

## Menschen gendern? Einige Gedanken über Datenmodellierung zur Erhebung von Geschlechterverteilung anhand der TEI2016 Abstracts App

**Hanneschläger, Vanessa**

vanessa.hanneschlaeger@oeaw.ac.at  
OEAW Österreichische Akademie der Wissenschaften,  
Österreich

**Andorfer, Peter**

peter.andorfer@oeaw.ac.at  
OEAW Österreichische Akademie der Wissenschaften,  
Österreich

### Ausgangspunkt: Die TEI2016 Abstracts App

Die Abstracts zur TEI Konferenz 2016 wurden via ConfTool eingereicht und anschließend in Microsoft Word und InDesign ediert, um ein gedrucktes Book of Abstracts herzustellen.<sup>1</sup> Die InDesign-Datei wurde dann als PDF exportiert und online verfügbar gemacht. Diese beiden Fassungen des Book of Abstracts wurden vor der Konferenz fertiggestellt. Nach der Konferenz wurde die InDesign Datei erneut exportiert, diesmal in XML Format. Diese Datei wurde dann teils manuell, teils automatisiert in Einzeldateien (jedes Abstract als eine Datei) zerlegt, bearbeitet und mit umfangreichen Metadaten versehen, um TEI-P5-Dateien herzustellen. Diese wurden auf GitHub veröffentlicht. Die jeweiligen <teiHeader> wurden mit umfassenden Informationen ausgestattet und detailliert codiert; etwa wurden innerhalb der mit Normdaten (VIAF) angereicherten <editor>- und <author>-Tags konsequent <persName>-, <forename>- und <surname>-Tags eingesetzt. Das stellte sich für die Weiterverarbeitung, Analyse und Visualisierung der Daten später als vorteilhaft heraus.

Dieses Corpus bildete die Basis für die TEI2016 Abstracts App, in der die Abstracts in der Erstversion über ein Inhaltsverzeichnis, das nach Abstract-Titel oder nach Verfassenden sortiert werden konnte, ansteuerbar waren. Schon in dieser rudimentären Form zeigte sich das Potential der Daten und der Applikation für Auswertungen, die über Struktur und Zusammensetzung der Konferenz und ihrer

Beiträgenden Auskunft geben würden. Die App wurde daher sowohl im Bereich der Funktionalität als auch auf der Ebene der enthaltenen Daten substantiell weiterentwickelt.

### Fragestellung und gender- theoretische Überlegungen

Wir trafen die Entscheidung, der weiteren Entwicklung des Prototyps der App eine Forschungsfrage zugrundezulegen. Wir entschieden uns für die Frage nach der Geschlechterverteilung unter den Beitragenden.

Aus Perspektive der Gendertheorie (in Butler'scher Tradition) bedarf die Entscheidung, diese Frage zu stellen, an sich eine Rechtfertigung, denn "a performative utterance (or practice) brings into being that of which it speaks." (Butler 2010, 150f) Das bedeutet, dass die Frage nach der Verteilung der Geschlechter die Unterscheidung der Geschlechter erst hervorbringt, und so die historische Situation, die unser Körper bedeutet, fortsetzt: "As an intentionally organized materiality, the body is always an embodying of possibilities both conditioned and circumscribed by historical convention. In other words, the body is a historical situation, as Beauvoir has claimed, and is a manner of doing, dramatizing, and reproducing a historical situation." (Butler 1988, 521) Berücksichtigt man außerdem die Konzepte des *Doing* bzw. *Undoing Gender* (West & Zimmerman 1987; Hirschauer 1994 & 2001), so wird deutlich, dass die oben formulierte Frage deutlich präzisiert werden muss, um die zu erhebenden Daten in einer Weise strukturieren zu können, die eine zeitgenössische Auffassung von Geschlechteridentitäten wiedergibt.

Die Frage nach dem Geschlecht der Konferenzbeitragenden wurde daher wie folgt präzisiert: Da sie aus einem Interesse an der Stellung bzw. Repräsentation der Geschlechter im gegenwärtigen Wissenschaftsbetrieb heraus gestellt wurde, richtet sie sich nicht auf die biologische Beschaffenheit der Körper der Konferenzbeitragenden, sondern auf ihre gesellschaftliche (Selbst-)Wahrnehmung (also auf *gender* und nicht *sex*). Folgt man Butlers Auffassung, nach der die performative Äußerung das, wovon sie spricht, hervorbringt, wären daher eigene Angaben der Betroffenen zu ihrem Geschlecht die ideale Datenquelle für diese Fragestellung gewesen. Solche Angaben standen allerdings nicht zur Verfügung.

### Datenmodellierung und - anreicherung

Nachdem eine Personenliste aus dem abstracts-Corpus extrahiert worden war, kamen wir zu dem Schluss, dass das Zuweisen von Geschlechtern an Personen unter Berücksichtigung der oben skizzierten geschlechtertheoretischen Überlegungen kein gangbarer

Weg sein konnte. Um die soziodeterministische Neugier, aus der heraus die Ausgangsfrage gestellt worden war, dennoch befriedigen zu können, entwickelten wir ein gendersensibles Workaround für die Modellierung der Daten, das uns dennoch eine Erhebung der Geschlechterverteilung erlaubte: Anstatt den <persName>s ein <sex> zuzuweisen, wie es die TEI-Richtlinien vorsehen, entschieden wir uns dafür, dem jeweiligen <forename> einen @type zuzuweisen, der eine Angabe zum (sozialen) Geschlecht enthalten sollte. Diese Vorgehensweise erlaubte es, die Daten nicht den eigentlichen Personen nachzubauen und damit annahmen über deren Selbstwahrnehmung zu treffen, sondern zu erheben, welchem Geschlecht sie von einer allgemeinen Gesellschaft oder Öffentlichkeit aufgrund ihrer Namen aller Wahrscheinlichkeit nach zugeordnet würden.

Um das Vorgehen weiter zu objektivieren, entschieden wir uns gegen die Zuordnung der Geschlechter zu den Vornamen auf Basis unseres eigenen Weltwissens. Stattdessen machten wir uns auf die Suche nach Namensdatenbanken, die Angaben zum Geschlecht enthalten. Im Natural Language Toolkit (NLTK) etwa ist eine solche Liste enthalten (Kantrowitz 1997), jedoch werden keine Angaben gemacht, wie diese Liste zustande kam und auf welcher Grundlage die Geschlechter zugeordnet wurden. Wir entschieden uns daher für genderize.io, eine Datenbank, die nach eigenen Angaben auf Daten von Social-Media-Plattformen basiert, wo Personen die Informationen zu ihrem Geschlecht selbst bestimmen können. Nach dem Abgleich unserer Personenliste mit genderize.io via deren API konnten wir 102 von 124 Namen ein Geschlecht zuordnen. Da genderize.io nur zwei Geschlechter kennt, setzten wir "male", "female" und "no-match" als @type-Werte ein. Die Namen, die nach diesem ersten Abgleich "no-match" waren, verglichen wir anschließend manuell mit der genderchecker.com-Datenbank, die ihre Geschlechtsangaben aus den "2001 and 2011 UK Census Data" bezieht, "together with multiple online sources and contributions from our 2m website visitors". Diese kommerzielle Datenbank konnte nicht für einen Gesamtvergleich genutzt werden, da nur die manuelle Verwendung kostenfrei ist. Ein interessanter Aspekt des Datenmodells von genderchecker.com ist die Berücksichtigung der Möglichkeit eines dritten Geschlechts: "If we see just one instance of a name appearing as both male and female, we categorise it as unisex." Aus gendertheoretischer Perspektive ist dieses Modell ein Schritt in die richtige Richtung, wenngleich die Benennung "unisex" diskutabel scheint. In unserer Namensliste kam kein solcher Fall vor - nach dem Abgleich der verbleibenden 22 Namen mit genderchecker.com blieben drei Namen "no-match".

Da Daten und Code, aus denen die TEI2016 Abstracts App gebaut ist, unter einer offenen Lizenz auf GitHub frei zugänglich sind, steht es den Beitragenden frei, ihr Geschlecht und damit unsere statistische Auswertung der Geschlechterverteilung zu verändern - das Ergebnis bleibt

also ein vorläufiges. Im geplanten Vortrag wird es im Rahmen einer live-Demonstration der App präsentiert werden.

## Fußnoten

1. Dieser Workflow ist, zugegebenermaßen, einer Konferenz im Bereich der Digital Humanities, und der TEI Konferenz im Besonderen, nicht ganz angemessen; eine Integration etwa des DH-Convalidators hätte den Prozess wesentlich gestrafft. Eingespielte Prozesse und Zeitdruck waren die Hauptgründe für diese Vorgehensweise.

## Bibliographie

**Peter Andorfer, Vanessa Hanneschläger** (2017): TEI Abstracts 2016. A little application to publish the abstracts of the TEI conference 2016. <http://tei2016app.acdh.oeaw.ac.at/>

**Judith Butler** (1988): Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. In: Theatre Journal 40:4, 519–531.

**Judith Butler** (2010): Performative Agency. In: Journal of Cultural Economy 3:2, 147–161.

**Genderchecker**: <http://genderchecker.com/>  
**genderize.io: Determine the gender of a first name.** <https://genderize.io/>

**Vanessa Hanneschläger, Daniel Schopper (Hg.)** (2017): TEI Conference and Members' Meeting 2016. Book of Abstracts [XML/TEI files]. <https://github.com/acdh-oeaw/TEI2016abstracts>

**Stefan Hirschauer** (1994): Die soziale Fortpflanzung der Zwei-Geschlechtlichkeit. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 46:4, 668–692.

**Stefan Hirschauer** (2001): Das Vergessen des Geschlechts: Zur Praxeologie einer Kategorie sozialer Ordnung. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie (Sonderheft 41), 208–235.

**Mark Kantrowitz** (1997): Name Corpus: List of Male, Female, and Pet names. CMU Artificial Intelligence Repository. <http://www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/nlp/corpora/names/>

**Natural Language Toolkit (NLTK).** <http://www.nltk.org/>

**Susanne Oelkers** (2003): Naming gender. Empirische Untersuchungen zur phonologischen Struktur von Vornamen im Deutschen. Frankfurt a. M.: Lang.

**TEI Consortium** (2017): TEI P5: Guidelines for Electronic Text Encoding and Interchange. <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>

**Virtual International Authority File (VIAF).** <https://viaf.org/>

**Candance West, Don Zimmerman** (1987): Doing Gender. In: Gender and Society 1:2, 125–151.