WEB BASED TAMIL LANGUAGE LEARNING

—————————————

A Thesis

Presented to the

Faculty of

San Diego State University

—————————————

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Computer Science

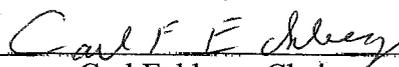—————————————

by

Vasumathi Senthil

Spring 2011

SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Vasumathi Senthil:
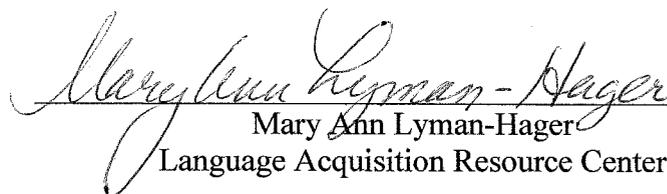
Web Based Tamil Language Learning

_____
Carl Eckberg, Chair
Department of Computer Science

_____
William Root
Department of Computer Science

_____
Mary Ann Lyman-Hager
Language Acquisition Resource Center

_____
Approval Date

DEDICATION

To Tarun, Tara, Senthil, Mom and Dad.

ABSTRACT OF THE THESIS

Web Based Tamil Language Learning
by
Vasumathi Senthil
Master of Science in Computer Science
San Diego State University, 2011

The main objective of this thesis is to create a web based learning tool that enables Tamil Language learning without learning the Tamil script. The goal is accomplished by using English script for teaching Tamil.

A web interface is provided for teachers to upload and maintain lessons that could contain text, audio or video. There is a reverse transliterator which converts any Tamil text to its English phonetic equivalent enabling users to read any Tamil content. An English to Tamil translation dictionary provides a huge database of translated words in the English phonetic form, along with an option for the users to add and approve words. Additionally the dictionary search function searches and displays words from the web. The online networking feature provided by the tool enables users to network with other native or non-native users.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The main objective of this thesis is to enable Tamil language learning without learning the Tamil script. It is based on the situational language learning concept that when a language is presented orally first, then in the written form, language skills can be learned more effectively. It also follows a communicative language learning approach by enabling communication between native and nonnative speakers in a social setting.

In order to follow all the above teaching methodologies the following issues have to be addressed:

- Find a medium for teaching and learning that does not involve the Tamil language script.

- Provide a setting where native and nonnative speakers can network and communicate.

- Provide a means for teachers to post lessons.

- Provide a means to read all available Tamil literature via the selected medium, to strengthen language skills and learn more about Tamil and Tamilian culture.

- Provide a knowledge base for self-directed learning.

Since all existing teaching sources teach Tamil using the Tamil script, this project takes a different approach and follows the methodology that teaches oral Tamil. The site targets learners who follow an analytical approach towards language acquisition where reading and analyzing texts builds language knowledge. The basic idea is to make the learning curve of Tamil short by bypassing learning the scripts. For someone who is only

interested in speaking Tamil, this site is of immense value because Tamil script is not a requirement.

Situational language learning is enabled by teaching the language in the English phonetic form. In a setting where face to face communication is not possible, when someone wants to learn to speak a language without learning the language script, there has to be a way (in this case, an already known script) to learn the language. The website uses English as a medium to teach Tamil.

Communicative language learning is made possible through social networking with other native and nonnative users of the site, via Twitter and instant messenger. The site uses Twitter API to emulate a subset of Tamil based Twitter community which makes all the users of the site visible as a community. The site eliminates the need for users to search for people who might be interested in Tamil to form a community.

The main barrier when learning a language in a multilingual community is native and non-native speakers of the language think and communicate in a different way based on their cultural norms. This is due to the difference in culture, ideologies and logic developed while growing up in their respective cultural environment. Research shows that in a multicultural setting, students communicate less with other students who don't know about their culture. Consequentially, a low-level of social interaction takes place between the native and nonnative speakers in the community. The site aims to overcome this barrier and enable nonnative students to learn about the culture of Tamil speakers by providing a gateway to all the Tamil literature available on the World Wide Web. The Tamil Reverse Transliterator converts any Tamil text to its English phonetic form. The Reverse Transliterator uses a complex set of rules to decode Tamil text to its English phonetic equivalent. By getting

access to all available Tamil literature in easy readable format, students can practice and strengthen Tamil language skills in addition to learning about Tamilian culture. The site also has an English-Tamil dictionary which helps strengthen vocabulary.

This Tamil learning site is targeted at the following audience:

- Self-learning learners where individuals have access to lessons and network with native and nonnative speakers.

- Learners following a comprehensive learning approach by providing them with tools that makes Tamil literature from the web readable.

CHAPTER 2

EXISTING TAMIL RESOURCES

Web Assisted Learning and Teaching of Tamil developed by the Penn Language Center follows the natural approach to teach Tamil. The site provides beginner, intermediate and advanced level Tamil lessons.

Padanool is another web based teaching resource offered by University of North Carolina. It takes a natural approach to teaching by offering structured lessons starting from Tamil basics to complete sentence formation.

National Translation Mission which is an initiative taken by the government of India to offer language translation in all Indian languages for dictionary words and phrases. Tamil is one of the languages for which translation has been completed.

There are a few Facebook groups and blogs intended for teaching Tamil. The owners of the groups and blogs are usually Tamil speakers who are interested in promoting the Tamil language. The way these groups and blogs teach is, teach one Tamil sentence at a time. This could be useful for someone who follows the learning method of memorizing words and phrases to learn a language.

There are a few tuition based teaching centers that teach Tamil. English to Tamil and Tamil to English dictionaries, translated phrases for a limited set of Tamil sentences are examples of resources offered by a few other websites.

CHAPTER 3

WEB BASED TAMIL LANGUAGE LEARNING

3.1 TAMIL LANGUAGE

Tamil is a Dravidian language predominantly spoken by the people in the state of Tamil Nadu. The following sections explain the basics of the language.

3.1.1 Tamil Letters

Tamil language consists of five short vowels (a, i, u, e, o), five long vowels (aa, ee, oo, ae, oe) and two diphthongs ai, ou.

There are 18 consonants consisting of which six surds k, s, t, th, p, r), six sonants (ny, ng, n, n, m, n) and six medials (y, r, l, v, z, l). There are no aspirated consonants like gha or cha in Tamil. Likewise the letter h is also absent in Tamil. But a corresponding letter q known as ayutam is used to soften the surds in Tamil. There are no distinct letters for surds and sonants in Tamil. The vowel consonants are pronounced like surds and sonants depending on their placement within a word. ka, sa, ta, tha and pa are pronounced as gha, cha, da, dha and bha according to their place in the word.

3.1.2 Morphology

There are three common types of words namely:

- **Verbs** which is classified into finite and infinite verbs. Finite verbs are formed with suffixes which indicate its gender or animate and inanimate qualities. The gender is not distinguished both in abstract nouns and in relative participles.

- **Nouns** which indicate animate and inanimate objects, gender, numbers and person. Uyarthinai and Akrinai are the two classifications of nouns. Humans, gods and

demons belong to Uyarthinai, whereas everything else that is animate and inanimate belongs to Akrinai.There are three genders in Uyarthinai: masculine, feminine and neuter. Palar paal or neuter plural gender indicates many in number. Masculine and feminine genders in Tamil indicate only singular number. Akrinai is classified into ondran paal (singular of the impersonal class) and palavin pal (plural of the impersonal class).There is three 'persons' in Tamil, namely, first person, second person and third person. Case inflexions are many in Tamil and their indicators form as suffixes in words. The meaning of words makes the distinction between animate and inanimate things, and masculine and feminine genders.

- **Particles** which have no meaning of their own but acquire meaning when added to other words and help to differentiate their meanings too.

Tamil is an agglutinative language where case, time or gender indicators are suffixed to root words. The suffixes are either derivational which changes the part of the speech, or inflectional which marks the gender, number, tense etc. Since there is no limit to the agglutination more than one suffix can be attached to a root to denote more than one marker. Tamil is a head-final language where verb comes at the end of a class typically. It is also a null subject language where sentences can be formed without a subject or an object or a verb.

3.2 FEATURES OF THE SITE

The four main components of the site are:

- English-Tamil Dictionary

- Lessons

- Social Networking

- Reverse Transliteration

Appendix A provides instructions for using the site and Appendix B provides instructions for maintaining the site.

3.2.1 English-Tamil Dictionary

The English to Tamil Dictionary provides a huge database of English words translated to Tamil and its English phonetic form. Words can be browsed alphabetically or can be searched using the 'Search' facility. Additionally there is a provision for native Tamil speakers to contribute words to the database. The site is peer moderated, so words added by one person usually needs approval by somebody before it goes into the system.

3.2.2 Lessons

The lessons module lets Tamil teachers or contributors to teach Tamil by posting articles, videos, audio etc. There is also an interface for users to maintain the lessons they added. Through that interface they can make changes to a lesson or even delete it. Lessons can also be searched based on the title, description or author.

3.2.3 Social Networking

Twitter API has been used to create a subset of Twitter just for the Tamil learning community. Every user of the site can create their own group by adding or removing members of the site. This feature offers the benefit of eliminating the process of searching for Tamil learners and teachers to form an exclusive Tamil learning community. Users can set up their group to receive updates from the members of their group. This feature can be used in both a classroom setting where a student can form a group with his teacher and classmates and communicate with them. Outside the classroom, individuals can randomly connect with other users of the site to enhance their Tamil skills.

3.2.4 Reverse Transliteration

Reverse transliteration is the process of converting Tamil text to an equivalent English phonetic text. It is accomplished by using a conversion algorithm in conjunction with a rules database. The conversion tool takes a Tamil text in Unicode format and converts it into a readable form. This tool is useful in cases where a student wants to read already available Tamil content without having to learn the script.

In addition to the above modules, there is also an instant messenger that shows all the users browsing the site. A chat conversation could be initiated with anyone online, which further enhances communication possibilities within the community. A random quiz generator tests vocabulary by offering multiple choice vocabulary questions.

3.3 TECHNICAL DETAILS

Reverse transliteration is accomplished by using a combination of database rules and a conversion algorithm. The following subsections discuss each of them in detail.

3.3.1 Algorithm for Reverse Transliteration

The stepwise algorithm for reverse transliteration is as follows:

1. All the reverse transliteration rules are in the form of database entries.

2. Every entry in the database is given a priority (1 to 20) and a sub-priority (1 to 20).

3. A Tamil letter array is formed first with high priority letters (priority = 20) at the top of the array and low priority letters (priority = 1) at the bottom of the array. For letters with the same priority but different sub-priorities, the high sub-priority takes an upper position than the low sub-priority.

4. Using the Tamil letter array, a key value array is formed with the Tamil letter as the subscript and its English phonetic equivalent as the value.

5. Letter replacement is done in a loop for every entry in the Tamil letter array.

3.3.2 Priority and Sub-Priority

The English phonetic equivalent of Tamil vowels and consonants are given in the Tables 3.1 and 3.2. Each consonant combines with each of the vowels to form a different letter. These letters are called 'uyirmei' letters. From the tables it can be observed that some consonants are associated with more than one English letter. This is because; in Tamil based on the placement of a letter within a word the pronunciation varies. Similarly certain combinations of letters are pronounced in a different way than the individual letters translated separately.

Table 3.1. Vowels Pronunciation Table

Vowel	English Mapping	Pronunciation
அ	a	sounds like the a in animal
ஆ	aa	sounds like the a in all
இ	i	sounds like the i in image
ஈ	ee	sounds like the e in easy
உ	u	sounds like the ou in could
ஊ	oo	sounds like the oo in zoo
எ	e	sounds like the e in edit
ஏ	ae	sounds like the a in ate
ஐ	ai	sounds like the i in idea
ஒ	o	sounds like the o in only
ஓ	oe	sounds like the o in ocean
ஔ	ou	sounds like the ou in out

Tamil letters are unique in a way that a letter can form another letter by adding a suffix or prefix or both to it. For e.g. 'க' forms 'கா' by adding a suffix, 'கை' by adding a

Table 3.2. English-Tamil Mapping

Tamil Consonant	English Mapping
க்	k, g, c
ங்	ng
ச்	s, ch
ஞ்	gn
ட்	t, d
ண்	n, nd
த்	th, dh
ந்	n, nh, ndh
ப்	p, b
ம்	m
ய்	y
ர்	r
ல்	l
வ்	v
ழ்	z
ள்	l
ற்	r, tr, dr
ன்	n, nh

prefix and 'கொ' by adding both a prefix and a suffix. Since the algorithm follows letter

replacement logic, letters should be replaced in a certain order to avoid partial translation. If

'க' is given higher priority than 'கா', then partial replacement takes place due to the suffix

in 'கா'. To avoid this partial replacement 'கா' is given a higher priority. Within the same

consonant group the following is the order of priorities assigned to avoid partial replacement.

Similarly when certain combination of letters needs to be translated first instead of individual letter translation, a priority and a sub-priority is assigned to that combination. Combination letters are assigned the highest priority (priority = 3) when compared to individual letters given in the Table 3.3. The combination letters are assigned a sub-priority equivalent to the above table. For example when the letter 'ச' or any 'ச' derived letter appears in combination with 'ச்' it is pronounced as 'ch' instead of 's' as given in the Table 3.4. There are numerous such combinations of letters that are pronounced differently in different combinations.

Table 3.3. Priority Table

Tamil	English	Priority
க	Ka	16
கா	Kaa	10
கி	Ki	15
கீ	Kee	14
கு	Ku	13
கூ	Koo	12
கெ	Ke	9
கே	Kae	8
கை	Kai	11
கொ	Ko	7
கோ	Koe	6
கௌ	Kou	5
க்	K	4

Table 3.4. Sub-Priority Table

Tamil	English	Priority	Sub-Priority
ச்ச	chcha	3	12
ச்சா	chchaa	3	11
ச்சி	chchi	3	10
ச்சீ	chchee	3	9
ச்சு	chchu	3	8
ச்சூ	chchoo	3	7
ச்செ	chche	3	6
ச்சே	chchae	3	5
ச்சை	chchai	3	4
ச்சொ	chcho	3	3
ச்சோ	chchoe	3	2
ச்செள	chchou	3	1
ச்ச்	chch	3	13

3.3.3 Sample Code

The code for reverse transliteration is given below:

```
/*********************************************************/
/* This function reverse transliterates a Tamil text to its English      */
/* phonetic equivalent                                                   */
/*********************************************************/
function translate($text)
{
$tamil_db = 'learntamil';
$tamil_user = 'learntamil';
$tamil_password ='Tamil123';
$con =
mysql_connect("learntamil.db.5023439.hostedresource.com","learntamil","Tamil
123");
if (!$con)  {  die('Could not connect: ' . mysql_error());  }
mysql_select_db("learntamil");
mysql_query("SET character_set_client=utf8", $con);
```

```php
mysql_query("SET character_set_connection=utf8", $con);
mysql_query("SET character_set_results=utf8", $con);

//Create Tamil letter priority array
$letters = array();
$i = 0;
for($priority=0;$priority < 20;$priority++){
 $sql = "select * from phonetics where priority = $priority order by subpriority";
 $result =  mysql_query($sql, $con);
 for($j = 0;$j < mysql_num_rows($result);$j++){
  $tamil_letters = mysql_result($result, $j,"tamil");
  $tamil = explode(",", $tamil_letters);
  for($k = 0;$k < count($tamil);$k++){
   $letters[$i] = $tamil[$k];
$i++;
  }
 }
}

//Create key value array with Tamil letter keys and English values
$sql = "select * from phonetics";
$result =  mysql_query($sql, $con);
$trans = array();
for($i = 0;$i < mysql_num_rows($result);$i++){
 $tamil_letters = mysql_result($result, $i,"tamil");
 $english = mysql_result($result, $i,"english");
 $tamil = explode(",", $tamil_letters);
 for($k = 0;$k < count($tamil);$k++){
$trans[$tamil[$k]] = $english;
 }
}

//Translate text
for($i = 0;$i < count($letters);$i++){
 $text = str_replace($letters[$i], $trans[$letters[$i]], $text);
}
}
```

CHAPTER 4

OBSERVATIONS AND ENHANCEMENTS

4.1 OBSERVATIONS AND SUGGESTIONS

The tool has been tested extensively with native speakers and a few non-native speakers. The reverse transliteration tool is 85% successful in reverse transliterating the Tamil text. The application could be fine tuned to get a better result by adding more rules to the rules database. This would however require the guidance of an expert in Tamil linguistics.

The following suggestions came up during testing and demonstration, and have been incorporated in the application:

- An option to add audio while adding new words to the dictionary.

- Including a column which displays the word in Tamil script, in the dictionary. This feature helps learners to associate how the word sounds to how the word looks.

4.2 ENHANCEMENTS

This web based application provides a lot of tools for Tamil language learning. However there is scope for further enhancements like the following:

- The whole application can be extended to support languages other than Tamil. The whole application is usable in that aspect with the exception of the reverse Transliterator. Since the reverse transliterator follows an algorithm based on the nature of the Tamil language, it might be unable to use the reverse transliterator for other languages. But different reverse transliterators could be developed and used with this application.

- Another possibility is to extend this application to learn Tamil or any other language using one's native language as the medium instead of English. This might require the developer to know the linguistics of both the languages involved.

- Voice support can be added as a feature. This feature enables the non-native language learner to learn the exact pronunciation of words.

WORKS CONSULTED

GitHub Inc. Abraham/twitteroauth, 2010. http://github.com/abraham/Twitteroauth, accessed Sep. 2010.

TamilCube.com. Homepage, 2010. http://www.tamilcube.com, accessed Sep. 2010.

Twitter. Twitter API Wiki, 2010. http://apiwiki.Twitter.com, accessed Sep. 2010.

APPENDIX A

USER'S MANUAL

USER ACCOUNT

Creating a new user:

1. Click on the 'Register' link on the top right hand corner of the page.

2. Fill out all fields in the main page and submit.

Making changes to user account

1. Click on the 'My Account' link on the top right hand corner of the page.

2. Click 'Modify' in the main page.

3. Make changes and submit.

LESSONS

Browsing and Searching Lessons:

1. Go to 'Lessons' by clicking the Lessons tab.

2. Click the 'Browse' tab in the main page to browse lessons.

3. A list of lessons with title and description is displayed below.

4. The content images on the right indicate what type of content the lesson has.

5. Lessons can be searched by typing the search text in the search box. By default the search is by 'title', but search by author or description can be done as well by selecting the appropriate option from the drop down list.

Adding a lesson:

1. Login with your username and password.

2. Go to 'Lessons' by clicking the Lessons tab.

3. Click the 'Add a Lesson' tab in the main page.

4. Fill in Title and Description.

5. To upload audio or video, use the appropriate upload buttons to upload the files.

6. Add text in the text editor.

7. It is not mandatory for a lesson to have all three types of content.

Modifying or Deleting a lesson:

1. Login with your username and password.

2. Go to 'Lessons' by clicking the Lessons tab.

3. Click the 'My Lessons' tab in the main page.

4. For deleting a lesson, click on the 'Delete' link corresponding to it.

5. For making changes to a lesson, click on the corresponding 'Edit' link.

6. Make necessary changes and submit.

COMMUNITY

Connecting with Twitter:

1. Login with your username and password.

2. Connect with Twitter by clicking the 'Sign in with Twitter' button in the right panel.

3. You will be taken to the Twitter site where you will have to use your Twitter username and password to sign in. If you don't have a Twitter account, create one and go to step 2.

Creating a group for networking:

1. Login with your username and password.

2. Go to 'Community' by clicking the Community tab.

3. In order to create a group you should first be connected with Twitter. See instructions for 'Connecting with Twitter'.

4. The main page lists all the registered users of the site.

5. Enter the name you want for your group in the first text box and click the 'Create' button.

6. Add members to you group by selecting the appropriate icon from the list.

7. Similarly members can be removed from the group.

Viewing messages from your group:

1. Add a member to your group by selecting the add icon corresponding to the member.

2. Go to 'Home' in the right panel to view messages from the member.

Viewing messages of a specific user:

1. Select the 'bird' icon to view messages from the user on the right panel.

2. Sending messages:

   1. Login with your username and password.

   2. Go to 'Community' by clicking the Community tab.

   3. Type your message in the text box and submit.

Viewing your messages:

1. In the right panel, select 'Tweets' to view your messages.

Mention a user in your message:

1. Click on the @ icon corresponding to the user and type your message in the text box and submit

Sending direct messages:

1. A mail icon will be visible for users who are following you. A direct message can be sent by clicking on the mail icon.
2. A text box will appear on the right panel where the message can be typed and send.

ENGLISH-TAMIL DICTIONARY

Browsing and Searching Words:

1. Click on the 'English Tamil Dictionary ' link in the top panel.
2. Click the 'Search' link to view words available in the dictionary. Clicking on the 'a' to 'z' hyperlinks lists words starting with the corresponding letters.
3. In order to search for words, type the desired word in the text box and submit. A list of words, both related and unrelated are displayed.

Adding a word:

1. Login with your username and password.
2. Click on the 'English Tamil Dictionary ' link in the top panel.
3. Click the 'Add a Word or Sentence' link.
4. Type the English word in the textbox and submit.
5. If the word already exists, an option is given to still add the word with an alternate meaning.
6. Fill all the required fields in the form and submit.

Approving words added by others:

1. Login with your username and password.
2. Click on the 'English Tamil Dictionary ' link in the top panel.
3. Click the 'Approve Words' link.

4. A list of words waiting for approval is listed.

5. Words can be approved or disapproved individually by selecting the appropriate icons or in bulk by selecting the checkboxes and selecting the 'Approve' or 'Delete' button.

QUIZ

1. Select an answer and submit.

2. Click 'more' link in the 'Take a Quiz' panel for more questions.

INSTANT MESSENGER

Initiating a chat session

1. Click the 'Online' link to see a list of online users.

2. Click on the user you want to chat with and invite them for chat.

REVERSE TRANSLITERATION

Reverse Transliterating text

1. Cut and paste any Unicode Tamil text in the text area and submit.

2. The reverse transliterated text appears below.

3. For help with pronunciation, click on the 'Need help with pronunciation' link.

APPENDIX B

ADMINISTRATOR'S MANUAL

TWITTER APPLICATION FOR COMMUNITY

Application name: Learn Tamil
Twitter username: larctamil
Twitter password: wordpass123

Twitter callback url can be set in any one of the following ways:

1. Can be set at the Twitter Application page

2. config.php (overrides the url set in the Application page)

INSTANT MESSENGER

Login link: /tamil/blabim/admin.php
ACP Key: 1gnu8pt7

Once logged in there are detailed instructions for maintaining the chat application

OTHERS

Database login details
Username: learntamil
password: Tamil123

By logging into the database, the following tasks can be performed by executing the corresponding queries

Deleting Users: delete from user where username = '<username>';

Deleting Lessons: delete from lessons where id = '<lessonid>';

Approving words directly in the db: set published = 1 to approve them without going through the application.

Location of the uploaded audio and video files: /tamil/uploads