

dehyph-expt1*

Experimentelle Trennmuster für die deutsche Sprache

Die deutschsprachige Trennmustermannschaft

31. März 2017

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`. More information can be found in the `Trennmuster-Wiki`¹ (in German).

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.

Dieses Paket richtet sich ausschließlich an Nutzer der Programme `TEX` und `pdfTEX`. `LuaTEX`- und `XeTEX`-Nutzer können experimentelle Muster aus dem Paket `hyph-utf8` verwenden und benötigen dieses Paket nicht.

Warnung!

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

*This document describes the `dehyph-expt1` package v0.41.

¹<http://projekte.dante.de/Trennmuster>

Inhaltsverzeichnis

	3.2. Traditionelle Rechtschreibung in der Schweiz	11
1. Einleitung	3.3. Reformierte Rechtschreibung .	12
2. Verwenden der Trennmuster	4. Trennfehler	14
2.1. Sprachvarietät und Recht- schreibung	A. Dateien und Installation	16
2.2. Aktivieren der Trennmuster . .	B. Fragen & Antworten	17
3. Trennregeln und Konventionen		7
3.1. Traditionelle Rechtschreibung in Deutschland und Österreich		8

1. Einleitung

Der in $\text{T}_{\text{E}}\text{X}$ 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 Trennmuster für die deutsche Sprache bei der Worttrennung in gewöhnlichen Texten eine akzeptable Fehlerrate erreichen, enthalten sie doch eine Reihe von Schwächen:²

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 T_{I} - sowie unvollständigen $\text{O}_{\text{T}_{\text{I}}}$ -kodierte Mustern. Mit Erscheinen von 16-Bit-fähigen $\text{T}_{\text{E}}\text{X}$ -Varianten werden sauber UTF-8 -kodierte Trennmuster nötig [MRo8].

²Diese Liste bezieht sich auf die Trennmusterdateien `dehypht.tex`, Version 3.2a vom 3. 3. 1999, und `dehyphn.tex`, Version 31 vom 7. 5. 2001.

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 Ausnahmeliste mit über 3500 korrigierten Trennungen einfacher Wörter [Lemo3, Lemo5].³
- Wird in der traditionellen Rechtschreibung *ß* durch *ss/SS* oder *sz/SZ* 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-ke!«, »Kus-shand«, »Massnah-me«, »mis-sach-ten« (nur Schweiz) und »Ex-presszug«, »Fit-nes-scen-ter«, »Fit-nesstrai-ner«.

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 [Sch98]. Aus diesem Grund ist die Worttrennung mit den Trennmustern für die reformierte Rechtschreibung etwas schlechter als mit den Trennmustern für die traditionelle Rechtschreibung.

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.

³CTAN:language/hyphenation/dehyph/dehyph.tex

⁴Die Trennung der herkömmlichen Muster 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.

Tabelle 1: Kompatibilität mit verschiedenen T_EX-Varianten

Programm	Quelle experimenteller Muster
T _E X	dehyph-expt1
pdfT _E X	dehyph-expt1
XeT _E X	hyph-utf8, dehyph-expt1
LuaT _E X	hyph-utf8
pT _E X	hyph-utf8

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

Dieses Paket stellt die experimentellen Trennmuster für die Verwendung mit L^AT_EX und dem Sprachenpaket Babel zur Verfügung. Das Paket richtet sich derzeit ausschließlich an Nutzer der Programme T_EX und pdfT_EX (siehe [Tabelle 1](#)).

Mit den nicht unterstützten T_EX-Programmen können experimentelle Trennmuster dennoch verwendet werden, da das Paket hyph-utf8 ebenfalls Muster dieses Projekts enthält. Werden die UTF-8-fähigen Programme XeT_EX und LuaT_EX oder pT_EX⁵ verwendet, so werden jene Muster standardmäßig aktiviert. Die Trennmuster dieses Pakets sollten dann nicht verwendet werden. Siehe [Anhang B](#) für Hinweise, wie die Version der im Paket hyph-utf8 enthaltenen Muster ermittelt werden kann.

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

⁵Eine in Japan populäre T_EX-Variante.

Tabelle 2: Die unterstützten Varietäten und Rechtschreibungen

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

die Zeichenkette `latest` zu ersetzen. [Anhang B](#) enthält Hinweise, wie das Datum ermittelt werden kann, falls die experimentellen Trennmuster in Ihrer \TeX -Verteilung schon vorinstalliert sind.

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
- Schweiz⁶

gebräuchlichen Hauptvarietäten der deutschen Standardsprache. [Tabelle 2](#) 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

⁶Für 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?

deutsche Anwender können daher dieselben Trennmuster verwenden. (Wie das auch schon bei den herkömmlichen Trennmustern der Fall war.)

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

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. Die Trennmusterbezeichner, die vom Paket Babel verwendet werden, werden dazu mit Hilfe des Pakets hyphsubst aus dem OBERDIEK-BÜNDEL so geändert, dass sie auf die experimentellen Trennmuster verweisen:

```
\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 3](#) zusammengefasst.

Tabelle 3: Trennvarianten

<i>traditionelle Rechtschreibung</i>		<i>reformierte Rechtschreibung</i>	
herkömmlich	experimentell	herkömmlich	experimentell
<u>lös</u> -te	lö-ste	<u>lös</u> -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- <u>stern</u>	Abend- <u>stern</u>	Abends- <u>tern</u>	Abend- <u>stern</u>
Mor-dop-fer	Mord-op-fer	Mor-dop-fer	Mord-op-fer

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

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 [Rat06, 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.⁷ 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:

⁷Im Ergebnis weicht in reformierter Rechtschreibung die Trennung zum Beispiel des Dudens (nach Sprechsilben) von der Trennung mit diesen Trennmustern (bevorzugt etymologisch) ab, siehe auch [Regel R2](#) und [Regel R3](#) sowie [Anhang B](#).

Tabelle 4: Unterschiedlich ermittelte Trennmöglichkeiten.

Quelle	Trennmöglichkeiten
pdf \LaTeX mit Babel 3.8	Me-tall(-l)egierung
pdf \LaTeX mit Babel 3.9	Me-tall(-l)e-gie-rung
erwünscht (vgl. Regel T6)	Me-tall(-l)egie-rung

- 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 R5](#)).
- Die von einem Programm aus diesen Mustern abgeleiteten möglichen Trennstellen können (u. a. durch Programmfehler) durchaus von denen der zugrundeliegenden Wortliste abweichen. So führt zum Beispiel die Eingabe `Meta"llegierung` (Dreikonsonantenregel in der traditionellen Rechtschreibung) mit dem Paket Babel zu den in [Tabelle 4](#) gezeigten Trennmöglichkeiten.
- Die von \TeX gewählte Trennung kann in Einzelfällen mit den \TeX - 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 Die minimal unterstützte Silbenlänge am Wortanfang und -ende beträgt zwei Buchstaben [[Wis91](#), R 178].

Beachte, die Mindestlänge abgetrennter Silben lässt sich in \TeX mit den Makros `\lefthyphenmin` und `\righthyphenmin` und in \LaTeX mit

dem Makro \germanhyphenmins (Babel) anpassen. Wird die Mindestlänge auf weniger als zwei Buchstaben verringert, so können fehlerhafte Trennungen auftreten.

T2 Sinnentstellende und irreführende Trennungen werden möglichst vermieden [Wis91, R 181] (siehe auch [Regel T6](#)):

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

Beachte, dass derzeit die Unterdrückung von solchen Trennstellen bis zu einem gewissen Grade willkürlich und subjektiv ist. Außerdem ist die Erfassung weit davon entfernt, vollständig zu sein. Um diesen Problemen abzuhelpen, ist für zukünftige Versionen der Trennmuster ein anderer, automatisierter Ansatz geplant, welcher die Anzahl manuell zu erfassender Fälle sehr stark reduzieren wird.

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

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

Beachte, die Trennstellen »Drucker-zeugnis« und »Musiker-leben« sind in den Interpretationen *Druck-Erzeugnis* und *Musik-Erleben* irreführend. Sie entfallen nach [Regel T2](#) und sind nicht als übereinstimmende Trennstellen anzusehen. Zur Spezialtrennung »Fußball-lehre« siehe auch [Regel T6](#).

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 *ß* auftreten, werden nicht berücksichtigt (vergleiche [Regel T4](#)).

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

T4 Wird der Buchstabe β durch *ss/SS* ersetzt, so bleibt die Trennung davon unberührt [[Wis91](#), R 179]:

GRÖSS-TE
GRÜ-SSE
MA-SSES

Für diese Regel gilt die folgende Einschränkung:

- Wenn durch den Ersatz von β an dieser Stelle keine eindeutige Trennung möglich ist, 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 mit `\MakeUppercase` durch *SS* ersetzt, so bleibt *SS* stets ungetrennt. Die Trennung richtet sich dann nach der Schreibweise mit β im Quelldokument.
- Existiert ein Wort in verschiedenen Varietäten in der Schreibweise mit β und mit *ss*, so wird aufgrund dieser Einschränkung *s-s* stets getrennt:

Ge-scho-ße (AT)
Ge-schos-se (D) GESCHO-SSE (AT)

- Wenn durch den Ersatz des β an entfernten Stellen keine eindeutige Trennung möglich wird, zum Beispiel an Wortfugen, so werden die betroffenen Trennungen gemäß [Regel T3](#) unterdrückt. In der Folge wird gegebenenfalls auch die Trennung von *ss/SS* unterdrückt.

BAHN-HOFSTRASSE BAHNHOF-STRA-SSE
BAHNHOFS-TRAS-SE

T5 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

T6 Spezialtrennungen (*engl.*: non-standard hyphenation), die nach Regeln erfolgen, die über das bloße Einfügen eines Trennstrichs hinausgehen,

wie die *ck*- oder die Dreikonsonantenregel, kann \TeX 82 nicht automatisch behandeln. Aus diesem Grund sind solche Trennstellen in diesen Trennmustern nicht berücksichtigt.

drucken	drucken	druk-ken
Zuckerbäcker	Zucker-bäcker	Zuk-kerbäk-ker
Brennessel	Brennes-sel	Brenn-nessel
Stoffetzen	Stoffet-zen	Stoff-fetzen

Die Dreikonsonantenregel birgt aufgrund des ausgefallenen Konsonanten die Gefahr irreführender und sinnentstellender Trennungen (siehe auch [Regel T2](#)). Trennstellen, die in einem Abstand von zwei Lauten auf eine Wortfuge mit Anwendung der Dreikonsonantenregel folgen, werden daher grundsätzlich unterdrückt.

Metalllegierung	Me-tallegie-rung	Metall-legierung	Metalle-gierung
schnellebige	schnellebi-ge	schnell-lebige	schnelle-bige
Stilleben	Stilleben	Still-leben	Stille-ben

auch:

Abfallager	Ab-fallager	Abfall-lager	Abfalla-ger
Zellstofffabrik	Zell-stofffabrik	Zellstoff-fabrik	Zellstoffa-brik

Das Paket `Babel` stellt verschiedene Kürzel zur Verfügung, u. a. "`ck`"⁸ oder "`ff` etc.", mit denen Spezialtrennungen im Quelldokument ausgezeichnet werden können (siehe auch [Tabelle 4](#)).

\LuaTeX ⁹ soll in einer zukünftigen Version Mechanismen zur automatischen Behandlung von Spezialtrennungen bereitstellen. Eine physische Auszeichnung im Quelltext ist dann nicht mehr erforderlich. Die entsprechenden Spezialtrennmuster für die deutsche Sprache werden ebenfalls im Rahmen dieses Projekts erstellt.

3.2. Traditionelle Rechtschreibung in der Schweiz

Die Trennmuster für die traditionelle Rechtschreibung in der Schweiz folgen weitgehend den Konventionen für die traditionelle Rechtschreibung in Deutschland (siehe [Unterabschnitt 3.1](#)). Die folgende Liste enthält daher nur

⁸Spezialtrennungen werden in \TeX mit Hilfe des Kommandos `\discretionary` kodiert. So wird zum Beispiel das `Babel`-Kürzel "`ck`" in der Eingabe während des Setzens durch `\discretionary{k-}{k}{ck}` ersetzt, wodurch *k-k*-Trennungen möglich werden.

⁹<http://www.luatex.org/>

Fälle, in denen davon abgewichen wird oder deren Beschreibung aus anderen Gründen sinnvoll erscheint.

TS1 Wörter mit β werden gemäß den Regeln für die traditionelle Rechtschreibung in Deutschland getrennt.

TS2 Abweichend von [Regel T4](#) wird *ss/SS* immer als Doppelkonsonant behandelt und gegebenenfalls getrennt:

grös-ste

Grüs-se

Mas-ses

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 (siehe [Regel TS1](#)).

3.3. Reformierte Rechtschreibung

R1 Die minimal unterstützte Silbenlänge am Wortanfang und -ende beträgt zwei Buchstaben [[Rato6](#), [Wiso6](#), § 107].

Beachte, die Mindestlänge abgetrennter Silben lässt sich in $\text{T}_{\text{E}}\text{X}$ mit den Makros `\lefthyphenmin` und `\righthyphenmin` und in $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ mit dem Makro `\ngermanhyphenmins` (Babel) anpassen. Wird die Mindestlänge auf weniger als zwei Buchstaben verringert, so können fehlerhafte Trennungen auftreten.

R2 Falls die Trennung nach Sprechsilben und die etymologische (sprachgeschichtliche) Trennung 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

R3 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		

R4 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

R5 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 R4 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
-------------	------------	-----------

R6 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 β 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.

4. Trennfehler

Mit den vorliegenden Trennmustern können sämtliche Wörter der zugrundeliegenden Wortliste fehlerfrei getrennt werden. Technisch gesprochen endet der letzte PATGEN-Lauf mit der Meldung

```
1266082 good, 0 bad, 0 missed  
100.00 %, 0.00 %, 0.00 %
```

(der Wert vor `good` ist vom Listenumfang abhängig). Trotz des großen Umfangs der Wortliste lassen sich Trennfehler in Wörtern, die nicht in der Liste enthalten sind, nicht vermeiden. Der Umfang der Wortliste kann allerdings nicht beliebig erweitert werden.¹⁰ In den folgenden Fällen sollten fehlerhafte Trennungen der Trennmuster jedoch gemeldet werden:

- A. Das Wort ist bereits in der Wortliste enthalten. Der Eintrag ist jedoch fehlerhaft.

Falls das Wort nicht in der Wortliste enthalten ist, bestehen sehr gute Chancen, dass es aufgenommen wird, wenn eines der folgenden Kriterien erfüllt ist:

- B. 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 in T_EX der folgende Aufruf verwendet werden (die Ausgabe erfolgt in der LOG-Datei):

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

- C. Es handelt es sich um eine orthographisch richtige, aber sinnentstellende oder irreführende Trennung. Berücksichtigt werden allerdings nur

¹⁰Liangs Schema sieht nur einen begrenzten Bereich für die Trennstellenbewertungen vor (0–9). Die derzeitigen Trennmuster verwenden bereits Bewertungen der Höhe 8.

Wörter, die aus höchstens zwei (gegebenenfalls prä- und suffigierten) Wörtern zusammengesetzt sind, zum Beispiel »Talent-wässerung«. Nicht berücksichtigt wird hingegen die »Talent-wässerungsanlage«.

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@dante.de,
- wl@gnu.org (Werner Lemberg).

Trennfehler, 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}
```

Die aktuelle und ältere Ausgaben der Trennmuster sind im Dateibereich des Trennmuster-Wikis erhältlich.¹¹ Im Entwicklerrepositorium¹² befindet sich ein Makefile, mit dem jederzeit neue Trennmuster erzeugt werden können.

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] Miklavac, 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/tb93miklavac.pdf>.

¹¹<http://projekte.dante.de/Trennmuster>

¹²siehe <http://projekte.dante.de/Trennmuster/Entwickler>

Tabelle 5: 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

- [Rato6] Rat für deutsche Rechtschreibung: *Deutsche Rechtschreibung*. <http://rechtschreibrat.ids-mannheim.de/download/regeln2006.pdf>, München, 2006.
- [Sch98] Schmidt, Walter: *T_EX und die neue deutsche Rechtschreibung*. Die T_EXnische Komödie, 10(2):35–37, September 1998.
- [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.

A. Dateien und Installation

Die eigentlichen Trennmusterdateien liegen in UTF-8-Kodierung vor (siehe [Tabelle 5](#), Endung pat). Sie werden von T_EX nicht direkt geladen, sondern durch Manteldateien, die ebenfalls Teil des Pakets sind (Endung tex). Wird eine 8-Bit-fähige T_EX-Variante erkannt, übernehmen diese Manteldateien die Konvertierung der Trennmuster in die T₁-Kodierung.

Bei der Installation werden die Manteldateien an die in [Tabelle 2](#) 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 verwendet werden.

B. Fragen & Antworten

I. Verwenden der Trennmuster

1. Die experimentellen Trennmuster waren in der T_EX-Verteilung bereits vorinstalliert. Wie kann das Datum ermittelt werden, das Teil des Trennmusterbezeichners ist, wie in Abschnitt 2 erwähnt?
2. Wie kann die Version der experimentellen Muster ermittelt werden, die im Paket *hyph-utf8* enthalten sind?

II. Rechtschreibung

3. 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«.
4. Gibt es dudenkonforme Trennmuster für die reformierte Rechtschreibung?
5. Weshalb werden noch Trennmuster für die traditionelle Rechtschreibung bereitgestellt?
6. Können Trennungen mit einer Silbenlänge von nur einem Buchstaben ermöglicht werden, zum Beispiel für den Satz in schmalen Kolumnen?

III. Mitarbeit

7. Wie kann ich mich über dieses Projekt informieren?
8. Wie kann ich helfen?
9. Sollten Trennfehler einzeln oder gesammelt eingereicht werden?

I. Verwenden der Trennmuster

1. Die experimentellen Trennmuster waren in der T_EX-Verteilung bereits vorinstalliert. Wie kann das Datum ermittelt werden, das Teil des Trennmusterbezeichners ist, wie in [Abschnitt 2](#) erwähnt?

Die Trennmusterbezeichner werden in der Datei `language.dat` definiert, wo auch die Verbindung zu den Manteldateien hergestellt wird. Da ein Teil der Trennmusterbezeichner bereits bekannt ist, `german-x`, vgl. [Tabelle 2](#), kann die Datei `language.dat` danach durchsucht werden. Zunächst muss der Ort der Datei mit Hilfe des Kommandos `kpsewhich` ermittelt werden. Es folgen die vollständigen Kommandos für unixähnliche Shells und die Windows-Kommandozeile `cmd.exe`. Achtung, bei den einfachen Anführungszeichen handelt es sich um Gravis („Backquotes“).

Listing 1: Shell

```
grep -i german-x 'kpsewhich language.dat'
```

Listing 2: cmd.exe

```
for /F "usebackq" %f in ('kpsewhich language.dat') do find /i "german-x" "%f"
```

Die Ausgabe dieser Kommandos sieht etwa wie folgt aus (das Datum kann abweichen):

```
german-x-2009-06-19 dehyphnt-x-2009-06-19.tex
=german-x-latest
ngerman-x-2009-06-19 dehyphn-x-2009-06-19.tex
=ngerman-x-latest
```

Die gesuchten Trennmusterbezeichner befinden sich in der ersten Spalte und lauten in diesem Beispiel `german-x-2009-06-19` und `ngerman-x-2009-06-19`. In der zweiten Spalte kann man die Namen der Manteldateien erkennen (vgl. [Tabelle 5](#)). Die mit einem Gleichheitszeichen beginnenden Zeilen definieren ein Synonym für den Trennmusterbezeichner der unmittelbar vorangehenden Zeile in der Datei `language.dat`.

2. *Wie kann die Version der experimentellen Muster ermittelt werden, die im Paket `hyph-utf8` enthalten sind?*

Das Vorgehen ähnelt dem der vorherigen Antwort. Zunächst wird der Ort einer bestimmten Datei ermittelt. Diese wird dann nach einer hilfreichen Zeichenkette durchsucht.

Listing 3: Shell

```
grep dehyph 'kpsewhich hyph-de-1901.lic.txt'
```

Listing 4: cmd.exe

```
for /F "usebackq" %f in ('kpsewhich hyph-de-1901.lic.txt') do find "dehyph" "%f"
```

Die Ausgabe dieser Kommandos sieht etwa wie folgt aus (das Datum kann abweichen):

```
% dehyphnt-x-2011-06-01.pat
\message{German Hyphenation Patterns (Traditional Orthography) 'dehyphnt-x' 2011-
06-01 (WL)}
```

II. Rechtschreibung

3. *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.

4. *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.

5. *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.

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

Mit diesen Trennmustern ist das nicht möglich, da die zugrundeliegende Wortliste solche Trennungen nicht enthält. Die Abtrennung einzelner Vokale, zum Beispiel »A-bend«, war nur kurzzeitig zulässig; die entsprechende Regelung

¹³Oder auch keine von beiden.

von 1996 wurde mit der Rechtschreibreform 2006 wieder zurückgenommen. Wird die Silbenmindestlänge für die Worttrennung auf eins verringert, können falsche Trennungen auftreten.

III. Mitarbeit

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

Zentrale Anlaufstelle ist das Trennmuster-Wiki, welches sich momentan jedoch noch im Aufbau befindet.¹⁴ Für Fragen und Hinweise kann nach Anmeldung die Mailingliste¹⁵; oder alternativ die Usenetgruppe `de.comp.text.tex` genutzt werden. 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 `README.wortliste` eine Beschreibung des Formats der Wortliste.

8. *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.
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.

¹⁴<http://projekte.dante.de/Trennmuster>

¹⁵trennmuster@dante.de

- Fragen, Hinweise und Ideen sind auf der Mailingliste immer willkommen!

9. *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. Die aktuellen Trennmuster sind im CTAN im Paket `dehyph-expt1` oder im Dateibereich unter der Projekt-URL erhältlich. Im Repositorium befindet sich auch ein Makefile, mit dem jederzeit neue Trennmuster aus der aktuellen Wortliste erzeugt werden können.

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 Korrekturen sinnvoll.

Betreff Der Betreff einer Fehlermeldung sollte aussagekräftig sein. Daher sollten mehrere Trennfehler, 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: [A] sonnendurchfluteten

für einen Trennfehler im Wortes *sonnendurchfluteten*, der durch einen Fehler in der Wortliste hervorgerufen wird.