

Das luxemburgischsprachige Oeuvre von Michel Rodange (1827-1876)

Editionsphilologische und korpuslinguistische Analyse



Texterfassung

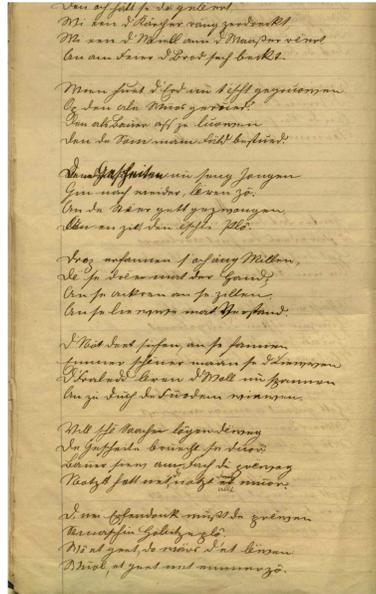
- ▶ Renert oder de Fuuß am Frack an a Ma'nsgrëßt, erschienen 1872 in Luxemburg, ca. 40.000 Tokens. Älteste Ausgabe vollständig eingescannt und erfasst.
- ▶ D'Lëerchen (Dem Leiwäckerche sei Lidd), ca. 5.000 Tokens. Das Manuskript wurde in der BNL fotografiert und im Rahmen dieses Projekts transkribiert.
- ▶ Dem Grow Sigfrid seng Goldkuommer, erschienen 1929 in Luxemburg, ca. 15.000 Tokens. Älteste Ausgabe vollständig eingescannt und erfasst.

Auszug aus der Transkription

Manuskript des Lerchenlieds, 3^v

- ▶ Mit historisch-kritischen Kommentaren versehen
- ▶ Mit früheren Editionen des Werks verglichen

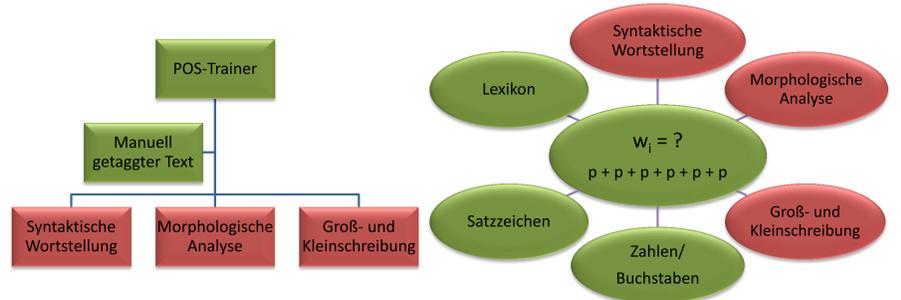
Aus den Ergebnissen wurde eine elektronische Edition erstellt:



POS-Tagging

- ▶ Tagger trainieren / Wahrscheinlichkeiten extrahieren

- ▶ Wahrscheinlichkeiten bei der Entscheidung verwenden

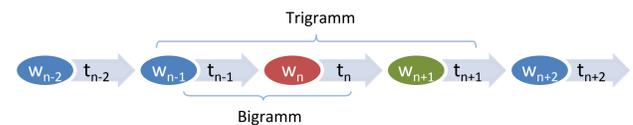


Syntaktische Wortstellung beim Taggen

- ▶ Die Bigramm- und Trigramm-Wahrscheinlichkeiten besagen, mit welcher Wahrscheinlichkeit eine Wortklasse auf eine andere folgt.

- ▶ Die 5 höchsten Bigramm-Wahrscheinlichkeiten:

INFTO	VERB	1,0
DET	NOUN	0,81
PREPA	NOUN	0,76
NUM	NOUN	0,66
VERBZ	PUNC	0,62



TEI-Auszeichnung

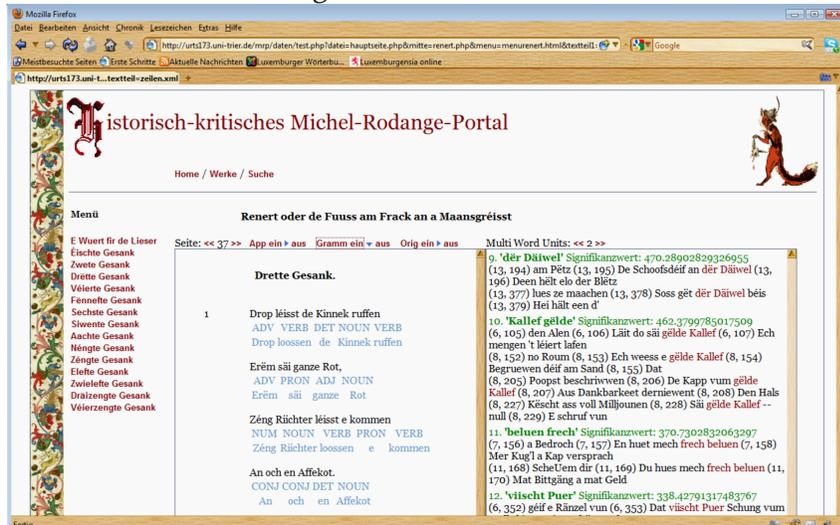
```
<lg type="stanza">
  <l>Wäll de Besch mam grengen Hittchen <app>
    <rdgGrp>
      <rdg><quote>d</quote> in <quote>Hittchen</quote> wurde
        zum ersten <quote>t</quote>
        verbessert</rdg>
      <rdg> Atten glaubt, <quote>Hutt</quote> sei zu
        <quote>Hittchen</quote>
        verbessert</rdg>
    </rdgGrp>
  </app>
</l> ....
</lg>
```

EDV-Technologien

- ▶ XML bzw. TEI; TUSTEP, Scriptsprache Perl, Reguläre Ausdrücke
- ▶ Programmiersprache Java für linguistischen Tools (z.B. POS-Tagger)
- ▶ Konvertierung XSLT, X-Path, Ausgabe ist beispielsweise HTML
- ▶ Für das Internet-Portal PHP und AJAX

Das Michel-Rodange-Portal

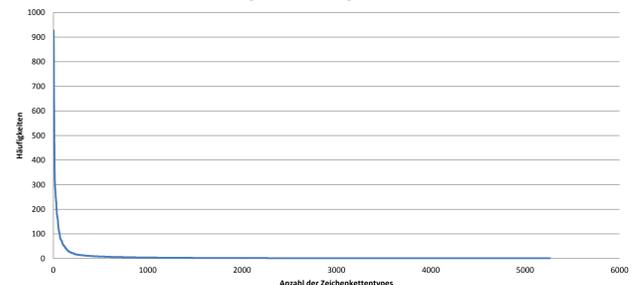
- ▶ Alle Ergebnisse dieser Analysen werden dynamisch und verlinkt in einem Online-Portal dargestellt:



Algorithmen in der phraseologischen Analyse

- ▶ Feste Wortverbindungen können mit statistischen Verfahren automatisch extrahiert werden, z.B. Chi-Quadrat-Test, Fischers Exakter Test, t-Test, z-Test, etc. Viele dieser Tests basieren auf der Annahme, dass die sprachlichen Zeichen nicht gleich verteilt sind:

Häufigkeitsverteilung im RENERT



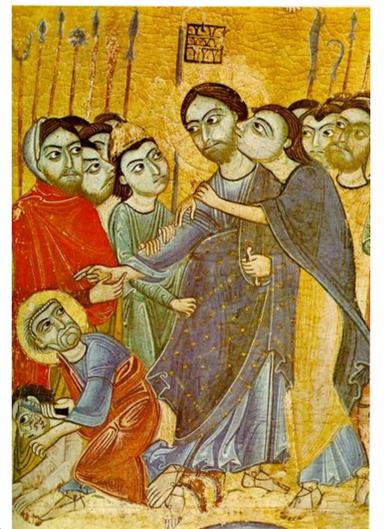
- ▶ Chi-Quadrat-Test: $\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$
- ▶ Fishers Exakter Test: $\varphi(a) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$

Kulturhistorische Implikationen

Viele der automatisch extrahierten Phraseologismen aus dem RENERT sind auch in anderen Bereichen von Kunst und Kultur zu finden:

- (6, 489)*Fuuss, o falsche Judas?
- (6, 490)Sou rífft mat haarder Stëmm
- (6, 491)Voll Rei a Leed de Kinnek,
- (6, 492)O hätt ech dech erëm!

- (13, 617)Jo, d'Steng an d'Wolke ruffen:
- (13, 618)Kuckt dee mam roude Baart,
- (13, 619)Ma kuckt deem falsche Judas
- (13, 620)Dajch endlech nun an d' Kaart!



*Die erste Zahl in Klammern verweist auf die Gesangsnummer, während die zweite Zahl die Versnummer innerhalb der einzelnen Gesänge darstellt.
*Die Abbildung: Der Kuss des Judas Ischariot; 12. Jahrhundert, unbekannter Künstler; heutiger Standort: Uffizien, Florenz