UNITED NATIONS

Group of Experts on Geographical names

Twelfth Session
Geneva, 29 September - 7 October 1986

WORKING PAPER No.57

1 October 1986

<u>Agenda Item 6(d)</u>

<u>An Automated Toponymic Transliteration/Transcription/Translation
System for World Atlases</u>

Presented by Mr. N. Kadmon (Israel)

AN AUTOMATED TOPONYMIC TRANSLITERATION/TRANSCRIPTION/TRANSLATION SYSTEM FOR WORLD ATLASES

Paper presented by Mr. N. Kadmon (Israel)

Perhaps no written product requires more individual investment in verbal transfer methodology than a world atlas. Producing a map of a single country with a linguistic/toponymic base other than that of the receiver language may require a single transliteration or transcription system only. The task is made more difficult if the scripts of the two countries are different, too, and even more so if they belong to different generic types such as alphabetic, syllabic and logographic. In a world atlas produced in a given receiver language all names, including those which in their original language are unwritten, must be transferred to a single writing and pronunciation system. This problem is well known to every atlas editor. But only in few cases are all names indeed transferred directly from the donor language to that of the receiver. In most instances "standardized" intermediate atlases are used for the production of a new atlas, and the more thorough the work which has been invested in the former, the higher can by the toponymic quality of the latter. If the number of published general atlases may be regarded as an indicator, it seems that the majority of work has been done on atlases in Roman script. This fact alone would justify the insistence of the UN Conferences on the Standardization of Geographical Names, as well as of the UN Group of Experts on Geographical Names, on the development of a single official romanization system for all non-Roman scripts. Although the English language does not appear to the present writer to be the single most efficient medium for phonetic transmission of sounds of different languages (Armenian, for example, has a much wider range of consonants), no language group seems to have produced more good, well-founded world atlases than the English-speaking countries.

Some years ago the first computerized bi-scriptual gazetteer of Israeli place names, carrying Hebrew and official romanized forms, was developed (Kadmon, document E/CONF.74/L.24, UN Conference on the Standardization of Geographical Names, 1982). The work, by a well-known Israeli publishing house with a tradition of map and and atlas production, described below, carries automated toponymy a very great step forward. This firm has lately acquired the translation rights of a major world atlas in the English language. A magnetic tape of the index of over 200,000 names has been supplied, and this serves as the starting point for the computer-assisted production of the new atlas which is being printed in Hebrew. This language has an alphabetic script in which consonants are written in full, but vowels are usually omitted, a practice followed also, e.g. by Arabic. However, foreign words, and in particular geographical names, require at least partial vocalization in order to avoid ambiguity.

A team of linguists and toponymists was assembled, with experts representing each of the linguistic blocks or regions, resembling to a certain degree the divisions of the UN Conferences on the Standardization of Geographical Names

Various methods of transfer apply, of course, to different languages and scripts in relation to Hebrew. Thus, Arabic, another Semitic language, has similar attributes and Arabic script can be referenced via a single system of transliteration. For Roman alphabets a phonetic transfer system has to be used, and one of the first acts required is setting up a basic table of equivalence. Since names in non-Roman scripts are ingested into the system through Roman characters with certain diacritics, this table is used also for toponyms in non-Roman scripts. A statistical analysis is then made, resulting in a basic "universal" transliteration system into Hebrew. On this the specific transliteration and transcription rules for various languages are then superimposed. This set of instructions is then incorporated in the processing software.

A computer-generated printout is now produced. Each record, i.e., each name, carries all relevant information pertaining to the name, including location, typeface and size, and, of course, the automatically transliterated name. All records of names relating to a given map, or country, or language block - as the case may be - are checked by the expert involved and corrected if necessary. On the checking document the latter can also add secondary names and decide whether these should appear only in the atlas index or in the maps, too. If word order of a name in the index has to be changed in relation to that in the maps, e.g. in complex names involving generic and specific terms (see N. Kadmon, Working Paper No.3, UNGEGN, 12th Session) a proper code is inserted.

Geographical terms forming part of toponyms in all languages, some 2,000 in all, were translated and inserted into a terminology table. The software recognizes these strings of characters in the donor language and automatically translates them into Hebrew. However, the human element - the linguistic-toponymic expert - retains overriding control at all stages and of all elements, both linguistic such as spelling, and cartographic, such as typeface and size.

After the corrections are inserted in the computer file and after checking, the magnetic tape is run on an optical typesetting system. This produces strip film which is duplicated as necessary for names appearing on more than one map. Names placement is carried out manually, in conformity with the views expressed in the Report of the 11th Session of UNGEGN, 1984 (document ESA/RT/C/GN/9, para.49).

The expertise gained enables the firm concerned to carry out map and atlas translation work in practically any language.