

Annotation of Non-Standard Varieties

Seltmann, Melanie E.-H.

melanie.seltmann@univie.ac.at
Universität Wien, Österreich

Das Dissertationsprojekt beschäftigt sich mit nachhaltiger Annotation sprachlicher Kategorien in der Variationslinguistik. Untersuchungsgegenstand ist gesprochene Sprache. Hierbei fokussiert das Projekt verschiedene thematische Bereiche: Hauptthemen sind dabei zum einen Standardisierungsversuche von linguistischen Annotationen, Annotation insbesondere von Nonstandardvarietäten und dabei auftretende Probleme, Annotation als Kategorisierung und die damit verbundenen Möglichkeiten des Erkenntnisgewinns sowie der Explikation theoretischer Grundannahmen, die Beeinflussung der Kategorien vorab anhand der Erhebungsmethode, technische Aspekte der Modellierung und Darstellung, Nachhaltigkeit der Annotation sowie eine kritische Reflexion verwendeter Annotationsmethoden.

Die Dissertation entsteht im Rahmen des Spezialforschungsbereichs „Deutsch in Österreich. Variation – Kontakt – Perzeption“ (FWF F60) (im Folgenden SFB) im Teilprojekt PP11 „Kollaborative Online-Forschungsplattform „Deutsch in Österreich““ und nimmt auch deswegen die Annotationen des SFB als Ausgangs- und Referenzgröße (vgl. Budin et al. 2018). Der SFB bietet die Möglichkeit ein geschlossenes Annotationssystem, das alle linguistischen „Systemebenen“ aus verschiedenen sprachwissenschaftlichen Perspektiven abdeckt, zu entwickeln bzw. iterativ dessen Entwicklung zu beobachten und zu untersuchen. Insbesondere im Bereich Standardisierung von Annotationen werden auch andere linguistische Projekte zum Vergleich herangezogen und nationale sowie internationale Forschungsverbände und ihre Best Practices sowie De-facto-Standards berücksichtigt.

Ziel der Arbeit ist es herauszuarbeiten, welche Vorteile und Möglichkeiten in einer standardisierten Annotation liegen, welche Möglichkeiten es für eine weitestgehende Standardisierung innerhalb der (Variations-)Linguistik gibt, aber auch welche Probleme dieser inne liegen. Zeitgleich wird reflektiert, inwiefern ein Versuch eines solchen standardisierten Annotationssystems (nämlich das des SFB) gelingen kann und wo seine Grenzen liegen. Dabei wird ein neuartiges Annotationssystem vorgeschlagen, das insbesondere auf die Standardisierung der Modellierung von Tag Sets abzielt, aber auch auf die Standardisierung des verwendeten Annotationsvokabulars (vgl. Breuer/Seltmann 2018: 147).

Der Dissertation liegt die übergeordnete Forschungsfrage zu Grunde:

Was sind Anforderungen an ein Annotationssystem für die Variationslinguistik, insbesondere zur Annotation von Nonstandardvarietäten? Inwiefern können Annotationssysteme zu einer guten Forschungspraxis beitragen und den Erkenntnisgewinn bzw. die -vermittlung erhöhen oder zumindest erleichtern? Welche Erfahrungen können im SFB „Deutsch in Österreich. Variation – Kontakt – Perzeption.“ mit dem dafür entwickelten Annotationssystem gemacht werden? Welche Stärken und Schwächen hat dieses?

Um diese Fragestellung beantworten zu können, orientieren sich die verschiedenen Papers an folgenden Unterfragen:

1. Kategorisierungsprozess
 - Inwiefern handelt es sich bei der Annotation um einen Erkenntnisprozess?
 - Welche Vor- und Nachteile bringt die Annotation als eine Methode der Digital Humanities dem Erkenntnisprozess?
 - Unter welchen Voraussetzungen können auch Citizens Analyses mittels Annotation betreiben?
2. Standardisierung
 - Welche Anforderungen werden an ein Annotationssystem für Nonstandardvarietäten gestellt?
 - Inwiefern ist ein solches Framework standardisierbar bzw. welche Aspekte davon?
 - Welche Vor- und Nachteile birgt ein standardisiertes Annotationssystem?
3. Variationslinguistische Variable in der Annotationsumsetzung
 - Wie ist im Hinblick auf linguistische Annotationen systemübergreifend eine Variable zu definieren?
 - Wie ist der Variablenbegriff auf verschiedene linguistische Systemebenen ansetzbar?
 - Wie ist ein solcher Variablenbegriff technisch für ein Annotationssystem implementierbar?
 - Inwiefern kann die technische Implementierung helfen, den Variablenbegriff für die Variationslinguistik greifbarer zu machen?
 - Welche technisch-formale Modellierung eignet sich für diesen Variablenbegriff: eher eine hierarchische oder eher eine relationale?
 - Inwiefern unterstützt die technische Abbildung der Variable den Forschungsprozess? Inwiefern beeinflusst sie die Ergebnisse?
 - Inwiefern werden die Ergebnisse des Forschungsprozesses vorab durch die Erhebungsmethode beeinflusst?

4. Nachhaltigkeit

- Was bedeutet „Nachhaltigkeit“ in Forschung und Infrastruktur bei (variations-) linguistischen Projekten im Hinblick auf Annotationen?
- Wie kann eine solche Nachhaltigkeit in linguistischen Projekten erreicht werden?
- Wie versucht konkret der SFB Nachhaltigkeit sowohl der Forschungsergebnisse als auch der technischen Erzeugnisse zu gewährleisten?
- Wie kann Nachhaltigkeit auch für die (interessierte) Öffentlichkeit gewährleistet werden?

5. Usability

- Welche Erfahrungen und Probleme sind mit der Umsetzung der Annotationsrichtlinien im SFB eingetreten?
- Inwiefern waren Abläufe und Teilaspekte des Annotationsprozesses zielführend und hilfreich, an welchen Punkten bedarf es Verbesserung?

Der Dissertation liegen die theoretisch-methodologischen Grundlagen verschiedener Paradigmen zugrunde, einerseits variationslinguistische Grundlagen, auf denen der SFB aufbaut (u.a. Auer 2005, Lenz 2003, Kehrein 2012), zweitens theoretische Konzepte der Korpus- und Computerlinguistik sowie der Digital Humanities (u.a. Chiarcos 2009, Ide/ Pustejovsky 2017), die für die Annotationen von Bedeutung sind, und drittens unterschiedliche linguistische Theoriekonzepte, die für die spezifischen Modellierungen der einzelnen untersuchten Phänomene auf den verschiedenen Systemebenen wichtig sind.

Die Dissertation wird als kumulative Dissertation geschrieben und strebt 5 Forschungspapers an (s.o.). Die Papers werden zuvor auf wissenschaftlichen Fachtagungen diskutiert. Dabei reichen die verschiedenen Papers in unterschiedliche Schwerpunkte hinein.

Die Datenbasis bildet das Korpus des SFB, insbesondere auditive Aufnahmen aus verschiedenen Erhebungssettings wie beispielsweise Sprachproduktionsexperimente, Übersetzungsaufgaben, Interviews oder Freundesgesprächen (vgl. Lenz 2018) von verschiedenen Orten (urban und rural) in Österreich.

Variation – Kontakt – Perzeption“ In: *Zeitschrift für germanistische Linguistik* 46(2) 300–308.

Chiarcos, Christian / Dipper, Stefanie / Götze, Michael / Leser, Ulf / Lüdeling, Anke / Ritz, Julia / Stede, Manfred (2009): A flexible framework for integrating annotations from different tools and tagsets. In: *Traitement Automatique des Langues*, 49 (2), 271–291.

Ide, Nancy / Pustejovsky, James (Hg.) (2017): *Handbook of Linguistic Annotation*. Dordrecht.

Kehrein, Roland (2012): *Regionalsprachliche Spektren im Raum – Zur linguistischen Struktur der Vertikale* (= *Zeitschrift für Dialektologie und Linguistik. Beihefte* 152). Stuttgart.

Lenz, Alexandra N. (2018): The Special Research Programme „German in Austria. Variation – Contact – Perception“. In: Ammon, Ulrich / Costa, Marcella (eds.): *Sprachwahl im Tourismus – mit Schwerpunkt Europa. Language Choice in Tourism – Focus on Europe. Choix de langues dans le tourisme – focus sur l’Europe*. Berlin / Boston: De Gruyter (*Yearbook Sociolinguistica* 32) 269–277.

Lenz, Alexandra N. (2003): *Struktur und Dynamik des Substandards*. Eine Studie zum Westmitteldeutschen (Wittlich/Eifel) (= *Zeitschrift für Dialektologie und Linguistik. Beihefte* 125). Stuttgart.

Bibliographie

Auer, Peter (2005): Europe's sociolinguistic unity, or. A typology of European dialect/standard constellations. In: *Perspectives on Variation, Sociolinguistic, Historical, Comparative*. Berlin, 7–42.

Breuer, Ludwig M. / Seltmann, Melanie E.-H. (2018): Sprachdaten(banken) – Aufbereitung und Visualisierung am Beispiel von SyHD und DiÖ. In: Börner, Ingo / Straub, Wolfgang / Zolles, Christian (eds.): *Germanistik digital*. Wien, 135–152.

Budin, Gerhard / Elspaß, Stephan / Lenz, Alexandra N. / Newerkla, Stefan M. / Ziegler, Arne (2018): Der Spezialforschungsbereich „Deutsch in Österreich (DiÖ).