hhu.

Advantages and Pitfalls of Machine Translation for Party Research: The Translation of Party Manifestos of European Parties Using *DeepL*

ECPR Summer School Political Parties, Aarhus University

Johanna I. Plenter, M.A.

Heinrich Heine University Düsseldorf

08.08.2022

Problem & Aims

- **Problem:** Parties typically publish their party manifestos only in the national language(s) → cross-national analyses are difficult, expensive and lengthy

- **Possible Solution:** Machine Translation (MT) of entire manifestos

© cristinn / Fotolia

- **Aims:**
  1. Discussion of both advantages and pitfalls of MT for party research
  2. Exemplary application to over 100 manifestos of European parties in order to analyse how salient the Gig Economy is in party competition
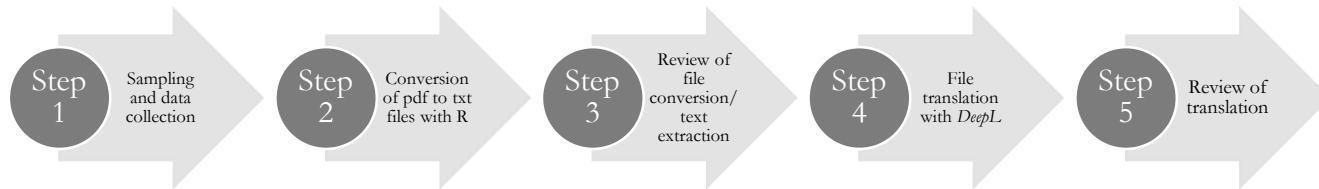
Why should we care?

© BIGSSS

→ MT algorithms are rapidly improving, enabling entirely new access to multilingual text data

→ Manifestos are the primary source for analyses of parties' policy positions or the (political) salience of an issue

→ Entire content of manifestos made accessible for analyses in a cost- and time-efficient way

→ Workers' strikes and legal procedures resulted in a high media salience which makes it reasonable for parties to develop a policy position

→ Gig work/the gig economy is representative of a range of new business models

© Paula Mitchell

Machine Translation with R and DeepL

- Practical guide for MT using R for data preparation and *DeepL* for translation
- *DeepL*: commercial translation AI currently offering 28 (mainly European) languages; file translation vs. API → depending on the amount of characters to be translated
- Workflow in 5 steps:

| Step 1 | Sampling and data collection | Step 2 | Conversion of pdf to txt files with R | Step 3 | Review of file conversion/ text extraction | Step 4 | File translation with *DeepL* | Step 5 | Review of translation |

© Johanna I. Plenter

- R packages: *rJava*, *tabulizer* (conversion of pdf to txt); if DeepL-API is used: *deeplr*

Data & Methods

- Sample: 137 parties, 123 election manifestos, 14 countries (Austria, Belgium, Denmark, Germany, Estonia, Finland, Ireland, Italy, Latvia, Poland, Spain, Sweden, Switzerland, UK)

- Analysis period: 2018-2021 → only most recent manifesto

- Party selection: every party with at least one seat in national parliament, entire ideological spectrum (depending on the election result)

- Quantitative text analysis: Dictionary Analysis & Targeted Sentiment Analysis

Concept	Dictionary entry/key
Gig economy/work	gig, crowdwork, clickwork, flexwork
Employment status	bogus, self-employment
Contract status	contractor, zero-hour, task-based, on-call, on-demand
Platform economy/work	platform

Salience of the Gig Economy in Party Manifestos

Keyword	# of matches	# of manifestos	Corpus coverage[1]	Party (Country)
Gig economy	10	7	5.69%	Labour (Ireland); Alliance, Cons, Green, LibDem, SDLP, SNP (UK)
Gig work	1	1	0.81%	Green (UK)
Platform economy	30	14	11.38%	CDV, Ecolo, Groen, NVA, spa (Belgium); SF (Denmark); PS, SDP (Finland); Greens, CDU/CSU, SPD (Germany); Pais, Podemos (Spain); GPS (Switzerland)
Platform work	11	5	4.07%	SPÖ (Austria); spa (Belgium); EK (Estonia); SDP (Finland); Left (Germany)
Crowdwork	5	4	3.25%	Greens, SPÖ (Austria); CDU/CSU, Greens (Germany)
Clickwork	1	1	0.81%	Left (Germany)
Contractor	9	8	6.50%	PVDA, spa (Belgium); Greens (Germany); FF, SF (Ireland); M5S (Italy); PiS (Poland); LibDem (UK)
Zero-hour	12	8	6.50%	spa (Belgium); SDP (Finland); Aontu, PBP (Ireland); Alliance, Labour, LibDem, SNP (UK)
Work on demand	1	1	0.81%	PdAS (Switzerland)
Work on call	4	3	2.44%	SPÖ (Austria); spa (Belgium); Left (Germany)

[1] = corpus coverage refers to the percentage of all manifestos containing the respective concept

Salience of the Gig Economy in Party Manifestos

- Gig economy an issue in party competition in 13 of the 14 countries (Latvia not represented) → corpus coverage: 27.64%
- Distinction between direct and indirect addresses; most important indirect phrase: "bogus self-employment" (19.51% coverage)
- "Platform" is the most used keyword → problem: many false positives
- Within the countries: several parties address the issue
- Keywords: country-specific usage → "gig economy"/"gig work" especially in UK
- Entire ideological party spectrum represented → extreme left (0.33 Workers' Party of Belgium) to extreme right (7.92 Belgian New Flemish Alliance); however, tendency towards left-wing parties → social democratic, socialist, ecological, left-libertarian
- Sentiment:
  - "gig" and "zero-hour" used in positive context
  - "bogus" used in negative context

Conclusion

1. Machine Translation in party research:
   - High-quality translations of (party) documents open entirely new avenues for textual analyses
   - Independent of databases or existing corpora
   - Cheap, fast, easy (data pre-processing can be problematic/lengthy depending on the file layout)

2. Gig economy as an economic niche issue within party competition:
   - ~28% explicit coverage can be regarded quite high for a very recent and niche phenomenon
   - ~20% indirect addresses through "bogus self-employment" shows that parties are aware of some features of the GE to be problematic
   - National contexts (i.e. workers' strikes or court rulings) seem to be the major predictor whether parties develop a policy position or not → focus on national patterns
   - Parties do not (yet) know how to address this issue: chance for labour markets vs. threat to workers' rights

Thank you for your attention!

✉@ Johanna.Plenter@hhu.de

🐦 @frau_jiw

R^G Johanna Plenter