Who Advertises in Newspapers?
Data Criticism in Mining Historical Job Ads

Klara Venglarova, Raven Adam, Wiltrud Mölzer, Saranya Balasubramanian,
Manfred Füllsack, Jörn Kleinert, Georg Vogeler
University of Graz

ABOUT

- **JobAds** Project (FWF P35783)
- Investigate evolution of **labor market**
- Extract and analyze job ads form **digitized newspapers**
- **Dataset:**
  - **ANNO Corpus** (Österreichische Nationalbibliothek, 2021)
  - Time span: **1850-1950**
  - 29 newspaper titles
  - Predominantly **German-language** job ads

HISTORICAL NEWSPAPERS AS A RESEARCH SOURCE

- **Several channels** to match job seekers and vacancies**:** Personal contact, asking around, centralized labor offices, commercial job agencies, and **job ads** in **newspapers**
  - Newspapers ads **30%** (Mölzer and Kleinert, 2024)
  - **Dominating** among **white-collars jobs**
  - **Blue-collar** jobs **underrepresented**
  - **Real-time data** on skills, vacancies, job market dynamics
- Newspapers and their **focus**
  - Political orientation, geographical area, temporal coverage, social focus, intended readers
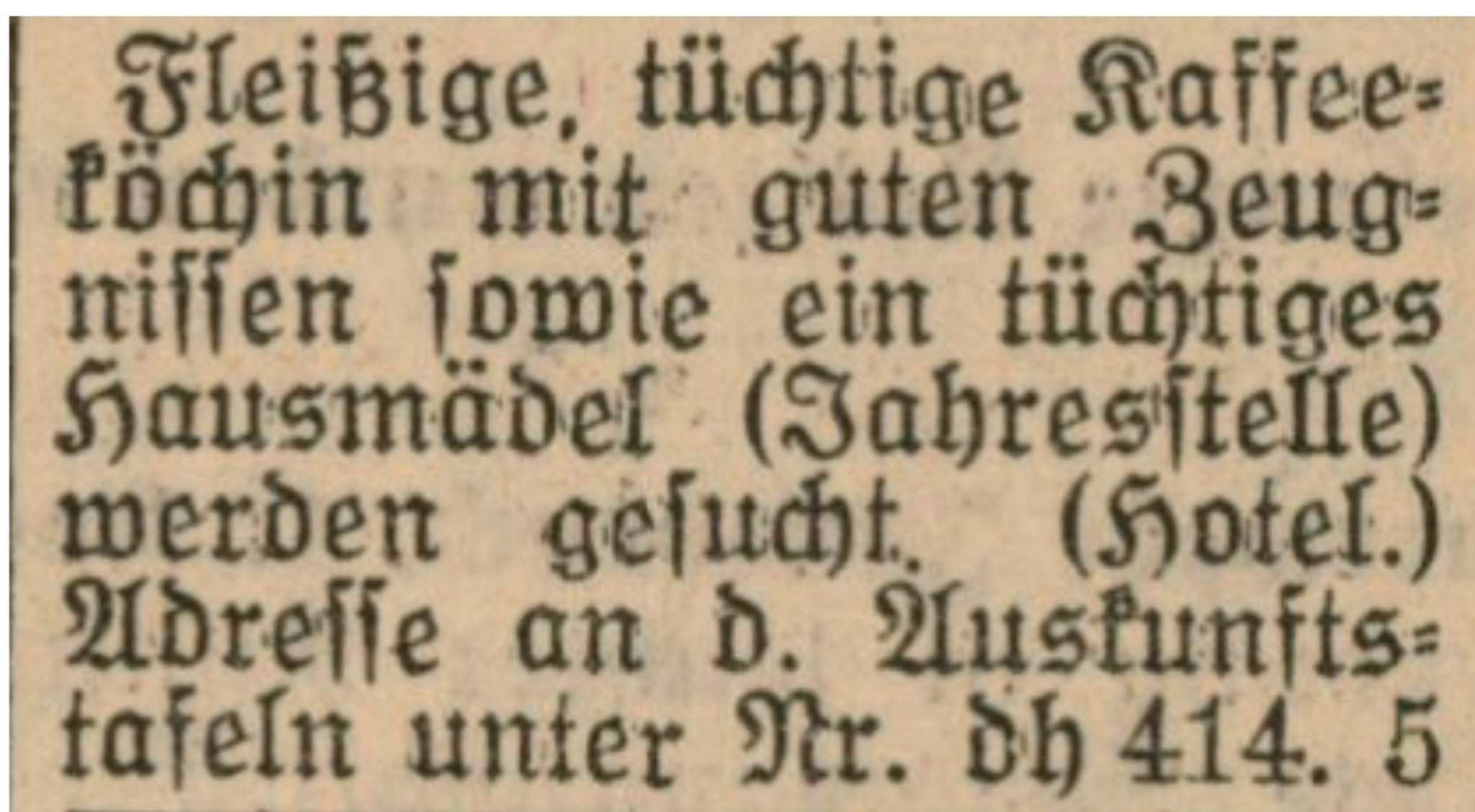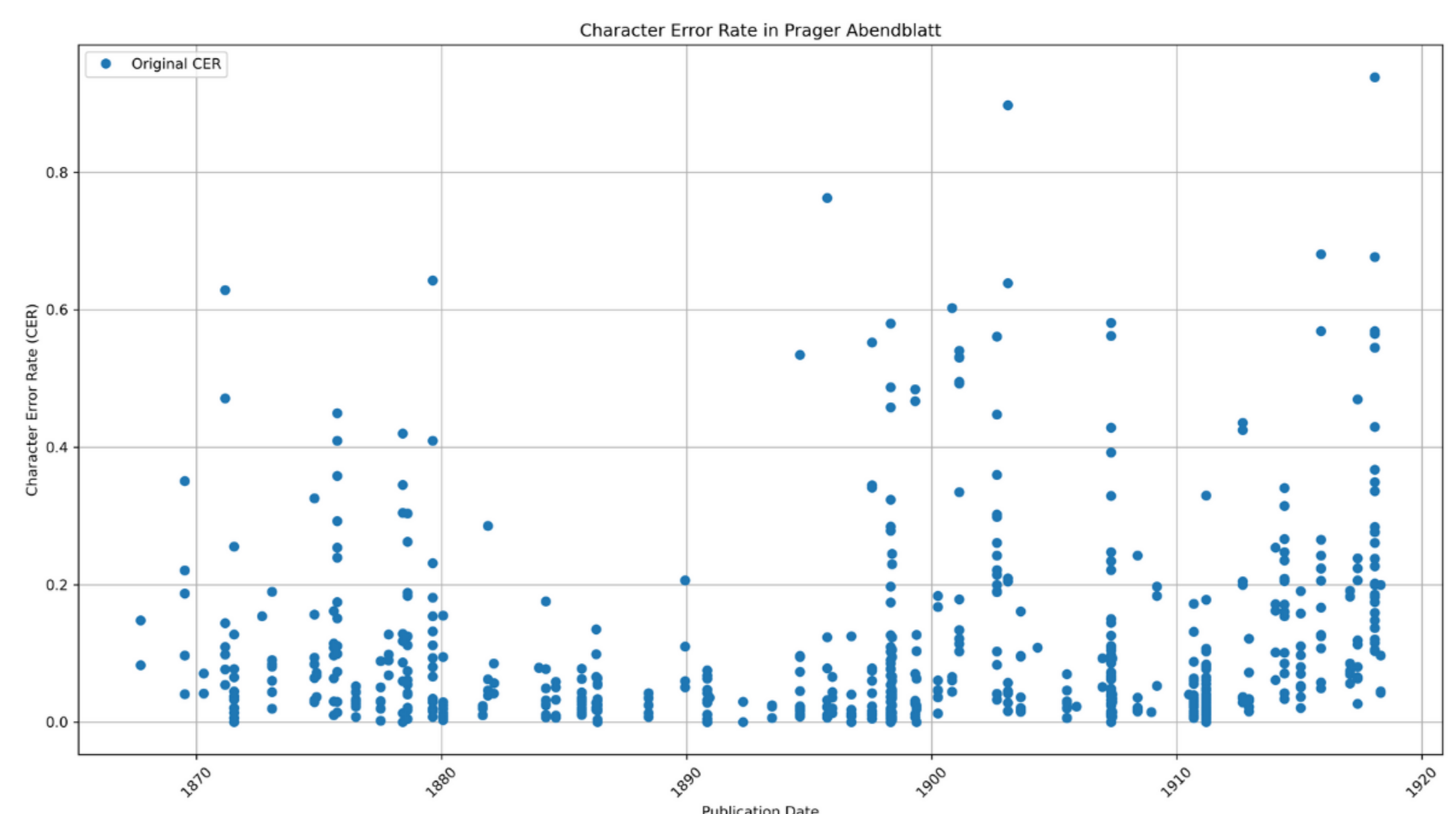

Image source: Innsbrucker Nachrichten, 19.5.1938, p. 10, https://anno.onb.ac.at/cgi-content/anno?aid=ibn&datum=19380519&seite=10

WHO ADVERTISES IN NEWSPAPERS?

- Where local announcements or asking around is not enough
- **Distance**: geographical, social, anonymity in urban society
- **Extrems**: high qualification, high demand
- Advertising is **costly**
- **Who reads** newspapers

CRITICISM IN DATA PROCESSING

- **Corpus creation** → **heterogenity**, avoid 'convenience sample' (Wijfjes 2017)
- **Preselection of relevant pages** → aiming for high **recall**
- **Segmentation process** → methodology for its **evaluation**
- **OCR** → **manually corrected** subset for evaluation
- **Post-correction** → **evaluation** using the gold standard

Example: Looking for Paperhanger

Absolute frequencies of positions found in the OCRed text between 1850-1900 and 1901-1950.

	1850-1900	1901-1950
Tapezierer (Paperhanger)	2	4
Stubenmädchen (Maid)	5	25
Verkäuferin (Shop Assistant f.)	7	14
Bäcker (Baker)	4	22
Vertreter (Agent/Representative)	6	18

Absolute frequencies of positions found in the manually corrected text between 1850-1900 and 1901-1950.

	1850-1900	1901-1950
Tapezierer (Paperhanger)	6	10
Stubenmädchen (Maid)	7	28
Verkäuferin (Shop Assistant f.)	8	21
Bäcker (Baker)	4	37
Vertreter (Agent/Representative)	10	21

Relative frequencies of positions found in the manually corrected text between 1850-1900 and 1901-1950.

	1850-1900	1901-1950
Number of job ads	1016	1763
Tapezierer (Paperhanger)	0.591%	0.567%
Stubenmädchen (Maid)	0.689%	1.588%
Verkäuferin (Shop Assistant f.)	0.787%	1.191%
Bäcker (Baker)	0.394%	2.099%
Vertreter (Agent/Representative)	0.984%	1.191%