

Inhalt

Vorwort — V

Konventionen und Abkürzungen — IX

Teil I: **Einleitung**

1 Vorhaben — 3

1.1 Motivation der Machbarkeitsstudie — 3

1.2 Beiträge der Arbeit — 10

2 Gliederung — 15

Teil II: **Technische Hinführung**

3 Einleitung — 19

4 Digital Humanities kurzgefasst — 20

4.1 Gewinn für die Geisteswissenschaften — 22

4.1.1 Daten (retro-)digitalisieren — 23

4.1.2 Daten digital produzieren — 24

4.1.3 Daten digital analysieren — 30

4.1.4 Daten digital publizieren — 33

4.2 Kritische Aspekte — 40

4.3 Fazit — 44

5 Der Datenzugang im World Wide Web und seine Defizite — 45

5.1 Defizite bei der wortbasierten Recherche — 48

5.2 Defizite bei der lexikalisch-semantischen Recherche — 51

6 Der semantische Datenzugang mithilfe von *Linked Open Data* — 59

6.1 Semantic Web, *Linked Open Data* und Ontologien — 59

6.1.1 Das Semantic Web — 60

6.1.2 (*Linguistic*) *Linked (Open) Data* — 62

6.1.3 Ontologien — 66

6.1.4 Resource Description Framework – RDF — 76

6.1.5 RDF Schema– RDFS — 83

6.1.6 RDF in attributes – RDFa — 84

6.1.7	SPARQL Protocol And RDF Query Language —	86
6.1.8	Simple Knowledge Organization System – SKOS —	90
6.1.9	Web Ontology Language – OWL —	91
6.1.10	Identifizierung der Sprachen und Sprachvarietäten —	92
6.2	Mehrwert im historisch-linguistischen Anwendungskontext —	105
6.3	Der lexikalisch-semantische Datenzugang —	108
6.4	Kritische Aspekte, erneut betrachtet —	115
6.5	Fazit —	118
7	Auswahl der Vokabulare und Ontologien —	120
7.1	Abbildung des sprachlichen Zeichens: OntoLex-Lemon —	122
7.1.1	Kernvokabular —	123
7.1.2	Zusatzmodule —	127
7.1.3	Erweiterung von OntoLex-Lemon für Etymologien: <i>lemonEty</i> —	134
7.1.4	Herausragende Merkmale aus lexikologisch-lexikographischer Sicht —	136
7.2	Metadatenstandards für außersprachliche Informationen —	137
7.3	Ontologien für Personen- und Ortsnamen —	139
7.4	Ontologien für linguistische Terminologie —	139
7.5	Realwelt-Ontologien —	141
7.5.1	Aspekt der Außersprachlichkeit —	141
7.5.2	Aspekt der Historizität —	142
7.5.3	Geeignete Ontologien —	143
7.5.4	Thesauri und semantische Netze —	151
8	Verwandte Arbeiten —	155
8.1	Identifizieren der Sprachen und Sprachvarietäten —	155
8.2	Modellieren von Bedeutungswandel und Bedeutungs differenzierungen —	159
8.3	Lexikalisch-semantisches Mapping —	163
8.4	Linguistische Ressourcen mit OntoLex-Lemon —	167

Teil III: Die linguistischen Ressourcen

9 Einleitung — 179

10 Die Sprachen des Mittelalters: Speichermedium der Kultur — 181

11 Wort — 184

- 11.1 Begriffsklärung — 184
- 11.1.1 Wörter und Bedeutungen — Lexien und Sememe — 184
- 11.1.2 Sprachliches Zeichen — Semiotisches Fünfeck — 185
- 11.2 Skriptafragen — Ausdrucksseite des sprachlichen Zeichens — 188
- 11.2.1 Lemmatisierung der Schreibformen — 192
- 11.2.2 Die Dimension der historischen Sprachstufe (1): Erweiterung des semiotischen Fünfecks um den textanalytischen Aspekt — 199
- 11.3 Lexikalische Semantik — Inhaltsseite des sprachlichen Zeichens — 201
- 11.3.1 Die Dimension der historischen Sprachstufe (2): Die Bedeutung — 204
- 11.3.2 Wörter und Sachen — 205
- 11.3.3 Semasiologie — Onomasiologie — 208
- 11.3.4 Zur Definition einer Bedeutung — 212
- 11.3.5 Hermeneutischer Zirkel (1): Die Bedeutung auf der Ebene von *langue* und *parole* — 216
- 11.3.6 Mosaiksteinchen — 220

12 Text — 238

- 12.1 Begriffsklärung — 238
- 12.1.1 Hermeneutischer Zirkel (2): Textphilologie und Geschichtswissenschaften — 239
- 12.1.2 Exkurs: Eine kurze Geschichte der Textphilologie — 240
- 12.2 Glossar — 243
- 12.3 Die *Grande Chirurgie* des Gui de Chauliac — 245
- 12.3.1 Das medizinisch-fachtextliche Umfeld der *Grande Chirurgie* — 249
- 12.3.2 Hermeneutischer Zirkel (3): Textphilologie und Lexikographie — 251
- 12.4 Digitale Editionen — 252
- 12.4.1 Zur Qualitätssicherung — 253
- 12.4.2 Die Chrétien-de-Troyes-Editionen des DÉCT — 257
- 12.4.3 Die Chrétien-de-Troyes-Editionen auf Wikisource — 263
- 12.4.4 Die Lyoner *Queste del saint Graal* — 269
- 12.4.5 Der Versuch einer Abstrahierung — 279
- 12.4.6 Das Glossar in einer digitalen Edition *versus* Die Auslagerung des Glossars in ein Wörterbuch — 284

12.4.7	Fazit: Inhalte erschließen —	286
--------	------------------------------	------------

13 Textkorpus — 288

13.1	Begriffsklärung —	288
13.2	Kriterien und Problematik der Korpuserstellung —	290
13.2.1	Die Dimension der historischen Sprachstufe (3): Die Verschärfung der generellen Problematik —	291
13.2.2	Das Problem der Korpusgrenzen —	293
13.2.3	Exkurs: Auf dem Weg zu einem Referenzkorpus des Altfranzösischen —	295
13.2.4	Integration einer Edition <i>versus</i> neue Handschriftentranskription —	297
13.2.5	Integration welches Editionstypus? —	299
13.2.6	Integration welcher Editionen? —	302
13.3	<i>Documents linguistiques galloromans</i> (DocLing) —	303
13.4	Korpusannotierung aus Sicht der historischen Lexikographie —	304
13.4.1	Semantische Annotierung —	305
13.4.2	Drei Herangehensweisen —	306
13.4.3	Aufwand —	307
13.4.4	Alternative — erster Schritt: Automatische Auswertung der Glossare —	308
13.4.5	Alternative — zweiter Schritt: Semi-automatische Auswertung der Einleitungen und Anmerkungen —	311

14 Wörterbuch — 316

14.1	Begriffsklärung —	317
14.2	<i>Dictionnaire étymologique de l'ancien français</i> (DEAF und DEAFél) —	319
14.2.1	Das Korpus des DEAF: Texte und digitales Zettelarchiv —	320
14.2.2	Publikationen —	321
14.2.3	Ein Korpus als «Abfallprodukt» der Redaktion von DEAF <i>plus</i> —	326
14.2.4	Der korpuslinguistische Ansatz des DEAF —	327
14.3	<i>Dictionnaire onomasiologique de l'ancien gascon</i> (DAG und DAGél) —	328
14.3.1	Lexikographischer Rahmen —	329
14.3.2	Projektgeschichte und Buchpublikation des DAG —	333
14.3.3	Elektronische Version DAGél —	334
14.3.4	Bibliographie —	336
14.3.5	Die Schwierigkeit der onomasiologischen Zuordnung: Kritik am Hallig-Wartburgschen <i>Begriffssystem</i> —	337
14.4	Integration von DocLing und DEAF —	339
14.4.1	Hintergrund: Lexikographie und Korpuslinguistik —	340
14.4.2	Workflow —	343

Teil IV: Umsetzung der Machbarkeitsstudie

15 Einleitung — 355

16 Beiträge zur Überwindung der Anwendungsgrenzen — 358

- 16.1 Sprachcodes und Sprachtags: Ein Framework zur Erweiterung von ISO 639 — **358**
 - 16.1.1 Hinzufügen von geographischen Informationen zum *privateuse*-Subtag — **359**
 - 16.1.2 Hinzufügen von zeitbezogenen Informationen zum *privateuse*-Subtag — **363**
 - 16.1.3 Eine Methode für Form und Struktur des *privateuse*-Subtags — **363**
 - 16.1.4 Web Application für das Erstellen und Dekodieren von Sprachtags — **374**
- 16.2 Spezifizierung der Sprachen und Sprachvarietäten: MoLA, eine Ontologie für die Sprachbeschreibung — **380**
 - 16.2.1 Entwicklung des Modells — **381**
 - 16.2.2 Anwendungsfall: Französisch – Varietäten und historische Entwicklung — **386**
 - 16.2.3 Diskussion der Schwierigkeiten — **391**
- 16.3 Von Hallig-Wartburg zu einer Ontologie — **393**
 - 16.3.1 Struktur des Hallig-Wartburg — **393**
 - 16.3.2 Vom *Begriffssystem* zu einer Ontologie — **396**
 - 16.3.3 Anwendungsbeispiele — **406**
 - 16.3.4 Diskussion und weiterführende Arbeiten — **408**
- 16.4 Abbilden von Verfahren des Bedeutungswandels: SemShift — **413**
 - 16.4.1 Assoziationsrelationen: Grundlage rhetorischer Stilfiguren — **415**
 - 16.4.2 Geeignete Mittel für die Abbildung der semantischen Relationen — **417**
 - 16.4.3 Semantische Relationen der linguistischen Ressourcen — **418**
 - 16.4.4 Verweisungen auf DBpedia — **423**
 - 16.4.5 Integrieren von Properties in die RDF-Datensätze — **425**
 - 16.4.6 Ein ontologisches Modell für die Repräsentation von Verfahren des Bedeutungswandels — **426**
 - 16.4.7 Probleme bei der automatisierten Repräsentation — **431**

17 Modellierung der Ressourcen — 434

- 17.1 Wörterbuch — **434**
 - 17.1.1 DEAF: Vom Zettel zum Tripel — **435**
 - 17.1.2 RDF-Modellierung der Artikel des DEAF*plus*: Generelle Problematik — **440**

17.1.3	Modellierung der Lemmata — 447
17.1.4	Modellierung der <i>Part-of-Speech</i> -Kategorien — 452
17.1.5	Hauptlemma-Unterlemma-Beziehung mit dem <i>Lexicog</i> -Modul — 455
17.1.6	Modellierung von <i>Etyma</i> — 457
17.1.7	Modellierung der <i>autres langues</i> — 464
17.1.8	Modellierung der graphischen Varianten — 468
17.1.9	Modellierung der Definitionen — 471
17.1.10	Modellierung der Fachsprachlichkeit — 475
17.1.11	Modellierung der Stilmittel — 477
17.1.12	Modellierung der Mehrwortverbindungen — 482
17.1.13	Modellierung der unstrukturierten Artikel und Artikelteile — 496
17.1.14	Modellierung der Artikel des <i>DEAFpré</i> — 501
17.1.15	Modellierung der Metadaten — 506
17.1.16	DAG: Vom Zettel zum Tripel — 508
17.1.17	Modellierung der Artikel des <i>DAGél</i> — 510
17.2	Evaluierung der Wörterbuchdaten — 517
17.2.1	Manuelle Erhebung von Soll- und Ist-Zustand — 518
17.2.2	Visualisierung der Ergebnisse — 522
17.2.3	Differenzen zwischen Soll- und Ist-Werten — 529
17.2.4	Fehlermeldungen — 543
17.2.5	Ergebnis — 544
17.2.6	Diskussion zukünftiger Arbeiten — 546
17.3	Textkorpus — 551
17.3.1	Abbildung der Einheit «DocLing-Edition» — 552
17.3.2	Abbildung der Einheit «DocLing-Zettel» — 554
17.4	Evaluierung der Korpusdaten — 561
17.5	Text — 562
17.5.1	GuiChaulMT: Datenformat <i>TeX</i> — 562
17.5.2	Semantische Annotation — 563
17.5.3	Transformation in <i>Linked Open Data</i> — 566
17.6	Evaluierung der Textdaten — 574
17.6.1	Verlinkung der Lemmata «GuiChaulMT – DEAF» — 575
17.6.2	Extrahierte RDF-Tripel — 582
17.6.3	Diskussion: Verlinkung der lexikalischen Einheiten — 582
18	Etablierung des lexikalisch-semantischen Datenzugangs — 586
18.1	Manuelles LexSemMapping — 587
18.2	Herausforderungen des LexSemMappings auf inhaltlicher Ebene — 590
18.2.1	Many-to-one- oder Many-to-many-Beziehung — 591

18.2.2	Die Dimension der historischen Sprachstufe (4): Moderne Ontologien — 592
18.3	Automatisierung I: Vier Methoden für Substantive — 593
18.3.1	Beschreibung der Herangehensweisen auf inhaltlicher Seite — 595
18.3.2	Umsetzung und Evaluation I.a: Händisch gesetzte Schlüsselwörter für das LexSemMapping — 600
18.3.3	Umsetzung und Evaluation I.b: Automatisch erzeugte Schlüsselwörter für das LexSemMapping — 605
18.3.4	Extrapolation auf die Gesamtdaten des DEAF — 612
18.4	Automatisierung II: Eine Methode für Nicht-Substantive — 614
18.4.1	Evaluation — 617
18.4.2	Extrapolation auf die Gesamtdaten des DEAF — 618
18.5	Anwendung auf reale RDF-Datensätze und Evaluation — 619
18.6	Gesamtergebnis des LexSemMappings — 623
18.7	LexSemMapping der Korpusdaten — 624
18.8	LexSemMapping der Textdaten — 624
19	Datenanfragen mit SPARQL — 626
19.1	Anfrage nach Wortformen — 627
19.2	Anfrage nach onomasiologischen Kategorien über Hallig-Wartburg — 629
19.3	Anfrage über SemShift — 632
19.4	Anfrage über das LexSemMapping — 633
19.5	Konklusion der SPARQL-Anfragen — 638
20	Resümee — 639
21	Literaturverzeichnis — 645

