

## Text2DEK<sup>1</sup>

transformiert Texte in DEK-Verkehrsschrifts<sup>2</sup>stenogramme, z. B.:

Vor dem Gesetz steht ein Türhüter. →

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

vor<sub>□</sub>dem<sub>□</sub>Gesetz<sub>□</sub>steht<sub>□</sub>ein<sub>□</sub>Türhüter<sub>□</sub>.

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  $(\text{,}ge)(\text{,}s)(\text{e},z)$  als Beispiel für  $\overline{p}j$  — dem Gesetz-Stenem.)  $K$  entspricht den Abstrichen und steht für Konsonanten (b, c, ...), Konsonantenfolgen (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 Diphtonge (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 <sup>3</sup>	Metaform	Stenem
.		$(\text{,period}_)$	.
Gesetz	ge <sup>^</sup> setz	$(\text{,}ge)(\text{,}s)(\text{e},z)$	
Türhüter	tür#hüt <sup>~</sup> er	$(\text{,}t)(\ddot{u},r)\#(\text{,}h)(\ddot{u},t)(\text{e},r)$	
dem		$(\text{,}dem)$	
ein	ein	$(\text{,}ein)$	
steht	steh <sup>~</sup> t	$(\text{,}st)(\text{e},t)$	
vor		$(\text{,}vor)$	

Der in Token zerlegte Text wird mit L<sup>A</sup>T<sub>E</sub>X 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. <- Proof=7

<sup>1</sup>S. J. Šarman: DEK-Verkehrsschrift mit METAFONT und L<sup>A</sup>T<sub>E</sub>X, Die T<sub>E</sub>Xnische Komödie, 1 (2009), 7–20

<sup>2</sup>Basisstufe der Deutschen Einheitskurzschrift

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