



# Computergesteuerte etymologische Wörterbücher

Konzeption und Umsetzung  
anhand des Französischen

Sylvia Weber

Sylvia Weber  
Computergesteuerte etymologische Wörterbücher



Sylvia Weber

# Computergesteuerte etymologische Wörterbücher

Konzeption und Umsetzung anhand des Französischen

**F** Frank & Timme  
Verlag für wissenschaftliche Literatur

Umschlagabbildung © Matthew Horner und Sylvia Weber

ISBN 978-3-7329-0710-6

ISBN E-Book 978-3-7329-9265-2

ISSN 1862-6149

© Frank & Timme GmbH Verlag für wissenschaftliche Literatur  
Berlin 2020. Alle Rechte vorbehalten.

Das Werk einschließlich aller Teile ist urheberrechtlich geschützt.  
Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar.  
Das gilt insbesondere für Vervielfältigungen, Übersetzungen,  
Mikroverfilmungen und die Einspeicherung und Verarbeitung in  
elektronischen Systemen.

Herstellung durch Frank & Timme GmbH,  
Wittelsbacherstraße 27a, 10707 Berlin.  
Printed in Germany.

Gedruckt auf säurefreiem, alterungsbeständigem Papier.

[www.frank-timme.de](http://www.frank-timme.de)

Diese Arbeit wurde 2020 als Dissertation mit dem Titel  
„Konzeption und Umsetzung eines computergesteuerten  
etymologischen Wörterbuchs des Französischen“ an der  
Philosophischen Fakultät der Friedrich-Alexander-Universität  
Erlangen-Nürnberg verteidigt.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>XI</b>
<b>Tabellenverzeichnis</b>	<b>XV</b>
<b>Quellcode-Verzeichnis</b>	<b>XIX</b>
<b>Abkürzungsverzeichnis</b>	<b>XXIII</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Gegenstand . . . . .	4
1.2 Nutzen . . . . .	5
1.3 Einordnung in die Fachgebiete: Das Verhältnis von Computerlinguistik und Sprachwissenschaft . . . . .	6
1.4 Ziele . . . . .	19
1.5 Vorbemerkungen zur Notation . . . . .	21
1.6 Aufbau . . . . .	21
<b>2 Stand der Forschung</b>	<b>23</b>
2.1 Überblick . . . . .	23
2.2 Burton-Hunter (1976) . . . . .	32
2.3 Eastlack (1977): <i>Iberochange</i> . . . . .	32
2.4 Hartman (1981–2018): <i>Phono</i> . . . . .	33
2.5 Schmuki (2000): <i>Etymo</i> . . . . .	36
2.6 Kritische Zusammenfassung . . . . .	38
<b>3 Praktische Umsetzung</b>	<b>43</b>
3.1 Allgemeine Betrachtungen . . . . .	44

3.1.1	Aufbau des Systems . . . . .	45
3.1.2	Merkmalstrukturen . . . . .	48
3.2	Wortformerkennung . . . . .	49
3.2.1	LA-Grammatik und JSLIM . . . . .	49
3.2.2	Aufbau und Funktionsweise der Wortformerkennungskomponente im <i>CEW</i> . . . . .	55
3.2.3	Unterschiede zu JSLIM . . . . .	63
3.2.4	Umsetzung im C#-Code . . . . .	65
3.2.5	Zusammenfassung . . . . .	67
3.3	Phonetisch-phonologische Vorverarbeitung und Nachbereitung . . . . .	68
3.3.1	Transkription . . . . .	69
3.3.2	Silbentrennung . . . . .	71
3.3.3	Betonung . . . . .	74
3.3.4	Umsetzung im C#-Code . . . . .	75
3.3.5	Zusammenfassung . . . . .	77
3.4	Lautentwicklungen . . . . .	78
3.4.1	Regelaufbau und Regelbedingungen . . . . .	80
3.4.2	Mapping von Lauten und Graphemen . . . . .	84
3.4.3	Weitere Änderungen . . . . .	97
3.4.4	Umsetzung im C#-Code . . . . .	98
3.4.5	Zusammenfassung . . . . .	99
3.5	Morphologiebasierte Entwicklungen . . . . .	100
3.6	Abgleich zwischen Quell- und Zielsprache . . . . .	103
3.7	Graphische Oberfläche . . . . .	104
3.8	Zusammenfassung . . . . .	108
<b>4</b>	<b>Die Umsetzung des <i>CEW</i> am Beispiel des Französischen (<i>CEW-fr</i>)</b>	<b>109</b>
4.1	Sprachgeschichtlicher Überblick . . . . .	109
4.1.1	Externe Sprachgeschichte . . . . .	110
4.1.2	Definitionen und damit verbundene Schwierigkeiten aus compu- terlinguistischer Sicht . . . . .	112
4.1.3	Interne Sprachgeschichte . . . . .	118
4.2	Implementierung der lateinischen Wortformerkennung . . . . .	122

4.2.1	Ausgangsdaten . . . . .	124
4.2.2	Substantive . . . . .	127
4.2.3	Adjektive . . . . .	133
4.2.4	Verben . . . . .	136
4.2.5	Unregelmäßige Paradigmen . . . . .	142
4.2.6	Weitere Wortarten . . . . .	143
4.2.7	Zusammenfassung . . . . .	144
4.3	Phonetisch-phonologische Vorverarbeitung der lateinischen Wortform . .	145
4.3.1	Transkription der lateinischen Grapheme . . . . .	145
4.3.2	Lateinische Silbentrennung . . . . .	154
4.3.3	Betonung lateinischer Wortformen . . . . .	156
4.3.4	Zusammenfassung . . . . .	158
4.4	Implementierung von Lautregeln . . . . .	158
4.4.1	Die Wahl einer geeigneten Datengrundlage . . . . .	158
4.4.2	Methodologie und Aufbau von Rheinfielders <i>Altfranzösische Gram-</i> <i>matik</i> . . . . .	162
4.4.3	Überblick über die Implementierung . . . . .	165
4.5	Morphologiebasierte Entwicklungen am Beispiel des Deklinationssystems	183
4.5.1	Vulgärlateinische Veränderungen . . . . .	183
4.5.2	Entwicklungen zum Altfranzösischen . . . . .	187
4.5.3	Entwicklungen zum Mittel- und Neufranzösischen . . . . .	196
4.5.4	Zusammenfassung . . . . .	197
4.6	Implementierung der französischen Wortformerkenung . . . . .	197
4.6.1	Ausgangsdaten . . . . .	198
4.6.2	Substantive . . . . .	200
4.6.3	Adjektive . . . . .	202
4.6.4	Verben . . . . .	204
4.6.5	Unregelmäßige Paradigmen . . . . .	211
4.6.6	Nicht flektierbare Wortarten . . . . .	211
4.6.7	Zusammenfassung . . . . .	212
4.7	Phonetisch-phonologische Nachbereitung der französischen Wortform: Transkription . . . . .	213
4.8	Abgleich der lateinischen und neufranzösischen Wortform . . . . .	235

4.9 Zusammenfassung . . . . .	236
<b>5 Evaluierung des Systems</b>	<b>237</b>
5.1 Quantitative Analyse . . . . .	237
5.2 Qualitative Analyse . . . . .	240
5.3 Schwierigkeiten . . . . .	248
5.3.1 Wortformerkenntung . . . . .	249
5.3.2 Phonetisch-phonologische Vorverarbeitung und Nachbereitung . .	250
5.3.3 Sprachwandel . . . . .	257
5.3.4 Abgleich der lateinischen und neufranzösischen Wortform . . . . .	266
5.4 Zusammenfassung . . . . .	266
<b>6 Zusammenfassung und Ausblick</b>	<b>269</b>
6.1 Umsetzung der Ziele . . . . .	269
6.2 Ausbaumöglichkeiten . . . . .	272
6.2.1 <i>CEW-fr</i> . . . . .	272
6.2.2 <i>CEW</i> . . . . .	273
<b>Literaturverzeichnis</b>	<b>277</b>
<b>Software und elektronische Datensätze</b>	<b>297</b>
<b>Anhänge</b>	
<b>A Morphologie der lateinischen Verben im <i>CEW-fr</i></b>	<b>299</b>
A.1 Erste Konjugation (Flexionsklasse <i>cantare</i> ) . . . . .	299
A.2 Zweite Konjugation (Flexionsklasse <i>tenere</i> ) . . . . .	300
A.3 Dritte Konjugation (Flexionsklasse <i>regere</i> ) . . . . .	301
A.4 Vierte Konjugation (Flexionsklasse <i>punire</i> ) . . . . .	302
<b>B Morphologie der französischen Verben im <i>CEW-fr</i></b>	<b>305</b>
B.1 Flexionsklasse <i>chanter</i> . . . . .	305
B.2 Flexionsklasse <i>appeler</i> . . . . .	306
B.3 Flexionsklasse <i>appuyer</i> . . . . .	306
B.4 Flexionsklasse <i>lever</i> . . . . .	307

B.5 Flexionsklasse <i>préférer</i> . . . . .	308
B.6 Flexionsklasse <i>changer</i> . . . . .	309
B.7 Flexionsklasse <i>commencer</i> . . . . .	309
B.8 Flexionsklasse <i>assiéger</i> . . . . .	310
B.9 Flexionsklasse <i>apiécer</i> . . . . .	311
B.10 Flexionsklasse <i>sortir</i> . . . . .	312
B.11 Flexionsklasse <i>agir</i> . . . . .	312
B.12 Flexionsklasse <i>apercevoir</i> . . . . .	313
B.13 Flexionsklasse <i>ouvrir</i> . . . . .	314
B.14 Flexionsklasse <i>rompre</i> . . . . .	315
B.15 Flexionsklasse <i>vendre</i> . . . . .	315
B.16 Flexionsklasse <i>battre</i> . . . . .	316
B.17 Flexionsklasse <i>prendre</i> . . . . .	317
B.18 Flexionsklasse <i>conduire</i> . . . . .	318
B.19 Flexionsklasse <i>coudre</i> . . . . .	318
B.20 Flexionsklasse <i>croire</i> . . . . .	319
B.21 Flexionsklasse <i>accroître</i> . . . . .	320
B.22 Flexionsklasse <i>atteindre</i> . . . . .	321
<b>C Liste der verwendeten Symbole im <i>CEW</i></b> . . . . .	<b>323</b>
C.1 Symbole in regulären Ausdrücken . . . . .	323
C.2 Allgemeine Symbole im <i>CEW</i> . . . . .	324
C.3 Sprachspezifische Symbole . . . . .	324
C.4 Verwendete X-SAMPA-Symbole . . . . .	325
<b>D Ausschnitte aus den einzelsprachlichen Textdateien</b> . . . . .	<b>327</b>
D.1 Lateinische Wortformerkenung . . . . .	327
D.1.1 Lexikodateien . . . . .	327
D.1.2 Allomorphiedateien . . . . .	330
D.1.3 Flexionsdateien . . . . .	331
D.1.4 Weitere Dateien . . . . .	336
D.2 Phonetisch-phonologische Vorverarbeitung und Nachbereitung . . . . .	337
D.3 Sprachwandel . . . . .	339

D.4	Französische Wortformerkennung . . . . .	349
D.4.1	Lexikondateien . . . . .	349
D.4.2	Allomorphiedateien . . . . .	351
D.4.3	Flexionsdateien . . . . .	353
D.4.4	Weitere Dateien . . . . .	358

# Abbildungsverzeichnis

1	Morphemanalyse von <i>déformons</i> . . . . .	9
2	Endlicher Automat für Verbformen bestehend aus den Allomorphen <i>dé, re, form, mont, ons, ez</i> und <i>er</i> . . . . .	10
3	Transduktor für Verbformen bestehend aus den Allomorphen <i>dé, re, form, mont, ons, ez</i> und <i>er</i> . . . . .	11
4	Transduktor für die Verbformen <i>gebetet</i> und <i>betest</i> (Trommer 2004: 204; adaptiert) . . . . .	13
5	Ausgabe der protoindoeuropäischen Form <i>ang<sup>u</sup>(h)i-</i> bei Smith (1969: 89)	24
6	Ausgabe der uralischen Form <i>kala</i> bei Bátori (1981: 139; gekürzt) . . . . .	26
7	Ableitungen der protoslawischen Form <i>golva</i> bei Pognan/ Panevová (2013: 66; gekürzt) . . . . .	26
8	Ausgabe für urgermanisch <i>sneiθ</i> bei Becker (1982b: 11) . . . . .	28
9	Umsetzungsregeln (a) und Wahrscheinlichkeitsmatrix (b) bei Klein (1976: 162f.; adaptiert) . . . . .	31
10	Ausgabe für lat. FACTUM bei Burton-Hunter (1976: 220) . . . . .	32
11	Ausgabe für lat. CAELUM bei Eastlack (1977: 83f; korrigiert) . . . . .	33
12	Ausgabe für lat. OCULUM in <i>Phono 4.1</i> . . . . .	35
13	Ausgabe für lat. HOMINEM in <i>Etymo</i> . . . . .	38
14	Ablaufdiagramm des Programms . . . . .	45
15	Komponenten des Systems . . . . .	46
16	Ausschnitt aus der Projektdatei <b>cew.project</b> . . . . .	48
17	Merkmalstruktur der Wortform RĪPĀ . . . . .	48
18	Definition einer LA-Grammatik für präfigierte Verben bestehend aus den Allomorphen <i>dé, re, form, mont, ons, ez, er</i> . . . . .	51
19	Ableitung von <i>déformons</i> nach der in Abb. 18 definierten LA-Grammatik	52

20	Einträge im Allomorphlexikon (dargestellt als Merkmalstruktur) . . . . .	55
21	Ablaufdiagramm der Wortformererkennung . . . . .	56
22	Beispiel einer Trie-Struktur . . . . .	59
23	Ebenen der Darstellung von Graphemen . . . . .	61
24	Attribute einer Wortform . . . . .	62
25	Wichtige Klassen der Allomorphgenerierung . . . . .	66
26	Wichtige Klassen der Wortformererkennung . . . . .	67
27	Ablaufdiagramm der phonetisch-phonologischen Vorverarbeitung . . . . .	68
28	Die Grapheme von $\bar{R}\bar{I}\bar{P}\check{A}\check{M}$ als Instanzen der Klasse <b>Sign</b> . . . . .	69
29	Die Laute von $\bar{R}\bar{I}\bar{P}\check{A}\check{M}$ als Instanzen der Klasse <b>Sign</b> . . . . .	70
30	Ausgabe der Wortform $\bar{R}\bar{I}\bar{P}\check{A}\check{M}$ nach der Transkription . . . . .	71
31	Die Silben von $\bar{R}\bar{I}\bar{P}\check{A}\check{M}$ als <b>Syllable</b> -Objekte . . . . .	73
32	Ausgabe der Wortform $\bar{R}\bar{I}\bar{P}\check{A}\check{M}$ nach der Silbentrennung . . . . .	74
33	Die Silben von $\bar{R}\bar{I}\bar{P}\check{A}\check{M}$ als <b>Syllable</b> -Objekte mit Betonung . . . . .	75
34	Ausgabe der Wortform $\bar{R}\bar{I}\bar{P}\check{A}\check{M}$ nach Anwendung der Betonungsregeln . . . . .	75
35	Wichtige Klassen der phonetisch-phonologischen Vorverarbeitung . . . . .	76
36	Ausgabe der Wortform $\bar{R}\bar{I}\bar{P}\check{A}\check{M}$ nach der phonetisch-phonologischen Vorverarbeitung . . . . .	77
37	Beispiele für die Syntax des <b>Betonung</b> -Attributs . . . . .	82
38	Mapping der Wortform $\bar{R}\bar{I}\bar{P}\check{A}\check{M}$ . . . . .	85
39	Mapping der Wortform $\bar{P}\check{U}\check{L}\check{C}\check{H}\check{E}\check{R}$ . . . . .	86
40	Remapping für vlat. <i>spina</i> > <i>espina</i> (Typ I) . . . . .	88
41	Remapping für vlat. <i>laridum</i> > <i>lardum</i> (Typ VI) . . . . .	91
42	Remapping für afrz. <i>ombre</i> [õmbrə] > [õbrə] (Typ VIII) . . . . .	91
43	Remapping für <i>fasie</i> [fasʲə] (lat. $\check{F}\check{A}\check{S}\check{C}\check{I}\check{A}$ ) > <i>faisse</i> [fajssə] (Typ XIa) . . . . .	93
44	Remapping für <i>bailare</i> [bajlarɛ] > <i>baillare</i> [balʲarɛ] (< lat. <i>bāiulāre</i> ; Typ XIId) . . . . .	96
45	Ausgabe nach Anwendung einer Lautregel . . . . .	98
46	Wichtige Klassen der Lautentwicklung . . . . .	99
47	Das vulgärlateinische Substantiv <i>cornu</i> vor und nach dem Wechsel der Deklinationsklasse . . . . .	101
48	Wichtige Klassen der morphologischen Entwicklung . . . . .	103
49	Graphische Oberfläche des <i>CEW</i> . . . . .	105

50	Ausgabe der Attribute . . . . .	107
51	Ausgabe in <i>Words</i> für die Substantive RĪPĀM und <i>camisia</i> . . . . .	125
52	Das Substantiv RĪPĀ in der Wortliste von Whitaker . . . . .	125
53	Singularformen der Deklinationsklasse 1.1 in <i>Words</i> . . . . .	128
54	Das Adjektiv BŌNŪS in der Wortliste von Whitaker . . . . .	133
55	Das Verb LAUDĀRĒ in der Wortliste von Whitaker . . . . .	136
56	Ergebnis der Wortformerkennung für RĪPĀM . . . . .	144
57	Möglichkeiten der lautlichen Darstellung von Diphthongen in der Sign- Repräsentation . . . . .	148
58	Beispiel einer geeigneten Grammatik . . . . .	161
59	Entwicklung des haupttonigen [e] in freier Silbe bei Rheinfelder (1976a: 20) . . . . .	170
60	Nasalierung des [e] bei Rheinfelder (1976a: 75) . . . . .	172
61	Entwicklung von lat. NĪGRŪM > afrz. <i>neir</i> . . . . .	178
62	Entwicklung von lat. PĚTRĀ > afrz. <i>piedre</i> > <i>piere</i> . . . . .	179
63	Anwendung der Regel MORPH_Analog_M1_1 auf lat. PĀNĒS . . . . .	191
64	Ergebnis der Wortformerkennung für <i>rive</i> . . . . .	213



# Tabellenverzeichnis

1	Einzel sprachliche Textdateitypen der Hauptkomponenten . . . . .	47
2	Lautregeltypen . . . . .	86
3	Subtypen zu Lautregeltyp XI . . . . .	92
4	Übereinstimmungsgrade . . . . .	103
5	Einzel sprachliche Textdateien der lateinischen Wortformerken- nung . . . . .	123
6	Flexionsklassen und Paradigmen der Substantive der ersten traditionellen Deklinationsklasse im <i>CEW-fr</i> . . . . .	129
7	Flexionsklassen und Paradigmen der Substantive der zweiten traditionel- len Deklinationsklasse im <i>CEW-fr</i> . . . . .	130
8	Flexionsklassen und Paradigmen der Substantive der dritten traditionel- len Deklinationsklasse im <i>CEW-fr</i> . . . . .	131
9	Flexionsklassen und Paradigmen der Substantive der vierten traditionel- len Deklinationsklasse im <i>CEW-fr</i> . . . . .	132
10	Flexionsklasse und Paradigma der Substantive der fünften traditionellen Deklinationsklasse im <i>CEW-fr</i> . . . . .	132
11	Flexionsklassen und Paradigmen der Adjektive der ersten Deklination- sklasse im <i>CEW-fr</i> . . . . .	134
12	Oberflächenformen und distinktive Kategorien von ĀLTŪS . . . . .	135
13	Flexionsklassen und Paradigmen der Adjektive der dritten Deklination- sklasse im <i>CEW-fr</i> . . . . .	136
14	Verbmorpheme im Präsens . . . . .	139
15	Einzel sprachliche Textdateien für die phonetisch-phonologische Vorverar- beitung der lateinischen Wortformen . . . . .	145
16	Graphem-Laut-Zuordnung . . . . .	146
17	Anzahl der Lautregeln im <i>CEW-fr</i> . . . . .	165

18	Regeln im <i>CEW-fr</i> mit den bei Rheinfelder (1976a) aufgeführten Zeit räumen . . . . .	175
19	Suffixe von CĂNTŪS beim Wechsel der Deklinationsklasse von <i>passus</i> zu <i>circus</i> . . . . .	184
20	Erste Deklinationsklasse der Maskulina im Altfranzösischen . . . . .	188
21	Zweite Deklinationsklasse der Maskulina im Altfranzösischen . . . . .	192
22	Dritte Deklinationsklasse der Maskulina im Altfranzösischen . . . . .	193
23	Erste Deklinationsklasse der Feminina im Altfranzösischen . . . . .	194
24	Zweite Deklinationsklasse der Feminina im Altfranzösischen . . . . .	195
25	Dritte Deklinationsklasse der Feminina im Altfranzösischen . . . . .	196
26	Einzel sprachliche Textdateien der französischen Wortformererkennung . . .	198
27	Flexionsklassen und Paradigmen der französischen Substantive im <i>CEW-fr</i> . . . . .	201
28	Flexionsklassen und Paradigmen der französischen Adjektive im <i>CEW-fr</i>	203
29	Französische Verbmorpheme im Präsens und <i>imparfait de l'indicatif</i> (nach Schpak-Dolt 2016: 60) . . . . .	205
30	Flexionsklassen und ausgewählte Formen der französischen Verben im <i>CEW-fr</i> . . . . .	206
31	Transkription von ⟨a⟩ . . . . .	214
32	Transkription von ⟨e⟩: Graphemgruppen . . . . .	217
33	Transkription von ⟨e⟩: Präfixe mit [e] . . . . .	218
34	Transkription von ⟨e⟩: einsilbige Wörter . . . . .	218
35	Transkription von ⟨e⟩: vor Doppelkonsonanten . . . . .	219
36	Transkription von ⟨e⟩: vor zwei oder mehr Konsonanten . . . . .	220
37	Transkription von ⟨e⟩: wortinitiale Silbe . . . . .	220
38	Transkription von ⟨e⟩ und ⟨eu⟩: wortauslautende Silbe . . . . .	221
39	Transkription von ⟨e⟩: restliche Fälle . . . . .	222
40	Transkription von ⟨eu⟩: vor Vokal . . . . .	223
41	Transkription von ⟨eu⟩: vor einem Konsonanten . . . . .	224
42	Transkription von ⟨eu⟩: vor mehreren Konsonanten . . . . .	225
43	Transkription von ⟨e⟩: Diakritika . . . . .	225
44	Transkription von ⟨i⟩ . . . . .	225
45	Transkription von ⟨o⟩: Graphemkombinationen und Diakritika . . . . .	226

---

46	Transkription von ⟨o⟩: Wortende . . . . .	227
47	Transkription von ⟨o⟩: vor Vokal . . . . .	228
48	Transkription von ⟨o⟩: restliche Fälle . . . . .	228
49	Transkription von ⟨œ⟩ und ⟨oe⟩ . . . . .	229
50	Transkription von ⟨u⟩ . . . . .	230
51	Transkription von ⟨b⟩, ⟨d⟩, ⟨f⟩, ⟨h⟩, ⟨j⟩, ⟨l⟩, ⟨m⟩, ⟨n⟩, ⟨v⟩, ⟨w⟩, ⟨z⟩ . . . . .	230
52	Transkription von ⟨k⟩, ⟨p⟩, ⟨r⟩ und ⟨t⟩ . . . . .	231
53	Transkription von ⟨c⟩ und ⟨g⟩ . . . . .	232
54	Transkription von ⟨q⟩ . . . . .	233
55	Transkription von ⟨s⟩ . . . . .	233
56	Transkription von ⟨x⟩ . . . . .	234
57	Transkription von ⟨y⟩ . . . . .	234
58	Lexikoneinträge im <i>CEW-fr</i> . . . . .	238
59	Regeln im <i>CEW-fr</i> . . . . .	238
60	Ergebnis der Analyse . . . . .	239
61	Analyseergebnis der ersten 100 Wortformen . . . . .	240
62	Einteilung der ersten 100 Wortformen nach ihrem Analyseergebnis . . . . .	241



# Quellcode-Verzeichnis

1	Regel bei Bátori (1981: 129) . . . . .	26
2	Programmausschnitt für das Vernersche Gesetz bei Becker (1982b: 19ff.)	29
3	Regel für die Spirantisierung im Spanischen in <i>Phono 4.1</i> . . . . .	35
4	Regel für die Synkope im Spanischen in <i>Etymo</i> . . . . .	37
5	Hierarchisierte, mehrwertige Merkmale in <i>Etymo</i> . . . . .	37
6	Eintrag im Elementarlexikon (adaptiert aus: Weber et al. 2010: 7) . . . . .	54
7	Allo-Tabelle in JSLIM (adaptiert aus: Weber et al. 2010: 8) . . . . .	54
8	Kombi-Tabelle in JSLIM (adaptiert aus: Weber et al. 2010: 10) . . . . .	55
9	Lexikondatei . . . . .	57
10	Tabelle einer Allomorphiedatei . . . . .	58
11	Tabelle einer Flexionsdatei . . . . .	60
12	Transkriptionen aus einer Datei des Typs <b>trans</b> . . . . .	69
13	Transkription mit Bedingungen aus einer Datei des Typs <b>trans</b> . . . . .	70
14	Variable <b>\$V</b> aus einer Datei des Typs <b>var</b> . . . . .	70
15	SAMPA-IPA-Zuordnung aus einer Datei des Typs <b>phon</b> . . . . .	71
16	Silbentrennungsregeln aus einer Datei des Typs <b>syll</b> . . . . .	72
17	Variablen <b>\$V</b> , <b>\$V2</b> und <b>\$K</b> aus einer Datei des Typs <b>var</b> . . . . .	72
18	Betonungsregeln aus einer Datei des Typs <b>acc</b> . . . . .	75
19	Regel aus einer Datei des Typs <b>ru1</b> . . . . .	80
20	Beispiele für die Lautwandelzeile . . . . .	81
21	Beispiel einer Regel aus einer Datei des Typs <b>ru1</b> . . . . .	85
22	Regel aus einer Datei des Typs <b>ru1</b> mit einer morphologiebasierten Entwicklung . . . . .	100
23	Kategorie beim Wechsel der Deklinationsklasse aus einer Datei des Typs <b>ncat</b> . . . . .	101

24	Attributzuordnung von Quell- zu Zielsprache aus einer Datei des Typs <code>match</code> . . . . .	104
25	XML-Eintrag für RĪPĀ in der <i>Perseus Digital Library</i> . . . . .	126
26	Lexikontabelle <code>L_aqua</code> (a), Allomorphietabelle <code>A_aqua</code> (b) und Flexionstabelle <code>F_aqua</code> (c) . . . . .	129
27	Lexikontabelle <code>L_altus</code> (a), Allomorphietabelle <code>A_altus</code> (b) und Flexionstabelle <code>F_altus</code> (c) . . . . .	133
28	Lexikontabelle <code>L_cantare</code> (a) und Allomorphietabelle <code>A_cantare</code> (b) . . .	137
29	Bildung der 2. und 3. Person Singular des Indikativ Präsens Aktiv in der Flexionstabelle <code>F_cantare</code> . . . . .	139
30	Bildung der 2. Person Singular Konjunktiv Präsens Aktiv in den Flexionstabellen <code>F_cantare</code> und <code>F_tenere</code> . . . . .	140
31	Lexikontabelle <code>L_esse</code> (a) und Allomorphietabelle <code>A_esse</code> (b) . . . . .	142
32	Lexikontabelle <code>L_bene</code> (a) und Allomorphietabelle <code>A_bene</code> (b) . . . . .	144
33	Beispiele aus der Datei <code>latein.trans</code> . . . . .	147
34	Beispiele aus der Datei <code>lat-frz.phon</code> . . . . .	147
35	Silbentrennungsregeln der Datei <code>latein.syll</code> . . . . .	154
36	Die Variablen <code>\$V_PH_KLAT</code> , <code>\$V_PH_KLAT_2</code> und <code>\$K_PH_KLAT</code> in der Datei <code>lat-frz.var</code> . . . . .	155
37	Die Betonungsregeln für ein-, zwei- und dreisilbige Wörter in der Datei <code>latein.acc</code> . . . . .	157
38	Das Attribut <code>Betonung</code> zur Gruppierung der Vokalentwicklungen . . . . .	168
39	Die Regel <code>V_Haupt_e_1</code> in der Datei <code>lat-frz.rul</code> . . . . .	171
40	Die Regel <code>MORPH_Dekl_4</code> in der Datei <code>lat-frz-morph.rul</code> in gekürzter Form . . . . .	184
41	Die Regel <code>MORPH_Genus_1</code> in der Datei <code>lat-frz-morph.rul</code> in gekürzter Form . . . . .	186
42	Die Regel <code>MORPH_Assim_1</code> in der Datei <code>lat-frz-morph.rul</code> in gekürzter Form . . . . .	188
43	Die Regel <code>MORPH_Analog_M1_1</code> in der Datei <code>lat-frz-morph.rul</code> in gekürzter Form . . . . .	189
44	Ausschnitt aus der Datei <code>lat-frz.ncat</code> . . . . .	190

45	Die Regel MORPH_Ersatz_1 in der Datei <code>lat-frz-morph.rul</code> in gekürzter Form . . . . .	192
46	Eintrag für frz. <i>rive</i> in <i>Morphalou</i> 2.0 . . . . .	199
47	Lexikontabelle <code>L_femme</code> (a), Allomorphietabelle <code>A_femme</code> (b) und Flexionstabelle <code>F_femme</code> (c) . . . . .	200
48	Lexikontabelle <code>L_joli</code> (a), Allomorphietabelle <code>A_joli</code> (b) und Flexionstabelle <code>F_joli</code> (c) . . . . .	202
49	Lexikontabelle <code>L_chanter</code> (a) und Allomorphietabelle <code>A_chanter</code> (b) . . . . .	204
50	Bildung der 2. und 3. Person Singular des <i>imparfait de l'indicatif</i> in der Flexionstabelle <code>F_chanter</code> . . . . .	205
51	Lexikontabelle <code>L_etre</code> (a) und Allomorphietabelle <code>A_etre</code> (b) . . . . .	211
52	Lexikontabelle <code>L_hier</code> (a) und Allomorphietabelle <code>A_hier</code> (b) . . . . .	212
53	Attributzuordnung für Präpositionen in der Datei <code>lat-frz.match</code> . . . . .	236
54	Die Regeln <code>V_tonl_Zwischen_2</code> und <code>V_tonl_Zwischen_5</code> in der Datei <code>lat-frz.rul</code> in gekürzter Form . . . . .	258



# Abkürzungsverzeichnis

**Akt.** Aktiv

**Gram. Kat.** Grammatische Kategorie

**Graph.** Graphem

**klat.** klassisch-lateinisch

**KNG** Kasus-Numerus-Genus-Affix

**Konj.** Konjunktiv

**Konjug.** Konjugation

**Li.** Linker Kontext

**NG** Numerus-Genus-Affix

**O.** Obliquus

**PN** Person-Numerus-Affix

**Pqp.** Plusquamperfekt

**Ps.** Person

**R.** Rektus

**Rad** Radikal

**Re.** Rechter Kontext

**SA** Stammauslaut

**SE** Stammerweiterung

**Segm.** Segmentierung

**splat.** spätlateinisch

**TM** Tempus-Modus-Affix

**tV** thematischer Vokal

**vlat.** vulgärlateinisch

**VSt** Verbstamm

**wgerm.** westgermanisch

# 1 Einleitung

Si mon livre peut servir pendant quelques années de point de départ à des études qui feront progresser et renouvelleront la science, fût-il ensuite condamné et abandonné, tout sera bien, car dans le grand travail collectif et anonyme qui a pour but d'édifier l'histoire, l'ambition de celui qui aime la vérité est, non pas de chercher à marquer son nom à une place éclatante, mais de contribuer à l'ensemble, et quand il a la joie de mettre à cet ensemble plus qu'une pierre, de donner un plan, grâce auquel de bons ouvriers montent rapidement et solidement de nouvelles assises, son but est atteint et sa peine récompensée. (Brunot 1905: XXIf.)

Mit diesen Worten schließt Ferdinand Brunot das Vorwort zu seinem mehrbändigen Werk *Histoire de la langue française* ab. Betrachtet man neben diesem Zitat den Umfang seines Werkes, so erahnt man die Komplexität der Geschichte der französischen Sprache. Ferdinand Brunot und Charles Bruneau, der Brunots unvollendetes Werk fortführte, beschreiben die französische Sprachgeschichte auf insgesamt mehr als 10.800 Seiten. Dass in den folgenden Jahrzehnten viele weitere Publikationen zu dieser Thematik entstanden sind, bestätigt Brunots Prognose, dass er mit seiner umfassenden Arbeit lediglich einen Grundstein setzte. Obwohl die Erarbeitung der Geschichte und der historischen Sprachstufen des Französischen seit dem Erscheinen von Brunots erstem Band auf 110 weitere Jahre Forschung zurückblicken kann, gibt es bis heute noch unbeantwortete Fragen in diesem Teilbereich der Romanistik.

Die letzten Bände von Brunot selbst entstanden in einer Zeit, in der in einem anderen Forschungsbereich revolutionäre Erfindungen gemacht wurden: 1941 baute Konrad Zuse den ersten funktionsfähigen Computer der Welt.<sup>1</sup> Dass dieser Innovation ein langer Entwicklungsprozess vorausging, zeigt die Forschungsliteratur zur Geschichte der Informatik, die meist den Weg vom ersten Abakus (ab 1100 v. Chr. im Fernen Osten) über die mechanischen Rechenmaschinen des 17. Jahrhunderts beschreibt (zum Beispiel

---

<sup>1</sup> Naumann (2001: 136) beschreibt Konrad Zuse „als Schöpfer der ersten vollautomatischen, programmgesteuerten und frei programmierbaren, in binärer Gleitpunktrechnung arbeitenden Rechenanlage der Geschichte“.

Naumann 2001 oder Bruderer 2018a und 2018b). Ein Artikel von Comrie in der Zeitschrift *Nature* aus dem Jahr 1946 veranschaulicht, dass bereits 100 Jahre vor Zuse die Forschung weit fortgeschritten war. Der Autor beschreibt die zwei Jahre zuvor gebaute, erste programmgesteuerte Großrechenanlage *Automatic Sequence Controlled Calculator* und verweist mit der Wahl des Artikeltitels *Babbage's dream comes true* auf den Erfinder der ersten Universalrechenmaschine. Charles Babbage begann 1833 mit dem Entwurf der *Analytical Engine* (Fuegi/ Francis 2003: 17f.), die mit Steuerwerk, Rechenwerk, Speicher und Programmsteuerung bereits die Merkmale eines modernen Computers aufweist – allerdings nie technisch realisiert wurde (Bruderer 2018b: 9, 53). Im Jahre 1843 veröffentlichte Ada Lovelace, die mit Babbage eng zusammenarbeitete, eine Übersetzung zu einem Artikel von Luigi Menabrea, der die Funktionen von Babbages Maschine beschreibt<sup>2</sup> (Menabrea 1843). Da Lovelace die Konstruktionspläne besser kannte, fügte sie zahlreiche Fußnoten hinzu und korrigierte Menabreas Fehler (Fuegi/ Francis 2003: 18f.). Eine dieser Fußnoten enthält eine Beschreibung, wie die *Analytical Engine* zur Berechnung der Bernoulli-Zahlen verwendet werden könnte, was Lovelace den Titel der ersten Programmiererin der Welt einbrachte (Huskey/ Huskey 1980: 299). Auch für weitere Einsatzmöglichkeiten für Babbages Erfindung zeigt sich Lovelace visionär, wenn sie in Anmerkung A schreibt:

Even in the utilitarian aspect, however, we do not doubt that very valuable practical results would be developed by the extended faculties of the Analytical Engine; some of which results we think we could now hint at, had we the space; and others, which it may not yet be possible to foresee, but which would be brought forth by the daily increasing requirements of science, and by a more intimate practical acquaintance with the powers of the engine, were it in actual existence. (Menabrea 1843: 699f.)

Man kann mit Sicherheit behaupten, dass Lovelace die Rolle einer solchen Maschine zur Lösung mathematischer Probleme vorhersah. Unsicher ist allerdings, inwieweit die Mathematikerin auch an sprachliche und sprachwissenschaftliche Aufgaben dachte. Als schließlich Zuse den ersten Computer baute, dauerte es nur kurze Zeit, bis erste Gedankenexperimente zur Sprachanwendung angestellt wurden. Im Jahre 1949 weist Weaver in einem Memorandum<sup>3</sup> auf den möglichen Einsatz des Computers für die maschinelle Übersetzung hin.

---

<sup>2</sup> Der Turiner Ingenieur kannte die *Analytical Engine* aus Vorträgen von Babbage und publizierte 1842 den Artikel *Notions sur la machine analytique* (Fuegi/ Francis 2003: 18).

<sup>3</sup> Ein Nachdruck ist in Weaver (1955) zu finden.

---

Den eigentlichen Grundstein für die Computerlinguistik legte letztlich Chomsky, indem er 1957 eine Formalisierungsmöglichkeit der Beschreibung der natürlichen Sprachen aufzeigte. In den nächsten Jahrzehnten sollten zahlreiche weitere Formalismen und Subformalismen folgen. Objekt dieser Bemühungen waren und sind in erster Linie die modernen Sprachen. Die wenigen Untersuchungen, die sich mit der synchronen Betrachtung historischer Sprachstufen beschäftigen, betreffen den Bereich der Korpuslinguistik. Pionierarbeit leistete hier Robert Busa, der mit der Hilfe von Thomas J. Watson von IBM das erste maschinenlesbare Korpus auf Lochkarten zusammenstellte (McEnery/ Wilson 2001: 20; siehe auch Busa 1951). Sein ursprüngliches Projekt – eine Untersuchung zum Wort *in* in den Texten Thomas von Aquins – erweiterte er so weit, dass er in den Jahren 1949 bis 1967 ein 10,6 Mio. Wortformen umfassendes Korpus der mittelalterlichen Philosophie erstellte (McEnery/ Wilson 2001: 20).

Die ersten Versuche diachroner Untersuchungen mit Hilfe des Computers wurden in den 1960er und 70er Jahren veröffentlicht. Für die diachrone Betrachtung kann meist zwischen zwei verschiedenen Perspektiven gewählt werden, die Hewson (1989: 576) als *upstream view* und *downstream view* bezeichnet.

Der *upstream view* versucht mit Hilfe mehrerer Tochtersprachen ältere Sprachstufen zu rekonstruieren. Laut Hewson (1989: 579) ist Henry Gleason der Erste, der von Experimenten zur Rekonstruktion mit Hilfe des Computers berichtet. Gleason (1961: 182f.) beschreibt, dass er bereits ab 1954 erste Versuche mit verschiedenen Sprachgruppen anstellte, wobei der Computer nicht für den Rekonstruktionsprozess verwendet wird, sondern lediglich bei der Suche nach verwandten Wörtern in einer Liste von Wortpaaren hilft. Ein ähnliches Verfahren beschreibt auch Frantz (1970), bei dessen Systemen der Benutzer mehrere Wortpaare mit ähnlicher Bedeutung eingeben muss, um als Output regelmäßige Entsprechungen der Buchstaben und einen Ähnlichkeitsfaktor für jedes Wortpaar zu erhalten. Hewson (1973) stellt ein Verfahren vor, bei dem der Computer auch für den Rekonstruktionsprozess verwendet wird. Hewsons Ziel war es, ein Wörterbuch des Proto-Algonkin zu erstellen. Durch die Verwendung eines Computerprogramms wollte er das stundenlange Suchen nach Protoformen beschleunigen. Hierfür generiert das Programm mögliche Protoformen für jede der vier verwendeten Tochtersprachen. Wenn mindestens zwei Protoformen identisch sind, dann werden sie zusammen mit den Wörtern der Tochtersprachen ausgedruckt. Auf dieser Basis kann anschließend ein Linguist nach der wahrscheinlichsten Protoform suchen. Das Resultat seiner Arbeit mit

diesem Programm veröffentlichte Hewson 1993 im *Computer-Generated Dictionary of Proto-Algonquian*.

Im Unterschied zum *upstream view* setzt der *downstream view* bei einer älteren Sprachstufe an und untersucht deren Evolution im Laufe der Zeit. Diese Sicht profitiert ebenso wie der *upstream view* davon, dass lautliche Veränderungen einer Sprache bis zu einem gewissen Grad regelmäßig erfolgen. Folglich kann die Entwicklung einer Wortform mit Hilfe von Lautveränderungen – in Form von zeitlich geordneten Regeln – von einer älteren zu einer jüngeren Sprachstufe am Computer reproduziert werden.

Eine Umsetzungsmöglichkeit des *downstream view* beschreibt die vorliegende Arbeit. Die im Folgenden ausgeführten Überlegungen wurden in einem in der Programmiersprache C# geschriebenen Programm, das Bestandteil dieser Arbeit ist, umgesetzt.

### 1.1 Gegenstand

Das Thema der vorliegenden Arbeit ist im Bereich der Computerlinguistik anzusiedeln, genauer im Teilbereich der Modellierung natürlichsprachlicher Phänomene auf dem Computer.<sup>4</sup> Die Arbeit zeigt, wie mit Hilfe des Computers die Entwicklung einzelner Wortformen einer älteren Sprachstufe simuliert werden kann. Das Ergebnis ist schließlich ein computergesteuertes etymologisches Wörterbuch, das die historisch ältere Sprachform als Lemma verwendet, wie es auch beim *FEW* der Fall ist. Dabei soll die Verknüpfung eines Etymons mit seiner modernen Entsprechung nicht abgespeichert sein – vielmehr soll ein potentiell Kontinuat durch das Anwenden von Lautregeln vom Programm automatisch gefunden werden. Es handelt sich also um eine Simulation der Vorgänge, die in der Sprachgeschichte stattgefunden haben. Theoretisch kann ein solches System auf jede beliebige Sprache angewandt werden. Im Rahmen dieser Arbeit wurde speziell ein System für die Entwicklung französischer Wortformen aus dem Lateinischen erstellt.

Ein beispielhaftes Anwendungsszenario legt dar, in welche einzelne Teilprobleme sich Konzeption und Erstellung des Systems aufspalten. Zunächst erwartet das Programm als Eingabe durch den Benutzer eine lateinische Wortform. Diese wird mit einem internen Lexikon des Programms abgeglichen, um das Eingabewort mit grammatischen und

---

<sup>4</sup>Die Methoden und Erkenntnisse der Linguistik fließen dennoch mit ein, so dass das Thema der Linguistik näher rückt, als dies in der Computerlinguistik oft der Fall ist (siehe dazu Abschnitt 1.3).

morphologischen Informationen zu verknüpfen. Diese Aufgabe übernimmt ein Wortformerkennungssystem. Für die Anwendung der Lautentwicklungsregeln ist es notwendig, die Eingabe – da sie in graphischer Form vorliegt – über Transkriptionsregeln in die lautliche Form zu transformieren, um anschließend die Silbengrenzen und die betonte Silbe zu bestimmen. Diese drei Teilschritte werden in einer Komponente zusammengefasst, die vorbereitende Aufgaben für die Modellierung der Lautentwicklung realisiert. Die graphische und die lautliche Form sind wiederum Eingabe für verschiedene Regeln, die über Zwischenstufen die neufranzösischen Ergebnisse ermitteln, wofür die Komponente der Lautentwicklung zuständig ist. Diese rekonstruierten neufranzösischen Wortformen werden dann mit einem internen Lexikon des Neuf Französischen abgeglichen, um festzustellen, ob sie tatsächlich Teil des französischen Wortschatzes sind. Dieser Teil wird ebenfalls von der Wortformerkennungskomponente durchgeführt.

## 1.2 Nutzen

Gegenüber der manuellen Analyse eines einzelnen Wortes bietet die hier beschriebene Methode der automatisierten Simulation der Lautentwicklung bedeutende Vorteile. Der offensichtliche Vorteil ist der der Geschwindigkeit. Diese kommt nicht nur bei einer Einzelwortanalyse zum Tragen, sondern vor allem auch dann, wenn es darum geht, verschiedene Hypothesen zu verifizieren. So kann zum Beispiel die zeitliche Einordnung von Phänomenen, für die es keine oder nur widersprüchliche Informationen gibt, leicht überprüft werden, indem die zeitlichen Attribute einzelner Regeln bearbeitet werden. Die Prüfung kann dann innerhalb kürzester Zeit anhand mehrerer Wortbeispiele vorgenommen werden.

Des Weiteren kann das System bei der Untersuchung von Wörtern, die sich nicht regelmäßig entwickelt haben, hilfreich sein. Es kann leichter festgestellt werden, auf Grund welcher Regel und zu welchem Zeitpunkt die simulierte Entwicklung nicht mehr der tatsächlichen Entwicklung der Wortform entspricht, so dass zu diesem Zeitpunkt nach Faktoren gesucht werden kann, die den Prozess beeinflusst haben.

Neben der Analyse einzelner Phänomene oder einzelner Wörter können auch umfangreichere statistische Untersuchungen vorgenommen werden. So ist es zum Beispiel mög-