

বাঙলা থেকে BangTEX: mkbangtex

(Version 1.1)

কৌশিক রায়

তাত্ত্বিক পদার্থবিজ্ঞান বিভাগ
ইন্ডিয়ান এসোসিয়েশন ফর দি কাল্টিভেশন অফ সায়েন্স
কলিকাতা ৭০০ ০৩২, ভারতবর্ষ

সূচী

১	প্রবেশক	১
১.১	দুটি ধারা: bengali ও foreign	১
২	নিয়মাবলী	২
২.১	ব্যঞ্জন বর্ণ	২
২.২	স্বরবর্ণ	৩
২.৩	নাসিক্য ধনি	৩
২.৪	খন্ড ৭, হসন্ত, বিসর্গ ও সানুনাসিক বর্ণ	৪
২.৪।১	আরও উদাহরণ	৪
৩	রোমান প্রতিবর্ণীকরণ	৪
৪	ত্রুটি সংশোধন	৫
৫	ancient অপ্শন	৫
৬	ব্যত্যয়	৫

১ প্রবেশক

`mkbangtex` একটি Python pre-processor। এটি `BangTeX`-এর সহায়িকা। এর মূল উদ্দেশ্য রোমান হরফে বাঙলা লেখার কিছু সুবিধা বিধান। `BangTeX`-এ টাইপ করার সময় যুক্তব্যঞ্জন এবং এ-কার, ই-কার, ঐ-কার ও ঔ-কার লেখার জন্য কিছু নিয়ম মানতে হয়। এই নিয়মগুলি আদ্যোপাত্ত যুক্তিনির্ভর, ব্যতিক্রমহীন, গাণিতিক নিয়মের মত। যেমন, যুক্তব্যঞ্জন লেখা হয় ভগ্নাংশ হিসাবে। র, ম, ল, য-ফলা লেখা হয় ইংরেজি ক্যাপিটাল হরফে। কিন্তু, এই রীতি বাঙলা উচ্চারণানুগ না হওয়ায় তাতে টাইপ করতে বেশি সময় লাগে। `mkbangtex` উচ্চারণানুগ রোমান হরফে টাইপ করা বাঙলা লেখা থেকে `BangTeX`-এর উপযোগী একটি `LATEX` ফাইল বানিয়ে দেয়। `mkbangtex` `--help` ব্যবহার করে সংক্ষিপ্ত সাহায্যসূত্র ও ব্যবহারবিধি পাওয়া যাবে:

```
This is mkbangtex Version 1.1.
```

```
It creates a BangTeX file from a Bengali document written phonetically in Roman script. It works in two modes, which are automatically detected:
```

```
bengali: converts anything but single-word latex commands not protected by enclosing with @fn's.
```

```
foreign: converts only the portions enclosed within a pair of @bn's.
```

```
CAUTION: Only ONE mode can be used in a file.
```

```
A piece of text enclosed within a pair of @b2e's is converted from Bengali to accented Roman script following the IAST (1912) rules. Works in both modes.
```

```
[See accompanying documentation for further details.]
```

```
---
```

```
Usage: mkbangtex [-h, --alphabet --transparent, --ancient, -o <out_file>] <in_file>
```

```
-h, --help      : print this page
```

```
--alphabet     : transliteration table
```

```
--transparent  : transparent symbols and default positioning
```

```
--ancient      : ancient Bengali --- without antastha ya
```

```
-o <out_file>  : specify output file (default: <in_file>.tex)
```

<out_file> নির্দিষ্ট করা না থাকলে <in_file>.tex নামে একটি `LATEX` ফাইল তৈরি হবে।

যদি কেবল `text` টাইপ করতে হয়, তাহলে এই pre-processor ব্যবহার করার সহজতম উপায় হল উচ্চারণানুগ বাঙলা `text` সংবলিত একটি `LATEX` ফাইল লিখে

```
mkbangtex -o <out_file> <in_file>
```

চালিয়ে নেওয়া। `mkbangtex` এক শব্দ-বিশিষ্ট `LATEX` command গুলি ছাড়া অন্য শব্দগুলিকে বদলে একটি `BangTeX` ফাইল তৈরি করে দেবে। ‘শব্দ’ অর্থে দুটি `whitespace` -এর মধ্যবর্তী অংশ। যেমন, `\section{prabeshak}` একটি শব্দ। অতএব, `\section` -এর নাম বাঙলায় লিখতে হলে `prabeshak` কথার আগে ও পরে ফাঁক দিয়ে `\section{ prabeshak }` লিখতে হবে। পক্ষান্তরে, `\begin{center}` একটি শব্দ, কাজেই তা’ অপরিবর্তিত থাকবে।

কিন্তু লেখার মধ্যে ইংরেজি শব্দ, ছবি বা গাণিতিক ফর্মুলা থাকলে কাজ কিছু বেশি। এবার তার আলোচনা।

১.১ দুটি ধারা: `bengali` ও `foreign`

`mkbangtex` -এর দু’টি ধারা বা mode আছে: `bengali` ও `foreign`। যে ধারায় ফাইলে মূলতঃ বাঙলায় লেখা আছে, সেটি `bengali` ধারা। এই ধারায় `mkbangtex` এক শব্দ-বিশিষ্ট `LATEX` command গুলি অপরিবর্তিত রেখে অন্যান্য শব্দগুলি পরিবর্তিত

করবে। এই ধারায় কোনও অংশ অপরিবর্তিত রাখার জন্যে (যেমন, বাঙলা লেখার মধ্যে ইংরেজি শব্দ লিখতে হলে) সেই অংশ @fn দ্বারা বেষ্টিত ক’রে দিতে হবে। উদাহরণস্বরূপ, নীচের সমীকরণটি,

$$a = b + cd/5! \quad (১)$$

এইভাবে লেখা যাবে:

```
\begin{equation}
@fn
a = b+cd/5!
@fn
\end{equation}
```

স্পষ্টতঃ, \begin{equation}, \end{figure} ইত্যাদি এক-শব্দ-বিশিষ্ট L^AT_EX command হওয়ায় এদের বেষ্টনীর বাইরে রাখা যায়। বিশেষতঃ mathmode, figure, verbatim ইত্যাদি environment -এর মধ্যে @fn বেষ্টনী অত্যন্ত উপযোগী।

foreign ধারা এর ঠিক বিপরীত। ধরা যাক একটি মূলতঃ বিদেশী ভাষায় (যেমন L^AT_EX!) লেখার মধ্যে কিছু বাঙলা শব্দ ব্যবহার করতে হবে। এর জন্যে bengali ব্যবহার করা অনর্থক — bengali ধারায় mkbangtex চলতে অপেক্ষাকৃত বেশি সময় লাগে। এই ক্ষেত্রে মূল ফাইলের মধ্যে যে যে অংশ বাঙলায় লিখতে হবে, সেই অংশগুলিকে @bn দ্বারা বেষ্টিত ক’রে দিলেই চলবে। এতে runtime অনেক কম লাগবে। foreign ধারায় L^AT_EX command গুলি @bn বেষ্টনীর বাইরেই থাকবে। ফলে, এদের কোনও পরিবর্তন হবে না। যেমন, \section -এর নাম লিখতে \section{@bn prabeshak @bn} লিখতে হবে। পাশাপাশি দুবার @bn লেখা যাবে না — এতে গড়গোল হবে। দুটি ধারার বিবরণ ও বেষ্টনীর তালিকা নীচে পুনরায় দেওয়া হ’ল:

ধারা (mode)	প্রকার (type)	বেষ্টনী (delimiter)
bengali	বাঙলা	@fn
foreign	বিদেশী	@bn

উভয় ধারাতেই ডলার দিয়ে গাণিতিক ফর্মুলা স্বাভাবিক ভাবেই লেখা যাবে: $a = bd + c$ । এইভাবে কম-সংখ্যক ইংরেজি শব্দও লেখা যায়: like this!

ফাইলে @bn অথবা @fn থাকলে mkbangtex সঠিক mode ব্যবহার করবে। সুতরাং, একই ফাইলে দু’টি ধারা ব্যবহার করা যাবে না। পক্ষান্তরে, @fn বা @bn -এর কোনওটিই না থাকলে bengali ধারা ব্যবহৃত হবে। এই ধারাই mkbangtex -এর পূর্বসূত্র default ধারা।

২ নিয়মাবলী

এবার উচ্চারণানুগ বাঙলা লেখার বিধি। bengali এবং foreign উভয় ধারাতেই এই বিধি প্রযোজ্য। উচ্চারণ অনুসারে লেখার সুবিধা হ’ল, নিয়মগুলি অনুমান ক’রে নেওয়া যায়। উপরন্তু, মূল লেখার পাঠযোগ্যতা বৃদ্ধি পায়। mkbangtex -এও সেই সুবিধা পাওয়া যাবে। অনুমান-নির্ভর লেখায় ভুল হবার সম্ভাবনা অত্যন্ত কম। হ’লে এই ডকুমেন্টেশনের পরবর্তী অংশ পড়বার প্রয়োজন হ’তে পারে। এছাড়া, --alphabet option ব্যবহার করে সমস্ত চিহ্নের তালিকা মিলবে।

২.১ ব্যঞ্জন বর্ণ

ব্যঞ্জন বর্ণগুলির জন্য দু’টিমাত্র ব্যতিক্রম ছাড়া সব চিহ্নই BangT_EX-এর ডকুমেন্টেশনে যেমন বলা আছে, তেমনই লেখা যাবে। ব্যতিক্রমী ও অতিরিক্ত চিহ্নগুলি,

c, ch	চ
chh	ছ
Sh, shh	ষ
x	ক্ষ
h, H	হ

অন্যান্য ব্যঞ্জন বর্ণের চিহ্নগুলি BangTeX-এর মতই। শুধু এখন আর ভগ্নাংশ দিয়ে যুক্তাক্ষর অথবা ক্যাপিটাল হরফে -ফলা লিখতে হবে না। ব্যঞ্জন বর্ণগুলি পরপর লিখলেই চলবে, যেমন, উচ্ছন্ন, গাঙ্গেয়, অশ্বথ, উচ্চিঙে, মন্দ্রা, পৌঞ্জবর্ধন, মর্মন্তুদ, এই শব্দগুলি লিখতে যথাক্রমে ucchhanna, gaangeya, ashwaththa, uchchiNGrhe, mandrya, pouNdrabardhan, marmantuda টাইপ করতে হবে। কিন্তু, এখন দু'টি ব্যঞ্জন বর্ণ পাশাপাশি অবিকৃত লেখার জন্য তাদের মাঝে a লিখতে হবে, যেমন chalabe: চলবে, আর chaalaabe: চালাবে কিন্তু chalbe: চম্বে। এখানে a -র ভূমিকা BangTeX-এর o -র মত। o দিয়ে এখন ও-কার বা ও লেখা হবে।

২.২ স্বরবর্ণ

স্বরবর্ণ ও তদ্বৃত্ত 'কার' লেখার নিয়ম এই রকম:

	a	a, aman	অ, অমন
	aa	aa, aami, maataa	আ, আমি, মাতা
	i	i, icchhe, pitaa	ই, ইচ্ছে, পিতা
	ii	ii, iishaan, riiti	ঈ, ঈশান, রীতি
	u	u, uT, baabu	উ, উট, বাবু,
	uu	uu, ruup	ঊ, রূপ
⇨	Ri	Ri, {RR}, Rishhi	ঋ, ঋ, ঋষি
⇨	Ri-kaar	_r, k_rshhak, baTak_rshhNa, h_rday	ৠ, কৃষক, বটকৃষ্ণ, হৃদয়
	e	e, bane, ebaar	এ, বনে, এবার
	oi	oi, oisab, boitaalik	ঐ, ঐসব, বৈতালিক
⇨	o+i	{o}{i}, {o}{i}sab	ওই, ওইসব
	o	o, bhaalo, obhaabe	ও, ভালো, ওভাবে
	ou	ou, outsukya, noukaa	ঔ, ঔৎসুক্য, নৌকা
	ai, aai	ai, bai, aai, Jaaiba	আই, বই, আই, যাইব
	ae, aae	ae, ataeb, aae, naae	আএ, অতএব, আএ, নাএ
	aai, aaii	aai, aaii, baaiijii	আঈ, আঈ, বাঈজী
	au, aau	mau, jhaaugaachh	মউ, ঝাউগাছ
	ao, aao	ao, aao, haao	আও, আও, হাওয়া
	iu, iiu	iu, iiu, piu, jiiu	ইউ, ঈউ, পিউ, জীউ
⇨	i+i	i{i}	একটিই
	io, iio	io, iio, kario, baaTiio	ইও, ঈও, করিও, বাটীও
	ei	ei, nei	এই, নেই
	eu	eu, bheu	এউ, ভেউ
	eo	eo, Jeo	এও, যেও
	ui, uii, uui, uuii	ui, uui, uii, uuii, bh~nuiNJaa, shyuui	উই, উই, উঈ, উঈ, উইঞা, শ্যুই

যে চারটি রূপ অনুমান করতে সামান্য বেগ পেতে হ'তে পারে তাদের চিহ্নিত করা আছে।

২.৩ নাসিক্য ধনি

নাসিক্য ধনির ক্ষেত্রে BangTeX-এর সুবিধা অপরিবর্তিত আছে। যেমন, হেলসিঙ্কি অঙ্ক শঙ্খ অঞ্জ অঞ্জল মনোবাহা ব্যঞ্জন বাণ্বা বণ্টন লুণ্ঠন কাণ্ড টেণ্টন দন্ত পান্থ বন্দর আন্ধার ম্ন্য় — কিন্তু চিহ্ন অপরাঙ্ক ব্রন্মা বাণ্ময় বাঙলা:

```
helasinki anka shankha anga anchal manobaanchhaa byanjan jhanjhaa
baNTan luNThan kaanDa DheNDhan danta paantha bandar aandhaar m_rnmay
--- chinha aparaNha bramhaa baanNGmay baanGalaa
```

২.৪ খণ্ড ৭ , হসন্ত, বিসর্গ ও সানুনাসিক বর্ণ

খণ্ড ৭ , হসন্ত, বিসর্গ ও সানুনাসিক বর্ণ লেখার জন্য এই সারণী ব্যবহার করতে হবে:

খণ্ড ৭:	_t	ৎ
হসন্ত:	_h	্
বিসর্গ:	~h	ঃ
অনুস্বার:	_m	ং
চন্দ্রবিন্দু:	~n	ँ

তবে, ৎ-এর পরে ক, খ, ফ, প, ফ কিংবা স থাকলে কেবল t -ও লেখা যাবে। যেমন, উৎকট, উৎখাত, উৎক্ষিপ্ত, উৎপল, উৎফুল্ল, উৎস এ'গুলি এমনিও লেখা যাবে। অর্থাৎ এই দু'টিই ঠিক: utkaT, u_tkaT, u_tpal, utpal ইত্যাদি। কিন্তু স্থানুবৎ: sthaanuba_t, জলভাত: jalabhaat।

২.৪।১ আরও উদাহরণ

bhoir~no - bhair~no - bhoiro~n, s~nodaraban - so~ndaraban: ভৈরৌ - ভইরৌ - ভৈরৌ, সৌদরবন - সৌদরবন
— চন্দ্রবিন্দুর অবস্থানিক পার্থক্য লক্ষ্যণীয়। তেমনই, সালংকারা, সিংহ, সিংগ্যাল: saala_mkaaraa, si_mha, sig_hnyaal।
বিসর্গ-যুক্ত শব্দের উদাহরণ: প্রাতঃরাশ, ওতঃপ্রোত, স্বভাবতঃ, মাভেঃ। এ'গুলি যথাক্রমে praata~hraash, ota~hprota, swabhaabata~h, maabhoi~h।

৩ রোমান প্রতিবর্ণীকরণ

বাঙলা বা সংস্কৃত শব্দ রোমান হরফে লেখার জন্য accented letters ব্যবহৃত হয়। এরজন্য পরিবর্তনীয় অংশের আগে ও পরে @b2e লিখতে হবে। যথা, আদি শ্লোক,

মা নিষাদ প্রতিষ্ঠাং স্বম গমাঃ শাস্তীসমা।
যৎক্রৌঞ্চমিথুনাদেকমবধি কামমোহিতম্॥

mā niṣāda pratiṣṭhāṃ twama gamāḥ śāśwatīsamā.
yatkrauñcamithunādekamabadhi kāmamohitam..

এইভাবে লিখতে হবে:

@b2e

```
{\rm maa nishhaada pratishhThaa_m twama gamaa~h shaashwatiisamaa.\\  
Ja_tkrounchamithunaadekamabadhi kaamamohitam_h..}
```

@b2e

International Alphabet of Sanskrit Transliteration (IAST), 1912 অনুযায়ী প্রতিবর্ণীকরণ-বিধি নীচের সারণীতে দেওয়া হ'ল।

অ আ ই ঈ উ ঊ ঋ এ ঐ ও ঔ

a ā i ī u ū ṛi e ai o au

ক খ গ ঘ ঙ

ka kha ga gha ṅ

চ ছ জ ঝ ঞ

c ch ja jha ṇ

ট ঠ ড ঢ ণ

ṭa ṭha ḍa ḍha ṇa

ত থ দ ধ ন

ta tha da dha na

প ফ ব ভ ম

pa pha ba bha ma

য র ল ব

ya ra la wa

শ ষ স হ

śa ṣa sa ha

ড় ঢ ঙ ঞ

Ṛ Ṛh ṁ ḥ ṅ

সংস্কৃত ভাষায় চন্দ্রবিন্দু-র অস্তিত্ব নেই। ফলে, IAST -র নিয়মে এর জন্যে কোনও চিহ্ন নির্দিষ্ট নেই। এখানে একটি চিহ্ন নির্দিষ্ট হ’ল। পক্ষান্তরে, বাঙলায় ওষ্ঠ্যবর্ণীয় ও অন্তঃস্থবর্ণীয় ‘ব’-এর কোনও পার্থক্য নেই। অন্তঃস্থবর্ণীয় ‘ব’-এর জন্য IAST -তে wa বা va লেখা সাব্যস্ত হয়েছে। এখানে প্রথমটি রাখা হ’ল।

৪ ত্রুটি সংশোধন

BangTeX-এ ক, ক্ষ, ঙ, চ, ব, ট, ড, ঢ, ফ এবং ভ — এই বর্ণগুলির ও য-ফলা যুক্ত বর্ণের উ-কার, উ-কার এবং ঞ-কার কিছুটা স্থানচ্যুত দেখায়। mkbangtex এই ত্রুটি মেরামত ক’রে দেয়। যেমন কুল, কুল, ক্ষুধার্ত, ঘুঙুর, চুল, বুলোবুলি, টুল, ডুমুর, ঢুলঢুলু, ফুলমালা, ভুলভাষ্টি, ভূত, ক্শ, ভ্গ্গার। ট-এ ও ক্ষ-এ ব-ফলার ত্রুটি BangTeX-এর ডকুমেন্টেশনে উল্লেখ করা হয়েছে। ট-এ যুগপৎ র-ফলা ও উ-কার থাকলেও সামান্য অবস্থানিক ত্রুটি হয়। এগুলিও মেরামত করা হ’ল: খট্টা, ইক্ষাকু, টুথ। পরিশেষে, BangTeX-এ ষ্ট ও ঠ লেখা হয় স্বচ্ছ চিহ্নে, কিন্তু ষ্ক, ষ্প, ষ্ফ অস্বচ্ছ চিহ্নে। mkbangtex প্রথম দু’টিকেও ‘অস্বচ্ছ’ ক’রে দেবে: যষ্টি, শৃঙ্ককাষ্ঠ, বাষ্প, নিষ্ফলা।

mkbangtex -এর --transparent option ব্যবহার ক’রে এই অংশের পরিবর্তিত চিহ্নগুলির পরিবর্তে BangTeX-এর default রীতি ব্যবহার করা যাবে।

৫ ancient অপ্শন

প্রাচীন বাঙলা লেখার জন্য এই option -টির ব্যবহার সীমিত। প্রাচীন বাঙলায় অন্তস্থ য-র ব্যবহার ছিল না। এই প্রাচীন রূপের ব্যবহার চর্চাপীতিতে দেখা যায়। উদাহরণস্বরূপ, এই রীতিতে কায় হই ‘কাআ’। mkbangtex --ancient option ব্যবহার করে এই প্রাচীন বাঙলা লেখা যায়। নীচের উদাহরণে --ancient option সহ ও তা ছাড়া mkbangtex চালালে এই পার্থক্য বোঝা যাবে:

```
kaayaa tarubara pancha bi Daal. \\  
chanchala chiiye paiTho kaal.. \\  
dirhha kari mahaasuha parimaaN. \\  
lui bhaNayi guru puchchhiya jaaN..\\  
sayala samaahiya kaahi kariyai. \\  
sukha dukhete~n nichita mariyai..\\  
erhi eu chhaandaka baandha karaNaka paaTera aas.\\  
sunupaakha bhiti lehu re paasa..\\  
bhaNai lui aameh_h jhaane diThaa.\\  
dhamaNa chamaNa beNi piNDi baiThaa..
```

৬ ব্যত্যয়

BangTeX-এ ‘অ্যা’ ছাপার জন্য AYa লিখলেই চলে। কিন্তু mkbangtex -এ AYa বা aYa লিখলে অ্যা, Aya বা aya লিখলে অয় ছাপা হয়। ফলে, এখন BangTeXও mkbangtex -এর যৌথ নিয়মে aYaa লিখতে হবে, aYaasiD: অ্যাসিড। পরন্তু, BangTeX-এ অ্যা-র ধনি ও তৎসংশ্লিষ্ট -কার লেখার জন্য নতুন চিহ্ন প্রস্তাবিত হয়েছে: AA: এ ও aa: ঃ। কিন্তু mkbangtex -এ AA=aa: আ। ফলতঃ, BangTeX-এর চিহ্নগুলির জন্য নতুন নিয়ম ধার্য হয়েছে। YA লিখলে ঐ চিহ্নগুলি ছাপা হবে: YAKaa kYAna ele? : একা কেন এলে? লক্ষ্যণীয় যে, এ এবং ঃ লেখার জন্য দু’টি পৃথক চিহ্ন লেখার প্রয়োজন নেই।

এছাড়া আমি একটি গোলমাল পেয়েছি, dhanyabaadaarha: ধন্যবাদাড়, ধন্যবাদার্ব লিখতে গিয়ে! এ’টি বেশ মজার। rh মানে ড, আবার হ-য়ে রেফ-ও। অন্য সব জায়গায় দ্বিতীয়টিই হবে, যেমন বর্ণ বা বর্ষ। কিন্তু, এই ক্ষেত্রে দু’টি নিয়ম clash করছে। এখানে BangTeX-এর নিয়ম মেনে প্রথমটিই default -এ রাখা হ’ল। ফলে, দ্বিতীয় ক্ষেত্রে BangTeXএর নিয়মমার্কিন H দিয়ে “ধন্যবাদার্ব” লিখতে হবে: dhanyabaadaarHa। অর্থাৎ, BangTeX-এর নিয়মাবলী বাদ দেওয়া যাবে না — সেইগুলিও Zান্টি হবে!

কৃতজ্ঞতা

mkbangtex প্রোগ্রামটি Python2.4.1 -এ পরীক্ষিত। এই প্রোগ্রামের একটি গুরুত্বপূর্ণ অংশ Xavier Defrang -এর Single-Pass Multiple Replace, Py Cookbook 81330 থেকে গৃহীত হয়েছে। এই প্রোগ্রামের ব্যাপারে অফুরন্ত সাহায্য করেছেন জয়দীপ মজুমদার।