Database: Grambank (2467 languages, 195 features, 362025 not null data entries)

Grambank is designed to be used to investigate the global distribution of features, language universals, functional dependencies, language prehistory and interactions between language, cognition, culture and environment. Although Grambank is part of Glottobank, a research consortium that involves work on complementary databases of lexical data, paradigms, numerals and sound patterns in the world's languages, Grambank is aimed to ultimately cover all languages for which a grammar or sketch grammar exists.

During the analysis of the data, the following methods were implemented:

- Total spreadsheets merge of languages, parameters, values, and their codes into a single table with the most essential data ('ID', 'Language_ID', 'Parameter_ID', 'Value', 'Code_ID', 'Language_Name', 'Macroarea', 'Latitude', 'Longitude', 'Family_name', 'level', 'Parameter_Name', 'Boundness', 'Flexivity', 'Gender_or_Noun_Class', 'Locus_of_Marking', 'Word_Order', 'Informativity', 'Code_Description')
- Data values reshape into Language indexes as rows and Parameters as columns, additional language features as macroaread and family name codes added as well, the resulting values pivot table has dimensions of 2467x(195+7) and consists mainly of the categorical data
- Correlation matrix calculation for the language parameters's space using Cramer's V measure of relation between two variables in categorical scale
- Top correlations extraction for a given number N of top correlation matrix entries that returns a selected subset of the most correlated parameters
- Sufficient (>0.5 on a scale from 0 to 1) correlations extraction for a given parameter ID from the correlation matrix
- Correlation matrix visualisation including parameter names resolution for the row indexes

The data analysis findings include:

- Empty and unique values count for the merged data table revealed that most of the categorical parameters are binary by nature, while some have up to 4 code value options; many parameters have missing values for more than a half of all languages; moreover, 10759 languages are actually dialects; finally, all data entries are spread across the 6 macro areas, Papunesia being the largest in size
- Top 16 correlated parameters (GB092', 'GB299', 'GB107', 'GB115', 'GB089', 'GB155', 'GB314', 'GB171', 'Family_name_Code', 'GB071', 'GB114', 'Macroarea_Code', 'GB090', 'GB051', 'GB074', 'GB091', 'GB020', 'GB133', 'GB030', 'GB072', 'GB113', 'GB131', 'GB186', 'GB082', 'GB073', 'GB130', 'GB083', 'GB070', 'GB022', 'GB170', 'GB315', 'GB172') were visualized on a heatmap of their correlation matrix that shows the most significant relations between the language's family and its other parameters.

Picks of the statistical (Cramer's V) correlations among typological features are the following:

- Can the A (agent-like argument of a transitive verb) argument be indexed by a prefix/proclitic on the verb in the simple main clause is highly (0.91) correlated with the same feature for S (single argument of an intransitive verb) arguments, that suggest a presence of the polypersonal agreement - a linguistic phenomenon where a single verb can show agreement markers not only for the subject (S) but also for the object (A) in the same verb form - indicating a high level of intricacy in the way verbs agree with their arguments, common for highly synthetic language with complex verb conjugation systems.
- Is a pragmatically unmarked constituent order verb-initial for transitive clauses is highly (0.82) correlated with what is the pragmatically unmarked order of S and V in intransitive clauses that indicates a pattern in which the word order for intransitive clauses (S and V) and transitive clauses (subject and verb-initial constituent) is closely related in terms of pragmatics, showing that there is a unified pattern for structuring sentences, which could be influenced by how speakers pragmatically convey information, and the language has a relatively fixed word order for different types of clauses, simplifying the task of sentence structure prediction in discourse
- Can an adnominal property word agree with the noun in gender/noun class is highly (0.75) correlated with the same features for an adnominal demonstrative and adnominal numeral, that indicates a well-developed system of noun-adjective agreement, where various adnominal elements adapt their forms to match the gender or noun class of the nouns they modify