**Database of Hokkien dictionaries and textbooks published from the early 17th century until the present day**

Project proposal for “Working with digital data for historians” course

Sebestyen Hompot (Global Studies M.A.)

The project I wish to propose for the course is related to my M.A. studies at the Xiamen University of China[[1]](#footnote-1), my first M.A. thesis[[2]](#footnote-2) and a recently finished article[[3]](#footnote-3), already peer-reviewed and accepted for publication in the journal *Crossroads – Studies on the History of Exchange Relations in the East Asian World.*[[4]](#footnote-4)Furthermore, the proposed project is also relevant to my ongoing M.A. studies at the University of Vienna[[5]](#footnote-5) and my planned M.A. thesis[[6]](#footnote-6) supervised by Susanne Weigelin-Schwiedrzik (Sinology department, University of Vienna).

During my three-year M.A. studies in Xiamen (earlier known as Amoy) I studied the local history of the city and its surrounding region (the southern part of China’s Fujian province), with a special focus on linguistic interaction between locals and foreigners between the mid-19th and mid-20th centuries. As it quickly became clear during my research, in the last ca. 500 years the local variety[[7]](#footnote-7) of Chinese known as Hokkien[[8]](#footnote-8) have spread to numerous other places around the South China Sea as a result of migrations due to maritime trade and colonialism. In this time period, Hokkien became the majority spoken language of Singapore and Taiwan, as well as of overseas Chinese communities in Indonesia, Malaysia and the Philippines.

Textbooks and dictionaries of Hokkien have been published throughout a wide time range (from the manuscripts of early 17th century Philippines-based Spanish missionaries until the present day), in various locations and various languages (Classical Chinese, Dutch, English, Hokkien, Japanese, Mandarin, Spanish) and for various purposes (work of American/British/Spanish Christian missionaries, colonial administration in British Malaya, the Dutch East Indies and Japanese-occupied Taiwan, the national reform movement of mainland China in the Late Qing [1840’s-1911] and Republican periods [1912-1949], the pro-Taiwanese identity movement of post-WWII Taiwan, among others). The research work I conducted shed light on the large number and complexity of interconnections between geographical locations, authors, their works and their affiliations.[[9]](#footnote-9) Influenced by new approaches in historiography emphasizing local, regional and global perspectives instead of the current mainstream of the nation state and by the methods learned at the “Working with digital data for historians” course, in my opinion a database of Hokkien linguistic materials and their authors would be a project relevant to both the field of global history and the contents of the present course.

The database created for the project would be based on a table-format collection of Hokkien linguistic materials published between 1837 and 2013 by Chiung (2015).[[10]](#footnote-10) This table is made up of 75 entries (rows) with the following columns: “Year Published, Titles, Editors, LG [Language] written, Given names[[11]](#footnote-11), Places [of publication], Publishers”. This table, under the name “publications” would be manually expanded with further columns as well as with further rows containing pre-1837 and post-2013 works and other works omitted in Chiung’s work .A separate “authors” table would be created (instead of the Editors column of Chiung) with corresponding foreign-key references in “publications”. A further table named “transcription\_methods” will be created since the large diversity in the transcription methods used for Hokkien (including various Romanizations, Chinese characters, modified Japanese Katakana[[12]](#footnote-12), extended Bopomofo[[13]](#footnote-13) and own inventions of certain authors) is one of the most important aspects of Hokkien linguistics. I will hereby provide a basic SQL outline of my proposed database with the addition of comments to explain why I believe certain columns are necessary.

DROP DATABASE IF EXISTS `hokkien`;

CREATE DATABASE `hokkien`;

USE `hokkien`;

CREATE TABLE `countries` (

id INT PRIMARY KEY AUTO\_INCREMENT NOT NULL,

name\_english VARCHAR(50) NOT NULL,

name\_former VARCHAR(255)

);

CREATE TABLE `cities` (

id INT PRIMARY KEY AUTO\_INCREMENT NOT NULL,

name\_english VARCHAR(50) NOT NULL,

name\_asianchar VARCHAR(50),

name\_former VARCHAR(50),

country INT NOT NULL,

FOREIGN KEY `country` REFERENCES `countries`(`id`)

);

CREATE TABLE `missionary\_societies` (

id INT PRIMARY KEY AUTO\_INCREMENT NOT NULL,

name VARCHAR(255) NOT NULL,

church ENUM(‘Catholic’, ‘Protestant’, ‘Other’),

`home\_country` INT,

FOREIGN KEY `home\_country` REFERENCES `countries`(`id`)

);

CREATE TABLE `publishers` (

id INT PRIMARY KEY AUTO\_INCREMENT NOT NULL,

name\_orig VARCHAR(255) NOT NULL,

name\_asianchar VARCHAR(255),

name\_translated VARCHAR(255),

city INT,

FOREIGN KEY `city` REFERENCES `cities`(`id`),

country INT,

FOREIGN KEY `country` REFERENCES `countries`(`id`)

);

CREATE TABLE `people` (

id INT NOT NULL PRIMARY KEY AUTO\_INCREMENT,

name\_orig VARCHAR(50) NOT NULL,

*# In this column names in Asian languages are written in Latin transcriptions only.*

name\_asianchar VARCHAR(50),

*# For those understanding the relevant Asian languages, it is important to provide names/titles written with original characters, since one word in Romanized transcription can correspond to various characters. However, the difficulties of displaying Asian characters in MySQL applications will probably be a problem here.*

nationality INT,

FOREIGN KEY `nationality` REFERENCES `countries`(`id`),

occupation ENUM(‘interpreter’, ‘missionary’, ‘scholar’, ‘unspecified’),

missionary\_society INT,

FOREIGN KEY `missionary\_society` REFERENCES `missionary\_societies`(`id`)

);

CREATE TABLE `transcription\_methods` (

id INT NOT NULL PRIMARY KEY AUTO\_INCREMENT,

name\_orig VARCHAR(255) NOT NULL,

name\_asianchar VARCHAR(255),

creator INT,

FOREIGN KEY `creator` REFERENCES `people`(`id`),

first\_appearance INT,

FOREIGN KEY `first\_appearance` REFERENCES `publications`(`id`),

based\_on VARCHAR(255),

type ENUM(‘Romanization’, ‘Extended Chinese characters’, ‘modified Katakana’, ‘modified Bopomofo’, ‘mixed Chinese characters-Romanization’, ‘extended Chinese characters + own invention’, ‘own invention’, ‘other’) NOT NULL,

);

CREATE TABLE `publications` (

id INT NOT NULL PRIMARY KEY AUTO\_INCREMENT,

title\_orig VARCHAR(255) NOT NULL,

title\_asianchar VARCHAR(255),

title\_translated VARCHAR(255),

author INT NOT NULL,

FOREIGN KEY `author` REFERENCES `people`(`id`),

year\_publication DATE

publisher INT NOT NULL,

type ENUM(‘dictionary’, ‘manuscript’, ‘phrasebook’, ‘rime chart’, ‘rime dictionary’, ‘textbook’, ‘unspecified’) NOT NULL,

language ENUM(‘Classical Chinese’, ‘Dutch’, ‘English’, ‘Hokkien’, ‘Japanese’, ‘Latin’, ‘Mandarin’, ‘Spanish’, ‘other’) NOT NULL,

*# Language used in the introduction and explanations throughout the book. Usually depends on the country of origin of the publishing organization.*

transcription\_method INT,

FOREIGN KEY `transcription\_method` REFERENCES `transcription\_methods` (`id`),

digitized\_at VARCHAR(255),

paper\_version\_available\_at VARCHAR(255)

*# To provide information (with references) on the location of works which have not yet been digitized. Especially important for the early Spanish manuscripts.*

`mentioned\_in` VARCHAR(255),

);

***#Displaying authors and publications in a join table:***

CREATE TABLE `authors` (

`author` INT NOT NULL,

FOREIGN KEY `author` REFERENCES `people`(`id`),

`publication` INT NOT NULL,

FOREIGN KEY `publication` REFERENCES `publications`(`id`)

);

1. M.A. in “Chinese Language and Scripts” at Xiamen University (2014-2017). [↑](#footnote-ref-1)
2. “Dong-xifang shiyu de Gulangyu Gonggong Zujie yuyan wenhua jiaoliu yanjiu 东西方视域的鼓浪屿公共租界语言文化交流研究” (Foreign and Chinese Perspectives on Linguistic and Cultural Interaction on the Gulangyu Joint Concession), M.A. Thesis, Xiamen: Xiamen University, 2017. [↑](#footnote-ref-2)
3. The title of the yet unpublished article is “Xiamen at the Crossroads of Sino-Foreign Linguistic Interaction during the Late Qing and Republican Periods: The Issue of Hokkien Phoneticization”. [↑](#footnote-ref-3)
4. Chief editor: Angela Schottenhammer (Prof. for Non-European and Global History, University of Salzburg), website of the journal: [www.eacrh.net/](http://www.eacrh.net/) (As of 2019-01-02 inaccessible due to technical problems), FB page of the Crossroads Research Center: <https://www.facebook.com/crossroads.research.centre/> (Accessed on 2019-01-02). [↑](#footnote-ref-4)
5. Global Studies and Global History as part of the Erasmus Mundus Global Studies M.A. program, (First year finished at Ghent University, Belgium). [↑](#footnote-ref-5)
6. As already agreed with Prof. Weigelin-Schwiedrzik, the thesis will focus on contemporary global historiography in China. [↑](#footnote-ref-6)
7. “Variety” is a neutral term used instead of “dialect” and “language” here. While many linguists consider the mutually unintelligible varieties of Chinese as separate languages of the Sinitic branch of Sino-Tibetan languages, due to historical, social and political reasons Chinese people usually refer to their own varieties as “dialects (of a common Chinese language)”. [↑](#footnote-ref-7)
8. Hokkien (meaning “Fujian [speech]” in Hokkien) is the variety of Chinese native to the southern part of China’s Fujian province and the majority spoken language of Singapore and Taiwan, as well as of overseas Chinese communities in Malaysia, Indonesia and the Philippines. Similarly to other southeast Chinese language varieties such as Cantonese (Yue) or Shanghainese (Wu), it has no mutual intelligibility with Mandarin, the main official language and lingua franca of both mainland China and Taiwan and one of the official language of Singapore (along with English, Malay and Tamil). Hokkien has been referred to by various other names, these include Amoy (Xiamen) dialect / vernacular, Southern Min (Min being another name for Fujian province), Minnan (meaning “Southern Min” in Mandarin) among others. In Taiwan, the local variety of Hokkien is usually simply referred to as Taiwanese. [↑](#footnote-ref-8)
9. To name one example, most Protestant missionaries arriving to Xiamen in the wake of the First Opium War (1839-1842) were members of larger missionary societies such as the LMS (London Missionary Society) and had already been active among Hokkien-speaking communities on Southeast Asian colonies prior to their arrival. [↑](#footnote-ref-9)
10. See Appendix 1 (pp. 24-29) of “Taiwanese or Southern Min? On the Controversy of Ethnolinguistic Names in Taiwan” (last update: 2014-07-27) <http://uibun.twl.ncku.edu.tw/chuliau/lunsoat/english/2015/taiwanese.pdf> (Accessed on 2019-01-02). [↑](#footnote-ref-10)
11. Name used in the given publication for referring to Hokkien. [↑](#footnote-ref-11)
12. Syllabic script used in contemporary Japanese for the transcription of non-Chinese foreign words. [↑](#footnote-ref-12)
13. The semi-syllabic script known as Bopomofo or Zhuyin is the main phonetic transcription and digital input method for Mandarin in Taiwan. [↑](#footnote-ref-13)