text in some specific font by underlining rules only (no text). For that purpose use a `.tfm` file like, for instance, `u-cmr10.tfm`. See the figure of this article for an example.

`dvi2dvi` also supports font emulation where output in one font is replaced by output in a different font when the document is printed. This capability is not shown in this article.

Concluding Remarks

I hope that I was able to demonstrate the usefulness of DFPs in general and `dvi2dvi` in particular. Note that I have *not* discussed all the features of `dvi2dvi`. A 60 page long document describing `dvi2dvi` contains the description of all features plus additional macros which should be useful in applications of `dvi2dvi`.

I hope that I can encourage people to buy my DFP (yes, it costs a little money), and to port it to other operating systems (give me a call in case you are interested). Contact me at the address below and I think we can work something out. `dvi2dvi` is written in "standard C" and runs currently on a SUN running OS 3.5 (BSD 4.2). There should be no problem to port it to other operating systems with a C compiler.

Finally I would like to thank Ron Whitney for his cooperation: he had to transfer the `.dvi` file for this article to my computer to process it by `dvi2dvi` and then back to the AMS's computer for printing, a little additional inconvenience.

◇ Stephan v. Bechtolsheim
2119 Old Oak Drive
W. Lafayette, IN 47906
317-463-0162
svb@cs.purdue.edu

Notes on Russian TEX

Dimitri Vulis

By combining the new Cyrillic fonts from the University of Washington in Seattle with my hyphenation patterns, I've been able to create a usable Russian-language version of TEX.

Coding Cyrillic letters

The customary way to represent Russian letters in an ASCII computer is to use 8-bit coding, with capital Russian letters A–Ya in 176–207, followed by lower case a-ya in 208–239. This scheme, commonly known as GOSTCII (pronounced GOST-ski), is formally defined by the standards ISO 8859 part 5 [2] and ECMA 113. I use GOSTCII to code Russian text on my personal computer.[1] I use this coding in my Russian TEX files, but a convenient way of entering transliterated Russian text using only 7-bit ASCII (unfortunately, different from the elegant AMS scheme that uses ligatures) is also available.

Hyphenation patterns

To create the patterns, I ran `PATGEN` on a dictionary of over 50,000 fully hyphenated Russian words with inflections. Remarkably, `PATGEN` found all the good breaks and no bad breaks, outputting 4204 patterns. I keyed in and hyphenated most of the dictionary by hand; some words were supplied by Alexander Samarin, for which I am grateful.

I started by keying in the Russian part of a pocket Russian-French dictionary, hyphenating the words manually. I then ran `PATGEN` to examine the patterns, and also tried them on Russian texts. I saw that the patterns did not handle inflected words well because I keyed in only the nominative/singular/masculine/infinitive (whichever are applicable) forms. Hence I inflected a number of words representative of different classes, and continued this practice when I added words later.

I also noted that a number of patterns were of the form

⟨*vowel*⟩ 1 ⟨*consonant*⟩ ⟨*vowel*⟩

Rather than seeking words containing all such combinations, I preloaded to `PATGEN` patterns of the form

[1] This can be achieved with any MS-DOS PC that supports code pages; the required software can be FTPed from `SIMTEL20.ARMY.MIL` as `PD1:<MSDOS.SCREEN>CYRILIC2.ARC`.

⟨*vowel*⟩ 1 ⟨*consonant*⟩ ⟨*vowel*⟩

for all vowels and all valid consonant-vowel combinations (i.e., no жя) and of the form

⟨*vowel*⟩ 1 ⟨*consonant*⟩ ь ⟨*vowel*⟩

for all vowels and all vowels that can follow ь, and also of the form

α 1 α where $\alpha \in$ ⟨*letter*⟩

for all the letters that can occur twice. In the few cases when these patterns would hyphenate dictionary words incorrectly, they are overridden by the PATGEN-generated patterns.

It was no longer necessary to add to the dictionary the words that exhibited no new patterns, like корова or дорога. I browsed through a number of dictionaries (including Ozhegov, Foreign (to Russian) Words, English-Russian Polytechnic, Mathematical, Geographical, Computer Science, Medical, Obscenities, Thieves' Jargon, and others) looking for words that exhibited new patterns (mostly unusual combinations of consonants) and added them to the dictionary, hyphenating by hand.

The rules of Russian hyphenation were first formulated by Yakov H. Grot [1] and then revised, and made less strict, in 1918 and 1956 [3; 5; 6]. Russians prefer to break their words at syllable boundaries: after a vowel and before a consonant or another vowel, except when this would result in a conspicuous-looking cluster of consonants at the beginning of the next line, as in ка-рман; in such cases they advance the break, taking care not to split certain consonant combinations, e.g., мест-ный. Derivation may take precedence over pronunciation when breaking off a prefix, e.g., вы-рвать, or when the word has been borrowed from a foreign language, e.g., дис-пе-псия. A simplified English rendering of the rules can be found in the US Government Printing Office Style Manual, the Chicago Manual of Style, and the like. The following list of words illustrates their application. (Note that some words have the last 2 letters broken off; TEX won't find these breaks.)

аб-зац-ный аб-сорб-ция адрес-ный айс-берг ал-ло-хтон-ный ан-глий-ский ар-хео-пте-рикс арк-функ-ция арт-об-стрел ас-фальт-ный асин-хрон-ный астро-навт ах-нуть без-дна бес-ком-про-мисс-ный блок-схе-ма бой-скаут-ский борт-про-вод-ни-ца бу-кварь бур-жуаз-ный бух-гал-тер взро-слый ви-део-уси-ли-тель во-жди во-сем-на-дцать воль-фрам вольт-метр вось-ми-раз-ряд-ный впол-ли-ствы все-гда все-общ-ность вы-жжен-ный вы-рвать глас-ные го-ло-во-тяп-ство го-мео-морф-ный го-мо-сек-су-а-ли-сты горш-ко-вый гос-цирк гра-мот-ный гро-мозд-кий гроз-дьях гросс-бух гуа-шью дву-языч-ный ден-знак дер-жать ди-влюсь диа-гно-сти-ка ду-плекс-ный жем-чуж-ный жуж-жать за-вши-веть за-мкну-тый за-мше-вый зав-лаб затх-лый злост-ный зоо-гео-гра-фия из-вест-ный из-да-тельств изо-ане-мо-на изы-скан-ный им-пло-зив-ный им-пульс-ных иму-ще-ство ин-верс-ный ин-дек-сный ин-клю-зив-ный ин-те-грал ин-тер-ак-тив-ный ин-фра-крас-ный ис-клю-чать ис-кро-уло-ви-тель-ный ис-тлеть их-тио-за-вры ка-мен-но-уголь-ный ка-пуст-ный каз-нью ква-дра-тич-ный квинт-эс-сен-ция ки-ло-ватт-метр класс-ный ко-манд-ный кол-хоз ком-пью-тер комс-орг конц-ла-герь кре-стьян-ский крест-цо-вый кри-пто-си-сте-ма крио-элек-трон-ный кросс-эму-ля-тор ку-плю ку-рьез-ный культ-про-свет-ра-бот-ник кунст-ка-ме-ра кур-орт ла-текс-ный ланд-шафт-ный ландс-кнех-тах ле-мнис-ка-той лег-ко-атле-ти-ка лек-се-ма лин-гви-сти-че-ский ло-га-риф-ми-че-ский ло-ги-че-ский ло-каль-ный лох-ма-тый луч-ший львом льня-ной ма-ну-скри-пты марк-сист-ский мас-штаб-ный мат-обес-пе-че-ние ме-тео-стан-ция ме-чта меж-атом-ный мест-ный микс-ту-ра мин-здрав мно-го-уголь-ник мо-крый мор-фем-ный на-взрыд на-гра-ждать на-из-усть на-име-но-ван-ный наи-выс-ший не-аде-кват-ный не-есте-ствен-ный не-льзя нем-цах нео-ло-гизм неф-тя-ной но-ябрь-ские обл-ис-пол-ком обо-льстить обо-рвыш общ-ность оглох-ший од-но-днев-ка ок-тябрь-скую орг-вы-во-ды осво-бо-жда-ет оскор-бле-ние осле-пли осу-ще-ствлял-ся от-мстить отра-вле-ние па-лео-био-гео-гра-фия па-три-ар-хат парт-ак-тив парт-съезд паст-би-ще пе-ре-жжешь при-льнуть при-мкнув-ший при-со-еди-нить про-бле-мой про-гно-зах про-грамм-ный про-образ про-цесс-ный про-чтут проф-вред-ность псев-до-ана-ли-ти-че-ский пя-ти-этаж-ный ра-дио-изо-топ ра-зы-гры-вать рав-ный разъ-ездной рай-он-ный рас-чет рас-ши-фро-вать ре-ко-гнос-ци-ров-ка ро-жде-ние ру-блей рявк-нуть са-мо-очист-ка сак-во-я-жик сап-фир сверх-опе-ра-тив-ный свое-образ-ный се-го-дняш-ний сем-на-дца-тый си-зи-гий-ный си-сте-мо-тех-ни-ка син-хро-сиг-нал сколь-зкий смеж-ник смерт-ность со-впа-де-ние со-еди-нить со-лжешь со-мна-бу-ла сов-ин-форм-бю-ро сот-ник соц-культ-быт спец-от-дел

срав-нить сред-ство сте-рео-угол стерж-нем су-ма-сшед-ший та-блич-ный те-ле-съем-ка те-тра-дью тек-сту-аль-ный тер-мо-ядер-ный тер-плю то-жде-ство транс-фи-нит-ную трех-адрес-ный три-фтонг уль-тра-основ-ной успеш-ный уст-ный устройств утвер-ждать уточ-нять фак-си-ми-ле фак-тор-ал-ге-бра фо-то-вспыш-ка фырк-нуть хаус-дор-фо-во хри-пло хряст-нуть ци-кло-и-да ци-фро-вых че-рес-чур че-ты-рех-уголь-ник чер-ствый честный чи-слен-ный чист-кой чув-ство шах-тер ши-фров-кой штрих-пунк-тир-ный эв-фе-мизм экзем-пляр элек-трон

The dictionary contains a large number of:

1. words borrowed from foreign languages (mostly technical and engineering terms); PATGEN is very good at understanding that one should break дур-шлаг, but марш-рут;
2. abbreviations (сложносокращенные слова), e.g., гос-за-каз, парт-учеба, комс-орг. These are not really part of the language, but occur often in newspapers and technical literature. The number of words usually abbreviated for such compounds is finite and the patterns are very good at identifying them. It is possible to construct abbreviations that these patterns will not hyphenate properly; when in doubt, one has to use \- in such words;
3. compound words. The patterns will not split a single vowel off a part of many compound words. Thus, прямоу-гольник, би-олог, нео-бычный, не-офашизм, etc., are suppressed. Such breaks are not strictly illegal, but they don't look good, and a better break is only a letter away. Once again, I took advantage of the fact that the number of words usually used in compounds is manageably small. A sufficient number of examples given to PATGEN eliminates the unwanted breaks in many compound words that are not in the dictionary.

A preliminary version of this work was presented in my M.A. thesis, "An Implementation of Liang's Algorithm for the Russian language", submitted to CCNY in October of 1988.

After the hyphenation patterns were complete, A. Samarin graciously sent me the paper [4] describing a non-TEX algorithm for hyphenating Russian text. Pavlova's algorithm produces the same hyphenations that I produced by hand, except for a few "special" words, like нововведение that it does not handle correctly; it breaks the words containing the letter combinations вн, сн, and x+consonant differently: ди-вный, ме-стный, ша-хта, which is unusual; and there are other minor differences.

In order to get PATGEN to run under MS-DOS, I had to concoct a very long .CH file, based on the Kellerman and Smith VAX change file. I will be happy to discuss it with anyone trying to get PATGEN to work. In order to trick PATGEN into processing GOSTCII input, I used the following:

```
xord[chr(208)]:="A";
xord[chr(209)]:="B";
xord[chr(210)]:="C";
 ...
xord[chr(231)]:="X";
xord[chr(232)]:="Y";
xord[chr(233)]:="Z";
xord[chr(234)]:="[";
xord[chr(235)]:="\";
xord[chr(236)]:="]";
xord[chr(237)]:="^";
xord[chr(238)]:="_";
xord[chr(239)]:="`";
 ...
xchr["A"]:=chr(208);
xchr["B"]:=chr(209);
xchr["C"]:=chr(210);
 ...
xchr["X"]:=chr(231);
xchr["Y"]:=chr(232);
xchr["Z"]:=chr(233);
xchr["["]:=chr(234);
xchr["\"]:=chr(235);
xchr["]"]:=chr(236);
xchr["^"]:=chr(237);
xchr["_"]:=chr(238);
xchr["`"]:=chr(239);
@z

@x
@d cmin="@@"
@d cmax="Z"
@d edge_of_word="@@"
@y
@d cmin="@@"
@d cmax="`"
@d edge_of_word="@@"
@z
```

Russian TEX

I used with my Russian TEX the Cyrillic fonts kindly mailed to me by Thomas Ridgeway, the director of the Humanities and Arts Computing Center at U. of Washington, Seattle. Their organization is similar to that of the Cyrillic fonts developed

by AMS in MF79. In particular, TEX's ligature mechanism is used to enter certain Russian letters. For example, to type the word жнец, one enters its MR transliteration `zhnets`. If the current font is Latin, the transliteration is printed out; and if the current font is Cyrillic, then the letters `zh`, taken as a ligature, produce character *'031*, which is ж in the font, while `ts`, as a ligature, produce ц. If the text is being set in Latin, then it produces its transliteration. A problem arises when a letter combination used for a ligature actually occurs in a word. For example, to enter the word отсев one has to type `ot{\cydot}sev`, where `{\cydot}` has to be defined as `kern0pt` for Cyrillic, to suppress the ligature, and as `$\cdot$` for Latin, to transliterate the word as "ot·sev". When hyphenation is desired, the explicit kern interferes with it. Moreover, one of the hyphenation patterns is `2t1s`, meaning that entering `otsenka` for оценка is liable to result in от-сенка being hyphenated. The fault, of course, lies with the transliteration scheme.

Thus, I had to abandon this elegant ligature scheme and to define the following control sequences to enter Russian letters that have no obvious Latin equivalents.

```
\chardef\Zh='021
\chardef\zh='031
\chardef\Ui='022
\chardef\ui='032
\chardef\Kh='110
\chardef\kh='150
\chardef\Ts='103
\chardef\ts='143
\chardef\Ch='121
\chardef\ch='161
\chardef\Sh='130
\chardef\sh='170
\chardef\Shch='127
\chardef\shch='167
\chardef\cDprime='137
\chardef\cdprime='177
\chardef\cPrime='136
\chardef\cprime='176
\chardef\Ee='003
\chardef\ee='013
\chardef\Yu='020
\chardef\yu='030
\chardef\Ya='027
\chardef\ya='037
```

I used `PLtoTF` and `TFtoPL`, TEXware programs, to delete all the ligatures in the Cyrillic fonts except those for quotes and dashes. The examples above would be entered as `{\zh}ne{\ts}`, `otsev` and `o{\ts}enka`. When Russian text is being transliterated, the control sequences need to be redefined:

```
\def\Zh{\t{Z}{h}}
\def\zh{\t{z}{h}}
\def\Ui{\u{I}}
\def\ui{\u{\i}}
\def\Kh{\t{K}{h}}
\def\kh{\t{k}{h}}
\def\Ts{\t{T}{s}}
\def\ts{\t{t}{s}}
\def\Ch{\t{C}{h}}
\def\ch{\t{c}{h}}
\def\Sh{\t{S}{h}}
\def\sh{\t{s}{h}}
\def\Shch{\t{S}{h}\t{c}{h}}
\def\shch{\t{s}{h}\t{c}{h}}
\def\cprime{$^\prime$}
\def\cPrime{$\underline{^\prime}$}
\def\cdprime{$^{\prime\prime}$}
\def\cDprime{$\underline{^{\prime\prime}}$}
\def\Ee{\'{E}}
\def\ee{\'{e}}
\def\Yu{\t{Y}{u}}
\def\yu{\t{y}{u}}
\def\Ya{\t{Y}{a}}
\def\ya{\t{y}{a}}
```

The control sequence `\cydot` is no longer needed and the tie accent indicates when a single Russian letter is transliterated by two Latin ones. The ability to change the transliteration scheme is an additional benefit:

```
\def\Zh{\v{Z}}
\def\zh{\v{z}}
\def\Ui{J}
\def\ui{j}
\def\Kh{Ch}
\def\kh{ch}
\def\Ts{C}
\def\ts{c}
\def\Ch{\v{C}}
\def\ch{\v{c}}
\def\Sh{\v{S}}
\def\sh{\v{s}}
\def\Shch{\v{S}\v{c}}
\def\shch{\v{s}\v{c}}
\def\cprime{\kern0pt'\kern0pt\relax}
\def\cPrime{\kern0pt'\kern0pt\relax}
\def\cdprime{\kern0pt'\kern0pt'\kern0pt\relax}
\def\cDprime{\kern0pt'\kern0pt'\kern0pt\relax}
\def\Ee{\'{E}}
\def\ee{\'{e}}
\def\Yu{\t{J}{u}}
\def\yu{\t{\j}{u}}
\def\Ya{\t{J}{a}}
\def\ya{\t{\j}{a}}
```

As with the ligature scheme, it is the user's responsibility to switch the meanings of control sequences together with the fonts.

There are 32 letters in the Russian alphabet, and only 26 in the English one. For this reason, before the `\patterns` command can be used in INITEX, it is necessary to extend the `\uccode` and `\lccode` tables for the letters ж, й, э, ю, ь, ъ, and я.

I use a short PASCAL program to translate GOSTCII letters into Latin letters and control sequences before it can be fed to TEX. The proposed version 3.0 of TEX will accept 8-bit input, making this preprocessor unnecessary; and its enhanced handling of ligatures and hyphenation may make it unnecessary to enter as many as 10 characters to produce a single Russian letter, when GOSTCII is not used. For Russian `\language`, these patterns should work with `\left` and `\righthyphenmin=2`.

There is a Bitnet mailing list dedicated to the discussion of the Russian TEX project. To subscribe, send the command

`SUB RUSTEX-L` ⟨*your name*⟩

to

Bitnet: `LISTSERV@UBVM`

To submit an article, mail it to

Bitnet: `RUSTEX-L@UBVM`

(Note that node `UBVM` on BITNET is the same as node `UBVM.CC.BUFFALO.EDU` on Internet.)

Acknowledgements

I would like to thank Barbara Beeton for the truly incredible amount of help, support, and warm encouragement, all rendered via Internet; Alexander Samarin for the voluminous advice and support, again rendered via Internet; Donald Knuth, both for his interest in this project, and for creating TEX in the first place. Last but not least, I would like to thank Frank Liang for his doctoral research; without him, we would still be breaking перес-тройка.

Bibliography

[1] Яков Грот = Yakov Grot. Русское Правописание: Руководство = Russian Orthography: a Guide.

[2] International Organization for Standardization, Standard 8859-5: 1989, Information Processing—8-bit Single-Byte Coded Graphic Character Sets—Part 5: Latin/Cyrillic Alphabet.

[3] В. Ф. Иванова = V. F. Ivanova. Современный русский язык = Modern Russian language. Moscow, Prosveshchenie, 1976.

[4] Ю. Г. Павлов, В. А. Павлова, А. П. Соколов = Yu. G. Pavlov, V. A. Pavlova, A. P. Sokolov. Алгоритм автоматизированного переноса русских слов = An algorithm for automatic hyphenation of Russian words, Institute of High Energy Physics, Preprint #83-72, Serpukhov, 1983.

[5] А. И. Кайдалова, И. К. Калинина = A. I. Kaĭdalova, I. K. Kalinina. Современная Русская Орфография = Modern Russian Orthography, p. 189. Moscow, Vysshaya Shkola, 1973.

[6] Д. Э. Розентал = D. É. Rozental, ed. Современный Русский Язык = Modern Russian language, v. 5, §100. Moscow, Vysshaya Shkola, 1986.

◇ Dimitri Vulis
Department of Mathematics
Graduate Center
City University of New York
33 West 42nd Street
New York, NY 10036-8099
Bitnet: `dlv@cunyvms1`

Hyphenation Exception Log

Barbara Beeton

This is the annual update of the list of words that TEX fails to hyphenate properly. The list last appeared in Volume 9, No. 3, starting on page 239. Everything listed there is repeated here. Owing to the length of the list, it has been subdivided into two parts: English words, and names and non-English words that occur in English texts.

This list is specific to the hyphenation patterns that appear in the original `hyphen.tex`, that is, the patterns for American English. In the future, if such information becomes available, exceptions to other patterns will also be listed. See below, "Hyphenation for languages other than English".

In the list below, the first column gives results from TEX's `\showhyphens{...}`; entries in the