

Muster von “you” und “thou”

Modellierung der Anrede im englischen Sonett

Rath, Brigitte

brigitte.rath@uibk.ac.at
Universität Innsbruck, Austria

Bekanntermaßen unterscheidet das Early Modern English zwei Reihen von Pronomina der zweiten Person Singular: V-Formen (you, your etc.) und T-Formen (thou, thy etc.) Die T-Formen gehen im Zuge des Sprachwandels zum Modern English im Verlauf des 17. Jahrhunderts verloren (vgl. z.B. Lass 1999: 153), bleiben jedoch in der Lyrik erhalten. So beginnt etwa ein berühmtes, 1850 von Elizabeth Barrett Browning veröffentlichtes Sonett mit diesem Vers: “How do I love thee? Let me count the ways.” Diese wie selbstverständliche Verwendung der T-Formen in der englischsprachigen Lyrik auch weit nach dem Wandel zum Modern English ist bisher nicht systematisch untersucht.

Dieses Projekt sucht daher mit Hilfe digitaler Methoden Antworten auf folgende zwei Forschungsfragen:

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

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

Die Hypothese zur Frage (1) lautet, dass die Verwendung von T-Formen und V-Formen sich als nicht synonym erweisen wird, weil es gerade in der Lyrik eine gesteigerte Sensibilität für die Anrede gibt, und so zu erwarten steht, dass die linguistisch gebotenen Möglichkeiten für Differenzierung voll ausgeschöpft werden. Diese Erwartung widerspricht dem in Gedichtkommentaren häufig anzutreffenden und üblicherweise nicht weiter belegten Hinweis, “thou” sei einfach eine in Gedichten anzutreffende Version von “you”.

Hypothesen für Faktoren, die eine Rolle bei der Frage (2) interessierenden Musterbildung spielen könnten, werden vor allem aus der historischen Soziolinguistik gewonnen: So kommen neben potentiellen individuellen Vorlieben von Autor:innen sowie der Entstehungszeit als plausible Faktoren nominale Anredeformen in Frage, weil Studien aus der historischen Soziolinguistik nahelegen, dass bestimmte Anreden (z.B. “terms of endearment”) mit der Verwendung von T-Formen verbunden sind (vgl. Nevala 2004: 2146; Mazzon 2010), sowie die jeweilige Kategorie des:der Angesprochenen, weil Studien einen Zusammenhang zwischen bestimmten Kategorien von Angesprochenen wie etwa Kinder, Tiere oder Geister und der Verwendung von T-Formen zeigen (vgl. Yang 1991: 258; Carter/McRae 2002: 120-121).

Basis für diese Untersuchung ist ein selbsterstelltes Korpus von (bisher) 1.611 englischsprachigen, auf den britischen Inseln zwischen 1530 und 1910 publizierten Sonetten, für das die Gedichttexte manuell in TEI-5 konformem XML transkribiert und mit Metadaten (Autor:in, Titel, Entstehungsjahr, Publikationsjahr) und Annotationen zu nominalen Anredeformen, Kategorie der Adressat:innen (Gott, Mensch, Tier, Naturphänomen etc.), intertextuellen Verweisen und Reimschemata angereichert werden.

Mit diesem Korpus wurde eine Reihe von Experimenten mit Machine Learning Prediction Modellen gemacht. Mit fünf verschiedenen Machine Learning Prediction Modellen (Naive Bayes, Support Vector Machine, Decision Tree, Random Forest und XG-

Boost) wurde jeweils der k-fold cross validation approach (vgl. z.B. Han, Pei, Kamber 2011) durchgeführt. Die jeweiligen Trefferquoten wurden mit drei Baseline-Modellen (ZeroR sowie zwei Modellen, die alle Sonette jeweils einer Klasse zuordnen, hier: AlwaysT und AlwaysV) auf der Basis üblicher Standardwerte für Machine Learning verglichen: Precision, Recall, FMeasure, Accuracy und Area Under the ROC Curve (AUC). (vgl. z.B. Han, Pei, Kamber 2011; Mohri, Rostamizadeh, Talwalkar 2012) Es zeigt sich, dass Machine Learning Modelle, insbesondere XG-Boost, bessere Ergebnisse als die Baseline-Modelle für die Vorhersage liefern können.

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	0.829
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

Abb. 1

Da dieses Modell einen hohen Anteil an Fällen korrekt zuordnet, folgt, dass das Modell Regeln in der Verteilung von T- und V-Formen erkennt, dass T- und V-Formen im Sonett also nicht austauschbar sind. Auf der Basis dieser Experimente kann so die erste Frage, “Ist die Verwendung von T-Formen und V-Formen in Sonetten nach dem Sprachwandel synonym?” tentativ mit nein beantwortet werden.

Für Hinweise auf mögliche Faktoren, die die Musterbildung beeinflussen, wurden mit den Machine Learning Prediction Modellen Ablation-Experimente durchgeführt: Input-Faktoren wurden individuell entfernt und die jeweilige Performanz des Modells erneut gemessen. Sinkt die Vorhersagekraft des Modells durch das Entfernen eines Faktors, so spielt dieser Faktor für die Vorhersage dieses Modells eine Rolle, was als ein erstes Indiz dafür gewertet werden kann, dass der entsprechende Faktor zur Musterbildung auch jenseits des Modells beitragen könnte. Es zeigt sich, dass dabei die Verwendung des Pronomens “ye”, das bisher bei der Entwicklung von V- und T-Formen kaum beachtet wird, eine wichtige Rolle spielt; als weiterer möglicher Faktor erweist sich die Kategorie der Angesprochenen.

Das Projekt bietet für die historische Linguistik einen Beitrag zur präziseren Beschreibung der Sprachentwicklung. Für die Literaturwissenschaft erlaubt diese erstmalige systematische Beschreibung der Verteilung von T-Formen und V-Formen in englischsprachigen Sonetten bessere Gedichtinterpretationen, weil sie erstens überhaupt ein Augenmerk auf die verwendeten Pronomen der Anrede legt und zweitens die im Einzeltext gewählten Formen nun vor dem Hintergrund eines Moders gelesen werden können. Das Projekt trägt so zur aktuellen Forschungsdiskussion zur Anrede in der Lyrik bei. (vgl. z.B. Culler 1981, Culler 2015, Hedley 2009, Keniston 2006, Pollard 2012, Waters 2012)

Dieses Projekt wurde vom Vizerektorat Forschung der Universität Innsbruck mit Mitteln aus der Aktion D. Swarovski und vom Forschungszentrum Digital Humanities der Universität Innsbruck durch Mittel aus dem DI4DH Programm unterstützt und so erst ermöglicht. Die Korpuserstellung übernahmen mit einem ebenso

scharfen Blick fürs Detail wie für das Gesamtprojekt Marina Höfler, Serena Obkircher und Teresa Wolf. Die Machine Learning Prediction Models wurden mit großer Umsicht von Ario Santoso und Mingzi Kong konzipiert, implementiert und trainiert.

Bibliographie

Carter, Ronald/McRae, John (2002). "Language Note: Changing patterns of 'thou' and 'you'". *The Routledge History of Literature in English. Britain and Ireland*. Second Edition. Routledge. 120-121.

Culler, Jonathan (1981). "Apostrophe". Jonathan Culler. *The Pursuit of Signs. Semiotics, Literature, Deconstruction*. Routledge and Kegan Paul. 135-154.

Culler, Jonathan (2015). *Theory of the Lyric*. Harvard UP.

Han, Jiawei / Pei, Jian / Kamber, Micheline (2011). *Data Mining: Concepts and Techniques*. Elsevier.

Hedley, Jane (2009). *I Made You to Find Me. The Coming of Age of the Woman Poet and the Politics of Poetic Address*. Ohio State UP.

Keniston, Ann (2006). *Overheard Voices. Address and Subjectivity in Postmodern American Poetry*. Routledge.

Lass, Roger (1999). "Phonology and Morphology". Roger Lass (ed.). *The Cambridge History of the English Language*. Volume III 1476-1776. Cambridge UP. 56-186.

Mazzon, Gabriella (2010). "Terms of Address". Andreas H. Jucker, Irma Taavitsainen (eds.) *Historical Pragmatics*. De Gruyter Mouton. 351-376. (Handbook of Pragmatics 8)

Mohri, Mehryar / Rostamizadeh, Afshin / Talwalkar, Ameet (2012). *Foundations of Machine Learning*. MIT Press.

Nevala, Minna (2004). "Accessing politeness axes: forms of address and terms of reference in early English correspondence". *Journal of Pragmatics* 36: 2125-2160.

Pollard, Natalie (2012). *Speaking to You. Contemporary Poetry and Public Address*. Oxford: Oxford University Press.

Waters, William (2012). Art. "Address". *The Princeton Encyclopedia of Poetry and Poetics*. Ed. Ronald Greene, Stephen Cushman. Fourth edition. Princeton UP. 6-8.

Yonglin, Yang (1991). "How to talk to the Supernatural in Shakespeare". *Language in Society* 20/2: 247-261.