

This page intentionally left for curious readers.

Alte und neue Tücken von Silbentrennung und semantischem Markup

3. August 2015

CC-by-sa 3.0 hg -r218

Was, wie, warum?

Nicht...

- Technisches der Silbentrennung
- Doku für babel/polyglossia, fontspec, oder die Silbentrenn-Heuristik von T_EX
- ...

...sondern

- »Was muss der *Fließtext* leisten, damit ich den Satz *typographisch automatisieren* kann?«
- »Semantisches« Markup auf Buchstaben-Ebene

T_EXnical Disclosure

Engine Lua_AT_EX beta-0.79.1 mit polyglossia

Schriften Merriweather Sans

Junicode

Inconsolata

Google Webfonts

SourceForge

Raph Levien

Trennstellen hervorheben showhyphens.sty

Nur mit Lua_AT_EX!

Einfaches Beispiel für Mehrdeutigkeit: Leerzeichen

- »ASCII 0x20« für Zwischenraum von Wörtern, Sätzen, Abkürzungen mit Groß-/Kleinbuchstaben...
- T_EX sucht davor nach Satzzeichen und Groß-/Kleinbuchstaben (Nota: `\nonfrenchspacing`)
- Bedeutung *mehrdeutig* und *kontextabhängig!*
- Manuell mit `\@`: Verarbeite vorangegangenes Zeichen wie einen Kleinbuchstaben

>>Guter<< Satz: mehr als ASCII-Quelle

Ligaturen und Trennstellen

Wachstube

Hoffnung offen

effizient

T_EX: **Heuristik** für Trennmuster¹ und Ligaturen (NB: aufs Englische optimiert)

⇒ Händische Um\bruch\stellen

¹<http://www.tug.org/docs/liang/>

Kollisionen Trennstelle–Ligatur

offen ↔ offen ↔ of|fen

- Abhängig von der Schrift (Verfügbare Ligaturen)...
- ...und der T_EX-Engine!

of\–fen

offen
pdfT_EX

offen
LuaT_EX

*ignoriert \– zugunsten
von Ligaturen!*

of" | fen

Linguistik und Mikrotypographie

Nicht...

- Technisches der Silbentrennung
- Doku für babel/polyglossia, fontspec, oder die Silbentrenn-Heuristik von T_EX
- ...

...sondern

- »Was muss der *Fließtext* leisten, damit ich den Satz *typographisch automatisieren* kann?«
- »Semantisches«
Markup auf
Buchstaben-Ebene

Ein bisschen Linguistik...

Gliederung von Wörtern nach:

1 Silben: nur nach Sprachrhythmus
wahr-haft wahr-haf-tig

2 Morphemen = Sinn-Einheiten (Morphologie)
wahr|haft|ig²

Repräsentation im Quellcode: Weder-noch!
wahrhaftig

²<http://wortschatz.uni-leipzig.de/>

Wozu soll Mikrotypographie dienen?

Design-Entscheidung...

Einsatz/Weglassen von Ligaturen für die Erfassung von

- 1 Klang, Rhythmus? (\Rightarrow Silben)
- 2 Sinneinheiten? (\Rightarrow Morpheme)



Probleme und Lösungen

1 Heft → Hefte / Hefte ?

Gleicher Wortstamm, unterschiedliche Glyphen?

offen

2 offen → of-⌞ fen ?

Zeilenende: Ligatur muss aufgebrochen werden

⇒ *Umbrechende Ligatur in den Quellcode*

>>Take home<<

- Umbruch-Resultate *qualitativ* abhängig von Schrift, T_EX-Engine, und Vorarbeit im Code
- Guter Satz erfordert mehr als (blankes) ASCII: Sinn-gehalt muss erraten werden, oder vorliegen



³Duden, 11. Aufl. (1937)

>>Lösung<<: *Umbrechende Ligatur*

Für Ligatur-Satz nach Sinneinheiten:

- Harte Ligatur-Unterbindung
- Für Revision: Suche nach Regexp `f[flirt]`

Meine Lösung:

- In T_EX via `\discretionary`
- Code: Unicode-Glyph `f` zweckentfremdet
- vim-Integration mit Shortcut

Backup-Folien und Beispiele I

Unicode U+00AD (soft hyphen): s-hy wird von
LuaTeX sichtbar dargestellt

Dampföffnung

lästig läftig

dadaistisch (keine st-Ligatur!)