

# Biographik in den Digital Humanities – Kritische Bestandsaufnahme und quantitative Analysemöglichkeiten am Beispiel des Österreichischen Biographischen Lexikons 1815–1950

**Schlögl, Matthias**

matthias.schloegl@oeaw.ac.at  
Österreichische Akademie der Wissenschaften, Österreich

**Bernád, Ágoston**

agoston.bernad@oeaw.ac.at  
Österreichische Akademie der Wissenschaften, Österreich

**Kaiser, Maximilian**

maximilian.kaiser@oeaw.ac.at  
Österreichische Akademie der Wissenschaften, Österreich

**Lejtovicz, Katalin**

katalin.lejtovicz@oeaw.ac.at  
Österreichische Akademie der Wissenschaften, Österreich

**Rumpolt, Peter**

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

Quantitative Auswertungen gewinnen zunehmend auch in den Geisteswissenschaften an Bedeutung (z. B. Harber 2011). Nicht selten wird mittlerweile auch der Begriff „Big Data“ in Zusammenhang mit den Humanities ganz allgemein verwendet (z. B. gral 2016). Für solche Analysen wird des Öfteren auf einheitliche Korpora zurückgegriffen (z. B. Wikipedia: Russo et al. 2015), die nicht nur einen standardisierten Zugriff auf tausende Datensätze, sondern auch homogene Metadaten und harmonisierte Vokabulare bieten. Während diese Korpora formal geradezu prädestiniert für Analysen mit den neuen digitalen Tools scheinen, sind die grundlegenden Daten durch die Entstehungsgeschichte der Datensammlungen nicht selten unausgewogen und können mitunter zu falschen Ergebnissen führen. Insbesondere trifft dies auf über

einen längeren Zeitraum entstandene Nationalbiographien zu. Trotz dieses offensichtlichen Dilemmas fokussieren Studien meist auf die Zuverlässigkeit der digitalen Tools selbst und weniger auf die Verlässlichkeit der Quellen (z. B. Reinert et al. 2015; Stotz et al. 2015).

Die Posterpräsentation bietet eine exemplarische Korpusanalyse der digitalen Fassung des 1946 gegründeten und seit 1954 (seit 2009 auch online) erscheinenden *Österreichische Biographische Lexikon 1815–1950*. Dieses enthält in 14 Bänden (68 Lieferungen) über 18.000 Biographien. Es wird dadurch die oftmals vernachlässigte Unausgewogenheit einer auf den ersten Blick einheitlich wirkenden Quelle diskutiert.

## Analyse der vorläufigen Ergebnisse

### 1. Informationsdichte

Anders als ursprünglich angenommen, bleibt die Informationsdichte (Named Entities pro tokens) mit steigender Länge der Biographietexte etwa gleich hoch. Biographien werden inhaltlich bereits ab Anfang der 1960er-Jahre umfangreicher. Ab den 1980er-Jahren zeigt die Kurve eine sehr deutliche Steigerung (Abb. 1.).

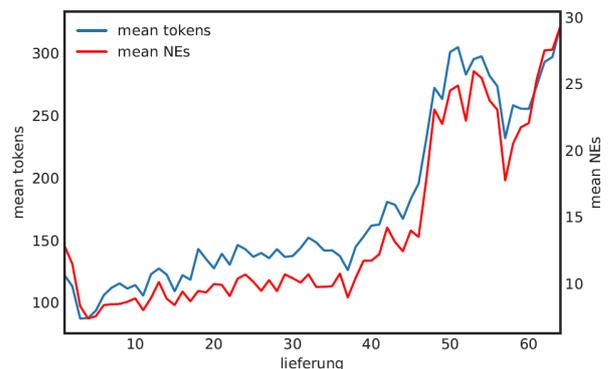


Abbildung 1: Durchschnitt Tokens und Named Entities nach Lieferung

Diese Entwicklung lässt sich anhand der Werkzeugenese bzw. der Geschichte des Lexikons erklären. Nach den ursprünglichen Plänen sollte das ÖBL drei bis vier Bände an Kurzbiographien umfassen. Die ab 1946 verfassten Biographien der ersten fünf Lieferungen – gesammelt im ersten Band, 1957 erschienen – sind allesamt nach diesem Konzept entstanden. Aus methodischen Gründen wurde der Ursprungsplan jedoch bereits in den 1950er-Jahren als unbrauchbar verworfen. Die Österreichische Akademie der Wissenschaften entschied sich für die Gestaltung eines umfangreichen Nachschlagewerkes (Reitterer 1998) wodurch die durchschnittliche Länge der Biographien deutlich anstieg.

## 2. Die geographische Verteilung der Geburtsorte der biographierten Personen

Die geographische Verteilung der Biographierten nach Geburtsorten ist nicht ausgewogen. Verglichen wurden die erste Lieferung des ÖBL (1954) mit den Lieferungen 62-66 (2010–2015); beide Stichproben enthalten etwa gleich viele Biographien (Abb. 2. und 3.).

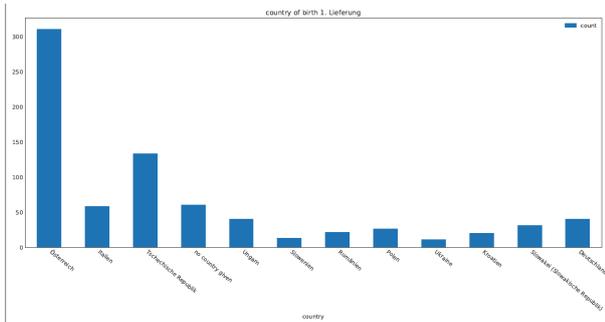


Abbildung 2: Verteilung der Geburtsorte in der 1. Lieferung des ÖBL (1954)

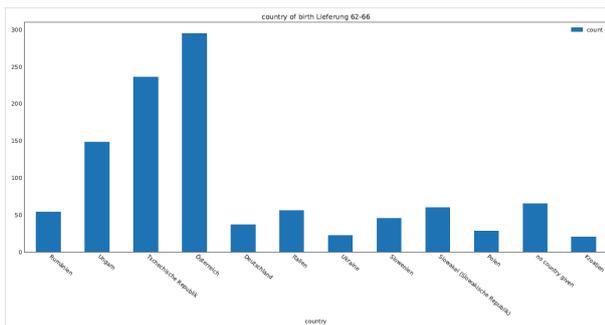


Abbildung 3: Verteilung der Geburtsorte in den Lieferungen 62–66 des ÖBL (2010–2015)

Die Kronländer der westlichen Reichshälfte der Monarchie (u. a. die Nachfolgestaaten Österreich und Tschechien) sind in der ersten Lieferung gegenüber der östlichen Reichshälfte (auf den Abbildungen die Nachfolgestaaten Ungarn, Slowakei, Rumänien [Siebenbürgen], Kroatien) deutlich bevorzugt. Die Verteilung der Geburtsorte in den Lieferungen 62–66 zeigt hingegen ein wesentlich ausgeglicheneres Bild. Die Diskrepanz resultiert aus den ursprünglichen Aufnahmekriterien des ÖBL, die zeitlich und räumlich, einerseits nach kulturhistorischen Gesichtspunkten, andererseits jedoch staatsrechtlich definiert wurden (Obermayer-Marnach 1957). Bei der Aufnahme von Personen fokussierte man zwar auf den gesamten Raum des ehemaligen Habsburgerreiches, allerdings mit der Einschränkung, dass nach dem Ausgleich (1867) auf dem Gebiet der Länder der Ungarischen Krone geborene

Personen – von wenigen Ausnahmen abgesehen – nicht in das Lexikon aufgenommen wurden.

## 3. Anteil weiblicher Persönlichkeiten im ÖBL

Der Anteil weiblicher Biographierter hat sich über die mehr als siebenzig Jahre lange Geschichte des Lexikons kaum verändert (Abb. 4.).

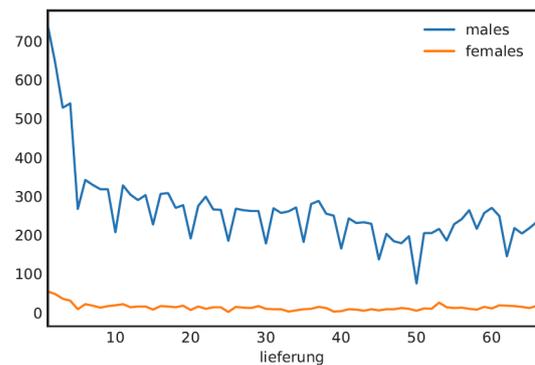


Abbildung 4: Männliche und weibliche Personen im ÖBL nach Lieferung

Unsere ursprüngliche Annahme, die ab den 1970er- und 1980er-Jahre sich entfaltende Frauen- und Geschlechtergeschichte hätte in diesem Zusammenhang eine positive Auswirkung auf die Anzahl der in das Lexikon aufgenommenen Frauen gehabt, konnte nicht bestätigt werden. Der geringe Anstieg ab der 50. Lieferung lässt sich auf den relativ hohen Frauenanteil in der Hauptberufsgruppe “Musik und darstellende Kunst” zurückführen (Abb. 5.).

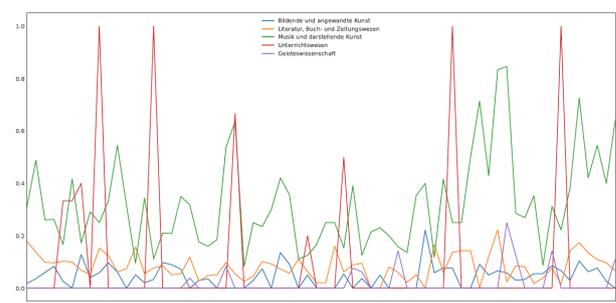


Abbildung 5: Verteilung weiblicher Personen im ÖBL nach Hauptberufsgruppen und Lieferung

Der relativ geringe Anteil an Frauen lässt sich durch den Berichtszeitraum des ÖBL, allen voran jedoch durch den Schwerpunkt 19. Jahrhundert erklären. Als Beispiel sei hier lediglich darauf verwiesen, dass in der Monarchie Frauen die Möglichkeit des Studiums großteils verwehrt blieb (Heindl, Tichy 1993). Anteilsmäßig findet man die

