University of British Columbia, Vancouver
Department of Computer Science

CPSC 304 Project Cover Page

Milestone #: 1

Date: February 10th 2025

Group Number: 87

Name	Student Number	CS Alias (Userid)	Preferred E-mail Address
Karli Winkler	69709426	r6y7o	karliswinkler@gmail.com
Mostafa Mostafa	42907808	y3x7z	mxstafa@student.ubc.ca
Arman Thariani	46081311	j5v0v	arman.thariani@gmail.com

By typing our names and student numbers in the above table, we certify that the work in the attached assignment was performed solely by those whose names and student IDs are included above. (In the case of Project Milestone 0, the main purpose of this page is for you to let us know your e-mail address, and then let us assign you to a TA for your project supervisor.)

In addition, we indicate that we are fully aware of the rules and consequences of plagiarism, as set forth by the Department of Computer Science and the University of British Columbia

University of British Columbia, Vancouver
Department of Computer Science

Application Domain

The domain of this project is Linguistics Field Research and Language Data Management. This domain involves documenting the use of language by various speakers and the sounds used by speakers of specific dialects and languages. In addition, it includes the classification of sounds and languages for use in phonetics and phonology research.

Aspects of Domain Modeled

This database will aid linguists and other enthusiasts in conducting research systematically, providing metrics and data to enhance their understanding of language. Here's how:

- The database transcribes words using IPA notation, helping track pronunciation variations and phonetic structures across different languages and their dialects respectively.
- Each word is organized by linking it to its respective language and dialect, enabling researchers to analyze how words evolve and compare lexical differences across languages. Additionally, each language is associated with its origin, speaker population, and geographical distribution, capturing demographic and regional variations.

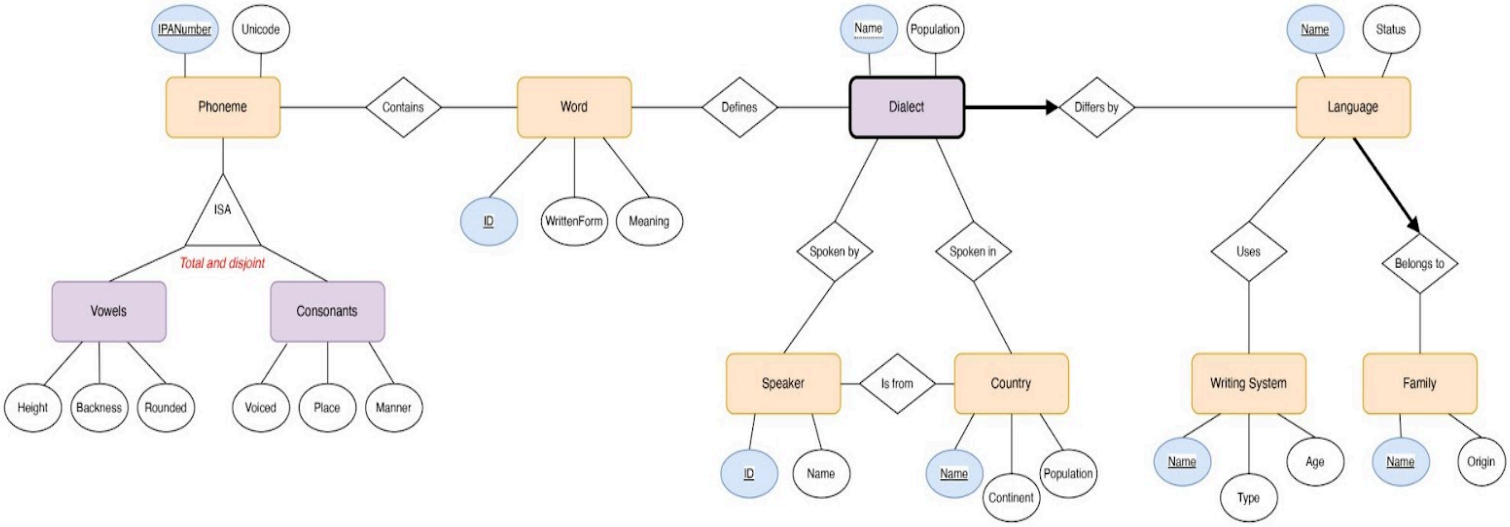
As a result, the database enables researchers to systematically document and analyze phonetic and lexical data, track language changes, and potentially discover and preserve endangered languages. Hence, a real-world application of this would be publishing phonetic research to aid in the revitalization of languages at risk of extinction.

Database Specifications

The database can be used as a tool for linguistic researchers to record data collected in the field when encountering new words, phonemes, and dialects of various languages. Researchers will also be able to record information upon meeting speakers of different languages, including their country of origin and any other relevant personal information. Outside of the field, linguistics professionals can refer back to this database for data on languages and phonetics.

University of British Columbia, Vancouver
Department of Computer Science

ER Diagram



The above diagram shows Vowels and Consonants separately as they are classified with different phonological features and both are different types of phonemes that are the base sound unit of what makes up a word so they are defined by an ISA relationship. The Dialect is a weak entity relation as it can only be defined with reference to a language. Each language belongs to exactly one language family and can use multiple writing systems. Similarly, a dialect can be spoken by multiple people in multiple countries.
Specialization of ISA relation: Every Phoneme is either a vowel or consonant and a vowel cannot be a consonant and vice versa, so overlapping constraints: disjoint and covering constraint: total (yes, every phoneme fits as either a vowel or consonant)

We have not used any generative AI tools in this milestone of the project.