

Muster von "you" und "thou": Modellierung der Anrede im englischen Sonett

Ausgangspunkt

Das Early Modern English unterscheidet **V-Formen** (you, your etc.) und **T-Formen** (thou, thy etc.). Die T-Formen gehen mit dem Sprachwandel zum Modern English im 17. Jahrhundert verloren (vgl. Lass 1999: 153).

In der Lyrik jedoch bleiben die T-Formen erhalten. So beginnt ein berühmtes, 1850 von Elizabeth Barrett Browning veröffentlichtes Sonett mit diesem Vers: "How do I love thee? Let me count the ways."

Die Verwendung der T-Formen in der englischsprachigen Lyrik auch weit nach ihrem Verlust in der Alltagssprache ist bisher nicht systematisch untersucht.

Forschungsfragen

(1) Ist die Verwendung von T-Formen und V-Formen in Sonetten nach dem Sprachwandel synonym?

Falls nein: (2) Welche Muster lassen sich beschreiben?

Relevanz für die historische Linguistik: Beitrag zur präziseren Beschreibung der Sprachentwicklung.

Relevanz für die Literaturwissenschaft: Systematische Aufmerksamkeit auf Anrede-Pronomen verbessert Einzelinterpretationen und erweitert Perspektive für die aktuelle Diskussion zur Anrede im Gedicht.

Korpus

1.611 Sonette: englischsprachig, auf den britischen Inseln zwischen 1530 und 1910 publiziert
TEI-5 konformes XML

```
4233 <sourceDesc>
4234 <author key="pp.PullingWilliam">William Pulling</author>
4235 <title>Sonnets Written Strictly in the Italian Style</title>
4236 <date type="published">1848</date>
4237 <idno type="BL">http://explorer.bl.uk/BLVUI:LSOOP-ALL:BL01017475848</idno>
4238 </sourceDesc>
4239 </fileDesc>
4240 <teiHeader>
4241 <met>
4242 <body>
4243 <div type="primaryText">
4244 <head type="section"></head>
4245 <head to="span type="nounOfAddress" xml:id="n_1394_h1">the Muse</span></head>
4246 <head type="subTitle"></head>
4247 <lg type="sonnet" rime="abababababab" ana="lud_intermediary #aj_poom #da_positive #ea_up #fc_distant">
4248 <l n="1">On my young cheek a smile was wont to play,</l>
4249 <l n="2">When tenderly, <span type="nounOfAddress" xml:id="n_1394_h1">I loved Muse</span> <span type="referent" xml:id="n_1394_h2">my mothe</span>
4250 <l n="3">--said <span type="nounOfAddress" xml:id="n_1394_h1">I lov'd thee, <span type="nounOfAddress" xml:id="n_1394_h2">winter, dear boy</span> now distinct </l>
4251 <l n="4">--said <span type="nounOfAddress" xml:id="n_1394_h1">I lov'd thee, <span type="nounOfAddress" xml:id="n_1394_h2">winter, dear boy</span> now distinct </l>
4252 <l n="5">She knew not that my bosom ow'd <ref target="n_1394_h1">thy</ref> <ref type="saw">smile</ref> </l>
4253 <l n="6">That to <ref target="n_1394_h1">thy</ref> <ref type="shade-emboss'd">sest I hid</ref> </l>
4254 <l n="7">When from my father's door I sped my way,--</l>
4255 <l n="8">She knew not why at home so oft I sigh'd</l>
4256 <l n="9"><ref target="n_1394_h1">Thy</ref> sylvan residence I eager sought,</l>
4257 <l n="10">Led by a spirit, thro' the murkiest night,</l>
4258 <l n="11">To be by <ref target="n_1394_h1">thee</ref>--the charms of Nature taught;--</l>
4259 <l n="12">All objects touch'd by <ref target="n_1394_h1">thee</ref> seem'd heav'nly bright,</l>
4260 <l n="13">A force unknown <ref target="n_1394_h1">thou</ref>--distinct type="thst"spavst</distinct> to ev'ry thought,</l>
4261 <l n="14">And <ref target="n_1394_h1">thou</ref>--distinct type="thst"spavst</distinct> Poesy my chief delight</l>
4262 </lg>
4263 </div>
```

Metadaten und Annotationen: Autor:in, Titel der Publikation, Entstehungsjahr, nominale Anredeformen, Kategorie der Adressat:innen (Gott, Mensch, Tier, Naturphänomen etc.), intertextuelle Verweise, Reimschemata

Projektleitung und Kontakt: Brigitte Rath, Universität Innsbruck. Email: brigitte.rath@uibk.ac.at

Team: Marina Höfler, Serena Obkircher, Teresa Wolf (Transkription und Annotation), Ario Santoso, Mingzi Kong (Machine Learning Prediction Models)

Experiment Synonymität

Einteilung der Sonette in Kategorie "Y" (Verwendung von V-Formen) oder "T" (Verwendung von T-Formen).

Machine Learning Prediction Models werden trainiert, Sonette (ohne Pronomina) einer Kategorie zuzuordnen.

Fünf **Machine Learning Prediction Models:**

Naive Bayes, Support Vector Machine, Decision Tree, Random Forest, XGBoost

Drei **Baseline-Models:** ZeroR, AlwaysT und AlwaysV

Vergleichsbasis: Precision, Recall, F-Measure, Accuracy, Area Under the ROC Curve (AUC) (vgl. Han, Pei, Kamber 2011; Mohri, Rostamizadeh, Talwalkar 2012)

Training: k-fold cross validation approach (vgl. Han, Pei, Kamber 2011)

Ergebnis: Machine Learning Modelle, insbesondere XGBoost, liefern für die Vorhersage bessere Ergebnisse als die Baseline-Modelle.

Tabelle: Ergebnisse für trainierte Machine Learning Prediction Models und Vergleichsmodelle

Model Name	Acc.	AUC	Macro-average			Weighted-average		
			Prec.	Rec.	F1	Prec.	Rec.	F1
ZeroR	0.797	0.5	0.399	0.5	0.444	0.636	0.797	0.707
Always Y	0.203	0.5	0.101	0.5	0.169	0.041	0.203	0.068
Always T	0.797	0.5	0.399	0.5	0.444	0.636	0.797	0.707
Naive Bayes	0.767	0.627	0.636	0.627	0.63	0.762	0.767	0.764
SVM	0.797	0.583	0.399	0.5	0.444	0.636	0.797	0.707
SVM - opt - acc	0.836	0.815	0.749	0.707	0.723	0.825	0.836	0.828
SVM - opt - fl	0.824	0.814	0.731	0.722	0.724	0.824	0.824	0.823
Decision tree	0.777	0.637	0.65	0.637	0.641	0.77	0.777	0.772
Decision tree - opt - acc	0.836	0.771	0.76	0.69	0.712	0.825	0.836	0.825
Decision tree - opt - fl	0.836	0.771	0.76	0.69	0.712	0.825	0.836	0.825
Random Forest	0.83	0.778	0.826	0.602	0.619	0.83	0.83	0.787
Random Forest - opt - acc	0.85	0.765	0.839	0.66	0.695	0.848	0.85	0.824
Random Forest - opt - fl	0.839	0.83	0.751	0.755	0.753	0.841	0.839	0.839
XGBoost	0.841	0.831	0.768	0.685	0.711	0.826	0.841	0.826
XGBoost - opt - acc	0.854	0.814	0.872	0.656	0.694	0.86	0.854	0.825
XGBoost - opt - fl	0.842	0.838	0.757	0.735	0.745	0.836	0.842	0.838

Interpretation: Das Modell erkennt Muster in der Verteilung von T- und V-Formen: T- und V-Formen im Sonett sind nicht austauschbar und so nicht synonym.

Experiment Muster

Ablation-Experimente: Input-Faktoren einzeln entfernt, erneute Messung der Performanz des Modells. Bei sinkender Vorhersagekraft spielt der entsprechende Faktor für die Vorhersage dieses Modells eine Rolle. Erstes Indiz, dass der Faktor zur Musterbildung beiträgt.

Erste Ergebnisse: Die Verwendung des Pronomens "ye", bisher bei der Entwicklung von V- und T-Formen kaum beachtet, spielt eine wichtige Rolle; Kategorie der Angesprochenen (Gott, Mensch, Tier, Naturphänomen etc.) ist ein weiterer möglicher Faktor.

Nächste Schritte: Ausdifferenzierung vor und nach dem Sprachwandel; detaillierte deskriptive Statistik.