

Steffen Hessler

Autorschafts- erkennung und Verstellungsstrategien

Textanalysen und -vergleiche im Spektrum forensischer
Linguistik, Informationssicherheit und Machine-Learning

TBL Tübinger Beiträge zur Linguistik

**narr/f
ranck
e\atte
mpto**

Inhalt

1	Fragestellungen und Aufbau der Arbeit	15
2	Aufgaben und Tätigkeitsbereiche der forensischen Linguistik und Fragestellungen im Bereich der Autorenerkennung	19
3	Forschungsüberblick und Anwendungsbereiche	25
4	Weitere relevante Forschungsdisziplinen	33
4.1	Sozio- und Varietätenlinguistik	33
4.2	Linguistische Semantik und Pragmatik	35
4.3	Sprachkontaktforschung	36
4.4	Medienlinguistik und Online-Kommunikation	37
4.5	Fehlerforschung	37
5	Textanalysen und Textvergleiche im Bereich Autorschaftsanalyse .	39
5.1	Individuelles Sprachverhalten vs. Gesamtpopulation	39
5.2	Unvoreingenommenheit bei Textanalysen und insbesondere Textvergleichen	40
5.3	Textanalysen	40
5.4	Textvergleiche	43
5.5	Wahrscheinlichkeitsaussagen und Nicht-Linearität bei Textvergleichen	45
5.6	Hypothesenbildung bei der Autorschaftsanalyse	47
5.7	Keine exklusive Nutzung eines Ausdrucks und Sprachwandel	48
5.8	Zusammenfassung	48
6	Strittige Punkte und Forschungsdebatten	51
6.1	Qualitative und quantitative Analysen	51
6.2	Individualstil und sprachlicher Fingerabdruck vs. Varietäten und sprachlicher Fingerzeig	52

6.3	Multiple Autorschaft	57
6.4	Muttersprachliches Selbstbewusstsein	60
7	Autorschaftserkennung in multimedialen Umgebungen	63
8	Datengrundlage und Hinweise zum Datenschutz	65
8.1	Kenntnis des Mediums und des Untersuchungsgegenstandes	66
8.2	Deutschsprachige Amazon-Rezensionen	67
8.3	Englischsprachige Amazon-Rezensionen	67
8.4	Online-Foren (Gaming und Börse)	68
8.5	Inkriminierte Textserie	68
8.6	Medial stilisierte Sprache	69
8.7	Datenschutz, Urhebererschaft, gute wissenschaftliche Praxis sowie Opferschutz im Zusammenhang mit Autorenanalyse	69
9	Textanalysen – Methoden der Fehler- und Stilanalyse	71
9.1	Fehlerdefinition – Normiertheit und Regelwerke	71
9.2	Fehlerforschung	74
9.3	Fehlerbeschreibung, sprachliche Ebenen und Erscheinungsbild von Fehlern	78
9.4	Fehlertypen und Fehlerschwere	81
9.5	Fehlerursachen	83
9.6	Muttersprachliche, lernersprachliche und fingierte Fehler ..	84
9.7	Markiertheitstheorie – Stil, Fehler und Markiertheit	85
9.8	Stilanalysen und stilistische Merkmale	87
9.9	Reliabilität und Validität stilistischer Merkmale	90
9.10	Kurzer Exkurs: Kritik am zu vereinfachten Umgang mit Metadaten	92
9.11	Merkmalsets und Merkmalsetmuster	93
9.12	Qualitative Analysen	96
9.13	Quantitative Analysen	97
9.14	Zusammenfassung	98
10	Kriminaltechnisches Informationssystem Texte (KISTE)	101
10.1	Befunde in KISTE – Befundkategorien	104
10.2	Zusammenfassung	111

11	Autorenerkennung in Online-Umgebungen	115
11.1	Der Medienbegriff in dieser Arbeit	115
11.2	Das Internet als Medium	116
11.3	Prototypische Eigenschaften der Internetkommunikation ..	117
11.4	Sprache in Internetforen und Online-Rezensionen – konzeptionelle Mündlichkeit?	120
11.5	Problematik internetbasierter Daten und Fluidity im Bedeutungsspektrum forensischer Linguistik	121
11.6	Zusammenfassung	123
12	Erster empirischer Teil – Probeerhebungen – Rezensionen der Produktkategorie Cola	125
12.1	Diatopische Variation	126
12.2	Analyse von Cola-Rezensionen auf amazon.de	128
12.3	Zusammenfassung	136
13	Zweiter empirischer Teil – Verdacht auf Verbreitung von Desinformation und Verstellungsstrategien in Social Media	137
13.1	Verdacht auf Verschleierungsstrategien und Verbreitung von Desinformation in einem Internetforum	138
13.2	Kursmanipulationen an der Börse	142
13.3	Fachsprachlichkeit in Börsenforen	143
13.4	Verstellungsstrategien im Zusammenhang mit Kursmanipulationen	146
13.5	Zusammenfassung	151
14	Erpresserschreiben	153
14.1	Anonymität und Nicht-Öffentlichkeit von Erpresserschreiben	153
14.2	Die Textfunktionen von Erpresserschreiben und Abgrenzung von Drohbriefen	154
14.3	Das Erpresserschreiben als Textsorte	155
14.4	Einzelne Komponenten der Textsorte Erpresserschreiben ..	158
14.5	Zusammenfassung	162
15	Verstellungsstrategien	163
15.1	Anonymität, fingierte Autorschaft und metasprachliches Bewusstsein	163

15.2	Sprachliche Verstellungsstrategien – Grundlegende Übersicht	164
15.3	Verstellungsstrategien als sprachlicher Stil	165
15.4	Grenzen der Verstellung	166
15.5	Sich verstellen bzw. lügen	168
15.5.1	„Fake News“ und „Lügenpresse“	169
15.5.2	Lügen im psychologischen und juristischen Sinn ...	170
15.5.3	Lügen im linguistischen Sinn in Abgrenzung zum laienhaften Verständnis von Lügen	171
15.5.4	Lügen als Veränderung des sprachlichen Stils	173
15.5.5	Verstellungsstrategien als eine bestimmte Variante des Lügens und Täuschens	174
15.5.6	Zusammenfassung	175
15.6	Dissimilatorische und simulatorische Verstellungsstrategien	175
15.6.1	Dissimilatorische Verstellungsstrategien – Verschleierung	177
15.6.2	Simulatorische Verstellungsstrategien – Imitation ..	177
15.6.3	Imitation als Verstellungsstrategie und die Darstellung von Autorität	178
15.6.4	Imitation als Verstellungsstrategie und varietätenlinguistische Fragestellungen	181
15.6.5	Zusammenfassung	182
15.7	Imitation von Nicht-Muttersprachlichkeit als Verstellungsstrategie	183
15.7.1	Die unspezifische Imitationsstrategie Foreigner Talk	189
15.7.2	Spezifik von Einzelsprachen vs. Nicht-Spezifität der Darstellung von Nicht-Muttersprachlichkeit	190
15.7.3	Die Bedeutung von Code-Switching für Verstellungsstrategien als Nicht-Muttersprachler/in	190
15.7.4	Tertiärer Ethnolekt, Foreigner Talk und Xenolekt ..	192
15.7.5	Der Weg ins Bewusstsein – Mediale Vorbilder, Verbreitung und Generierung	193
15.7.6	Vorsicht bei Rückschlüssen auf Mehrsprachigkeit bei der Analyse einzelner Merkmale	195

	15.7.7 Untersuchungen zur Imitation von Nicht-Muttersprachlichkeit als Verstellungsstrategie	196
	15.7.8 Zusammenfassung	198
15.8	Grad der Verstellung und Detektion	199
	15.8.1 Fingierte Fehler im Rahmen von Verstellungsstrategien	201
	15.8.2 (In-)Konsistenz von Merkmalen bei Verstellungen und Spracherwerb	205
	15.8.3 Zusammenfassung	206
16	Korpuslinguistische Analysen von Verstellungsstrategien	209
16.1	Stilanalyse	210
	16.1.1 Autorenstilisierungen	210
	16.1.2 Datumsangaben, Grußformeln und Anreden	211
	16.1.3 Groß- und Kleinschreibung	212
	16.1.4 Interpunktion	214
	16.1.5 Fachsprachen	216
	16.1.6 Register	219
	16.1.7 Lexikalische Komplexität	222
16.2	Fehleranalyse	224
	16.2.1 Getrennt- und Zusammenschreibung	224
	16.2.2 Interpunktion	225
	16.2.3 Alte und neue deutsche Rechtschreibung	225
	16.2.4 Orthographie	225
	16.2.5 Syntax	227
	16.2.6 Morphologie	230
	16.2.7 Orthographie oder Morphologie	232
	16.2.8 Lexik	235
	16.2.9 Konstanz der Fehlerhaftigkeit und Plausibilität	236
16.3	Zusammenfassung	237
17	Dritter empirischer Teil – Analyse von Verstellungsstrategien im Rahmen einer Textserie inkriminierter Schreiben	241
17.1	Verstellungsstrategien – Verschleierung, Imitation und Stilisierung	243
17.2	Aspekte von Verstellungsstrategien	243
	17.2.1 Saliente, sprecheridentifizierende Merkmale und Gesamtpopulation	244

17.2.2	Konsistenz und Plausibilität	244
17.2.3	Stufen schriftsprachlicher Kompetenz	245
17.2.4	Errors / Mistakes / Tippfehler – Das Auftreten von korrekten neben falschen Formen	245
17.2.5	Zugänglichkeit – Oberflächliche und tiefer liegende sprachliche Ebenen	246
17.2.6	Inkonsistenzen im Verlauf eines Textes und Unterschiede zwischen mehreren Texten	246
17.2.7	Sprachmischungen	246
17.2.8	Xenolekt vs. Sprachkontaktphänomene	247
17.2.9	Der Konflikt der Verständnissicherung	247
17.3	Anmerkungen zur Arbeitsweise bei der Analyse	247
17.4	Heterogene schriftsprachliche Kompetenz und keine eindeutige Verstellung in den Texten X1–X10	248
17.5	Verschleierung der eigenen schriftsprachlichen Kompetenz und Imitation von Nicht-Muttersprachlichkeit in den Texten X11–X21	252
17.5.1	Orthographie	252
17.5.2	Morphologie	269
17.5.3	Lexik	278
17.5.4	Syntax	283
17.5.5	Darstellung von Fremdsprachlichkeit – Fremdsprachliche Lexik	284
17.5.6	Zwischenfazit zur Darstellung von Fremdsprachlichkeit	293
17.5.7	Darstellung von Fremdsprachlichkeit – Pseudofremdsprachliche Endungen	294
17.5.8	Hoher Grad der Verstellung	298
17.5.9	Widersprüche bei der schriftsprachlichen Kompetenz	298
17.5.10	Nicht-Plausibilität und Inkonsistenz anhand konkreter Textvergleiche	308
17.5.11	Vergleich der unterschiedlichen Realisationen eines Lexems in einem Text	310
17.5.12	Sonstige Auffälligkeiten	310
17.6	Fallenlassen der bisherigen Verstellungsstrategie in den Texten X22–X36	311
17.6.1	Analyse von Text X22	311
17.6.2	Analyse der Texte X23–X26	313

	17.6.3 Analyse der Texte X27–X30	314
	17.6.4 Analyse der Texte X31–X36	315
17.7	Aspekte der Stilisierung und Analyse der Stilisierungsstrategie	316
	17.7.1 Mediale Stilisierung	317
	17.7.2 Kommunikationsziele der Verstellungsstrategien im Vergleich	323
	17.7.3 Die Stilisierungsstrategie im Rahmen der inkriminierten Textserie im Vergleich mit medial stilisierten Texten	325
	17.7.4 Medial stilisierte Sprache als Vorbild für die Stilisierungsstrategie	338
17.8	Zusammenfassung	340
18	Teil-automatisierte Verfahren für Textvergleiche	343
	18.1 Quantitative, statistische und automatisierte Ansätze	344
	18.2 Datengrundlage	348
	18.3 Interne Struktur von AdHominem	349
	18.3.1 Trainingsset	350
	18.3.2 Test-Set	351
	18.4 Anpassungen der Texte durch AdHominem	352
	18.5 Markierungen bzw. Attentions in AdHominem	357
19	Vierter empirischer Teil – Analysen der Textvergleiche von AdHominem	359
	19.1 Probeerhebungen	360
	19.2 Einzelanalysen von Merkmalen innerhalb der Textvergleiche von AdHominem	362
	19.2.1 Zeichensetzung	364
	19.2.2 Orthographie	367
	19.2.3 Diatopische Variationen & Fremdsprachen	376
	19.2.4 Stilistik	378
	19.2.5 Syntax	379
	19.2.6 Kombinierte Merkmale	381
	19.3 Zusammenfassung, Herausforderungen und Ziele	385
20	Fazit und Ausblick	389