

dehyph-expt1*

Experimentelle Trennmuster für die deutsche Sprache

Die deutschsprachige Trennmustermannschaft

24. Juni 2009

Abstract

This package provides experimental hyphenation patterns for the German language, covering traditional and reformed orthography for several varieties of Standard German. The patterns can be used with packages `Babel` and `hyphsubst` from the `OBERDIEK BUNDLE`. Project-URL is <http://groups.google.de/group/trennmuster-opensource>.

Zusammenfassung

Dieses Paket enthält experimentelle Trennmuster für die deutsche Sprache. Die Trennmuster decken das in Deutschland, Österreich und der Schweiz gebräuchliche Standarddeutsch in der traditionellen und reformierten Rechtschreibung ab und können mit den Paketen `Babel` und `hyphsubst` aus dem `OBERDIEK-BÜNDEL` verwendet werden.

Inhaltsverzeichnis			
		3.2. Traditionelle Rechtschreibung in der Schweiz	9
1. Einleitung	2	3.3. Reformierte Rechtschreibung . .	9
2. Verwenden der Trennmuster	3	4. Fehltrennungen	11
2.1. Sprachvarietät und Rechtschrei- bung	4	A. Dateien und Installation	13
2.2. Aktivieren der Trennmuster . . .	5	B. Die Wortliste	13
3. Trennregeln und Konventionen	6	C. Fragen & Antworten	14
3.1. Traditionelle Rechtschreibung in Deutschland und Österreich .	7		

*This document describes the `dehyph-expt1` package vo.20.

1. Einleitung

Der in \TeX implementierte Trennalgorithmus arbeitet musterbasiert [Lia83]. Prinzipiell können mit einem solchen Algorithmus nicht alle möglichen Wörter korrekt getrennt werden. Die Qualität der Worttrennung einer Sprache wird jedoch maßgeblich von der Qualität der Wortliste beeinflusst, aus der die verwendeten Trennmuster berechnet wurden.

Obwohl die herkömmlichen Trennmuster¹ für die deutsche Sprache bei der Worttrennung in gewöhnlichen Texten eine akzeptable Fehlerrate erreichen, enthalten sie doch eine Reihe von Schwächen:

traditionelle Rechtschreibung

- Die herkömmlichen Trennmuster für die traditionelle deutsche Rechtschreibung können mit `PATGEN` nicht reproduziert werden, da die zugrundeliegende Wortliste verschollen ist. Die Pflege der Trennmuster ist daher schwierig bis unmöglich. Für freie Software ist dies kein zufriedenstellender Zustand.
- Umfang und Qualität der ursprünglichen Wortliste lassen sich nicht mehr einschätzen. Für die Trennmuster in traditioneller Rechtschreibung existiert jedoch inzwischen eine Ausnahmeliste² mit über 3500 korrigierten Trennungen einfacher Wörter [Lem03, Lem05].
- Wird in der traditionellen Rechtschreibung β durch *ss/SS* oder *sz/SZ*³ ersetzt, so bleibt die Trennung davon unberührt. Die herkömmlichen Trennmuster berücksichtigen diese Regel nicht und trennen häufig den Ersatz: »GRÖS-SE«, »GRÜS-SE«, »MAS-SES«. ⁴
- Abweichende Schreibweisen, die in der traditionellen Rechtschreibung in Österreich und der Schweiz verwendet werden, werden mangelhaft getrennt: »Gros-son-kel«, »Kus-shand«, »Massnah-me«, »mis-sach-ten« (nur Schweiz) und »Ex-presszug«, »Fit-nes-scen-ter«, »Fit-nesstrai-ner«.

¹Diese Liste bezieht sich auf die Trennmusterdateien `dehyph.t.tex`, Version 3.2a vom 3. 3. 1999, und `dehyphn.tex`, Version 31 vom 7. 5. 2001.

²`CTAN:language/hyphenation/dehyph/dehyph.tex`

³Im Versalsatz sollte auf Trennungen jedoch möglichst ganz verzichtet werden.

⁴Das Verhalten entspricht den Regeln der deutschen Standardsprache in der Schweiz, obwohl diese Sprachvarietät vom Paket `Babel` nicht offiziell unterstützt wird. `Babel` versucht den Mangel mit Hilfe des Kürzels "S zu kompensieren.

reformierte Rechtschreibung

- Die Trennmuster für die reformierte deutsche Rechtschreibung wurden nicht mit PATGEN aus einer Wortliste erstellt. Stattdessen wurden die Trennmuster für die traditionelle Rechtschreibung von Hand an die reformierten Regeln angepasst. Aus diesem Grund ist die Worttrennung mit den Trennmustern für die reformierte Rechtschreibung etwas schlechter als mit den Trennmustern für die traditionelle Rechtschreibung.

traditionelle und reformierte Rechtschreibung

- In zusammengesetzten Wörtern treten häufig Trennfehler an Wortfugen auf.
- Fremdwörter mit akzentuierten Buchstaben werden mangelhaft getrennt: »Café«, »Ci-troën«, »Façon«, »voilà«.
- Die Trennmusterdateien enthalten eine Mischung aus T1- sowie unvollständigen OT1-kodierten Mustern. Mit Erscheinen von 16-Bit-fähigen T_EX-Varianten werden sauber UTF-8-kodierte Trennmuster nötig [MRo8].

Das Projekt *Freie Wortlisten und Trennmuster für die deutsche Sprache* hat sich das Ziel gesetzt, neue Trennmuster hoher Qualität für die deutsche Sprache zu erstellen, die die genannten Probleme ausräumen.

Den experimentellen Trennmustern dieses Pakets liegt eine Wortliste mit den etwa fünfhunderttausend häufigsten Wörtern der deutschen Sprache zugrunde. Vermutlich ist diese Liste erheblich umfangreicher als die ursprüngliche Wortliste, in der Worthäufigkeiten wahrscheinlich überhaupt nicht berücksichtigt wurden. Die verwendete Wortliste deckt das in Deutschland, Österreich und der Schweiz gebräuchliche Standarddeutsch ab.

Mit den vorliegenden Trennmustern sollte für nicht-fachsprachliche Wörter eine sehr gute Trennqualität erreicht werden. Insbesondere sollte sich die Trennung häufig auftretender zusammengesetzter Wörter verbessern.

2. Verwenden der Trennmuster

Zur Installation der experimentellen Trennmuster siehe [Anhang A](#) und die Datei INSTALL. Beachte, in den folgenden Abschnitten ist <datum> durch das bei der Installation angegebene Datum in ISO-Notation (JJJJ-MM-TT) oder die Zeichenkette latest zu ersetzen!

Tabelle 1: Die unterstützten Varietäten und Rechtschreibungen

Trennmusterbezeichner	Sprachvarietät
german-x-<datum>	<i>traditionelle Rechtschreibung</i> Deutschland, Österreich
gswiss-x-<datum>	<i>traditionelle Rechtschreibung</i> Schweiz
ngerman-x-<datum>	<i>reformierte Rechtschreibung</i> Deutschland, Österreich, Schweiz

2.1. Sprachvarietät und Rechtschreibung

Dieses Paket stellt Trennmuster für die Worttrennung der deutschen Sprache in der traditionellen und der reformierten Rechtschreibung bereit. Die Trennmuster unterstützen zur Zeit die drei in

- Deutschland,
- Österreich und der
- Schweiz⁵

gebräuchlichen Hauptvarietäten der deutschen Standardsprache. [Tabelle 1](#) zeigt die Trennmuster, die abhängig von gewünschter Varietät und Rechtschreibung in einem Dokument aktiviert werden können.

Varietäten, die sich nur in der Verwendung und Schreibung einzelner Wörter voneinander unterscheiden, können durch gemeinsame Trennmuster unterstützt werden. Die Eingabewortliste für PATGEN ist dann eine Vereinigung der den Varietäten entsprechenden Wortlisten. Zum Beispiel werden Besonderheiten der österreichischen Standardsprache in den Trennmustern für die Standardsprache Deutschlands berücksichtigt. Österreichische und deutsche Anwender können daher dieselben Trennmuster verwenden.⁶

Aufgrund unvereinbarer Trennregeln in der traditionellen Rechtschreibung der Standardsprachen Deutschlands/Österreichs und der Schweiz werden für die letztere Varietät eigene Trennmuster bereitgestellt.

⁵Für Texte in Schweizer Standarddeutsch fehlt zur Zeit die Babel-Unterstützung. Zum Beispiel muss das Kürzel "s stets zu ss expandieren. In der traditionellen Rechtschreibung wird außerdem ein zusätzliches Kürzel "ss benötigt (Dreikonsonantenregel). Wer kann helfen?

⁶Wie das auch schon bei den herkömmlichen Trennmustern der Fall war.

Mit der Rechtschreibreform 1996 wurden die Trennregeln aller drei Standardsprachen so weit angeglichen, dass für die reformierte Rechtschreibung einheitliche Trennmuster für alle drei unterstützten Sprachvarietäten bereitgestellt werden können.

2.2. Aktivieren der Trennmuster

Das folgende Beispiel zeigt eine L^AT_EX-Präambel für die Aktivierung der experimentellen Trennmuster für die reformierte Rechtschreibung.

```
\RequirePackage[ngerman=ngerman-x-<datum>]{hyphsubst}
% \RequirePackage[ngerman=ngerman-x-latest]{hyphsubst}
\documentclass{article}
\usepackage[T1]{fontenc}
\usepackage[ngerman]{babel}
```

Die folgende Variante erleichtert das schnelle Umschalten zwischen verschiedenen Trennmustern im Editor. Weitere Hinweise können der Dokumentation des Pakets `hyphsubst` entnommen werden.

```
\RequirePackage{hyphsubst}
\documentclass{article}
\usepackage[T1]{fontenc}
% \HyphSubstLet{german}{german-x-<datum>}
% \usepackage[german]{babel}
\HyphSubstLet{ngerman}{ngerman-x-<datum>}
\usepackage[ngerman]{babel}
```

Ob die experimentellen Trennmuster korrekt aktiviert werden, kann mit dem folgenden Beispiel getestet werden. Die Ausgabe für die traditionelle und reformierte Rechtschreibung mit herkömmlichen und experimentellen Trennmustern ist in [Tabelle 2](#) zusammengefasst.

```
\begin{document}
\showhyphens{löste Fassade modernste Abendstern Mordopfer}
```

Diese Trennmuster befinden sich im experimentellen Status. Sie können jederzeit durch umbruchinkompatible Versionen ersetzt und vom CTAN oder aus T_EX-Verteilungen entfernt werden. Sie sind daher nicht für Anwendungen geeignet, die einen dauerhaft stabilen Umbruch erfordern. Ausgiebige Tests sind erwünscht!

Tabelle 2: Trennvarianten

<i>traditionelle Rechtschreibung</i>		<i>reformierte Rechtschreibung</i>	
herkömmlich	experimentell	herkömmlich	experimentell
<u>lös</u> -te	lö-ste	lös-te	lös-te
Fas-sa-de	Fas-sa-de	Fassa-de	Fas-sa-de
mo-d-ern-ste	mo-dern-ste	mo-d-erns-te	mo-derns-te
Abend-stern	Abend-stern	Abend <u>s</u> -tern	Abend-stern
Mor-dop-fer	Mord-op-fer	Mor-dop-fer	Mord-op-fer

3. Trennregeln und Konventionen

Die Trennmuster für die traditionelle Rechtschreibung in Deutschland und Österreich orientieren sich an den verbindlichen Regeln des Dudens in der Fassung von 1991 [Wis91]. Dasselbe gilt für die Trennmuster für die traditionelle Rechtschreibung in der Schweiz, jedoch mit einer unten beschriebenen Abweichung. Die Trennmuster für die reformierte Rechtschreibung orientieren sich an den amtlichen Regeln für die Rechtschreibung der deutschen Sprache in der Fassung von 2006 [Rato6, Wiso6].

Die Regeln lassen gewisse Freiheiten bei der Schreibung und Trennung von Wörtern zu. Da sich solche Freiheiten nicht ohne weiteres auf die maschinelle Worttrennung übertragen lassen, wurden die im folgenden beschriebenen Konventionen getroffen. Hauptsächlich betreffen diese die reformierte Rechtschreibung, die zusätzliche Freiheiten eingeführt hat. Beziehen sich die Konventionen für die reformierte Rechtschreibung auf die traditionelle Rechtschreibung, so werden die entsprechenden Regeln etwas ausführlicher dargestellt. Die folgenden Abschnitte enthalten jedoch keine vollständige Aufstellung der Silbentrennregeln. Diese sind den entsprechenden Regelwerken zu entnehmen. Es folgen zunächst einige allgemeine Hinweise:

- In Liangs Trennalgorithmus werden Groß- und Kleinschreibung nicht unterschieden [Lia83]. Die Schreibweisen *Nachtritt* und *nachtritt* werden aus Sicht des Trennalgorithmus gleich behandelt (siehe auch [Regel T3](#) und [Regel R3](#)).
- Konsonantenänderungen, die bei der Trennung nach traditioneller Rechtschreibung auftreten, können nicht in Trennmustern berücksichtigt werden, sondern erfordern derzeit zusätzliche Auszeichnung im Quelltext

Tabelle 3: Konsonantenänderungen in der traditionellen Rechtschreibung

Eingabe	Trennung
drucken	drucken
dru"cken	druk-ken
Metallegierung	Me-talle-gie-rung
Meta"llegierung	Me-tall-legierung

mit Hilfe des Pakets Babel. [Tabelle 3](#) stellt Eingaben und deren möglichen Trennungen gegenüber. Beachte, im letzten Beispiel sind die Trennstellen im zweiten Wortteil verloren gegangen. Es handelt sich dabei um einen bekannten Fehler im Paket Babel.⁷

- Die von T_EX gewählte Trennung kann in Einzelfällen mit den T_EX- und Babel-Kürzeln \- und "- geändert werden. Für dokumentweite Änderungen der Trennung eignet sich das Kommando \hyphenation.
- Die Datei CHANGES beschreibt bekannte, systematische Fehler der Trennmuster.
- In den Beispielen zeigt die linke (grüne) Spalte jeweils die Trennung mit den experimentellen Trennmustern, die rechten (roten) Spalten zeigen alternative oder unerwünschte Trennungen.

3.1. Traditionelle Rechtschreibung in Deutschland und Österreich

T1 Sinnentstellende und irreführende Trennungen werden möglichst vermieden [[Wis91](#), R 181]:

An-alpha-bet	Anal-phabet
Kaf-ka-kenner	Kafkaken-ner
Tal-entwäs-se-rung	Talent-wässerung

T2 Wird der Buchstabe β durch *ss/SS* ersetzt, so bleibt die Trennung davon

⁷<http://www.latex-project.org/>, Kennzeichen babel/3458

unberührt [[Wis91](#), R 179]:

GRÖSS-TE
GRÜ-SSE
MA-SSES

Für diese Regel gilt die folgende Einschränkung:

- Wenn durch den Ersatz von β eine Mehrdeutigkeit auftritt, so wird zugunsten der Bedeutung des Wortes in der normalen Schreibweise getrennt (siehe auch [Regel T3](#)).

FLÖS-SE (wegen flös-se) FLÖ-SSE
MAS-SE (wegen Mas-se) MA-SSE

Beachte, wird β jedoch mit `\MakeUppercase` durch `SS` ersetzt, so bleibt `SS` stets ungetrennt. Die Trennung richtet sich dann nach der Schreibweise mit β im Quelldokument.

T3 In mehrdeutigen Wörtern werden Trennungen nur an übereinstimmenden Trennstellen zugelassen.

Druckerzeug-nis Druck-erzeugnis Drucker-zeugnis
Fuß-balleh-re Fußball-ehre
Mu-sikerle-ben Musi-kerleben Musik-erleben Musiker-leben
nachtritt nach-tritt Nacht-ritt
Wachstu-be Wach-stube Wachs-tube

Beachte, die Trennstellen »Drucker-zeugnis« und »Musiker-leben« sind in den Interpretationen *Druck-Erzeugnis* und *Musik-Erleben* irreführend. Sie entfallen nach [Regel T1](#) und sind nicht als übereinstimmende Trennstellen anzusehen.

Für diese Regel gelten die folgenden Einschränkungen:

- Bei mehrdeutigen Wörtern endend auf *-ende*, *-enden*, *-endes* wird stets die Trennung der Partizipform des Verbs verwendet.
- Mehrdeutigkeiten, die durch die Ersatzschreibweise von Wörtern mit β auftreten, werden nicht berücksichtigt (siehe auch [Regel T2](#)).

spie-len-de Spiel-ende spielende
Mas-se MA-SSE MASSE

*T*₄ In Ableitungen von Namen auf *-ow* wird die Nottrennung der Ableitungssilben *-er*, *-ern*, *-ers* unterdrückt [[Wis91](#), R 180]:

Tel-tower Teltow-er
Trep-towern Treptow-ern
Pan-kowers Pankow-ers

*T*₅ Die minimal unterstützte Silbenlänge am Wortanfang und -ende beträgt zwei Buchstaben.

Beachte, die Mindestlänge von Silben lässt sich mit `\germanhyphenmins` (Babel) oder `\lefthyphenmin` und `\righthyphenmin` (T_EX) anpassen. Wird die Mindestlänge auf weniger als zwei Buchstaben verringert, so können fehlerhafte Trennungen auftreten.

3.2. *Traditionelle Rechtschreibung in der Schweiz*

*TS*₁ Abweichend von [Regel T₂](#) wird *ss/SS* stets als Doppelkonsonant behandelt und gegebenenfalls getrennt:

grös-ste
Grüs-se
Mas-ses

3.3. *Reformierte Rechtschreibung*

*R*₁ Sinnentstellende und irreführende Trennungen werden möglichst vermieden [[Rato6](#), [Wiso6](#), § 107]:

An-alpha-bet Anal-phabet
Kaf-ka-kenner Kafkaken-ner
Tal-entwäs-se-rung Talent-wässerung

*R*₂ Wird der Buchstabe *ß* durch *ss/SS* ersetzt, so wird *s-s* getrennt [[Rato6](#), [Wiso6](#), §§ 25 E₃, 110]:

GRÖS-STE
GRÜS-SE
MAS-SES

Beachte, wird *ß* mit `\MakeUppercase` oder in Kapitälchen in *SS* gewandelt, so bleibt *SS* ungetrennt. Dies ist kein Fehler in den Trennmustern, sondern im L^AT_EX-Kern fest implementiert.

R3 In mehrdeutigen Wörtern werden Trennungen nur an übereinstimmenden Trennstellen zugelassen.

Druckerzeug-nis	Dru-ckerzeugnis	Druck-erzeugnis	Drucker-zeugnis
Mu-sikerle-ben	Musi-kerleben	Musik-erleben	Musiker-leben
nachtritt	nach-tritt	Nacht-ritt	
Wachstu-be	Wach-stube	Wachs-tube	

Beachte, die Trennstellen »Drucker-zeugnis« und »Musiker-leben« sind in den Interpretationen *Druck-Erzeugnis* und *Musik-Erleben* irreführend. Sie entfallen nach [Regel R1](#) und sind nicht als übereinstimmende Trennstellen anzusehen.

Für diese Regel gilt die folgende Einschränkung:

- Bei mehrdeutigen Wörtern endend auf *-ende*, *-enden*, *-endes* wird stets die Trennung der Partizipform des Verbs verwendet.

spie-len-de Spiel-ende spielende

R4 Falls die Trennung nach Sprechsilben und die etymologische Trennung (nach Wortherkunft) kollidieren, wird weitgehend die etymologische Trennung gewählt [[Rato6](#), [Wiso6](#), § 113]:

in-ter-es-sant	inte-ressant	
Lin-ole-um	Li-noleum	Lino-leum
Päd-ago-ge	Pä-dagoge	Päda-goge

R5 In Fremdwörtern bleiben die Buchstabengruppen *bl*, *pl*, *fl*, *gl*, *cl*, *kl*, *phl*; *br*, *pr*, *dr*, *tr*, *fr*, *vr*, *gr*, *cr*, *kr*, *phr*, *thr*; *chth*; *gn*, *kn* im allgemeinen ungetrennt, nicht jedoch *str* [[Rato6](#), [Wiso6](#), § 112] i. V. m. [[Wis91](#), R 179]:

Ar-thri-tis	Arth-ritis
Di-plom	Dip-lom
igno-rie-re	ig-noriere
In-te-gral	Integ-ral

aber:

In-dus-trie	Indu-strie	Indust-rie
de-struk-tiv		
sub-lim		

R6 Die minimal unterstützte Silbenlänge am Wortanfang und *-ende* beträgt zwei Buchstaben.

Beachte, die Mindestlänge von Silben lässt sich mit `\ngermanhyphenmins` (Babel) oder `\lefthyphenmin` und `\righthyphenmin` (T_EX) anpassen. Wird die Mindestlänge auf weniger als zwei Buchstaben verringert, so können fehlerhafte Trennungen auftreten.

4. Fehltrennungen

Mit den vorliegenden Trennmustern können sämtliche Wörter der Wortliste fehlerfrei getrennt werden. Für fehlerhafte (falsche, ausgelassene oder unerwünschte) Trennungen gibt es zwei mögliche Ursachen:

1. Die zugrundeliegende Wortliste enthält einen Fehler.
2. Das betreffende Wort ist in der zugrundeliegenden Wortliste nicht enthalten.

Da der Umfang der Wortliste nicht beliebig erweitert werden kann, sollten Fehltrennungen nur dann gemeldet werden, wenn eines der folgenden Kriterien erfüllt ist:

- A. Das betreffende Wort wird mit den herkömmlichen Trennmustern für die traditionelle oder reformierte Rechtschreibung korrekt getrennt. Korrekt bedeutet hier: Nicht alle möglichen Trennstellen müssen erkannt werden; es werden jedoch in keinem Fall falsche Trennstellen ermittelt.

Zum Testen kann der folgende Aufruf verwendet werden (die Ausgabe erfolgt in der LOG-Datei):

```
\showhyphens{durch Leerzeichen getrennte Wörter}
```

- B. Es handelt sich um eine sinnentstellende oder irreführende Trennung eines Wortes, das nicht aus mehr als zwei prä- und suffigierten Wörtern zusammengesetzt ist, zum Beispiel »Talent-wässerung«. Nicht berücksichtigt wird hingegen die »Talent-wässerungsanlage«.
- C. Das Wort ist bereits in der Wortliste enthalten (siehe [Anhang B](#)) und [Punkt 1](#) trifft zu.

Einige bekannte Fehler in den Trennmustern sind in der Datei CHANGES verzeichnet. Noch nicht bekannte falsche, fehlende und unerwünschte Worttrennungen können an die folgenden E-Mail-Adressen gerichtet werden:

- trennmuster-opensource@googlegroups.com (Anmeldung erforderlich),
- wl@gnu.org (Werner Lemberg).

Fehltrennungen, die in den Trennmustern nicht korrigiert werden können, können mit Hilfe einer privaten Ausnahmeliste behandelt werden:

```
\hyphenation{Tal-entwäs-se-rungs-an-la-ge Kaf-ka-kenner-klub}
```

Aktuelle Trennmuster sind im Dateibereich unter der Projekt-URL⁸ oder im CTAN erhältlich.

Happy T_EXing!

Die deutschsprachige Trennmustermannschaft

Literatur

- [Lem03] Lemberg, Werner: *Hyphenation Exception Log für deutsche Trennmuster*. Die T_EXnische Komödie, 15(2):28–31, Mai 2003.
- [Lem05] Lemberg, Werner: *Hyphenation Exception Log für deutsche Trennmuster, Version 1*. Die T_EXnische Komödie, 17(2):24–51, Mai 2005.
- [Lia83] Liang, Franklin Mark: *Word Hy-phen-a-tion by Com-put-er*. Dissertation, Stanford University, 1983. <http://www.tug.org/docs/liang/>.
- [MR08] Miklavec, Mojca und Arthur Reutenauer: *Putting the Cork back into the bottle : Improving Unicode support in T_EX*. TUGboat, 29(3):1001–1004, 2008. <http://www.tug.org/tex-hyphen/tb93miklavec.pdf>.
- [Rat06] Rat für deutsche Rechtschreibung: *Deutsche Rechtschreibung*. <http://rechtschreibrat.ids-mannheim.de/download/regeln2006.pdf>, München, 2006.
- [Wis91] Wissenschaftlicher Rat der Dudenredaktion (Herausgeber): *Duden : Rechtschreibung der deutschen Sprache*, Band 1 der Reihe *Der Duden in 12 Bänden*. Dudenverlag, Mannheim, 20. Auflage, 1991.
- [Wis06] Wissenschaftlicher Rat der Dudenredaktion (Herausgeber): *Duden : Die deutsche Rechtschreibung auf der Grundlage der neuen amtlichen Rechtschreibregeln*, Band 1 der Reihe *Der Duden in 12 Bänden*, Seiten 1161–1216. Dudenverlag, Mannheim, 24. Auflage, 2006.

⁸<http://groups.google.de/group/trennmuster-opensource?hl=de>

Tabelle 4: Trennmuster- und Manteldateien

Trennmusterdatei	Manteldatei
dehypht-x-<datum>.pat	dehypht-x-<datum>.tex
dehyphts-x-<datum>.pat	dehyphts-x-<datum>.tex
dehyphn-x-<datum>.pat	dehyphn-x-<datum>.tex

A. Dateien und Installation

Die eigentlichen Trennmusterdateien liegen in UTF-8-Kodierung vor. Sie sind somit für die Verwendung mit den 16-Bit-fähigen T_EX-Varianten XeT_EX und LuaT_EX geeignet.⁹ Für die 8-Bit-fähigen T_EX-Varianten T_EX und pdfT_EX werden Manteldateien bereitgestellt, die die UTF-8-Dateien laden und in T₁-Kodierung umwandeln (Tabelle 4).

Bei der Installation für 8-Bit-T_EX werden die Manteldateien an die in Tabelle 1 gezeigten Trennmusterbezeichner gebunden. Diese Schritte werden für verschiedene T_EX-Verteilungen in der Datei INSTALL beschrieben. Nach der Installation können die experimentellen Trennmuster wie in Unterabschnitt 2.2 gezeigt mit den Paketen Babel und hypsubst aus dem OBERDIEK-BÜNDEL aktiviert werden.

B. Die Wortliste

Bezug Die Wortliste ist über das öffentliche Entwicklerrepositorium¹⁰ des Projekts¹¹ erhältlich. Eine Kopie kann mit

```
$git clone git://repo.or.cz/wortliste.git
```

bezogen werden.¹²

Hinweis: Ältere GIT-Versionen als v1.5.6.5 enthalten einen Fehler im Netzwerkzugriff, der dazu führt, dass statt eines Satzes von Änderungen häufig die komplette Wortliste übertragen wird.

⁹Tatsächlich verwendet XeT_EX bereits Trennmuster, die aus diesem Projekt hervorgegangen sind, jedoch unter anderen Namen.

¹⁰<http://repo.or.cz/w/wortliste.git>

¹¹<http://groups.google.de/group/trennmuster-opensource?hl=de>

¹²<http://repo.or.cz/>

Für einen reibungslosen Betrieb sollte nach dem Klonen dieses Repositoriums die Konfigurationsvariable `core.DeltaBaseCacheLimit` vom Standardwert 16 MB auf mindestens 64 MB erhöht werden. Mit älteren GIT-Versionen können auch deutlich höhere Werte sinnvoll sein. Die Konfiguration kann beispielhaft so erfolgen:

```
$git config core.DeltaBaseCacheLimit 64m
$git config --get core.DeltaBaseCacheLimit
```

Das letzte Kommando sollte nun den Wert `64m` zurückliefern.

Dateiformat Das Format der Wortliste wird in der Datei `dateikopf` beschrieben.

Version Der `SHA1`-Commit-Hash der Repositoriumversion, die den Musterdateien

```
dehypht-x-<datum>.pat
dehyphts-x-<datum>.pat
dehyphn-x-<datum>.pat
```

zugrundeliegt, sowie eine URL zum direkten Herunterladen der Wortliste (ca. 15 MB) kann dem Kopf der Dateien entnommen werden.

C. Fragen & Antworten

I. Rechtschreibung

- 1. Verlag, Prüfer o. ä. bemängeln die Trennung der Trennmuster für die reformierte Rechtschreibung. Zum Beispiel wird »In-dus-trie« getrennt, der Duden trennt jedoch »In-dust-rie«.*
- 2. Gibt es dudenkonforme Trennmuster für die reformierte Rechtschreibung?*
- 3. Weshalb werden noch Trennmuster für die traditionelle Rechtschreibung bereitgestellt?*
- 4. Können Trennungen mit einer Silbenlänge von nur einem Buchstaben ermöglicht werden, zum Beispiel für den Satz in schmalen Kolumnen?*

II. Mitarbeit

- 5. Wie kann ich mich über dieses Projekt informieren?*
- 6. Wie kann ich helfen?*

7. Sollten Trennfehler einzeln oder gesammelt eingereicht werden?

I. Rechtschreibung

1. *Verlag, Prüfer o. ä. bemängeln die Trennung der Trennmuster für die reformierte Rechtschreibung. Zum Beispiel wird »In-dus-trie« getrennt, der Duden trennt jedoch »In-dust-rie«.*

Die amtlichen Regeln für die Rechtschreibung der deutschen Sprache lassen für viele Wörter mehrere Trennvarianten zu. Die Trennmuster und ebenso Wörterverzeichnisse legen sich aus praktischen Gründen auf eine Trennvariante fest. Sie können daher unterschiedliche Trennungen verwenden, ohne dass eine von beiden falsch ist. Aus diesem Grund sind Wörterverzeichnisse nicht geeignet, eine bestimmte Trennung auf Richtigkeit zu prüfen. Verbindlich sind einzig die amtlichen Regeln für die Rechtschreibung der deutschen Sprache [Rato6]. Häufig sind diese im Anhang eines Wörterbuchs abgedruckt. Die von den Trennmustern befolgten Konventionen können [Abschnitt 3](#) entnommen werden.

Wenn Unsicherheit darüber herrscht, wie die Rechtschreibung geprüft wird, sollte dies frühzeitig geklärt werden. Nicht jedem Redakteur oder Prüfer ist bewusst, dass der Duden seine normative Stellung mit der Rechtschreibreform 1996 eingebüßt hat.

2. *Gibt es dudenkonforme Trennmuster für die reformierte Rechtschreibung?*

Zur Zeit nicht, es ist auch nicht geplant. Dieses Projekt ist jedoch offen für Vorschläge und Mitarbeit.

3. *Weshalb werden noch Trennmuster für die traditionelle Rechtschreibung bereitgestellt?*

Die amtlichen Regeln für die Rechtschreibung der deutschen Sprache in der Fassung von 2006 sind nur für öffentliche Einrichtungen und Behörden verbindlich. Im privaten Schriftverkehr kann man wahlweise die traditionelle oder die reformierte Rechtschreibung verwenden.¹³ Daher erfreut sich die traditionelle Rechtschreibung weiterhin großer Beliebtheit.

Für Texte in gebrochener Schrift ist die traditionelle Rechtschreibung sogar vorzuziehen.

4. *Können Trennungen mit einer Silbenlänge von nur einem Buchstaben ermöglicht werden, zum Beispiel für den Satz in schmalen Spalten?*

Mit diesen Trennmustern ist das nicht möglich, da die zugrundeliegende Wortliste solche Trennungen nicht enthält. Die Abtrennung einzelner Vokale, zum

¹³Oder auch keine von beiden.

Beispiel »A-bend«, war nur kurzzeitig zulässig; die entsprechende Regelung von 1996 wurde mit der Rechtschreibreform 2006 wieder zurückgenommen. Wird die Silbenmindestlänge für die Worttrennung auf eins verringert, können falsche Trennungen auftreten.

II. Mitarbeit

5. *Wie kann ich mich über dieses Projekt informieren?*

Zentrale Anlaufstelle ist die Projekt-URL.¹⁴ Nach Anmeldung¹⁵ kann die Mailingliste für Fragen und Hinweise genutzt werden, ersatzweise auch die Newsgruppe `de.comp.text.tex`. Die folgenden Dateien enthalten weitergehende Informationen:

- Neben der Dokumentation zu den Trennmustern enthält dieses Paket eine Projektbeschreibung (`projektbeschreibung.pdf`).
- Die Datei `CHANGES` enthält bekannte, systematische Fehler der Trennmuster.
- Das Entwicklerrepositorium enthält in der Datei `dateikopf` eine Beschreibung des Formats der Wortliste.
- Die HTML-Präsenz des Projekts enthält einen Dateibereich mit weiteren Dateien, etwa einem Archiv der Diskussion im `MBOX`-Format.

6. *Wie kann ich helfen?*

Die deutschsprachige Trennmustermannschaft ist eine offene Gruppe und benötigt dringend weitere Mithilfe. Interessenten sind daher hoch willkommen! Zum Mitmachen gibt es mehrere Möglichkeiten:

- Die einfachste ist, die experimentellen Trennmuster ausgiebig zu testen und Fehler zu melden (siehe [Abschnitt 4](#)).
- Besonders hilfreich wäre Mitarbeit am Projekt. Als Einstieg kann die (grobe) Aufgabenliste in der Projektbeschreibung dienen.

¹⁴<http://groups.google.de/group/trennmuster-opensource?hl=de>

¹⁵Falls Sie die Anmeldung bei GOOGLE scheuen, können Sie uns vielleicht beim Umzug auf einen anderen Host helfen? Die Anforderungen an einen alternativen Projekthost können auch privat oder in `de.comp.text.tex` besprochen werden.

Obwohl schon experimentelle Trennmuster veröffentlicht wurden, steht dieses Projekt noch ziemlich am Anfang. Ziel ist, die verwendete Wortliste von möglichst vielen Menschen in verteilter Arbeit auf Richtigkeit zu prüfen. Bis dahin bleibt jedoch noch viel zu tun.

- Mittelfristig – nach Fertigstellung der Prüfmaske – kann auch durch systematische Durchsicht eines Teils der Wortliste geholfen werden.
- Fragen, Hinweise und Ideen sind auf der Mailingliste immer willkommen!

7. Sollten Trennfehler einzeln oder gesammelt eingereicht werden?

Das ist egal. Es sollten allerdings die folgenden Hinweise beim Einreichen von Fehlern beachtet werden:

Aktualität Wenn Sie ältere Trennmusterdateien verwenden, prüfen Sie zunächst, ob der Fehler auch mit aktuellen Trennmustern auftritt. Aktuelle Trennmuster sind im CTAN im Paket `dehyph-expt1` oder im Dateibereich unter der Projekt-URL erhältlich.

Informationen In einem Fehlerbericht sollte für das betreffende Wort die richtige und bei einzelnen Worteinreichungen auch die falsche Trennung angegeben werden. Außerdem die genaue Version der verwendeten Trennmuster (traditionelle oder reformierte Rechtschreibung, Datum der Trennmusterdatei).

Listenformat Es erleichtert die Korrektur, wenn Berichtigungsvorschläge in Form einer Liste eingereicht werden, die automatisch mit Skripten bearbeitet werden kann. Die folgenden Konventionen sollten dabei eingehalten werden:

- Die Spalten werden mit einem Semikolon ; getrennt.
- Die erste Spalte enthält das betreffende Wort in ungetrennter Form.
- Die zweite Spalte enthält das Wort in der Trennvariante nach traditioneller Rechtschreibung.
- Falls die Trennung nach reformierter Rechtschreibung davon abweicht, steht diese in der dritten Spalte.
- Ab der vierten Spalte *können* weitere Trennvarianten folgen, etwa die falsche Trennung. Eine kurze Erklärung sollte dann darauf eingehen.

- Unerwünschte Trennungen werden mit einem Punkt . markiert.
- Anfang und Ende der Liste sollten klar erkennbar sein.

Eine Liste könnte beispielhaft so aussehen:

```
sonnendurchfluteten;son-nen-durch-flu-te-ten
Talentwässerung;Tal-ent.wäs-se-rung
Fensterplatz;Fen-ster-platz;Fens-ter-platz
```

Dieses Format ist auch für einzeln eingereichte Fehltrennungen sinnvoll.

Betreff Der Betreff einer Fehlermeldung sollte aussagekräftig sein. Daher sollten mehrere Fehltrennungen, sofern möglich, thematisch zusammengefasst werden. Außerdem kann zur einfacheren Zuordnung das entsprechende Kriterium aus [Abschnitt 4](#) dem Betreff der E-Mail in eckigen Klammern vorangestellt werden, zum Beispiel:

```
Betreff: [C] sonnendurchfluteten
```

für eine Fehltrennung des Wortes *sonnendurchfluteten*, die durch einen Fehler in der Wortliste hervorgerufen wird.