

Inhaltsverzeichnis

Abbildungsverzeichnis	ix
Tabellenverzeichnis	xi
1 Einleitung	I
2 Sprach- und Wortschatzwandel	II
2.1 Das PIOTROWSKI-Gesetz	14
2.2 Sprachwandel mit „chinesischen Besonderheiten“	16
2.3 Sprachwandel im Chinesischen am Beispiel der <i>zhengshi</i> 正史	20
3 Linguistische Datierung	35
3.1 Computerlinguistische Datierung von Texten	40
3.2 Datierung als Kategorisierungsproblem: DE JONG, RODE und HIEMSTRA	43
3.3 Systematisierung der bestehenden Ansätze	45
4 Computerlinguistische Methoden für schriftsprachliches Chinesisch	59
4.1 Forschungslandschaft	60
4.2 Korpora	62
4.3 Vorverarbeitung und Normalisierung	69
4.4 Tokenisierung & <i>Part-of-Speech</i> Tagging	73
4.5 Tokenisierung für schriftsprachliches Chinesisch: eine Feldstudie	77
4.5.1 Gesamtvergleich der getesteten Segementer	89
4.5.2 <i>n</i> -Gramm Zerlegung	91
4.5.3 Zurück zur <i>Bag of Words</i>	94
4.6 <i>ChronLex</i> – ein Segementer-Experiment	95
4.7 <i>Named Entity Recognition (NER)</i> und die <i>China Biographical Database (CBDB)</i>	97
4.8 <i>Temporal Expressions</i> und die <i>Time Authority Database</i>	103
5 Das <i>Hanyu da cidian</i> 漢語大詞典 als Datenquelle	107
5.1 Eine kurze Geschichte des <i>HYDCD</i>	109
5.2 Das Vorbild: <i>Oxford English Dictionary</i>	111
5.3 Aufbau und Inhalt des <i>HYDCD</i>	113
5.4 Digitale Ausgaben des <i>HYDCD</i>	115
5.4.1 Qualitätssicherung: Abgleich mit der gedruckten Ausgabe	116

5.5	Erzeugung einer diachronen Lexemdatenbank	120
5.5.1	Datenstruktur	121
5.5.2	Verwendung der Metadaten aus dem <i>DHYDCD</i>	127
5.5.3	Gewinnung von Daten aus der <i>China Biographical Database</i>	132
5.5.4	Ergänzung um frühere Belegstellen	134
5.6	Das <i>DHYDCD</i> als diachrones Behelfskorpus	137
5.7	<i>HYDCD-Data Science</i> : Erkenntnisse aus der Datenbank	138
5.7.1	Genauigkeit der gewonnenen Daten	139
5.7.2	Lexikalisierung pro Jahrhundert	142
5.7.3	Mono- und Polysyllabizität	146
5.7.4	Lexikalisierung nach <i>Locus classicus</i>	150
6	Textdatierung für schriftsprachliches Chinesisch	155
6.1	Datierung als Kategorisierungsproblem	156
6.1.1	Datierung mit <i>difangzhi</i> 地方誌 Sprachmodellen	158
6.1.2	Co-Datierung von Dokumenten	169
6.1.3	Datierung mit <i>DHYDCD</i> -Sprachmodellen	171
6.1.4	Sprachwandel im Sprachmodell	177
6.2	Datierung mit Neologismusprofilen	179
6.2.1	Erzeugung von Neologismusprofilen	182
6.2.2	Temporale Textprofile: Erweiterung um Namen und Zeitausdrücke	189
6.2.3	Interpretation temporaler Textprofile	190
6.2.4	Das <i>Zhongjing</i> 忠經 als Anwendungsbeispiel	192
6.2.5	Automatisierte Datierung mit temporalen Textprofilen	197
6.3	Datierung mit dem „durchschnittlichen Lexemalter“ von Texten	210
6.3.1	Ein optimiertes <i>AYL</i> -Regressionsmodell	220
6.3.2	<i>AYL</i> mit unterschiedlichen Korpora	224
6.4	Untersuchte Datierungsmethoden im Überblick	229
6.4.1	<i>VisualTime — user interface</i> für Datierungsmethoden	232
7	Ergebnisse und Ausblick	243
7.1	Ausblick	248
	Literaturverzeichnis	251
	Stichwortverzeichnis	281
	Epilog	285
	Abstract	287
	論文提要	289