

# Abbildungsverzeichnis

1.1	Klassisches und schriftsprachliches Chinesisch . . . . .	6
2.1	Formalisierung von Sprachwandel als s-Kurve . . . . .	15
2.2	<i>Zhengshi</i> 正史 nach Jahr der Veröffentlichung . . . . .	22
2.3	Abnehmende Nutzung schriftsprachlicher Partikel in den <i>zhengshi</i> 正史 . . . . .	24
2.4	<i>Zhi</i> 之 als Pronomen und subordinierende Partikel in den <i>zhengshi</i> . . . . .	26
2.5	Verwendung einiger Amtstitel in den <i>zhengshi</i> . . . . .	28
2.6	Vorkommen von <i>fo/bi/fu</i> 佛 und <i>seng</i> 僧 in den <i>zhengshi</i> -Texten . . . . .	30
2.7	Vorkommen buddhistischer Lexeme in den <i>zhengshi</i> -Texten . . . . .	33
3.1	PCA, 1.000 häufigste Lexeme – <i>zhengshi</i> 正史, LOEWE und <i>Xiaoshuo</i> 小說 . . . . .	42
3.2	Kumulative Häufigkeit von „Archaismen“ und „Neologismen“ . . . . .	49
4.1	<i>F-Scores</i> der <i>Jieba-Segmenter</i> . . . . .	85
4.2	<i>Recall</i> der <i>Jieba Modi</i> , diachrone Goldstandards . . . . .	86
4.3	<i>Precision</i> der <i>Jieba Modi</i> , diachrone Goldstandards . . . . .	86
4.4	<i>F-Scores</i> der im Test besten Tokenizer für alle Goldstandard-Texte . . . . .	90
4.5	<i>n</i> -Gramm Effizienz am Beispiel von <i>Hong lou meng</i> 紅樓夢 . . . . .	92
4.6	<i>Recall</i> der getesteten Tokenizer vs. Verwendung von 1–4-Grammen . . . . .	95
4.7	<i>F-Score</i> der getesteten Tokenizer vs. <i>ChronLex</i> . . . . .	97
4.8	Länge unterschiedlicher Namen in der <i>CBDB</i> in Zeichen, anteilig . . . . .	101
4.9	2–3 Gramme einzigartiger Namen in der <i>CBDB</i> nach Jahrhundert . . . . .	102
5.1	Eintrag <i>shiyou</i> 石油 („Steinöl“, Erdöl) in der Originalausgabe des <i>HYDCD</i> . . . . .	120
5.2	Beispielzeilen aus den Tabellen <i>the_words</i> , <i>the_books</i> , <i>the_citations</i> . . . . .	126
5.3	Neologismusprofil des <i>Shiji</i> 史記, ohne Korpusbelegstellen . . . . .	135
5.4	Neologismusprofil des <i>Shiji</i> 史記, mit Korpusbelegstellen . . . . .	136
5.5	Genauigkeit der Lexemdatierung . . . . .	140
5.6	„Unterschiedliche“ Bedeutungen in <i>HYDCD</i> -Einträgen . . . . .	141
5.7	Lexikalisierte Zeichen im <i>DHYDCD</i> nach Anzahl ihrer Lesungen . . . . .	142
5.8	Lexikalisierung im <i>HYDCD</i> nach Jahrhundert . . . . .	143
5.9	Lexikalisierung mit zusätzlichen Korpusbelegstellen . . . . .	145
5.10	Lexikalisierung im <i>DHYDCD</i> nach Jahrhundert (ohne zusätzliche Belegstellen) . . . . .	146
5.11	Lexikalisierung neuer Schriftzeichen im <i>DHYDCD</i> . . . . .	147
5.12	Lexikalisierung 3- und 4-silbiger Wörter . . . . .	147
5.13	Chronologisierbare Lexikalisierung im <i>DHYDCD</i> nach Länge der Lexeme . . . . .	149

5.14	Länge der Belegstellen im <i>DHYDCD</i> nach Jahrhundert . . . . .	149
5.15	Lexikalisierung im <i>DHYDCD</i> nach Jahrhundert – häufigste <i>Locus classicus</i> -Texte . .	152
6.1	Performance mit 1–2-Gramm <i>difangzhi</i> 地方誌-Sprachmodell . . . . .	160
6.2	Performance von 1–2-Zeichen <i>difangzhi</i> -Lexem-Sprachmodellen . . . . .	161
6.3	Performance von <i>difangzhi</i> -Sprachmodellen mit temporalen Ausdrücken . . . . .	162
6.4	Performance von <i>difangzhi</i> -Sprachmodellen mit temporalen Ausdrücken (Details) .	163
6.5	Experiment mit 105 zufälligen Texten aus dem <i>XXSKQS</i> . . . . .	176
6.6	KULLBACK-LEIBLER-Divergenz zum <i>chronon</i> des vorangegangenen Jahrhunderts .	177
6.7	KULLBACK-LEIBLER-Divergenz zum vorigen <i>chronon</i> . . . . .	178
6.8	JACCARD-„Divergenz“ zum <i>chronon</i> des vorigen Jahrhunderts . . . . .	179
6.9	Neologismusprofil <i>Meng xi bi tan</i> 夢溪筆談 mit Gewichtungskorrektur . . . . .	180
6.10	2–4 Zeichen Lexeme im <i>MXBT</i> chronologisch vs. Häufigkeit im Text . . . . .	181
6.11	Profile des <i>MXBT</i> (nur <i>DHYDCD</i> -Belege; + LOEWE-/ <i>zhengshi</i> Belege; + <i>DFZ</i> -Belege)	183
6.12	Neologismusprofile des <i>Shiji</i> ohne und mit zusätzlichen Korpus-Belegen . . . . .	184
6.13	Neologismusprofil für „鄜延境内有石油，舊說高奴縣出脂水，即此也。“ . . . . .	185
6.14	Neologismusprofil für das <i>Meng xi bi tan</i> vs. Lexikalisierung im <i>HYDCD</i> . . . . .	185
6.15	Neologismusprofile des <i>MXBT</i> (ohne Korpus-Belegstellen) . . . . .	186
6.16	Neologismusprofil für das <i>MXBT</i> (s-Gewichtungskorrektur) . . . . .	187
6.17	Kumulative Neologismusprofile des <i>Meng xi bi tan</i> (ohne Korpusbelegstellen) . . .	189
6.18	Temporales Textprofil für das <i>Meng xi bi tan</i> . . . . .	190
6.19	Temporale Profile des <i>Shiji</i> 史記 . . . . .	191
6.20	Temporale Profile des <i>Qingshi gao</i> 清史稿 . . . . .	191
6.21	Temporales Profil des <i>Zhongjing</i> 忠經, ohne Gewichtungskorrektur . . . . .	192
6.22	Anteil <i>false positives</i> (zu neu datierte Lexeme) nach Textveröffentlichung . . . . .	198
6.23	Korrelationsmatrix: <i>types</i> vor und zur Veröffentlichung und Gesamtanzahl <i>types</i> . .	198
6.24	Korrelation Lexem- <i>types</i> zur und vor Veröffentlichung und Gesamtanzahl <i>types</i> . .	199
6.25	Temporales Profil des <i>Guide fu zhi</i> 歸德府志 von 1754 . . . . .	201
6.26	Namen in den Trainingsdaten . . . . .	202
6.27	Performance profilbasierter Datierung, <i>Difangzhi</i> , 2–3 Zeichen-Lexeme . . . . .	204
6.28	Performance profilbasierter Datierung, <i>XXSKQS</i> , 2–3 Zeichen-Lexeme . . . . .	206
6.29	Performance Profildatierung, <i>zhengshi</i> , 2–3 Zeichen-Lexeme / <i>temporal expressions</i> .	206
6.30	Temporales Profil des <i>Yuan shan</i> 原善 . . . . .	209
6.31	Korrelation Veröffentlichung <i>zhengshi</i> , AYL bei 100 % 2–4 Zeichen <i>types</i> . . . . .	212
6.32	Vergleich linearer AYL-Modelle . . . . .	216
6.33	Korrelation Veröffentlichung <i>zhengshi</i> , AYL mit 15 % 2–4 Zeichen Lexem- <i>types</i> . . .	221
6.34	Diagnoseplots (Residuen, <i>Normal Q-Q</i> , <i>Cook's distance</i> , <i>Leverage</i> ) . . . . .	222
6.35	Korrelation Veröffentlichung LOEWE / <i>zhengshi</i> , AYL . . . . .	225
6.36	Vergleich linearer Modelle, 80 % häufigste 2–3-Zeichen-Lexeme . . . . .	227
6.37	Datierungsergebnis <i>Difangzhi</i> mit 80 % häufigsten 2–3-Zeichen-Lexemen . . . . .	228
6.38	<i>VisualTime</i> Startseite – Datei auswählen und hochladen . . . . .	233
6.39	<i>VisualTime</i> Ergebnisseite . . . . .	234
6.40	<i>VisualTime</i> – Temporales Profil des <i>Sanguo zhi</i> mit flexibel aktivierbaren <i>types</i> . . . .	235
6.41	<i>VisualTime</i> – Erkannte Namen im <i>Sanguo zhi</i> (Ausschnitt) . . . . .	235
6.42	<i>VisualTime</i> – Anzeige aller Textstellen mit HUANG Zhongtong 黃中通 im <i>Sanguo zhi</i>	236
6.43	<i>VisualTime</i> – Anzeige <i>NLLR</i> des <i>Sanguo zhi</i> zu einzelnen <i>DHYDCD</i> - <i>chronons</i> . . . . .	237

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