

Boundary Character

Erkenntnisse, Möglichkeiten, Probleme

Philipp C. MÜLLER

DANTE-Frühjahrstagung 2024

Fragestunde/Motivation

Wer weiß, was der Boundary Character ist?

Wer hatte schon einmal damit zu tun?

Wer weiß, wie der Boundary Character funktioniert?

Was ist der Boundary Character?

Das Ursprungsproblem: In T_EX gibt es keine Leerzeichen!

Wenn T_EX den Quellcode interpretiert werden die Leerzeichen zu Abständen umgewandelt (nachdem die Makros ersetzt wurden).

TUGboat 10, No.3 (1989) oder DTK 1990, Heft 1:

Every consecutive string of ‘characters’ read by T_EX in horizontal mode (after macro expansion) can be called a ‘word’. (Technically we consider a ‘character’ in this definition to be either a character whose `\catcode` is a letter or otherchar, or a control sequence that has been `\let` equal to such a character, or a control sequence that has been defined by `\chardef`, or the construction `\char<number>`.)

Beispiel: `ein_ flie"stext.` → `|ein|↔|fließtext.|`

Um Anfangs- oder Endbuchstaben eines Wortes besonders zu behandeln, braucht es den Boundary Character (BC).

The new T_EX now imagines that there is an invisible “left boundary character” just before every such word, and an invisible “right boundary character” just after it. These boundary characters take effect if the font designer has specified ligatures and/or kerning between them and the adjacent letters. Thus, the first or last character of a word can now be made to change its shape automatically.

Wie funktioniert der Boundary Character?

Zwei Arten:

- ▶ linker BC für Initialbuchstaben, ohne eigenen Char Code
- ▶ rechter BC für Finalbuchstaben, mit eigenem Char Code

Beide BC sind in TFM-Datei ($\text{T}_{\text{E}}\text{X}$ font metric) definiert.

Mit `tftopl` → PL-Datei (property list) ...

Grau ist alle Theorie!

Beispiel I – Das Sigma

Aus dem Mathesatz bekannt:

`\Sigma` `\sigma` `\varsigma`

Σ σ ς

textgreek: Sigma am Wortende ist offen: $\sigma \rightarrow \varsigma$

Bsp.: $\iota\sigma\sigma\varsigma$

Umsetzung z.B. in „Greek Modern“ (`grmn1000.tfm`):

(BOUNDARYCHAR 0 375) = D 253 → rechter Boundary Character

(LIGTABLE

⋮

(LABEL C s) = D 115 = kleines Sigma → σ

(LIG/ 0 56 C c) = D 46 = Punkt $\sigma \rightarrow \varsigma$.

⋮

→ weitere Satzzeichen

(LIG/ 0 375 C c) = rechter BC $\sigma| \rightarrow \varsigma|$

⋮

(STOP)

Beispiel I – Das Sigma

Griechische Schrift:

σ ↔ ζ

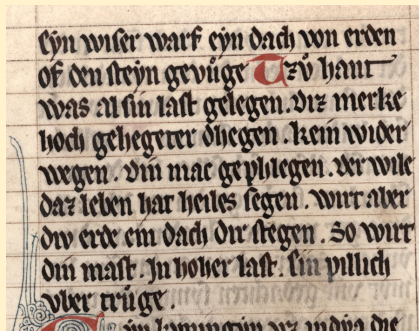
Gebrochene lateinische Schriften (Fraktur):

ſ ↔ ʒ

→ Aber meist noch mehr typographische Feinheiten, auch mit dem BC

Beispiel II

Jenaer Liederhandschrift von 1330 → jena1330-Paket (2012)

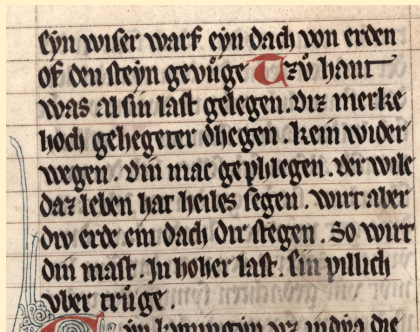


```
\textjena{
```

```
Eyn wiser warf eyn dach von erden  
of den steyn gev\r{u}ge Tz\r{v} hant  
was Al sin last gelegen. Diz merke  
hoch gehegeter dhegen. Kein wider-  
wegen. Din mac gephtegen. Der wile  
daz leben hat heiles seggen. wirt aber  
dir erde ein dach dir stegen. So wirt  
din mast. In hoher last. Ein pillich  
vber tr\r{u}ge.}
```


Beispiel II

Jenaer Liederhandschrift von 1330 → jena1330-Paket (2012)



Syn wiser warf eyn dach von erden
of den steyn gevügte **Vzv** hant
was al sin last gelegen. Dvz merke
hoch gehegeter dhegen. kein wider-
wegen. Din mac gephtegen. Der wile
daz leben hat heiles seggen. Wirt aber
dw erde ein dach dir stegen. So wirt
din mast. In hoher last. sin pillich
vber trüge.

Initial: v, w → **V/v** **W/w**
(linker BC)

Final: f, g, s, t → **f/f** **g/g** **s/s** **t/t**
(rechter BC)

Beispiel II

Auszüge aus der Jena1330.tfm:

(BOUNDARYCHAR 0 377)	= D 255	→ rechter BC (identisch mit ß)
(LIGTABLE		
(LABEL BOUNDARYCHAR)	= linker BC	
:		
(LIG C w 0 30)	= D 24 = Initial-w	→ ŵ (statt w)
(LIG C v 0 27)	= D 23 = Initial-v	→ ŵ (statt v)
(STOP)		
:		
(LABEL C s)	= D 115 = langes s	→ ſ
:	(≠ D 27 = rundes s	→ ſ)
(LIG/ 0 56 0 33)	= D 46 = Punkt	ſ. → s.
:	→ weitere Satzzeichen	
(LIG/ 0 377 0 33)	= rechter BC = ß	ſ → s
(STOP)		ſß → sß

→ Wahl des Char Codes kann Probleme machen!

Beispiel III

Optischer Randausgleich

Johann Wolfgang von Goethe:
„Die Liebe zum Wunderbaren
gehört eigentlich dem Poeten
und die Lust Schwierigkeiten
zu überwinden dem Mathema-
tiker.“ — Brief vom 9.6.1814

Johann Wolfgang von Goethe:
„Die Liebe zum Wunderbaren
gehört eigentlich dem Poeten
und die Lust Schwierigkeiten
zu überwinden dem Mathema-
tiker.“ — Brief vom 9.6.1814

Zeichen mit viel Weißraum werden in den Rand geschoben.

Wie kann man das umsetzen? → microtype-Paket

Character Protrusion → je Zeichen: links und rechts %o-Wert der Breite

Oder besser in der Schriftart selbst? → negatives Kerning mit BC

Beispiel III

Umsetzung in TFM-Datei:

(BOUNDARYCHAR 0 40)	= D 32 → rechter BC
(LIGTABLE	
(LABEL BOUNDARYCHAR)	= linker BC
(KRN 0 22 R -0.1)	= D 18 = „(\glqq) ¼ em nach links
:	
(STOP)	
:	
(LABEL 0 55)	= D 45 = Bindestrich
(KRN 0 40 R -0.125)	= rechter BC ⅛ em nach rechts
:	
(STOP)	

Anmerkung: Schiebung nicht nur in den Rand, auch in Leerzeichen.

→ Hier OK, aber „hängende Interpunktion“ geht damit nicht!

Beispiel III

Auffälligkeit: Bindestrich (D 45) = Trennstrich (`\hyphenchar`)

	←3 em→	
zu trennen	zu tren-	✓
	nen	
zu tren-nen	zu tren-	!
	nen	
zu tren-\discretionary{}{}{}nen	zu tren-	?
	nen	
zu tren-nen (mit pos. kerning zw. - und n)	zu tren-	!!
	nen	

→ Kerning mit `\hyphenchar` geht nicht!

Beispiel III

Alternative: Anderes Zeichen als \hyphenchar
(Bindestrich ≠ Trennstrich)

Z.B. Zeichen 127 (O 177) im Cork encoding (T1)

'177 is the hyphen character (that may be different from the dash ('055)). '202, '210, '242, and '250 are A's and E's with ogonek accents.

TUGboat 11, No.4 (1990):

	'0	'1	'2	'3	'4	'5	'6	'7	
'00x	`	´	ˆ	˜	¨	˘	◊	˙	"0x
'01x	˘	˙	˚	˛	˜	˝	<	>	
'02x	“	”	„	«	»	-	—	<cwmm>	"1x
'03x	o	ı	j	ff	fi	fl	ffi	ffl	
'04x	□	!	"	#	\$	%	&	'	"2x
'05x	()	*	+	,	-	.	/	
'06x	0	1	2	3	4	5	6	7	"3x
'07x	8	9	:	;	<	=	>	?	
'10x	@	A	B	C	D	E	F	G	"4x
'11x	H	I	J	K	L	M	N	O	
'12x	P	Q	R	S	T	U	V	W	"5x
'13x	X	Y	Z	[\]	ˆ	-	
'14x	‘	a	b	c	d	e	f	g	"6x
'15x	h	i	j	k	l	m	n	o	
'16x	p	q	r	s	t	u	v	w	"7x
'17x	x	y	z	{		}	˘	(hyph.char) -	
'00x	ł	ł	ć	ć	ń	ń	ę	ę	

Beispiel III

In Latin Modern zum Teil schon angelegt (ec-lmr10.tfm):

(LIGTABLE

:

(LABEL 0 55) = D 45 = Bindestrich (dash)


(LIG 0 55 0 25) zwei Bindestriche → Gedankenstrich

(LIG 0 177 0 177) = D 127 Bindestrich vor „hyphen“ entfällt!

(STOP)

:

D.h. mit `\hyphenchar=127` wird bei Trennung am Bindestrich ein Trennstrich eingefügt und der Bindestrich entfernt.

D 45:  D 127: 

Beispiel III

Ohne Änderung (`\hyphenchar\font=45`):

	$\leftarrow 3\text{ em} \rightarrow$	
Zu trennen	Zu tren-	✓
	nen	
Zu tren-nen	Zu tren-	!
	nen	

Beispiel III

Nur mit `\hyphenchar\font=127`:

Zu trennen	$\leftarrow 3\text{ em} \rightarrow$	
	Zu tren-	✓
	nen	
Zu tren-nen	Zu tren-	✓
	nen	

Beispiel III

In Latin Modern zum Teil schon angelegt (ec-lmr10.tfm):

(LIGTABLE

:

(LABEL 0 55) = D 45 = Bindestrich (dash)

(LIG 0 55 0 25) zwei Bindestriche → Gedankenstrich

(LIG 0 177 0 177) = D 127 Bindestrich vor „hyphen“ entfällt!

(STOP)

(LABEL 0 25)

(LIG 0 55 0 26) ← (LIG 0 177 0 25) fehlt noch

(STOP)

:

← (LABEL 0 26)

) (LIG 0 177 0 26) fehlt noch

+ Kerning mit dem Boundary Character

→ optischer Randausgleich fehlerfrei

Schlussfolgerungen

Was habe ich gelernt?

- ▶ zwei BC (links und rechts)
- ▶ technisch unterschiedlich umgesetzt!
- ▶ nützlich
- ▶ „Mach kein Kerning mit `\hyphenchar!`“
- ▶ optische Randausgleich mit BC möglich, wenn `\hyphenchar` ein zusätzliches Zeichen ist.

Weitere Frage: Wie außerhalb T_EX?

- ▶ Optischer Randausgleich im Satzprogramm (z.B. InDesign)
- ▶ In OpenType? → Feature `opbd` (optical boundary) ist **obsolet**, `lfbd` und `rtbd` sind spezifiziert, aber kaum implementiert!
→ Umsetzung in (lua)L^AT_EX steht noch aus.