level of documentation and, in the case of natural human spoken languages, with encoding and graphisation. The level of standardisation of a language reflects the conventionalised conformity of usage.

Definition 6. Standardisation

Applicable to Question 6 (spoken languages) and Question 5 (sign languages)

Modern standard language: Legally supported codified norms developed over a long period of time.

Young standard language: Legally supported codified norms created in the post-colonial area.

Quasi-standard language: Generally accepted codified norms; So-called archaic standard languages as Latin, Sanskrit, etc. are part of this group. They have legally unregulated but conventionalised norms that are generally used.

Standardised language: Codified norms accepted by a majority of users.

Semi-standardised language: Flexible norms with user-specific variation.

Unstandardised language: No norms.

Norm: The aggregate of conventionalised linguistic means in common use in a given language, including the rules governing the use of such means that have become generally accepted by a specific linguistic community during a specific historical period.

Part III. Users of the language

The questions in this part of the questionnaire refer specifically to information on users of the language. Data for this part of the questionnaire will be preferentially drawn from a census (if available). Alternatively, approximations based on reports, surveys, qualitative analyses and others are accepted (please provide source when applicable).

Definition 7. Demographics

Applicable to Question 7 (spoken languages) and Question 6 (sign languages)

User: Anyone who uses the language regularly in everyday life, irrespective if as a first or second language.

Territory: Federal state, province, region, district, town or village.

Rural setting: Users living mainly in the countryside in villages.

Urban setting: Users living mainly in towns and cities.

Definition 8. Reference community

Applicable to Question 10 (spoken languages) and Question 9 (sign languages)

Reference community (spoken language): Potential linguistic community that is defined by shared ethno-cultural norms, ideologies and practices by its members.

Reference community (sign language): Community of potential sign language users, which consists of all deaf people (except those deafened due to old age), hearing relatives of deaf people and other groups of hearing people in regular contact with deaf signers.

The UN's social indicators differentiate generations into three categories.

Definition 9. Generational distribution[4]

Applicable to Question 11 (spoken languages) and Question 10 (sign languages)

Young generation: Under the age of 15, range from children to young adults.

Middle generation: Aged 15 to 65 years, range from young adults to the established generation and include the so-called parental generation.

Older generation: Over the age of 65, range from older adults to the elders and include the so-called grandparent and great-grandparent generations.

Definition 10. Education[5]

Applicable to Questions 12 and 22 (spoken languages) and Questions 12 and 20 (sign languages)

Educational attainment: The highest education level completed by an individual, which is typically certified by a recognised qualification.

Early childhood education: Follows a holistic approach to support children's early cognitive, physical, social and emotional development for participation in school and society (also referred to as pre-primary education).

Primary education: Provides fundamental literacy skills and establishes a solid foundation for understanding core areas of knowledge, as well as personal and social development.

Continued...

[4] Based on UN's social indicators definitions: https://unstats.un.org/unsd/demographic/products/socind/default.htm

[5] Based on ISCED 2011: http://uis.unesco.org/en/topic/international-standard-classification-education-isced

Lower secondary education: Lays the foundation for lifelong learning and human development to pursue further educational opportunities. It may offer vocational education programmes.

Upper secondary education: Typically designed to complete secondary education in preparation for tertiary education or to provide skills relevant to employment, or both.

Tertiary education: Provides learning at a high level of complexity and specialisation, includes not only academic but also advanced vocational or professional education.

The International Labour Organization (ILO) defines several categories of occupational categories through their International Standard Classification of Occupations (ISCO-08).

Definition 11. Occupations[6]

Applicable to Question 14 (spoken languages) and Question 13 (sign languages)

Managers plan, direct, coordinate and evaluate the overall activities of enterprises, governments and other organizations, or of organizational units within them, and formulate and review their policies, laws, rules and regulations.

Professionals increase the existing stock of knowledge, apply scientific or artistic concepts and theories, teach about the foregoing in a systematic manner or engage in any combination of these activities.

Technicians and associate professionals perform technical and related tasks connected with research and the application of scientific or artistic concepts and operational methods, and government or business regulations.

Clerical support workers record, organize, store, compute and retrieve information, and perform a number of clerical duties in connection with money-handling operations, travel arrangements, requests for information and appointments.

Services and sales workers provide personal and protective services related to travel, housekeeping, catering, personal care, protection against fire and unlawful acts or demonstrate and sell goods in wholesale or retail shops, as well as at stalls and markets.

Skilled agricultural, forestry and fishery workers grow and harvest field or tree and shrub crops; gather wild fruits and plants; breed, tend or hunt animals; produce a variety of animal husbandry products; cultivate, conserve and exploit forests; breed or catch fish; and cultivate or gather other forms of aquatic life in order to provide food, shelter and income for themselves and their households.

continued...

[6] Based on ISCO-08: http://www.ilo.org/public/english/bureau/stat/isco/isco08/

Craft and related trades workers apply specific technical and practical knowledge and skills to construct and maintain buildings; form metal; erect metal structures; set machine tools or make, fit, maintain and repair machinery, equipment or tools; carry out printing work; and produce or process foodstuffs, textiles, wooden, metal and other articles, including handicraft goods.

Plants and machine operators and assemblers operate and monitor industrial and agricultural machinery and equipment on the spot or by remote control; drive and operate trains, motor vehicles and mobile machinery and equipment; or assemble products from component parts according to strict specifications and procedures.

Elementary occupations involve the performance of simple and routine tasks, which may require the use of hand-held tools and considerable physical effort.

Definition 12 – Literacy (spoken language only)

Applicable to Question 16 (Spoken languages)

Literacy: ability to read and write and to use language related numbers, images, as well as other dominant language related symbol systems to understand, to communicate, and to gain useful knowledge.

Definition 13. Digital use

Applicable to Question 17 (spoken languages) and Question 15 (sign languages)

Digital user (spoken languages): Person who uses the language in question with digital devices in text production and written electronic communication.

Digital user (sign languages): Person who uses the language in questions with digital devices in electronic communication (example: video recordings).

Part IV. Use of the language

The questions in this part of the questionnaire refer specifically to information on how the language is used. Data should be drawn from a census, if possible. Alternatively, approximations based on reports, surveys, qualitative analyses and other are accepted (please provide the source when applicable).

Definition 14. Economic dimensions

Applicable to Question 19 (spoken languages) and Question 17 (sign languages)

Primary sector extracts or harvests products from the earth, including the production of raw materials and basic foods: agriculture (subsistence and commercial), mining, forestry, farming, grazing, hunting and gathering, fishing, quarrying, etc.

Secondary sector manufactures finished goods, including metal working and smelting, automobile and textile production, chemical and engineering industries, aerospace manufacturing, energy utilities, engineering, construction, shipbuilding, etc.

Tertiary sector refers to the service industry, including retail and wholesale sales, transportation and distribution, entertainment (films, television, radio, music, theatre, etc.), restaurants, clerical services, media, tourism, insurance, banking, healthcare, law, etc.

Quarternary sector covers intellectual activities, including government, culture, libraries, scientific research, education, information technology, etc.

Quinary sector covers the highest levels of decision-making, including top executives or officials in government, science, universities, non-profit, healthcare, culture, media, etc.

The functionality of a language is specified by its domain-specific distribution.

Definition 15. Functionality of language use

Applicable to Question 20 (spoken languages) and Question 18 (sign languages)

Domain: A set of communicative situations that share a certain setting, specific roles relationships between the interlocutors and typical topics, e.g. the educational domain, the home domain, the work domain, etc.

As functional varieties are characterised by their degree of publicness, which ranges from situations of private intimate to public formal language use, the domain continuum of a repertoire may be divided into three layers:

Private domains: Language used in the social microcosm and in informal situations of private life with interlocutors of the closer social environment, e.g. with a partner, children, family, close friends, etc.

Everyday domains: Language used in the social macrocosm and in informal situations of social encounters in everyday life with interlocutors of the wider social environment, e.g. friends, acquaintances, colleagues, service providers, strangers, etc.

Public domains: Language used in the public sphere (of mainly institutional settings) and in formal situations of administration, at court, in education, in the media, at the political level, etc.

Nature of language use

Exclusive: The language in question is exclusively used in the particular domain cluster/social sphere.

Equal: The language in question and others have equivalent/parallel functionalities in the particular domains; all the languages function in all domains of the social sphere.

Complementary: The language in question and others have complementary functionality; they have different functions regarding their use in particular domains of the social sphere.

Stability of language use

Shifting decreasing: The language in question is losing function and may be replaced by others in the particular domain cluster/social sphere.

Shifting increasing: The language in question is gaining function and may replace others in the particular domain cluster/social sphere.

Definition 16. Administration levels

Applicable to Question 21 (spoken languages) and Question 19 (sign languages)

Administrational language use can play out on 4 territorial political levels:

International: Language is used in the administration of international organizations.

National: Language is used in the administration of the whole territory of a state.

Regional: Language is used in the administration of one or more political/administrative unit of a state, e.g. federal states, provinces, counties, districts, etc.

Local: Language is used in the administration of cities, municipalities, etc. of a state.

Census data on language: Data on languages are collected through a census.

Language on census surveys: Language is used for census surveys or census forms.

Definition 17. Ethnoculture

Applicable to Question 23 (spoken languages) and Question 21 (sign languages)

Traditions and expressions encompass a variety of spoken/signed forms including proverbs, riddles, tales, nursery rhymes, legends, myths, epic songs and poems, charms, prayers, chants, songs, dramatic performances and more. Traditions and expressions are used to pass on knowledge, cultural and social values and collective memory.

Knowledge and practices concerning nature and the universe include know-how, skills, practices and representations developed by communities by interacting with the natural environment. These ways of thinking are not only expressed through language and oral traditions but also through memories, spirituality and worldview.

Traditional medicine is the sum total of the knowledge, skills and practices based on the theories, beliefs and experiences indigenous to different cultures, whether explicable or not, used in the maintenance of health as well as in the prevention, diagnosis, improvement or treatment of physical and mental illness.

Customary law is a system of rules and regulations administered by a forum of community elders, mostly presided by a traditional leader that meets to resolve any dispute, which has arisen between members of the community.

Social practices, rituals and festive events (including religious practices) are habitual activities that structure the lives of communities. They are significant because they reaffirm the identity of those who practice them as a group or a society and, whether performed in public or private, are closely linked to important events.

Performing arts range from vocal (spoken languages) and instrumental music, dance and theatre to pantomime, sung verse (spoken languages) and beyond. They include numerous cultural expressions that reflect human creativity.

Traditional craftsmanship mainly refers to traditional skills and knowledge. Safeguarding attempts should concentrate on encouraging artisans to continue to produce craft and to pass their skills and knowledge onto others within their own communities.

Informal learning is defined as forms of learning that are intentional or deliberate but are not institutionalised. It may include learning activities that occur in the family, workplace, local community and daily life, on a self-directed, family-directed or socially-directed basis.

Definition 18. Healthcare

Applicable to Question 24 (spoken languages) and Question 22 (sign languages)

Public healthcare: A system/programme by which healthcare is made available to the population and financed by government, private enterprise or both.

The International Standard Industrial Classification Revision 4 (ISIC Rev. 4) provides five distinct categories that refer to information, communication and cultural production.

Definition 19. Information, communication and cultural production[7]

Applicable to Question 25 (spoken languages) and Question 23 (sign languages)

Publishing activities include the production of books, brochures, leaflets, dictionaries, encyclopedias, atlases, maps and charts; the publishing of newspapers, journals and periodicals; directory and mailing list and other publishing, as well as software publishing.

Motion picture, video and television programme production includes theatrical and non-theatrical motion pictures whether on film, videotape, DVD or other media, including digital distribution; supporting activities such as film editing, cutting, dubbing etc.

Sound recording and music publishing activities (spoken language only) include the production of original (sound) master recordings, such as tapes, CDs, etc. and sound recording service activities in a studio or elsewhere, including the production of taped (i.e. non-live) radio programming.

Programming and broadcasting activities include creating and broadcasting content, such as radio, television and data programmes of entertainment, news, talk, etc.; broadcasting can be performed using different technologies, over-the-air, via satellite, via a cable network or via Internet.

Information service activities include the activities of web search portals, data processing and hosting activities, as well as the provision of infrastructure for hosting and data processing services, search facilities and other portals for the Internet.

[7] Based on ISIC Rev.4 code: https://unstats.un.org/unsd/classifications/unsdclassifications

Definition 20. Periodical publications (spoken languages only)

Applicable to Question 26 (spoken languages)

Newspaper: Daily, weekly or biweekly publication.
Magazine: Weekly, bimonthly, monthly, quarterly or biannual publication.
Yearbook: Annual publication.

Nature of use

Monolingual: Use of only one language.
Bilingual: Language is used together with another language.
Print run: Number of copies of newspaper, magazine or yearbook printed in each edition.

Definition 21. Language use in radio (spoken languages only) and television

Applicable to Questions 27 and 28 (spoken languages) and Question 24 (sign languages)

Permanent programme: Regularly scheduled programme always present in schedule.
Regular programme: Regularly schedule programme not necessarily always present in schedule.
Irregular programme: Programme that is not regularly scheduled.

Definition 22. Legal system

Applicable to Question 30 (spoken languages) and Question 25 (sign languages)

Testimonies: Spoken/signed language use in accordance with the human rights of defendants and plaintiffs for testimonies in their mother tongue (with translation into the dominant language).
Written submissions: Written language use for legal submissions (and translation into the dominant language).
Court proceedings: Use of spoken/signed language in court proceedings.
Legal transactions: Language used for all kinds of legal transactions.
Legal provisions: Language used for all legal provisions and laws.