

6.44	<i>VisualTime</i> – Profil des <i>Sanguo zhi yanyi</i>	238
6.45	<i>VisualTime</i> – <i>NLLR</i> -Werte des <i>Sanguo zhi yanyi</i> für die einzelnen <i>chronons</i>	239
6.46	<i>VisualTime</i> – Temporale Ausdrücke im <i>Sanguo zhi yanyi</i> (Ausschnitt)	240
6.47	<i>VisualTime</i> – Lexeme im <i>Sanguo zhi yanyi</i> (Ausschnitt)	240

Tabellenverzeichnis

2.1	<i>Ershisi shi</i> 二十四史 und <i>Qing shi gao</i> 清史稿	23
4.1	Übersicht aller verwendeten Korpora	65
4.2	Getestete Tokenizer	77
4.3	Goldstandard-Texte für den Tokenizer-Vergleich	79
4.4	Ranking der durchschnittlichen Performance aller getesteten Tokenizer	91
5.1	Qualität der digitalen Ausgabe – Ergebnisse der Stichprobenanalyse	118
5.2	Ergänzt Dynastiesystem des <i>HYDCD</i> , chronologisch nach Anfangsjahr	129
5.3	30 häufigste <i>Locus classicus</i> -Angaben im <i>DHYDCD</i>	150
5.4	30 meistzitierte Werke im <i>DHYDCD</i>	153
6.1	Ergebnisse der beschriebenen Experimente mit <i>difangzhi-SLM</i>	165
6.2	Test von <i>Smoothing</i> -Methoden mit unterschiedlichen Parametern	166
6.3	Ergebnisse mit <i>DFZ</i> Co-Datierung vs. <i>chronon-SLM</i>	170
6.4	Ergebnisse mit <i>XXSKQS</i> Co-Datierung vs. <i>chronon-SLM</i>	171
6.5	Ergebnisse mit <i>zhengshi</i> und mit <i>DHYDCD-SLMs</i>	172
6.5	(Fortsetzung)	173
6.6	Ergebnisse mit <i>Difangzhi</i> und <i>HYDCD-SLMs</i>	174
6.7	Ergebnisse mit <i>Difangzhi</i> 1300–1925 und <i>HYDCD-SLMs</i>	174
6.8	Ergebnisse mit <i>Xu xiu si ku quan shu</i> und <i>HYDCD-SLMs</i>	175
6.9	Lexikalisierung und Gewichtungskorrekturfaktoren nach Jahrhundert	188
6.10	Lexikalisierte 2–4-Zeichen-Kombinationen im <i>Zhongjing</i> 忠經	193
6.11	2–4-Zeichen-Kombinationen im <i>Zhongjing</i> (mit <i>Zhongjing</i> belegt)	195
6.12	Ergebnisüberblick der Datierungsexperimente aus 6.2.5	207
6.13	Vergleich linearer Modelle	215
6.14	Korrelationskoeffizient (<i>R</i>) für Anteile häufigster 2–4-Gramme	219
6.15	Datierung unterschiedlicher Texte mit <i>AYL</i>	223
6.16	Vergleich der in Kapitel 6 vorgestellten Methoden anhand des <i>DFZ</i> -Korpus	231

Vorbemerkungen

Chinesische Begriffe und Namen von Personen, Werken und Orten werden in dieser Arbeit in der *Hanyu Pinyin* 漢語拼音 Transkription wiedergegeben, gefolgt von traditionellen Schriftzeichen (*fantizi* 繁體字). Ausnahmen bilden Namen von Autor:innen, die selbst eine abweichende westliche Schreibweise gewählt haben, z. B. HUANG Chu-ren [Juren] 黃居仁. Um eine gute Lesbarkeit der einzelnen Unterkapitel zu gewährleisten, werden die Schriftzeichen stets bei der ersten Nennung im jeweiligen Abschnitt mit angegeben. Dasselbe gilt für Zeitangaben zu chinesischen Dynastien. Als Referenz sind die wichtigsten Dynastien zusätzlich auf der folgenden Seite aufgeführt.

Wörtliche Zitate aus anderen Texten werden originalgetreu übernommen, auch wenn sie in Kurzzeichen (*jiantizi* 简体字) gesetzt sind, sowie andere Umschriften oder Rechtschreibfehler enthalten.

FAMILIENNAMEN, sowie NAMEN VON ORGANISATIONEN werden in KAPITÄLCHEN gesetzt, *Transkriptionen* und *fremdsprachige Fachbegriffe kursiv*. Quellcode, sowie Namen von Variablen, Funktionen und Datenbankspalten werden mit fester Zeichenbreite gesetzt.

Dynastien

Shang 商	ca. 1600–1045 v. u. Z. ¹
Zhou 周	ca. 1045–256 v. u. Z.
Westliche Zhou (<i>Xi Zhou</i>) 西周	11. Jh.–771 v. u. Z.
Chunqiu 春秋	770–476 v. u. Z.
Streitende Reiche (<i>Zhanguo</i>) 戰國	475–221 v. u. Z.
Qin 秦	221–206 v. u. Z.
Han 漢	202 v. u. Z.–220 u. Z.
Frühere Han (<i>Qian Han</i>) 前漢	202 v. u. Z.–9 u. Z.
Xin 新	9–23
Östliche Han (<i>Dong Han</i>) 東漢	25–220
Drei Reiche (<i>Sanguo</i>) 三國	220–280
Wei 魏	220–265
Shu 蜀	221–263
Wu 吳	222–280
Jin 晉	265–420
Südliche und Nördliche Dynastien (<i>Nanbei chao</i>) 南北朝	420–589
Nördliche Wei (<i>Bei Wei</i>) 北魏	386–534
Nördliche Liang [16 Reiche] (<i>Bei Liang</i>) 北涼	401–439
Südliche Song (<i>Nanchao Song</i> / Liu Song) 南朝宋 / 劉宋	420–479
Südliche Qi (<i>Nan Qi</i>) 南齊	479–502
Südliche Liang (<i>Nanchao Liang</i>) 南朝梁	502–557
Östliche Wei (<i>Dong Wei</i>) 東魏	534–550
Nördliche Qi (<i>Bei Qi</i>) 北齊	550–577
Nördliche Zhou (<i>Bei Zhou</i>) 北周	557–581
Chen [Südliche Dynastien] (<i>Nanchao Chen</i>) 南朝陳	557–589
Sui 隋	581–618
Tang 唐	618–907
Fünf Dynastien (<i>Wudai</i>) 五代	907–960
Liao 遼	916–1125
Song 宋	960–1279
Jin 金	1115–1234
Yuan 元	1279–1368
Ming 明	1368–1644
Qing 清	1644–1912
Republik China (<i>Zhonghua Minguo</i>) 中華民國	1912–
VR China (<i>Zhonghua Renmin Gongheguo</i>) 中華人民共和國	1949–

¹ Angaben aus Endymion WILKINSON 2000: *Chinese History. A Manual*. Revised and Enlarged. Cambridge, MA & London: Harvard University Asia Center, Harvard University Press, S. 10–12.

Abkürzungen

Seitenangaben zur Verwendung der gelisteten Fachbegriffe sind im Stichwortverzeichnis angegeben, die verwendeten Ausgaben der genannten Werke im Literaturverzeichnis.

ACL	<i>Association for Computational Linguistics</i>	MAE	<i>Mean average error</i>
API	<i>Application Programming Interface</i>	MXBT	<i>Meng xi bi tan 夢溪筆談</i>
AYL	<i>Average Year of Lexicalization</i>	NER	<i>Named Entity Recognition</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>	NLLR	<i>Normalized Log-Likelihood-Ratio</i>
BoW	<i>Bag of Words</i>	NLP	<i>Natural Language Processing</i>
CBDB	<i>China Biographical Database</i>	OCR	<i>Optical Character Recognition</i>
CKIP	<i>Chinese Knowledge and Information Processing</i>	OED	<i>Oxford English Dictionary</i>
CS	<i>Cosine Similarity (Kosinus-Ähnlichkeit)</i>	PCA	<i>Principal Component Analysis</i>
CTB	<i>Chinese Treebank</i>	PoS	<i>Part-of-speech</i>
DDBC	<i>Dharma Drum Buddhist College Time Authority Database</i>	RegEx	<i>Regular Expression (regulärer Ausdruck)</i>
DFZ	<i>Difangzhi 地方誌</i>	RMRB	<i>Renmin ribao 人民日報</i>
DH	<i>Digital Humanities</i>	SAYL	<i>Standardized Average Year of Lexicalization</i>
DHYDCD	Digitale Ausgabe des » HYDCD	SKQS	<i>Si ku quan shu 四庫全書</i>
GB(K)	<i>Guojia Biaozhun (Kuo-zhan) 國家標準 (擴展)</i>	SLM	<i>Statistical Language Model</i>
HHS	<i>Hou Han shu 後漢書</i>	SQL	<i>Structured Query Language</i>
HMM	<i>Hidden-MARKOV-Modell</i>	TE	<i>Temporale Entropie</i>
HYDCD	<i>Hanyu da cidian 漢語大詞典</i>	TEI	<i>Text Encoding Initiative</i>
idf	<i>inverse document frequency</i>	tf-idf	<i>term frequency inverse document frequency</i>
KLD	<i>KULLBACK-LEIBLER Divergenz</i>	UTF-8	<i>8 Bit Unicode Transformation Format</i>
KPCh	<i>Kommunistische Partei Chinas</i>	u. Z.	<i>unserer Zeitrechnung</i>
LCM	<i>Largest chronon minimum</i>	v. u. Z.	<i>vor unserer Zeitrechnung</i>
		FWAYL	<i>Frequency Weighted » AYL</i>
		XML	<i>Extensible Markup Language</i>
		XXSKQS	<i>Xu xiu si ku quan shu 續修四庫全書</i>