

## Building a Tamil-Telugu-English-Portuguese Dictionary using FLEx

**Meena Kumari**

Research Scholar,

Dept of English, Dravidian University,

Kuppam, Andhra Pradesh

---

### Abstract

FLEx Fieldworks Language Explorer (FLEx) 3.0 is software for organizing and analyzing linguistic data and is produced for free download by SIL International (SIL)<sup>1</sup>. Similar to the earlier versions of the program (from SIL) viz. Shoebox and Toolbox, FLEx is designed to organize lexicon, grammar, wordlist, etc. The components in the tool consist of many options which include lexicon, grammar, texts, words, etc. FLEx is also very useful for under-resourced languages and helps to document languages. In addition, FLEx allows field linguists to perform language documentation tasks.

In this paper, an attempt is made to develop a Tamil-Portuguese-English-Telugu dictionary using the FLEx tool. The paper will give a brief introduction to the tool and also elaborate on the building of a multilingual dictionary by providing the steps involved in the construction of the dictionary.

**Key Words:** FLEx, Language, Toolbox

### 1. Introduction

Lexicography is the science and art of compiling a dictionary. The word dictionary was used as Dictionarius in the 13<sup>th</sup> century by a British man John Garland and it has become dictionarium in 14<sup>th</sup> century. A dictionary is a collection of words or repositories of words in one or more specific languages. Often these dictionaries are listed in alphabetical order with their definitions, phonetic transcription, pronunciation, grammatical categories and etymologies. The first

---

<sup>1</sup><http://fieldworks.sil.org/flex/>

dictionary published under the English title Latin-English dictionary by Thomas Elyot 1538. In early days, the purpose of dictionary was interpreting texts and providing synonyms. Later dictionaries were compiled to meet different practical purposes based on the needs of the users. Since lexicon changes constantly, new words are added by dropping old words and modifying their significance to serve the purposes.

In the present scenario, one can make a distinction among dictionaries into two: printed dictionary and machine readable dictionary or online dictionary. Online dictionary are available through internet or web browser. These are available either for free or with a paid subscription for an extended period. In order to make a dictionary either in print or online one needs to look for four facts viz. planning, preparation, editing and press copy. Along with the four facts mentioned above, there are issues concerned with the time involved in making dictionary and funds to create dictionaries. Oxford dictionary which was started in 1886 was completed in 1928. Tamil lexicon which was started in 1913 was completed in 1938. Similarly Malayalam lexicon started in 1953 but only four volumes have come out so far. In making dictionaries written languages are given more important comparing to the unwritten languages. We hardly find any Indian dictionary that gives information about extinct or unwritten languages which will have enormous knowledge system embedded in them. In documenting and preserving such kind of minor and unscheduled language FLEX is the most useful tool.

FLEX, is a Fieldwork Language Explorer developed by SIL comes in handy for lexicographers and scholars who are working in the field of endangered languages in compiling dictionaries for major and minor languages. Cerquiglini (cited in Pruvost 2000:188) says about the use of computer and its related tools in lexicography. He distinguishes three phases, in the field of lexicography viz. computer-assisted (paper) lexicography, transfer of existing paper dictionaries to electronic medium, and electronic dictionaries in their own right, conceived afresh for the electronic environment.

## **2. Importance of archiving language**

One of the recent advancements that took place in the latter half of the 20<sup>th</sup> century is archiving of unwritten languages or extinct languages. An extensive work is going on all over the world in preserving and documenting the endangered or extintive languages by using different modes

of language documentation. This archiving of languages ranges from long term preservation to web based dissemination (Birds and Simon's 2003). Unfortunately in India many languages are in different stages of language extinction but less effort is put forward in archiving the minor and unwritten languages.

In this paper we try to discuss about one software, i.e. Language Explorer (shortly known as FLEx) which is one of the tools that is useful in archiving the extinct languages. With the help of this tool one can make a bilingual or trilingual or even multilingual dictionary with a minimum of 2000 basic words and maximum of one lakh words in no less time by covering all the semantic domains. Since dictionary is a reference for the information of that language it can be used in making the primers for any community children or can be used in revitalizing the language which are in the state of extinction. Once dictionary is made, the language can be reintroduced even in the absence of the speakers as it was introduced in Miami tribe of Oklahoma (add footnote here). Since dictionary making of an extinct language is one of the ways of language archiving, it will help in preserving and documenting unwritten languages and also preserving their knowledge system available in the language under consideration. So FLEx is very accessible and useful tool in serving this purpose.

In this paper we have made an attempt to develop a proto-type Tamil-Telugu-English-Portuguese dictionary using the FLEx tool covering 100 basic words drawn from Tamil (as source language) and populated for Telugu, English and Portuguese languages. The interest for the selection of two languages from the South Dravidian family of languages (Tamil and Telugu) and (English and Portuguese) comes from the author's native knowledge (Tamil and Telugu) and working knowledge in the other two languages. The rest of the paper is organized as follows: In the following section, a brief overview of FLEx tool is provided; Section 2 elaborates on the building of a multilingual dictionary by providing the steps involved in the construction of the dictionary; Section 3 lists the merits of the FLEx tool in the context of language documentation and the final section summarizes the work. An appendix of the dictionary is also presented at the end.

### **3. Overview of FLEx**

Fieldworks Language Explorer (FLEx) is an integral part of the SIL (Summer Institute in Linguistics) Field Works suite of tools. It is a software which is meant for organizing and analyzing linguistic data at various levels such as Lexicon, Grammar, etc. Similar to the earlier versions of the program from SIL viz. Shoebox and Toolbox, FLEx is also designed to organize dictionary, providing grammatical rules, etc. The components in the tool consist of multiple options which include lexicon, grammar, texts, words, etc. It is also very useful for under-resourced languages and helps to document languages. In addition, it allows field linguists to perform language documentation tasks. FLEx is specifically designed to assist with the analysis of any of the world's languages including those which have no defined orthography (Baines, 2009).

As already mentioned above, FLEx includes components such as Lexicon, Texts & Words and grammar which are designed to organize information for any given language. Each of these component parts consists of a number of options and data fields, most of which can be customized to the needs of a given language project. When they are taken together the program's components can be used to produce outputs useful in language conservation, such as: wordlists and dictionaries, collections of texts and examples sentences. So FLEx has support for many things including, dictionary development process, stem-based or root-based dictionaries, printing of a draft dictionary, plug-in for dictionary typesetting facility, choice of user interface languages, and so on. It has around 10 classes and 88 fields for describing lexical information (Baines, 2009).

### **4. The Lexicon Component in FLEx**

The Lexicon component is integrated into the Language Explorer of SIL Field Works. It has the following panes for building a dictionary, viz. Lexicon Edit, Browse, Dictionary, Collect words, Classified Dictionary, Bulk Edit Entries, Reversal Indexes and Bulk Edit Reversal Entries. Though, the Lexicon component has as many as eight panes, we have used only few of them namely Lexicon Edit, Browse, Dictionary, Classified Dictionary and Bulk Edit Entries for the creation of the multilingual dictionary involving Tamil, Telugu, English and Portuguese languages. In the following section we will exemplify each one of them with examples.

Fig: 1 – Lexicon Component

Headword	Lexeme Form	Glosses	Grammatical Category
அழகு	அழகு	the qualities that give pleasure to the senses	Noun
ஆசிரியை	ஆசிரியை	a person(female) whose occupation is teaching	Noun
இரண்டு	இரண்டு	the cardinal number that is the sum of one and one	Noun
உணவு	உணவு	any substance that can be metabolized by an organism to give energy and build tissue	Noun
எப்பொழுதும்	எப்பொழுதும்	at all times	Adverb
ஏற்றுக்கொள்	ஏற்றுக்கொள்	consider or hold as true	Verb
ஒவ்வொரு	ஒவ்வொரு	to, for, or by every one of a group	Adverb
கண்	கண்	the organ of sight	Noun
கணக்கு	கணக்கு	a statement of recent transactions and the resulting balance	Noun
கணிணி	கணிணி	an electronic device	Noun
கருப்பு	கருப்பு	being of the achromatic color of maximum darkness	Adjective
கிட்டத்தட்ட	கிட்டத்தட்ட	used as an intensive especially to indicate something unexpected	Adverb

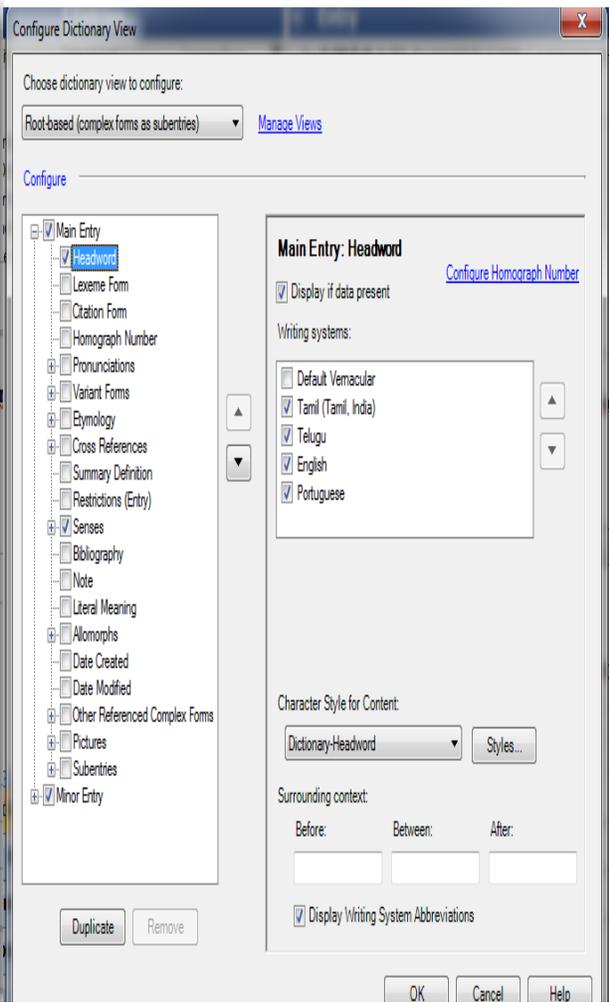
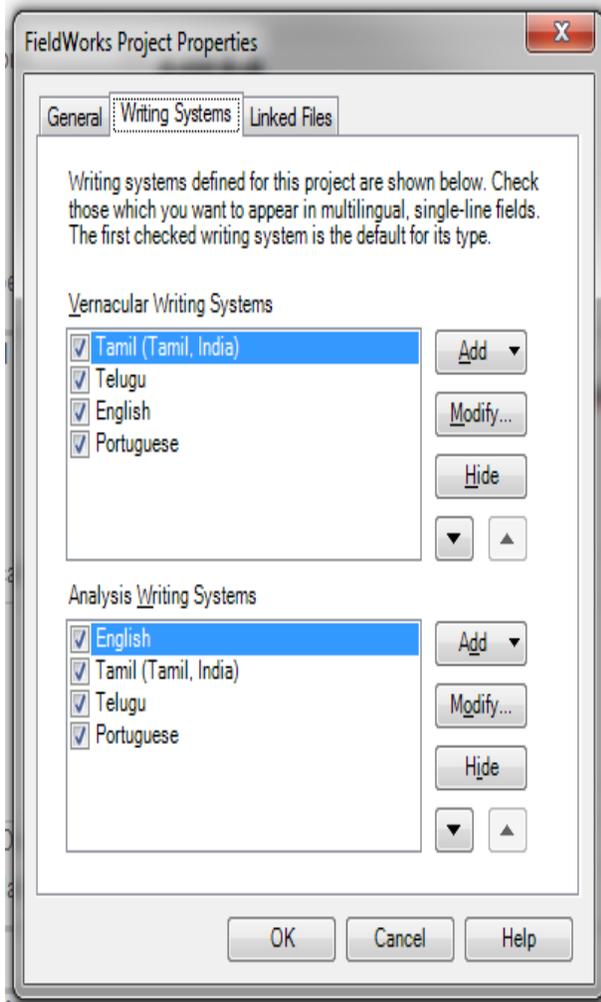
Figure 1 Lexicon Component

In the above lexicon pane, we have eight categories which already mentioned above. When we click on first one lexicon edit it will display four items viz. head word, lexeme form, glosses and grammatical category under the entries. Head word is a root word which can be taken from any language and enter it under the head word. The next one which is a lexeme form is almost similar to root word. Under glosses we get the meaning or different meanings of the head words and under the fourth one grammatical category we get all the information that, whether they are noun, adjective, verb, adverb and other grammatical categories.

The second step is as we see in the fig: 2 is set up writing. When we want to start writing we have to click on the option writing system. When we click on it

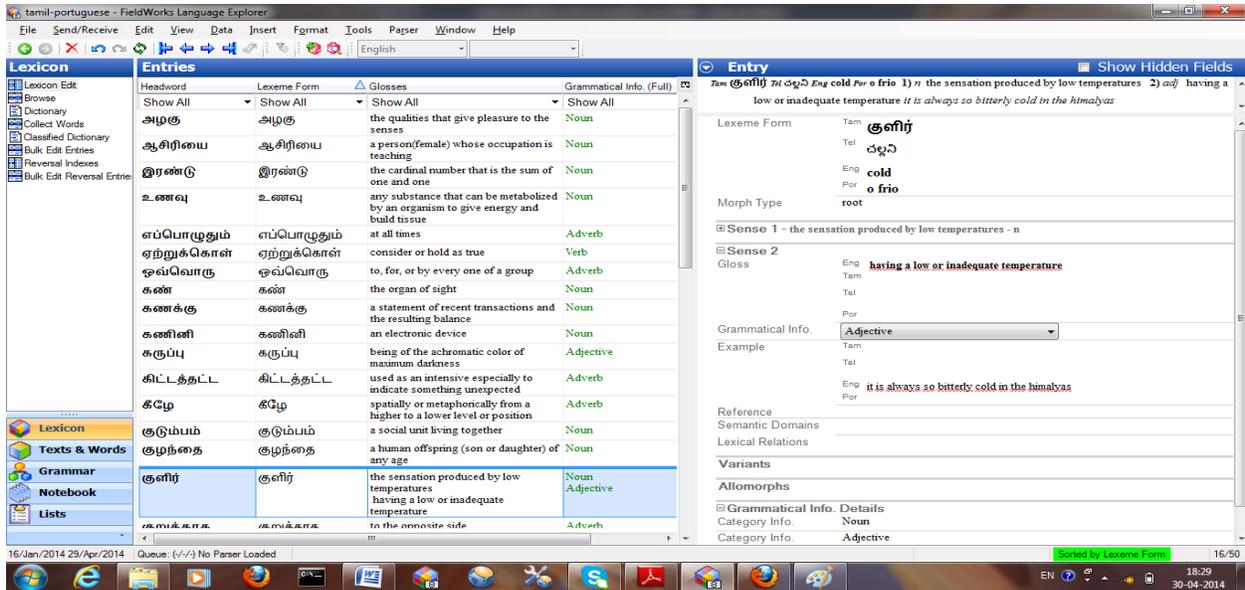
Fig: 2 - Setup Writing

Fig: 3 – Configure Dictionary



## 5. Lexicon Edit

Fig: 4 – Lexicon Edit



There are three basic types of views in Language Explorer--edit view, browse view, and publication view. An edit view separates each field in a record and places each field on a separate line. This makes it easy to see and edit each field in a record. A browse view places each record on a separate line and puts each field in a separate column. This makes it easy to compare the data in several records at once. A publication view shows how the data will look (or could look) when it is published. The primary view of the data is called the **Lexicon Edit** view. It is a combination of each of the three basic types of views. The left pane is the **Entries** pane which consists of a browse view. The right pane is the **Entry** pane which displays a single entry at a time. When you select an entry in the **Entries** pane, it will be displayed in the **Entry** pane. The **Entry** pane consists of a dictionary view at the top and an edit view at the bottom. The dictionary view enables you to see how the entry will appear in print with the settings you have currently chosen. The edit view enables you to edit each field of the entry. There is also a very specialized view, **Categorized Entry** that was designed to enable you to enter lots of new words efficiently using the Dictionary Development Process list of semantic domains. All of these views are views of a single database. The data is the same in each view. Language Explorer

merely presents it differently in order to facilitate a variety of tasks. The rest of this paper describes how to use the tools and views in Language Explorer to get the job done.

## **6. Lexeme**

In lexicography we use the term lexeme to refer to a unit in the mental lexicon. Sometimes the term is defined as the minimal unit of the semantic and sometimes it is used to refer to a combination of a form and a single sense. But in Language Explorer it is used to refer to the mental equivalent of a dictionary entry. So a lexeme can have more than one sense. A lexeme can be a morpheme, word, or lexical phrase, because each is a form with an associated meaning. If a lexeme is a stem, all the inflected forms of the stem are included. So a lexeme is a stem and all its inflected forms. If the lexeme has allomorphs or variants, they too are included. So a lexeme is an abstract unit. A lexeme can have more than one pronunciation and more than one meaning. It can be inflected in different ways and even belong to more than one grammatical category. In Language Explorer we create an entry for each lexeme. The basic form of the lexeme is put in the **Lexeme Form** field. We record the allomorphs and variants of the lexeme, create a sense for each meaning, and describe all aspects of its form, meaning, and usage.

## **7. Lexeme form**

Each lexeme will have different senses, grammatical information, Morph type, gloss grammatical information and all their semantic domains. With the inbuilt semantic domains and grammatical domains, it becomes easy for the people even who do not have the knowledge in dictionary making. A week of training in the tool will help even a lay man to build a concise dictionary that will help the minor languages spoken by the minor communities of the country.

Entry
Show Hidden Fields

Tam குளிர் Tel చల్లని Eng cold Por o frio 1) *n* the sensation produced by low temperatures 2) *adj* having a low or inadequate temperature *it is always so bitterly cold in the himalyas*

---

Lexeme Form	Tam <b>குளிர்</b>
	Tel చల్లని
	Eng <b>cold</b>
	Por <b>o frio</b>
Morph Type	root

---

☑ Sense 1
[Insert Example](#) [Insert Sense](#)

Gloss

Eng	<u>the sensation produced by low temperatures</u>
Tam	
Tel	
Por	

Grammatical Info. Noun

Example

Semantic Domains

Lexical Relations

---

☑ Sense 2

Gloss

Eng	<u>having a low or inadequate temperature</u>
Tam	
Tel	
Por	

Grammatical Info. Adjective

Figure 5 Lexical Relations

## 8. Lexical relations

Some lexemes are related not by form but by meaning. These meaning based relationships are called 'lexical relations' in language explorer. Some of them are: synonym, antonym, generic: specific, part: whole. In the example provided in Fig: below, the Tamil lexeme form 'cat' is linked to 'animal' by the semantic relation generic: specific

**Entry** Show Hidden Fields

Tam: பூனை *Tel* పిల్లి *Eng* cat *Por* gato *n* 1) feline mammal usually having thick soft fur and no ability to roar *Any closer, and they stop peeking altogether and watch the cat closely.* *gen:* விலங்கு. (sem. domains: 1.6.1.1 - Mammal.) 2) an informal term for a youth or man

---

Lexeme Form

Tam: பூனை  
Tel: పిల్లి  
Eng: cat  
Por: gato

---

Morph Type: root

---

**Sense 1**

Gloss

Eng: **feline mammal usually having thick soft fur and no ability to roar**  
Tam:  
Tel:  
Por:

---

Grammatical Info. **Noun**

---

Example

Tam:  
Tel:  
Eng: Any closer, and they stop peeking altogether and watch the cat closely.  
Por:

---

Reference

Semantic Domains: 1.6.1.1 - Mammal

---

Lexical Relations

Generic: விலங்கு (a living organism characterized by voluntary movement)

Figure 6 Entry

### 9. Browse view

Headword	Lexeme Form	Glosses	Definition	Grammatical Info. (Full)
அழகு	அழகு	the qualities that give pleasure to the senses		Noun
ஆசிரியை	ஆசிரியை	a person(female) whose occupation is teaching		Noun
இரண்டு	இரண்டு	the cardinal number that is the sum of one and one		Noun
உணவு	உணவு	any substance that can be metabolized by an organism to give energy and build tissue		Noun
எப்பொழுதும்	எப்பொழுதும்	at all times		Adverb
ஏற்றுக்கொள்	ஏற்றுக்கொள்	consider or hold as true		Verb
ஒவ்வொரு	ஒவ்வொரு	to, for, or by every one of a group		Adverb
கண்	கண்	the organ of sight		Noun
கணக்கு	கணக்கு	a statement of recent transactions and the resulting balance		Noun
கணினி	கணினி	an electronic device		Noun
கருப்பு	கருப்பு	being of the achromatic color of maximum darkness		Adjective
கிட்டத்தட்ட	கிட்டத்தட்ட	used as an intensive especially to indicate something unexpected		Adverb
கீழே	கீழே	spatially or metaphorically from a higher to a lower level or position		Adverb
குடும்பம்	குடும்பம்	a social unit living together		Noun
குழந்தை	குழந்தை	a human offspring (son or daughter) of any age		Noun
குளிர்	குளிர்	the sensation produced by low temperatures		Noun
		having a low or inadequate temperature		Adjective
குறுக்காக	குறுக்காக	to the opposite side		Adverb
கூட்டு	கூட்டு	make an addition by combining numbers		Verb
கெட்ட	கெட்ட	having undesirable or negative qualities		Adjective
கொடைக்காலம்	கொடைக்காலம்	the warmest season of the year		Noun
சகோதரன்	சகோதரன்	a male with the same parents as someone else		Noun
சாதி	சாதி	to gain with effort		Verb

Figure 7 Browse View

As given above, in browse view you often want to sort on the headword. Language Explorer displays a "virtual" **Headword** field in the browse views to enable you to sort on the headword. If there is something in the **Citation Form** field, it is used as the headword. Otherwise the

contents of the **Lexeme forms** will be used. You cannot edit the **Headword** field, since it is a virtual field. If you want to edit the headword, you must edit the **Citation Form** field or the **Lexeme Form** field. Language Explorer generates the headword in **Dictionary** view in the same way, by using the contents of the **Citation Form** field or, if there is nothing in the **Citation Form** field, by using the contents of the **Lexeme Form** field. The following screen shot shows three entries.

### 10. The Final Dictionary

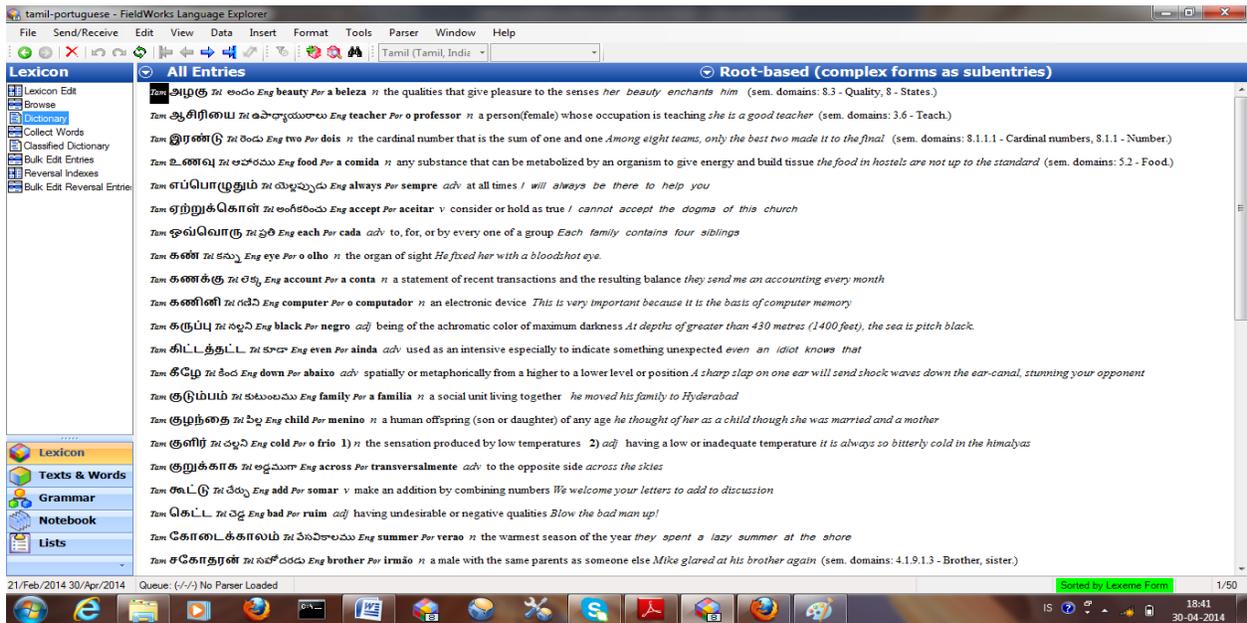


Figure 7 Final Dictionary

### 11. Conclusion

The FLEx is a boon in the area of dictionary making developed by SIL in the united states of America. The components of the tool will have different senses, grammatical information, Morph type, gloss grammatical information and all their semantic domains. With the inbuilt semantic domains and grammatical domains, it becomes easy for the people even who do not have the knowledge in dictionary making. A week of training in the tool will help even a lay man to build a concise dictionary that will help the minor languages spoken by the minor communities of the country.

## References

- Arulmozi, S. & Bapuji, M. 2014. Building a Tamil-Portuguese-English Dictionary using FLEx. Presented a paper at Dravidian University, Kuppam.
- Singh, Ram Adhar. 1982. An Introduction to Lexicography. Mysore: CIIL.
- Jensen, Joshua and Fitzgerald, Fitzgerald. 2011. The University of Texts at Arlington. Fieldwork Language Explorer Training.
- Mishra, Anupama. 2017. Electronic Tri-lingual Dictionary (Awadhi, Hindi, English). Unpublished Masters Dissertation. Varanasi: Banaras Hindu University.