Chapter 1

INTRODUCTION

1.1 Background

Human language or natural language or simply language, has its significance in human life. Language got evolved along with human life (Christiansen and Kirby, 2003). Communication is the prime property of the language which is used to exchange ideas, share feelings, exhibit the emotions and disseminate the knowledge. Initially gesture or sign oriented form of language was there and later it evolved to have phonetic and script forms. Even the visually challenged people have their own language to express their thoughts (Susanne, 2000, Knut et al., 2001).

Language has its Omni presence in every arena of human life, starting from day-to-day life, education, entertainment, politics, technology etc. In order to meet the demands, they have different forms like poetic form, dramatic form, literary form etc (Classical Tamil, 2005). Aesthetics of human life has overloaded the language. In brief, human can not think without languages since the language has become the representational form of his thoughts.

Beyond the communication task, language is treated as a social identity by the people. This helps the people to bind as a society or community with its own culture. Even it can be seen in the history, that there were language wars and language agitations to protect and establish their own language (Galindo, 1997). These cultural and communal aspects introduced new dimension to the language.

In brief, a language is a system of symbols, generally known as lexemes and grammar or system of rules by which they are manipulated (Tom, 1998, Nelleke, 2003). The classification of languages or language taxonomy can be performed on the basis of different underlying principles like genetic relatedness of languages, internal structure of languages and geographical closeness and contacts between language speaking communities (Smith, 1983, Mair, 1991).
In early times, when most people were members of small language communities, it was necessary to know two or more languages for trade or any other dealings outside one's own town or village, and this holds true today also in places of high linguistic diversity. Multilingualism at human level is defined as a person with communicative skills in more than one language (Holmes, 2006). Moreover, globalization has introduced newer market space with different languages and cultures to the business. It also enforces multilingualism. Linguists classified the multilingualism as coordinate, compound and subordinate. In coordinate form, the linguistic elements (words, phrases) in the speaker's mind are all related to their own unique concepts. In contrast, in the compound type the linguistic elements of different languages are attached to the same concepts. In subordinate form, the linguistic elements of one of the speaker's language is available only through elements of the speaker's other language.

On the other hand, languages got evolved due to social, political, cultural and technological influences. In order to meet the demands of the society, the language has to get newer forms. In particular, technological innovations significantly influence the change in the language as reformation of symbols, symbol set and rules in order to introduce printing technology (Schiffman, 1998). Similarly, Information and Communication Technology (ICT) demanded improvements over the language aspects like change in the character set, grammars, keyboard layouts etc, (Carr, 2007).

Innovations of information technology have given a new face to the language. Digital form of language has got its own significance. English is used as the digital medium language. Other languages are also reformed in order to fit in the digital form. Ample research is put into this domain on the following topics:

- Computational linguistics
- Language Engineering
- Localization and Internationalization
- Multilingual Software Development
These research areas are distinct with their objectives and also they have their roles separately and cooperatively in the structure of software, in order to enable the software as mono/multilingual capable.

1.2 Motivation

Multilingualism got its significance in software due to globalization (Fraber-consulting, 2007; Luo and Shenkar, 2006, Dietz, 2007). In the present scenario, each human has to know more than one language for their official life and/or social life. Therefore, the gadgets designed have to exhibit multilingualism (Dougnac, 1995). Human’s key gadget, i.e. computer, also has to meet this requirement. In order to make computer multilingual, appropriate software has to be designed. Multilingual software is the one which exhibits multilingualism in one or more aspects like Input/Output (IO), User Interface (UI), storage, process etc. (Dougnac, 1995; Acharya, 2007; Schmitt, 2001; M17N, 2007) The terms multilingual software and multilingual based software are there in the literature to represent a software with multilingualism. In order to avoid too many terminologies, the term multilingual software is used in this thesis.) The development of multilingual software is driven by stakeholders of multilingual software namely software engineers, computational linguists, language engineers and end users. The stakeholders’ requirements have been captured and realized in multilingual software.

The multilingual software development efforts were made in software industries (Schmitt, 2001, Indiainteractive, 2007), academic institution (Acharya, 2007) and research and development organization (M17N, 2007). Requirements of multilingual software can be categorized based on scenarios presented in (Schmitt, 2001):

Requirement 1:

The user A works with a workstation in his native language. Another user B wants to work with his own native language which is different from A’s language in the same machine. It can be categorized as same machine with multiple users.
**Requirement 2:**

The user A works with an application in his native language. Another user B wants to work with his own native language which is different from A's language in the same application. It can be categorized as *same application with multiple users*.

**Requirement 3:**

The user wants to work with an application in many languages by choosing languages dynamically or configuring the application. It can be categorized as *same application handling multilingual data with a user*.

Research (Dougnac, 1995; Acharya, 2007, Bateman et al., 1999) and commercial (Schmitt, 2001; M17N, 2007, Indianative, 2007) initiatives for developing multilingual software is being carried out in innumerable proportions. These initiatives stimulated the multilingual software development to shift to a better level to achieve multilingual digital world in the near future.

According to Multilingualization group (M17N, 2007), multilingualization in most of the software is peripheral, that is, multilingual facilities can be isolated from other (main) parts of the software. At the same time, most of the software have common requirement for their multilingual interfaces. A general library is being proposed that fulfils those requirements to make software development more efficient and inexpensive.

Domain level language requirements have a lot of significance in multilingual software development as seen in (Fraber-consulting, 2007, Luo and Shenkar, 2006). In (Luo and Shenkar, 2006), a framework was proposed for analyzing the multinational corporation (MNC) as a multilingual community, in which parent functional language and subunit functional languages are concurrently used and recursively linked through an intra corporate communication network.

Also, a group of developers in the Department of Computer Science, Pondicherry University have designed and developed a multilingual operating environment, named as PONN, with set of tools (Kuppuswami et al., 2003) and the
same is extended for the visually handicapped, named as SPECS (Kuppuswami et al., 1999).

The above mentioned efforts are few among the lot to show the glimpse of multilingual software development. Like any other horizontal domain, multilingual software also has architecture either implicitly or explicitly. Architecture based development approaches ensure not only the functional qualities but also the non-functional qualities like reusability, modifiability etc, than the traditional approaches (Bass et al., 2002; Clements and Northrop, 2002).

So it is clearly evident that, software architecture for multilingual software is important similar to other software domains. Architecture based development enables to achieve the qualities which are the concerns for multilingualism.

1.3 Problem Definition

In order to understand the multilingual software development, a review of the literature in this area was carried out. This literature review generates the following questions:

- What are the functional/non-functional qualities of multilingual software?
  
  Like other software domains, multilingual software should also possess certain qualities. These qualities have to be derived from the literature.

- Are there any models / methods for multilingual software development?
  
  There are no explicit models available for multilingual software development. Few of the existing multilingual software development process followed inherent models. Still, it is required to mine the existing multilingual software for the models.

- Will these models provide the qualities of multilingual software?
  
  These models have to be analyzed for the expected multilingual software qualities. The designer has to analyze the mined multilingual software models and find the inadequacies in it. These inadequacies have to be addressed by proposing a model for multilingual software.
1.4 Contributions

This research work tries to answer the above stated questions. The research contributions towards the multilingual software development are stated below and depicted as research work flow in Figure 1

1. A study on the development of multilingual software is carried out and models are mined for the same. The mined models are categorized as wrapper, monolithic and library based models. The background, concept, merits and demerits of the mined models are documented. In addition, a discussion on the mined models focusing on the concerns of stakeholders and qualities expected from multilingual software is carried out.

2. The language components of these models have been analyzed using design space approach. Clearly, it shows that there are numerous alternatives of a language component due to the language aspects which are demanded by the stakeholders of multilingual software. These alternatives introduce the design complexity in the multilingual software development. Also, these alternatives reduce the non-functional qualities like modifiability, reusability and understandability. To overcome the above said shortcomings, a new model for multilingual software is proposed and it is represented using algebraic structure. This is achieved by separating the language aspects and constructing the language component using the separated aspects. This model overcomes the limitations of the mined models and also meets the non-functional qualities of multilingual software.

3. An architectural reference model for multilingual software (ARMMS) has been derived using unit operations, based on the new model. This model ensures the non-functional qualities of multilingual software.
Figure 1.1: Research Workflow
1.5 Organization of the Thesis

Rest of the thesis is organized in the following manner.

Chapter 2 presents the definition and qualities of the multilingual software. The main focus is on the multilingual software development approaches and a comparative study of the development approaches based on the characteristics and qualities of multilingual software.

Qualities of the multilingual software are depicted in chapter 3. Research problem and the research approach are the main focus of this chapter.

Formal models of the multilingual software are mined from the existing multilingual software and presented in chapter 4. The inadequacies of these models based on the expected qualities of multilingual software are also specified.

In chapter 5, requirements of multilingual software are considered and analyzed using design space approach. It reveals the need for a new model and it is represented using algebraic structure.

Chapter 6 is dedicated for the design of architectural reference model for multilingual software (ARMMS) by applying the unit operations. Also, conceptual view and logical view of the proposed reference model are detailed.

Development of different multilingual software, by applying the ARMMS, is presented in chapter 7 and the architectural experiences are also shared in this chapter.

In Chapter 8, conclusion and future research directions on multilingual software development are presented.