

Text2DEK¹

transformiert Texte in DEK-Verkehrsschrifts²stenogramme, z. B.:

Vor dem Gesetz steht ein Türhüter. →

Der Eingabetext wird zuerst in einzelne, etwa eine Wortform umfassende, Token zerlegt, hier:

vor_□dem_□Gesetz_□steht_□ein_□Türhüter_□.

Als nächstes wird die Metaform eines Tokens festgestellt: entweder wird sie einem Wörterbuch entnommen (so bei Zahlen, Satzzeichen und Verkehrsschriftkürzeln) oder aus einer morphologischen Wortzerlegung regelbasiert ermittelt. DEK-Steneme sind eine wechselnde Folge von Abstrichen (Konsonanten) und Verbindungsaufstrichen (Vokalen) und haben die folgende Metaform:

$$(\text{,}K)\{(V,K)\}^*$$

(Mit $(\text{,}ge)(\text{,}s)(e,z)$ als Beispiel für $\overline{p}j$ — dem Gesetz-Stenem.) K entspricht den Abstrichen und steht für Konsonanten (b, c, ...), Konsonantfolgen (ch, sp, str u.ä.) bzw. Morpheme (Präfixe wie ge-, ... und Suffixe wie -heit, -ung, ...). V entspricht den Aufstrichen und steht für Vokale (a, ä, e, ...) und Diphthonge (au, äu, ...).

Anhand der Metaform wird für jedes Token on the fly in einem METAFONT-Lauf ein Stenem — ein METAFONT-Zeichen pro Wort — generiert.

Übersichtstabelle zu unserem Beispiel:

| Token | Wortzerlegung ³ | Metaform | Stenem |
|----------|----------------------------|---|--------|
| . | | $(\text{,period}_)$ | . |
| Gesetz | ge [^] setz | $(\text{,}ge)(\text{,}s)(e,z)$ | |
| Türhüter | tür#hüt [~] er | $(\text{,}t)(\ddot{u},r)\#(\text{,}h)(\ddot{u},t)(e,r)$ | |
| dem | | $(\text{,}dem)$ | |
| ein | ein | $(\text{,}ein)$ | |
| steht | steh [~] t | $(\text{,}st)(e,t)$ | |
| vor | | $(\text{,}vor)$ | |

Der in Token zerlegte Text wird mit L^AT_EX und dvips übersetzt, bevor die Seiten des entstandenen Stenogramms als gif-Bilder zum Browser geschickt werden.

Falls notwendig (bei Fremwörtern, falscher morphologischen Zerlegung), kann man die Metaform korrigieren.

¹S. J. Šarman: DEK-Verkehrsschrift mit METAFONT und L^AT_EX, Die T_EXnische Komödie, 1 (2009), 7–20

²Basisstufe der Deutschen Einheitskurzschrift

³mit \wedge , \sim bzw. $\#$ als Präfix-, Suffix- bzw. Kompositatrennzeichen; bei Kürzeln fehlt der Eintrag