

# Marathi as a Formal Language and development of a Marathi Compiler

(i.e. Marathitun Adnyavali Vikasit Karane)

Prof. Mukund Dhaygude

Email : mukunddhaygude1@yahoo.co.in

The paper deals with the following topics  
i.e.

1. Science, Technology and Management aspects of "Sanganak Ani Marathi"
2. Review of the present scenario
3. Mathematical Marathi or Marathi as a Formal Language
4. Computational Marathi and Marathi Compiler
5. Marathi as a Natural Language and Marathi Corpora Creation and Management

## 1. The Science, Technology and Management aspects of "Sanganak and Marathi"

Sanganak comes from the backgrounds of science, technology and management. Science deals with proving concepts mathematically / experimentally. Technology takes off from where science leaves by producing a prototype of the concept is more like the sense of 'The proof of pudding lies in having it'. Technology demonstrates the production on a mass scale of the product which was prototyped after scientists proved it. Management concerns itself with value addition in the society. Management begins where Science and Technology leaves. Science and technology thus mass produce several products. Management introduces these products in to the society and does a value addition to the society.

## 2. Review of the present scenario

We hypothesized last time that we should define Marathi as a formal language and write programs in Marathi since a Marathi

Basic interpreter/compiler is now available. The Mbasic.exe is available on [www.mbasic.8m.com](http://www.mbasic.8m.com) website. One can have it from us by contacting address given above (by phone or by email).

To familiarize the use of Marathi, we are proving Marathi to be mathematically and computationally complete. Later we will develop system softwares in Marathi and then we will develop application softwares in Marathi. The present system softwares are in English. Operating systems are now getting converted to show Marathi icons. We will then develop interpreter/compiler that will enable us to program in Marathi and that is what this white paper talks about.

## 3. Mathematical Marathi or Marathi as a formal language

An alphabet A of a language L comprises of a set of all possible letters/characters that are fundamentally or derivable in the Language. A language L comprises of a grammar G. The Grammar G uses the alphabet A. A Grammar G comprises of set of rules of the grammar. These rules are defined in terms of tokens, symbols, operators, precedence, and mathematical \Boolean\digital operators, statements, computational statements, conditional statements, loops, programming constructs etc. The rules are structured from elementary linguistic formations such as letters/ characters, words, sentences, paragraphs, stories, essays, news, dialogues, pages etc.

## 4. Computational Marathi and Marathi Compiler

Having defined the mathematical Marathi we then represent it using the language

representation schemes (such as Unicode etc) in the keyboard, printers, sanganak, operating system etc and then proceed to see the Mbasic.exe software that demonstrates the ability to interpret the simple sample programs that are distributed with it. Having executed the user programs the control is returned to operating system.

Marathi Basic or just MBasic for short is an experimental version of an interpreter created to consider the possibility of a full-fledged programming language based on the Marathi Language. This language will define its own syntax and statements that will be extremely close to the spoken Marathi language. MBasic being experimental, we have experimented at converting the entire interface to Marathi. But rather than forcing Windows Operating System itself to change the system font, which is done by a good many softwares, MBasic only keeps its own interface in Marathi while leaving other applications in control of their own. Hence, MBasic respects other applications privacy and will work in co-operation with accompanying software.

The name Marathi Basic was coined from a Marathi-to-basic transliterator that we had created to help Marathi medium students learn QBasic. But what resulted were new syntax, statements and entire interfaces in Marathi. MBasic has been designed to be extensible as a language and new statements should be easily integrated as and when required. This is a task for which I require your comments and suggestions. Please send me ideas and suggestions on how to improve, increase or change MBasic so that it can be a better piece of work.

MBasic 5.0 is a complete remake of the entire MBasic engine. Right from the compiler, interpreter, to the display, everything has been internally revamped. MBasic 5.0 is very slow. Unlike its predecessors, it provides features which cannot be compiled or even pseudo-compiled. They have to be interpreted. In a windows environment this costs system resources such as memory and more

importantly, the processor's attention. MBasic 5.0 has more debugging and error identification support. It checks and double-checks your code for errors. Yet, at this time it is not able to suggest solutions to the errors.

One may use, distribute and copy MBasic as many times as you wish, though you may not in any way modify MBasic.

We are desperately in need of someone writing a good help file on MBasic. Please help me. For now, you are provided with sample programs. I have tried to provide sample programs that use almost every variant of the MBasic syntax. You should be aware that a person with some programming knowledge must read the sample files and provide explanations for someone with absolutely no programming knowledge. Some reading, experimentation, interpolation and creativity should be enough to give you the hang of MBasic. As the author, I must say that programming in MBasic can be quite fun.

Modifications, Undocumented features, changes, additions, restrictions and things you should know about:

1. One may have to enclose a syllable consisting completely of a vowel in {} as the transliterator program is unable to determine when a syllable has ended. For example, to enter the word 'file' in devnagri, if one types only 'phaaila', the transliterator will not understand that 'phaa' is one syllable while 'i' itself is an independent syllable since no consonant comes between the two. Hence, one will have to type in 'phaa {} la' to get the proper transliteration. For more information on these conversion rules, please visit Omkarananda Ashram website [www.omkarananda-ashram.org](http://www.omkarananda-ashram.org). MBasic uses the transliterator developed by them for its conversion routines. For complex devnagri entry, it would be a good approach to visit and read about the ITrans transliteration scheme that is used by the translators.

2. Interlaced devnagri and roman text in the program is allowed. Even strings may

contain interlaced Roman and Devnagri text. Input and display functions support this feature completely. One may enter Roman as well as Devnagri text mixed together. String comparisons will take this into consideration when comparing two strings. To interlace roman text within devnagri text, just enclose it in ##. This feature is useful for entering special characters such as '{' and '}' which are recognized by the transliterator and need to be explicitly placed within ##s in order to show up.

3. Obviously, there is absolutely no way to add a # in the program. Sorry about this. Also, there is yet no facility to add a " within a string. Sorry again. Though the input function provides this facility, the editor will not allow this. We are working on schemes to provide these facilities. Such special features increase the load on the CPU exponentially and will be added only in case a lot of people demand it.

4. Provision for making stand-alone executables is provided. Of course, these executables still need support dll's and also the true type fonts. But they can run completely independent of the MBasic IDE. They also contain a pseudo-compiled code that is very difficult to decipher.

5. Graphics support for drawing lines, circles, boxes and points. These routines are the most fun to use. They have a very flexible syntax. They are almost natural Marathi language instructions. Very easy to read and use once you get the hang of them.

6. The find and replace utility searches only the upper text box which takes in phonetic user input. It does not search in the transliterated text box displayed below it. The transliteration displayed in the find and replace utility is provided for your reference and is not considered by the MBasic Editor when searching for text.

7. Sometimes the transliterator can display garbage in the output window at the

end of a line. In this case, just add some extra spaces at the end of the line to correct this problem. We still have to analyze why this happens. A possible cause might be that MBasic uses dll's compiled from a lot many languages including VC++, Delphi, VB, etc. It may be the transfer of strings from so many dll's that causes this.

8. The garbage collection in MBasic is very bad. It hogs a lot of memory. So it is advisable for one to save one's work regularly. Please save thy work before running any program. If it gives any problems, just restart it and it will work fine (since MBasic only runs on windows, you should be quite familiar with this practice).

MBasic has been a voluntary effort to develop MKeyboards, MBios, Mcomputers, Mprinters, and other Msoftwares so that we can check them out and fill up their weaknesses. We also have our hardware development projects in Marathi. If you like MBasic, please contact us for further development.

5 Marathi as a Natural Language and Marathi Corpora Creation and Management

We can now read Marathi Dailies i.e. Newspapers (Lokmat, Maharashtra Times, and Loksatta, Sakal etc) online on the Internet. We may need to download fonts if required. Unicode addresses this issue regarding the aesthetics of the written textual information.

We are in process of development of Marathi *Sanganakiya Paribhashik Shabdkosh*. Some of the words that we have found suitable are already put in use in our communications. We also have found that Janbharati group of CDAC/NCST has released a Linux version with Marathi icons/symbols. VPM Thane has also developed and released e-books on works of *Vishnushastri Chiplunkar. Sanganakiya Marathi Paribhashik Shabdkosh* development is also in progress.

☆☆☆