Power numerals in the general counting systems of Micronesia

Author, contact, & funding information

Authors

Andrea Bender & Sieghard Beller

Contact

Andrea Bender
Department of Psychosocial Science
& SFF *Centre for Early Sapiens Behaviour* (SapienCE)
University of Bergen, N-5020 Bergen, Norway

e-mail: Andrea.Bender@uib.no

Folder description

The folder contains four documents, one of which in two different formats:

- this **overview**,
  - once as formatted text, labelled   0a_ReadMe.pdf,
  - and once as plain text: 0b_ReadMe.txt
- a **data file** in comma-delimited format: 1_dat_MicPowNum_2021_04.csv
- a **source file** in UTF-8 encoded text-based format for bibliographic entries:
                                          2_sources_2021_04.bib
- a **transcription sheet** as formatted text: 3_transcrRules_2021_03.pdf

The data file contains the data on power numerals that are discussed in the manuscript entitled "Ways of counting in Micronesia", authored by Andrea Bender & Sieghard Beller and to be published in *Historia Mathematica*. The source file contains bibliographic information on the sources used to generate the data file. And the transcription sheet details how specific signs were transcribed.

Data file, source file, and transcription sheet are described in the following.

(1) Data file

Data file description

The data file contains information on the power numerals in the languages spoken in Micronesia, as collected from grammars and vocabularies of those languages, linguistic analyses, and ethnographic descriptions.

Data file structure

Data are entered in a single data sheet consisting of 20 columns. The first five of these columns contain information for identifying and classifying the languages or language variants:

- The column "ID" provides a unique alphanumeric identifier for each language variant for which data on power numerals was obtained. In two cases, this includes two distinct descriptions each of systems in the same language variant, provided by different authors.
- The columns "lang_Code" and "lang_Name" contain the Glottocodes and the names, respectively, for each of the analysed languages, following the *Glottolog* conventions; see https://glottolog.org/ (last accessed on 2021-03-27).
- The column "lang_Var" specifies the language variant for those instances, for which separate descriptions for distinct islands are available (i.e., Puluwatese, Woleaian, Ulithian, and Sonsorol), with either the name of the respective island or "main" in all those cases for which only one variant is reported.
- The column "lang_Br" contains values for whether or not the language belongs to the Micronesian sub-branch, with "1" = Micronesian and "2" = non-Micronesian (again according to *Glottolog*).

The subsequent ten columns "10^1" to "10^10" contain information on power numerals from 10^1 to 10^{10} (which is the highest power denoted in any of the languages under scrutiny). Data entered in these columns consist of the number word(s) reported in the main source (see "Source" column below), in line with the following conventions:

- The original orthography of the source is adopted as accurately as possible. Where it was not possible to find a single sign to represent the exact combination of letter and diacritics used in the source, a different sign was chosen (as documented in the *transcription sheet*, see (3) below) so that the same new sign was used consistently for all words using the same combination of letter and diacritics. Such changes are flagged in a separate column, "Orth" (see below). Ligature markings were occasionally made in the original sources, but ignored for the transcription.
- Square brackets indicate an optional part of the numeral or a placeholder for a prefix ("[-]") (only inserted when explicated in the source).
- In cases, in which more than one numeral for the same power is reported in the source, all numerals are given in the order in which they are reported, separated by "/".
- Loan words for numerals are included if reported in the source, but set in round brackets (e.g., "(tausin)" in kosr1238). The same convention is applied to power terms that are composed from several numerals according to a 'loaned structure', typically from a European language (e.g., "(ata arańan)" in naur1243). For further discussion, see the manuscript.
- Lexemes listed as power numerals in the source, but translated in a way that indicates the end of counting (e.g., "limit of counting", "finished", "outside", or "infinite"), are affixed with an asterisk (e.g., "*manutu" in cham1312).
- "?" indicates that it is unclear whether the numeral does not exist or just wasn't collected (e.g., Quackenbush, 1968, deliberately collected numerals only up to 1000).

The final set of columns contain various types of comments on the data:

- The column "Orth" indicates whether, for the numerals of that language (in columns from 10^1 to 10^{10}), orthography was fully retained = "0" or had to be adjusted = "1".
- The column "Source" specifies the main source for the reported data, with author name(s), year of publication, and page number. If several sources are available, as a general rule the

earliest most coherent report was chosen (see the manuscript, for more detailed information and discussion). Full bibliographic data is provided in a separate *source file* (see (2) below).

- The column "**source_Key**" provides the key to the entry names in the source file.
- The column "**comm_Source**" contains comments on sources, such as notes on inconsistencies in the source used (if such occur) and references to additional sources where available, or information on broader patterns, including notes on specific counting systems.
- The final column "**comm_Forms**" contains explanatory information on specific numerals, including notes on when to use which alternative for the same number. Entries begin with the form of the numeral concerned, followed by ":"; comments on different forms in the same language variant are separated by ";".

Data usage

The power numerals reported in columns "**10^1**" to "**10^10**" constitute the main data of Figure 6 in the manuscript, which illustrates the distribution of power numerals in the Micronesian languages, indicating for each language the range of power numerals (Fig. 6a) and the roots of which numerals are reflexes, detailed for base 10 (Fig. 6b) as well as for the next two powers of the base, 100 and 1000 (Fig. 6c). For further discussion, see the manuscript.

In addition, Pohnpeian power numerals are presented in Figure 8, and Woleaian power numerals in Figure 10 of the manuscript.

(2) Source file

The source file contains bibliographic information on the sources used to generate the data file – mostly grammars and vocabularies of these languages, linguistic analyses, and ethnographic descriptions – as well as all additional sources referred to in the comments.

Entry names consist of the last name of the first author and the year of publication (of the edition used). The bibliographic information specified follows BibTech conventions.

(3) Transcription sheet

File description

The transcription sheet specifies how orthography is handled when transferring numerals from the written sources to the data file. As explained above, the original orthography of the source was adopted as accurately as possible. Where it was not possible to find a single sign to represent the exact combination of letter and diacritics used in the source, a different sign was chosen so that the same new sign was used consistently for all words using the same combination of letter and diacritics.

The transcription sheet details which specific signs were retained, and which had to be replaced. In the latter case, a scan of the letter in original orthography was inserted. As noted earlier, ligature markings were ignored.

File structure

The overview contains two parts: a complete list for vowels with all diacritics used, and a list of the non-standard consonants used. Those shaded blue were retained, those shaded red had to be modified. A box with notes explains the colour coding.