Verfeinerung der romanischen Morphologieanalyse mit Wapiti

Modell	Trainingskorpus		Analysekorpus		
	Inhalt	Tokens	Inhalt	Tokens	korrekt
modell2	Bundi (1)	~500	Bundi (2)	494	64%
modell3	+ Bundi (2)	~1000	Capricorn	506	68%
modell4	+ Capricorn	~1500	Sclavs	508	75%
modell5	+ Sclavs	~2000	Danemarc (1)	490	82%
modell6	+ Danemarc (1)	~2500	Danemarc (2)	497	84%
modell7	+ Danemarc (2)	~3000	Dardin	492	87%
modell8	+ Dardin	~3500	DRG	504	81%
modell9	+ DRG	~4000	Dinosaur	505	84%

modell7 und Foma kommen für das Korpus „Dardin“ zusammen auf 91.7% - mit Input mit Satzgrenzen ist der Score 92.7%

Trainingskorpus: Jeweils Ausschnitte aus der romansichen Wikipedia

Verfahren beim Kombinieren von Wapiti und Foma:

- Sammeln des Outputs von Foma
- Sammeln der drei besten Analysen von Wapiti
- Wortweise die beste Analyse übernehmen:
  - 1. Kriterium: wie viele einzelne Tags im Wapiti-Output stören i. Vgl. zum Foma-Output
    ('`+Verb+3P+Sg`' passt in '`+Verb+IndImp+3P+Sg`'; '`+Adv`' passt nicht in '`+Prep`')
  - 2. Kriterium: Score pro Wort von Wapiti
  - 3. wird bei von Foma nicht erkannten Formen die Wapiti-Analyse übernommen