1 CS/CSX Encoding

During the 8th World Sanskrit Conference, Vienna 1990, a panel was held to discuss the standardization of Sanskrit for electronic data transfer. Dominik Wujastyk presented a paper, titled "Standardization of Sanskrit for Electronic Data Transfer and Screen Representation," which outlines the Classical Sanskrit and Classical Sanskrit Extended Encoding. The original paper is available in a .zip archive, named "iass.zip". Use archie to locate FTP sites that store iass.zip, if you need the original paper, as of this writing, it was available at: `ftp.bcc.ac.uk:/pub/users/ucgadkw/indology`.

ITRANS now supports CS/CSX, in a limited form.

To turn on recognition of CS/CSX encoded text, use this command before the CS encoded text:
`#usecsx`

If necessary, you can turn off recognition of CS/CSX encoded text (and revert to ITRANS only encoding), by using this command:
`#ignorecsx`

Both `#usecsx` and `#ignorecsx` apply to the text that follows in the file.

1.1 Avoid Capitals or Uppercase Characters

CS/CSX is a case-insensitive encoding, both uppercase and lowercase characters have the same meaning, but ITRANS is case-sensitive. Thus, an uppercase character in ITRANS is encoded to a different devanagari character from its lowercase counterpart.

This means that when you create input text in CSX, do not use uppercase characters. ITRANS differentiates between lower case and uppercase characters. For example, entering `la` results in ल, but entering `La` results in a different character: ळ.

(If you already have input text that is in CS/CSX encoding and it also uses uppercase characters, use some text editor or word processor to convert all the uppercase characters to lowercase. This can be done easily using the text editors such as emacs, or programs such as `tr` (available on UNIX machines, and also available for MS-DOS systems).)

Table 2 provides the complete input encoding accepted by ITRANS. Note that technically, this uppercase character limitation applies only to the ASCII character codes (those less than 128). CS/CSX character codes over 128 are handled correctly in ITRANS (by correct, I mean handling them with CS/CSX meanings). But, it is simpler to just avoid using any uppercase characters at all. The tables in this document should clarify all this.

Table 1 lists the non-ASCII character codes (128 to 255) accepted by ITRANS.

1.2 Incomplete CS/CSX Support

There are many characters in CS/CSX that are currently not mapped by ITRANS, simply because I do not know what devanagari characters (if any) they stand for.

---

[0]Last modification: December 4, 2009

devanagari	cs/csx char	cs/csx char code
आ	ā	224
आ	Ā	226
ऊ	ū	229
ऊ	Ū	230
ऋ	ṛ	231
ऋ	Ṛ	232
ॠ	ṝ	233
ॠ	Ṝ	234
ऌ	ḷ	235
ऌ	Ḷ	236
ॡ	ḹ	237
ॡ	Ḹ	238
ङ	ṅ	239
ङ	Ṅ	240
ञ	ñ	164
ट	ṭ	241
ट	Ṭ	242
ठ	ṭh	241 + h
ठ	Ṭh	242 + h

devanagari	cs/csx char	cs/csx char code
ड	ḍ	243
ड	Ḍ	244
ढ	ḍh	243 + h
ढ	Ḍh	244 + h
ण	ṇ	245
ण	Ṇ	246
श	ś	247
श	Ś	248
ष	ṣ	249
ष	Ṣ	250
ఴ	ḻ	215
	ṃ	252
	Ṃ	253
	ṁ	167
:	ḥ	254
:	Ḥ	255
ఴ (telugu)	ṟ	159

Table 1: **CS/CSX Mapping** (shows non-ASCII char codes only).

For example, l̃ is present in the CS table, but I don't know what it represents, so, it is currently unrecognized by ITRANS. If you find such gaps and know what devanagari or tamil or telugu character it stands for, please send me e-mail, so that I can include it in the next release of ITRANS. My e-mail address is available elsewhere in this document. Feel free to send me mail on any other issues regarding CS/CSX support in ITRANS, too.

1.3   Breaking the lexical scan in ITRANS

ITRANS always matches the largest possible input sequence when scanning the input. Thus, in devanagari, `sha` will produce श even when usecsx is in effect. If instead you need ITRANS to recognize it as two different consonants, you should use the lexical scan break character, the underscore. Thus, typing `s_ha` will result in स्ह .

This is a thing to watch out for in all cases where some character has a multiple letter mapping, and each letter by itself also represents some other character. In the above example, `sh` is the two letter map, and both `s` and `h` represent other characters.

2   ITRANS Encoding Accepted along with CS/CSX

Even when `#usecsx` is in effect, the other ITRANS mappings are still accepted. Actually, even without the usecsx command, ITRANS will still accept all the CS/CSX character codes. The only reason the usecsx command is needed is to disambiguate the च and छ characters. ITRANS by

default uses `cha` and `chha` to denote च and छ , but CS/CSX uses `ca` and `cha`. To handle this difference, the `#usecsx` command had to be introduced.

Though this document shows only the Devanagari script, CS/CSX can be used in ITRANS with every Indic Script that ITRANS supports. For more details on those languages, consult the documentation for that specific language. For Devanagari, consult dnvc.itx or dvng.itx, for Tamil, tamil.itx, for Bengali, beng.itx, for Telugu tlgutx.itx, for Punjabi pundoc.itx, for Romanized Sanskrit romancsx.itx.

2.1 Converting CSX Encoding to ITRANS Encoding

Sometimes, it is useful to convert input text from the 8-bit CS/CSX encoding to the 7-bit ITRANS Encoding (ASCII text). Many computers still do not have good 8-bit editing tools, and even if such tools are present, the CS/CSX font may not be available, in which case editing CSX texts becomes cumbersome.

To convert input text from CS/CSX encoding to ITRANS encoding, the program **csx2i** has been provided. It is present in the `src/` directory of the ITRANS distribution. For x86 MS-DOS compatible machines, an executable is also provided. Given a file called `input.csx` which contains text in CS/CSX encoding, to convert it to ITRANS encoding in a file called `output`, run it like this:

```
csx2i < input.csx > output
```

C-source code to csx2i is provided, in file csx2i.c, and it should compile easily on every platform.

Once the file `output` is checked over for manual corrections, you can embed that text in between ITRANS commands, and process it like any other ITRANS input file.

2.2 Author

ITRANS has been developed by Avinash Chopde.
E-mail: `avinash@aczoom.com`.
Home page: `http://www.aczoom.com/`

Vowels

अ	a
आ	aa *or* A
इ	i
ई	ii *or* I
उ	u
ऊ	uu *or* U
ऋ	RRi *or* R^i
ॠ	RRI *or* R^I
ऌ	LLi *or* L^i
ॡ	LLI *or* L^I
ए	e
ऐ	ai
ओ	o
औ	au
अं	aM
अः	aH

Digits

०	0
१	1
२	2
३	3
४	4
५	5
६	6
७	7
८	8
९	9

Consonants

क	ka
ख	kh$a$
ग	ga
घ	gh$a$
ङ	~Na *or* N^a
च	ca
छ	ch$a$ *or* chh$a$
ज	ja
झ	jh$a$
ञ	~na *or* JNa
ट	Ta
ठ	Tha
ड	Da
ढ	Dha
ण	Na
त	ta
थ	th$a$
द	da
ध	dh$a$
न	na
प	pa
फ	ph$a$
ब	ba
भ	bh$a$
म	ma
य	ya
र	ra
ल	la
व	va *or* wa
श	sh$a$
ष	Sha *or* shh$a$
स	sa
ह	ha
ळ	ld$a$ *or* La
क्ष	kSha *or* xa
ज्ञ	j~na *or* GYa

Specials/Accents

क़	qa
ख़	Ka
ग़	Ga
ज़	Ja *or* za
फ़	fa
ड़	.Da
ढ़	.Dha
ॐ	AUM or OM
ऱ्ग	Rga
र्ग	rg$a$ *or* ga^r
गं	ga.n
ऑ	aa.c
ड़	$D a$.N
ड़	D.h
दुः	duH
ऽ	.a

Table 2: **ITRANS: Devanagari to English with #usecsx** (in addition to Table 1).