Multilingual-Dictionary

TUSHAR MOHANPURIA

March 27, 2023

Contents

0.1 Introduction

The project focuses on creating a Multilingual Dictionary web applica-tion. The user inputs a word in English and the dictionary will give output as Syllable, Pronunciation, Scientific-name-if-noun, Image, Meaning, Examples, Antonyms and synonyms if any as well as translation in other languages.

0.1.1 Objective

The is to show the meaning of the given word with its *depiction* in other languages and cite its use with example for better understanding.I have considered the **"Hindi", "Marathi", "Tamil" and "Telugu"** as languages for the Database.

0.2 Given Problem Statement

Mr. Amar and Mr. Akbar were friends from childhood. Recently they started their own venture with the name i3Tech Pvt. Ltd., where CEO is Mr. Amar and CTO is Mr Akbar. One customer, Mr. Anthony approached Amar and Akbar to develop a dictionary like application (for example: https://www.merriam-webster.com, https://dictionary.cambridge.org) with the following description. Help Amar-Akbar and Anthony in developing the application. The application will store the following information

1. Meaning, example sentence(s) against each word and phrase of minimum three language- L1, L2 and L3, where Li will be any Indian language
2. Part-of-speech-tag, syllable, pronunciation, gloss (other language words written using Roman script)
3. Image, and scientific names against living entities, such as, tree, animals.

General Syntax

Searched-word

(*Syllable breakup, Pronunciation, Scientific-name-if-noun, Image*)

Meaning 1: (*part-of-speech-tag*) Description
Example:
Synonym: W_{1i} , $W_{1(i+1)}$
Antonym: W_{1j}, $W_{1(j+1)}$

Meaning N: (part-of-speech-tag) Description
Example:
Synonym: W_{Ni}, $W_{N(i+1)}$
Antonym: W_{Nj}, $W_{N(j+1)}$

In other languages
L_1 -
L_2 –
L_N –

Where, W_i, W_j are member of the word-list/dictionary

Key Points from General Syntax

---

1. Syllable, Pronunciation, Scientific-name and Image are represented on same level with Searched-word as there is only one entry for each per word.
2. Part-of-speech-tag helps to define a use of the word. storing description and example as well under a given meaning.
3. The input to the application will be always in English.

4. a word can have many synonym and antonym under a given meaning for a word. where words from stored dictionary list only can be used.

5. we can store multiple languages for a single word that too a language can have many interpretations for a particular word.

0.3 DATA MODEL

To achieve our objective, below procedure will be followed:

0.3.1 Define the Data Model

The first step is to define the data model. This involves determining the entities involved in the system, such as words, definitions, synonyms, antonyms, and example sentences. The next step is to identify the relationships between these entities, such as strong and weak entities, and to define the attributes for each entity. For example, the attributes for a word entity could be spelling, part of speech, and pronunciation.

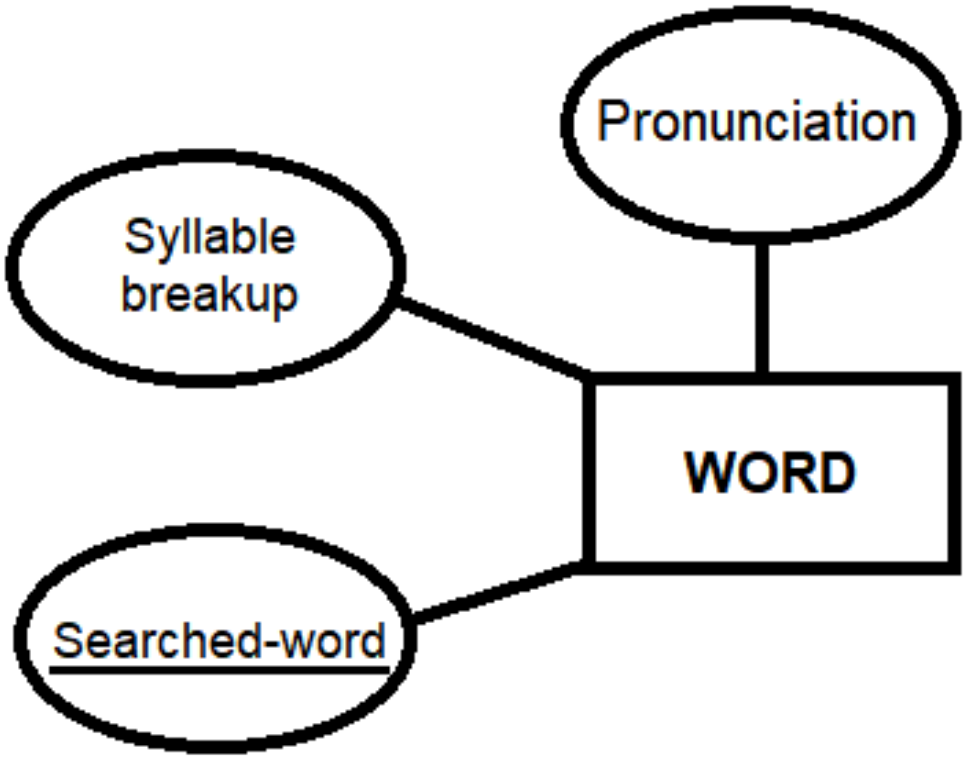
0.3.2 Assumption

These are my assumptions while creating the database:

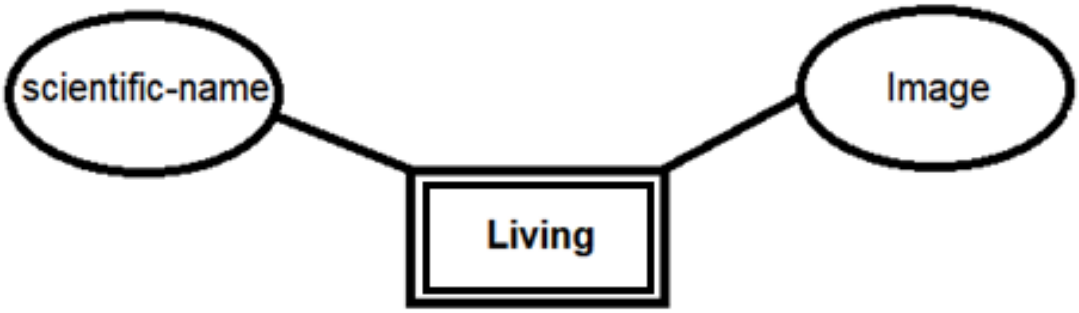
1. Syllable, Pronunciation, Scientific-name and Image are to be stored under Searched-word as the key attribute.
2. Part-of-speech-tag helps to define a use of the word. storing description and example as well as meaning.
3. The input to the application will be always in English.
4. a word can have many synonym and antonym under a given meaning for a word. where words from stored dictionary list only can be used.
5. we can store multiple languages for a single word that too a language can have many interpretations for a particular word.

0.4 ER Diagram

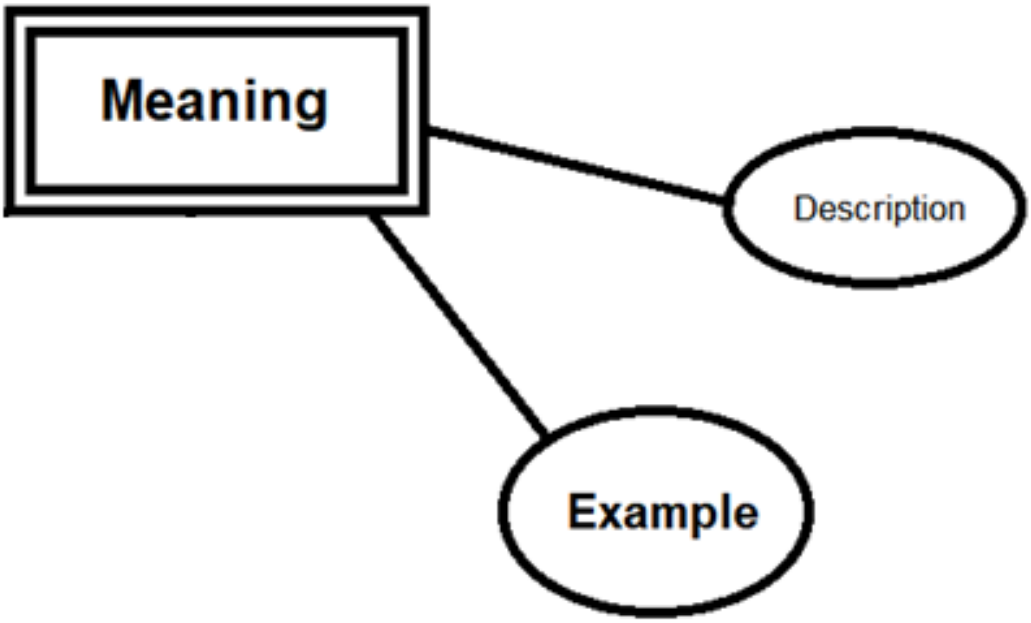
Entities :-
As shown above we have marked syllable breakup and pronunciation are marked under searched – word but as for scientific-name and Image they have a condition for the word to be a living Entity.



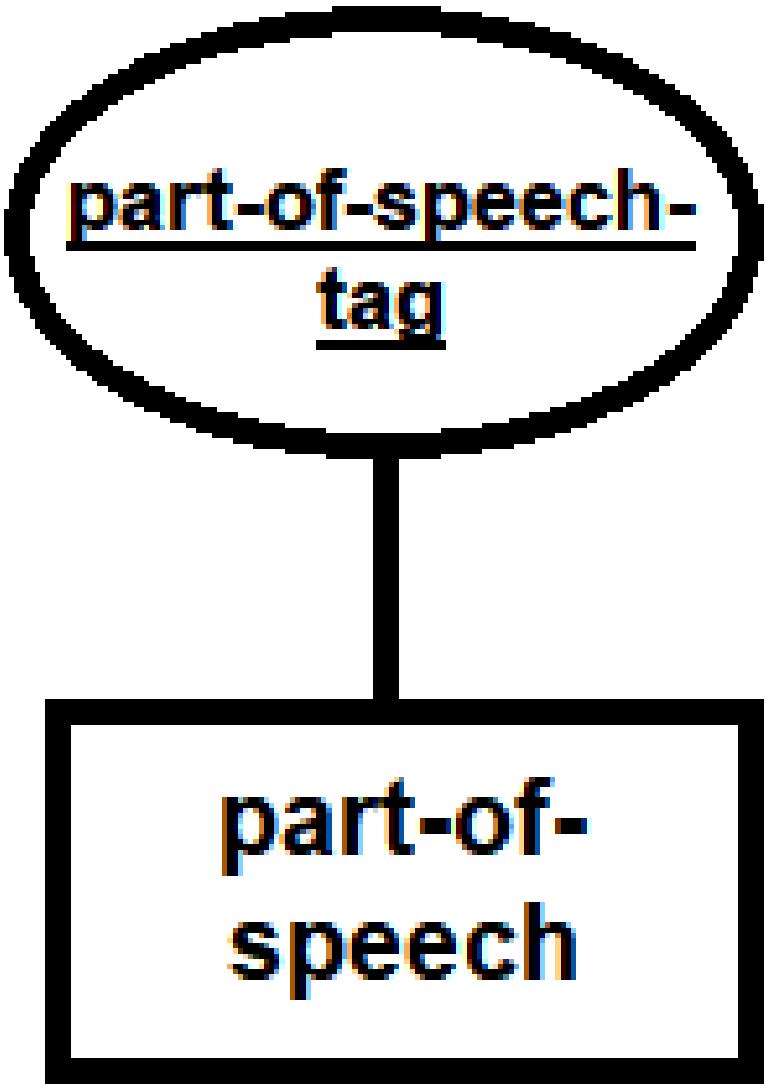
So as for scientific-name and Image we will mark then under an entity called living. And as it is dependent on word to determine the word whose scientific-name and image is stored, it will be a weak entity.



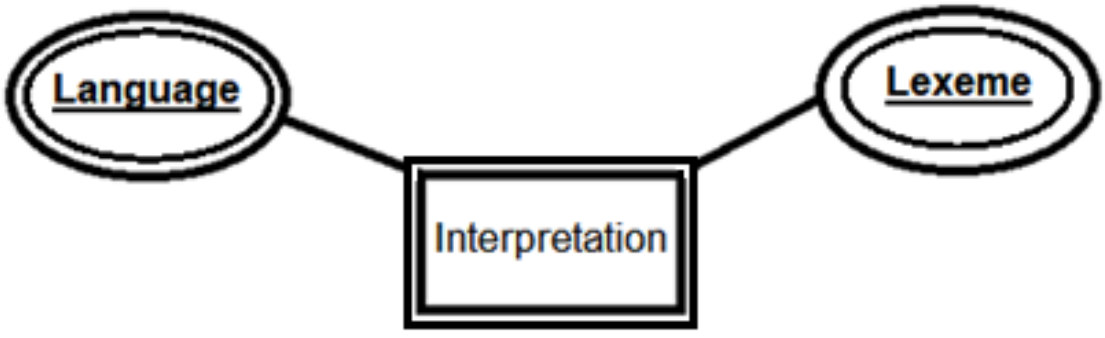
Now as mentioned Meaning is an attribute of word but as it has a dependency on Part-of-speech-tag to generate its meaning we must make it as its own entity with its description, example as its attributes.(As it has a dependency on word to express it self as its meaning we make it a weak entity.)



As there is a dependency on part-of-speech-tag to generate the meaning so we will have an entity between Word and meaning as part-of-speech.

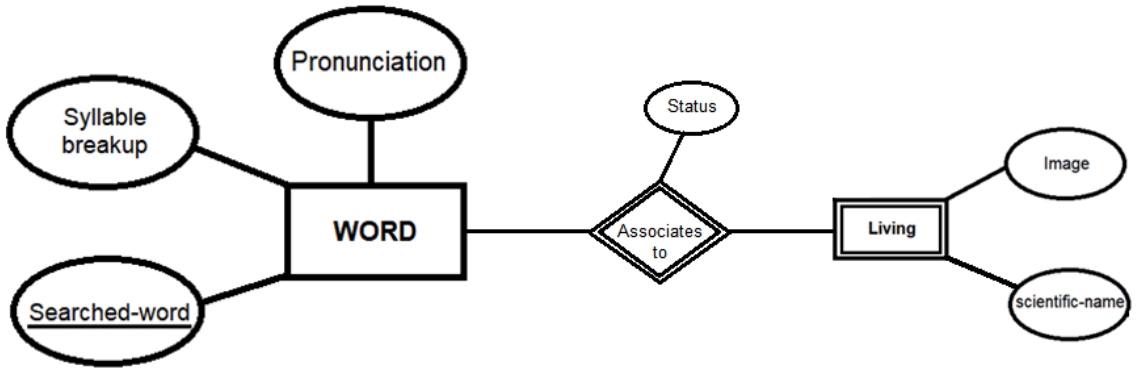


Now as for the translation of word to other languages that will be an entity on its own with languages and lexeme as its attributes. Where both attributes are multi-valued. And as it is dependent on word to determine the word to find its counter-parts.

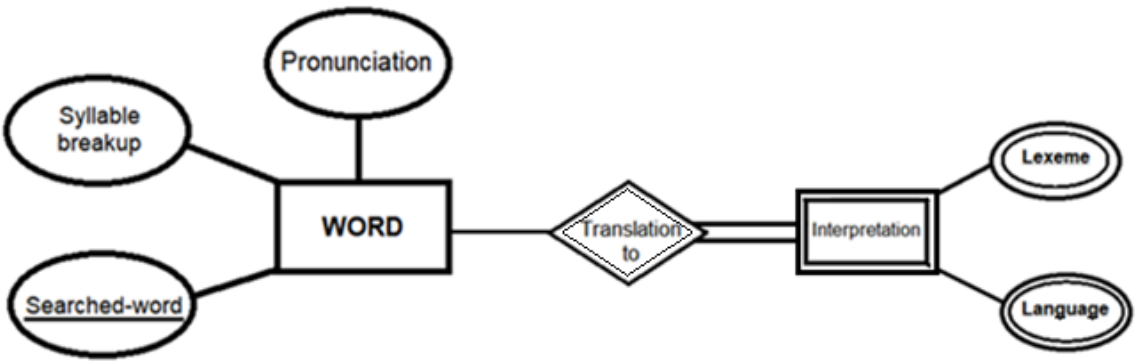


With these entities I conclude the entities to be created.
Now let's move on to relations.

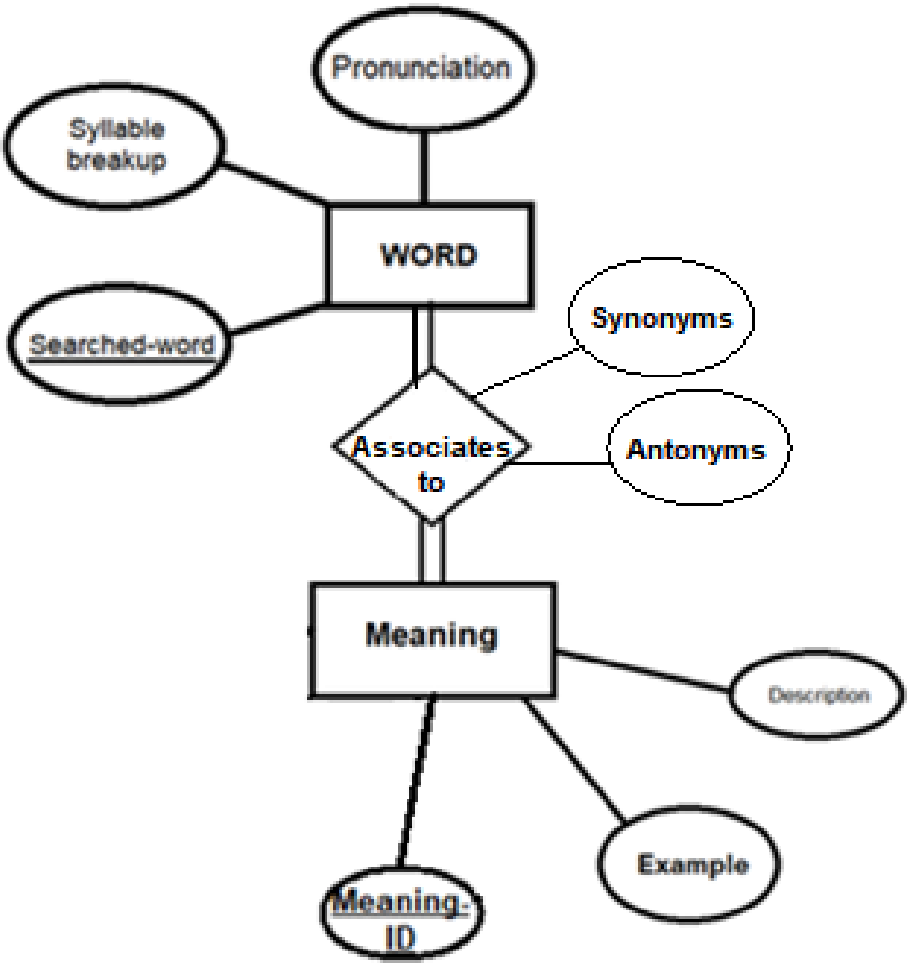
Relations :-
Relation between Word and Living where Word connects to living to check whether a given entity is alive or not and if so it can retrieve Image and Scientific-name from the relations.



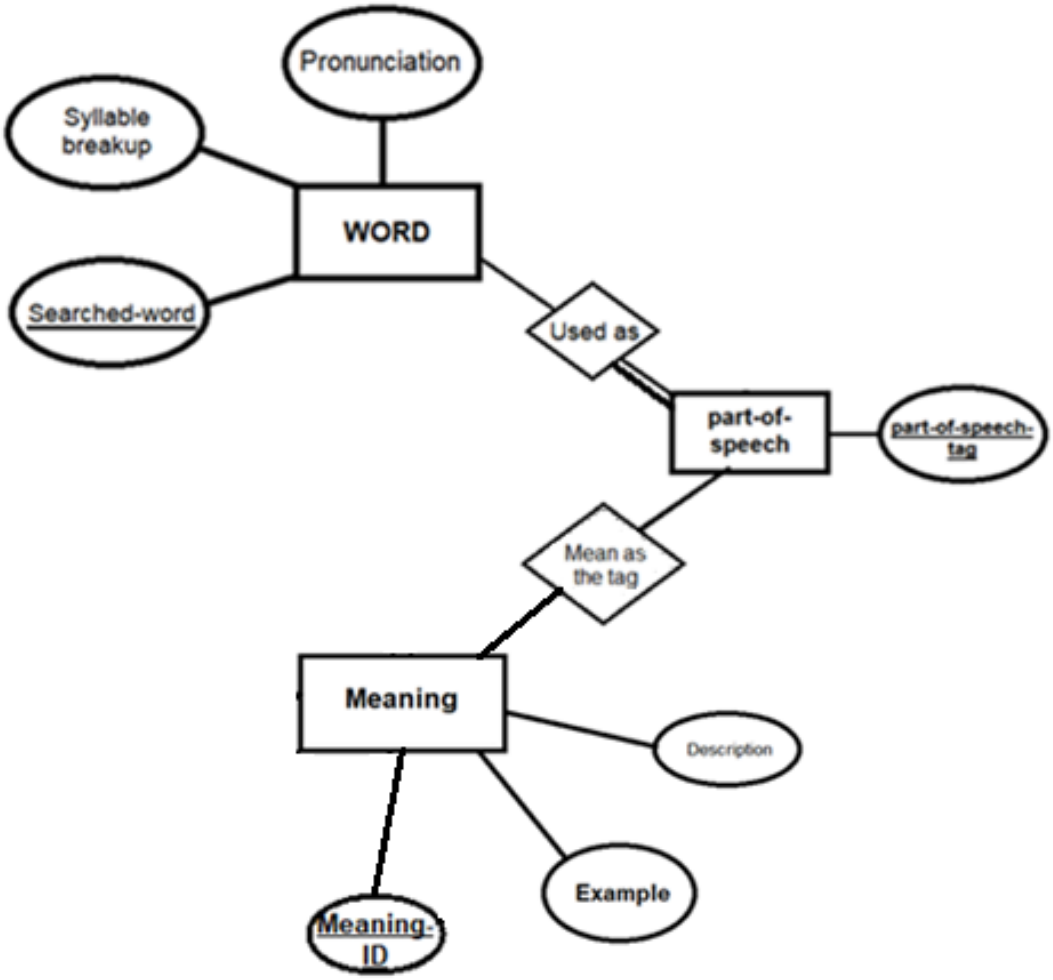
Then if we talk about the relation between Word and Interpretation. The word is required to retrieve its interpretation in other languages but as languages and their lexeme are stored together they express there meaning as which word belongs to which language. As we can have multiple translations of a word in a different language we have a one to many relation.



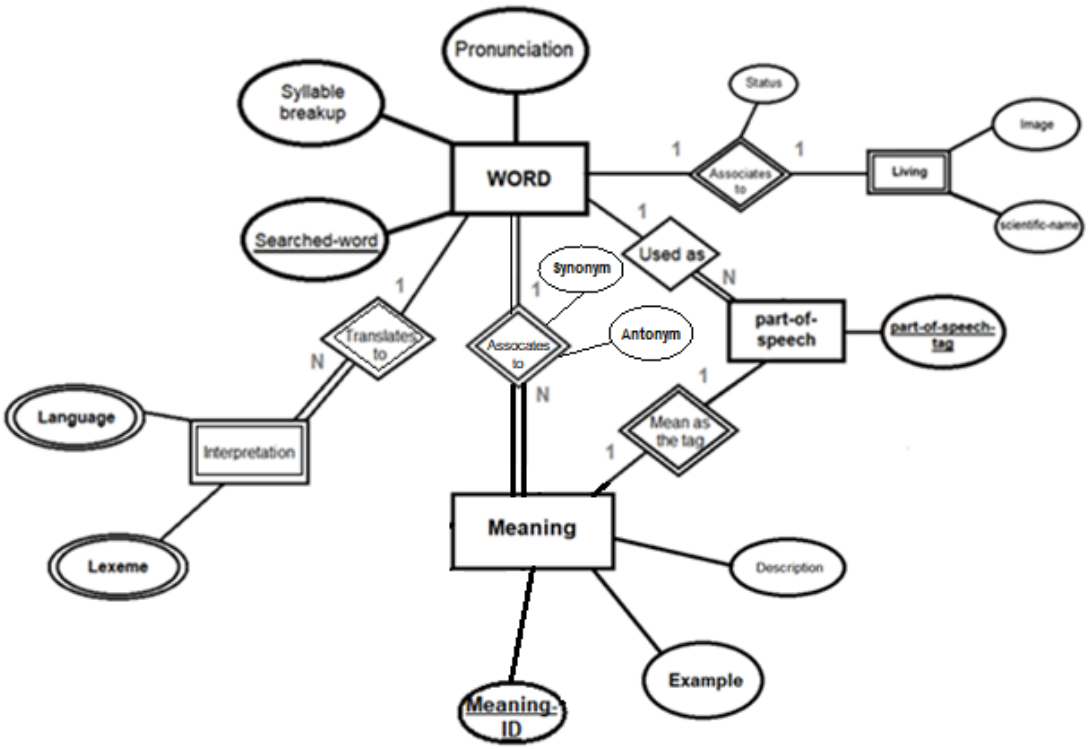
Next, we have the meaning which is connected to word with a relation to express the word's meaning. But as we can have multiple examples for same word the will be a one to many relation in this case. With Meaning-ID as any other attribute as they will be to long to store as a key efficiently. Also to store the Synonyms and antonyms as they should also be a part of word-list.



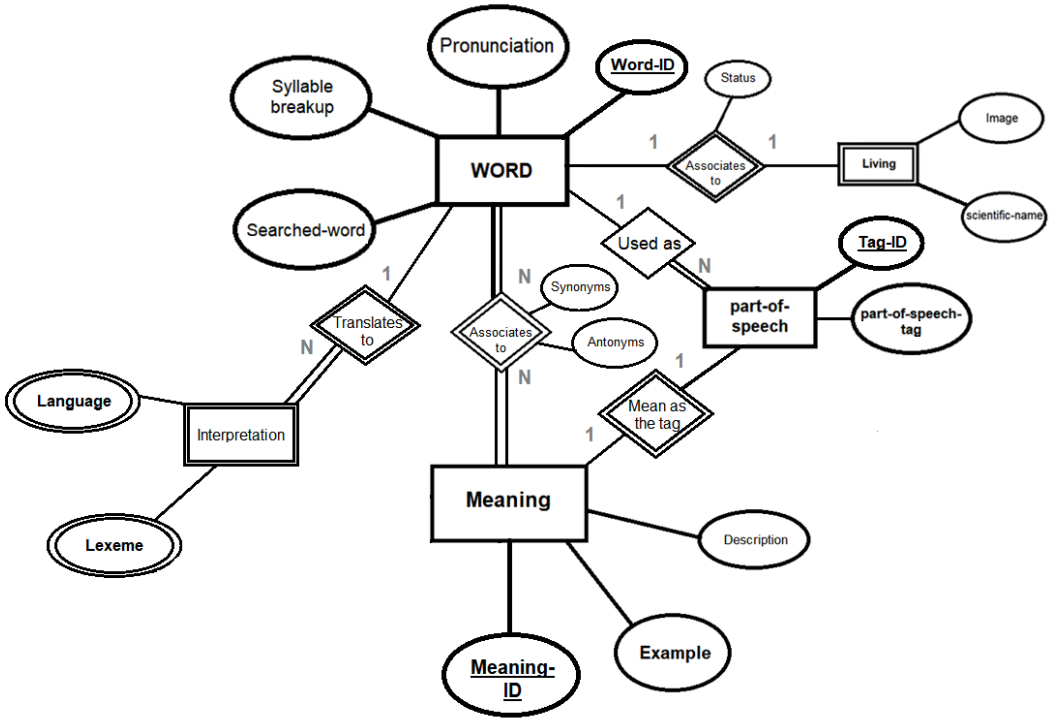
Lastly, we have part-of-speech which has two relations one with Word and one with meaning where meaning and part-of-speech has a one to one relation but word and part-of-speech has one to many relation. Because for each word as a part of speech we have only one meaning but as for a word we can use it as a noun ,verb etc. As meaning requires to take part-of-speech in account to give the meaning in that context so we have a dependency on part-of-speech to give a meaning.



Now if we connect all relations together and convert it into a single ER-diagram then we will get



Now converting ER Diagram to Relational Schema : -



0.5 Tables From ER

Word Table

Word-ID	Word	Pronunciation	Syllable

Living Table

As it is a weak entity it will have the related entity's primary key as a reference and as well as the attribute of the relation.

Word-ID	Image	Scientific-name	Status

Part-of-Speech Table

For 1 to many relation add primary key of 1-side as foreign key to N-side. Include relational attributes to the N-side entity/relation if there is any.

Word-ID	Tag-ID	Part-of-speech-tag

Interpretation Table

As it is a weak entity it will have the related entity's primary key as a reference and as well as the attribute of the relation.

Word-ID	Language	Lexeme

Meaning Table

For each binary M:N relationship, create a new relation R. The primary keys attributes of the participating entities will be add as foreign keys in R.

Meaning-ID	Description	Example

Association Table

Word-ID Synonym-ID	Meaning-ID	Antonym-ID

Meaning and pos relation Table

Tag-ID	Meaning-ID

0.6 Normalization

0.6.1 1NF

A relation is in 1NF if it contains an atomic value.
As you can see we have all the attributes on atomic level so need for any conversion in above tables as they already are in 1NF.

0.6.2 2NF

A relation will be in 2NF if it is in 1NF and all non-key attributes are fully functional dependent on the primary key.
As you can see we have all the attributes of any table are fully functional dependent on respective primary key of that table.

0.6.3 3NF

A relation will be in 3NF if it is in 2NF and no transition dependency exists.
Here we see a conversion of Part-of-Speech Table to its subset tables.

Word and pos relation Table

Tag-ID	Word-ID

Part-of-Speech Table

Tag-ID	Part-of-speech-tag

After changes we have no transition dependency.

0.6.4 BCNF

A stronger definition of 3NF is known as Boyce Codd's normal form.
So already in BCNF.

0.6.5 4NF

A relation will be in 4NF if it is in Boyce Codd's normal form and has no multi-valued dependency.
So as Interpretation table has multi-valued language and lexeme so we will have a relation table and 2 new table for multi-valued attributed that will replace the earlier table and similarly for association table will be divided into synonyms and antonyms.

Interpretation Table

Word-ID	Language-ID	Lexeme-ID

Language Table

Language-ID	Language

Lexeme Table

Lexeme-ID	Lexeme

Antonyms Table

Antonym-ID	Meaning-ID

Synonyms Table

Synonym-ID	Meaning-ID

0.6.6 5NF

A relation is in 5NF. If it is in 4NF and does not contain any join dependency, joining should be lossless.
As we can see we do not have any join dependencies.

0.7 All Tables after Normalization

Word Table

Word-ID	Word	Pronunciation	Syllable

Living Table

Word-ID	Image	Scientific-name	Status

Word and pos relation Table

Tag-ID	Word-ID

Part-of-Speech Table

Tag-ID	Part-of-speech-tag

Interpretation Table

Word-ID	Language-ID	Lexeme-ID

Language Table

Language-ID	Language

Lexeme Table

Lexeme-ID	Lexeme

Antonyms Table

Antonym-ID	Meaning-ID

Synonyms Table

Synonym-ID	Meaning-ID

Meaning and pos relation Table

Tag-ID	Meaning-ID

Meaning and Word relation Table

Word-ID	Meaning-ID

0.8 Data Dictionary

0.8.1 Word

```
mysql> desc word;
+---------------+-------------+------+-----+---------+----------------+
| Field         | Type        | Null | Key | Default | Extra          |
+---------------+-------------+------+-----+---------+----------------+
| Word_ID       | int         | NO   | PRI | NULL    | auto_increment |
| Searched_Word | varchar(46) | NO   |     | NULL    |                |
| Pronunciation | varchar(120)| NO   |     | NULL    |                |
| Syllable      | varchar(120)| NO   |     | NULL    |                |
| Status        | varchar(3)  | NO   |     | NO      |                |
+---------------+-------------+------+-----+---------+----------------+
5 rows in set (0.01 sec)
```

0.8.2 Living

```
mysql> desc living;
+-----------------+-------------+------+-----+---------+-------+
| Field           | Type        | Null | Key | Default | Extra |
+-----------------+-------------+------+-----+---------+-------+
| Word_ID         | int         | NO   | PRI | NULL    |       |
| Image           | longblob    | YES  |     | NULL    |       |
| Scientific_name | varchar(80) | YES  |     | NULL    |       |
+-----------------+-------------+------+-----+---------+-------+
3 rows in set (0.00 sec)
```

0.8.3 part-of-speech

```
mysql> desc part_of_speech;
+-------------------+-------------+------+-----+---------+----------------+
| Field             | Type        | Null | Key | Default | Extra          |
+-------------------+-------------+------+-----+---------+----------------+
| Tag_ID            | int         | NO   | PRI | NULL    | auto_increment |
| Part_of_Speech_Tag| varchar(20) | NO   | PRI | NULL    |                |
+-------------------+-------------+------+-----+---------+----------------+
2 rows in set (0.01 sec)
```

0.8.4 word-tag-rel

```
mysql> desc word_tag_rel;
+---------+------+------+-----+---------+-------+
| Field   | Type | Null | Key | Default | Extra |
+---------+------+------+-----+---------+-------+
| Word_ID | int  | NO   | PRI | NULL    |       |
| Tag_ID  | int  | NO   | PRI | NULL    |       |
+---------+------+------+-----+---------+-------+
2 rows in set (0.01 sec)
```

0.8.5 interpretation

```
mysql> desc interpretation;
+-------------+------+------+-----+---------+-------+
| Field       | Type | Null | Key | Default | Extra |
+-------------+------+------+-----+---------+-------+
| Word_ID     | int  | NO   | PRI | NULL    |       |
| Language_ID | int  | NO   | PRI | NULL    |       |
| Lexeme_ID   | int  | NO   | PRI | NULL    |       |
+-------------+------+------+-----+---------+-------+
3 rows in set (0.01 sec)
```

0.8.6 language

```
mysql> desc language;
+-------------+-------------+------+-----+---------+----------------+
| Field       | Type        | Null | Key | Default | Extra          |
+-------------+-------------+------+-----+---------+----------------+
| Language_ID | int         | NO   | PRI | NULL    | auto_increment |
| Language    | varchar(12) | NO   |     | NULL    |                |
+-------------+-------------+------+-----+---------+----------------+
2 rows in set (0.01 sec)
```

0.8.7 lexeme

```
mysql> desc lexeme;
+-----------+-------------+------+-----+---------+----------------+
| Field     | Type        | Null | Key | Default | Extra          |
+-----------+-------------+------+-----+---------+----------------+
| Lexeme_ID | int         | NO   | PRI | NULL    | auto_increment |
| Lexeme    | varchar(80) | NO   | UNI | NULL    |                |
+-----------+-------------+------+-----+---------+----------------+
2 rows in set (0.00 sec)
```

0.8.8 meaning

```
mysql> desc meaning;
+-------------+--------------+------+-----+---------+----------------+
| Field       | Type         | Null | Key | Default | Extra          |
+-------------+--------------+------+-----+---------+----------------+
| Meaning_ID  | int          | NO   | PRI | NULL    | auto_increment |
| Description | varchar(400) | NO   |     | NULL    |                |
| Example     | varchar(240) | NO   |     | NULL    |                |
+-------------+--------------+------+-----+---------+----------------+
3 rows in set (0.01 sec)
```

0.8.9 antonyms

```
mysql> desc antonyms;
+-------------+------+------+-----+---------+-------+
| Field       | Type | Null | Key | Default | Extra |
+-------------+------+------+-----+---------+-------+
| Meaning_ID  | int  | NO   | PRI | NULL    |       |
| Antonym_ID  | int  | NO   | PRI | NULL    |       |
+-------------+------+------+-----+---------+-------+
2 rows in set (0.01 sec)
```

0.8.10 synonyms

```
mysql> desc synonyms;
+-------------+------+------+-----+---------+-------+
| Field       | Type | Null | Key | Default | Extra |
+-------------+------+------+-----+---------+-------+
| Meaning_ID  | int  | NO   | PRI | NULL    |       |
| Synonym_ID  | int  | NO   | PRI | NULL    |       |
+-------------+------+------+-----+---------+-------+
2 rows in set (0.00 sec)
```

0.8.11 word-mean-rel

```
mysql> desc word_mean_rel;
+------------+------+------+-----+---------+-------+
| Field      | Type | Null | Key | Default | Extra |
+------------+------+------+-----+---------+-------+
| Word_ID    | int  | NO   | PRI | NULL    |       |
| Meaning_ID | int  | NO   | PRI | NULL    |       |
+------------+------+------+-----+---------+-------+
2 rows in set (0.00 sec)
```

0.8.12 tag-mean-rel

```
mysql> desc tag_mean_rel;
+------------+------+------+-----+---------+-------+
| Field      | Type | Null | Key | Default | Extra |
+------------+------+------+-----+---------+-------+
| Tag_ID     | int  | NO   | PRI | NULL    |       |
| Meaning_ID | int  | NO   | PRI | NULL    |       |
+------------+------+------+-----+---------+-------+
2 rows in set (0.01 sec)
```

0.8.13 Stored Words

Word_ID	Searched_Word
1	Tiger
2	Human
3	Dictionary
4	Banana
5	Book
6	Lexicon
7	Volume
8	Tome
9	Time
10	Hour
11	Late
12	Punctual
13	Prompt
14	Go
15	Idiot
16	Intelligent
17	Smart
18	Dumb
19	Sharp
20	Bright
21	Brainless
22	Proceed
23	Stand
24	Halt
25	Stop
26	Dog
27	Cat
28	Fox
29	Elephant
30	owl
31	I
32	Under
33	And
34	Alas!

0.9 BIBLIOGRAPHY

1. Merriam Webster : https://www.merriam-webster.com/
   To search for the meaning of the words and synonyms in English.

2. Shabdkosh Hindi : https://www.shabdkosh.com/dictionary/english-hindi/
   To search the words, Hindi translation from English .

3. Shabdkosh Marathi : https://www.shabdkosh.com/dictionary/english marathi/
   To search the words, Malayalam translation from English .

4. Shabdkosh Tamil : https://www.shabdkosh.com/dictionary/english-tamil/
   To search the words, Bengali translation from English .

5. Shabdkosh Telugu : https://www.shabdkosh.com/dictionary/english-telugu/
   To search the words, Bengali translation from English .