Text-As-Data Analysis of Preferential Trade Agreements

Julia Seiermann (UNCTAD & IHEID)
Joint work with Wolfgang Alschner (University of Ottawa) and Dmitriy Skougarevskiy (IHEID & European University at St. Petersburg)

Geneva Trade and Development Workshop (GTDW), 23 May 2017
This paper...

• Introduces a new textual database of preferential trade agreements
• Presents first insights based on text-as-data analysis, on agreement scope and textual similarity between agreements
• Provides simple examples how text-as-data methods offer new possibilities to study trade agreements from different disciplinary perspectives:
  • Economics
  • Political science
  • Law
Motivation

- Preferential trade agreements (PTAs) are increasingly common since the 1990s
- There are currently (at least) 291 agreements in force, many of which are hundreds of pages long
- Number, length and complexity makes their content difficult to access
- Text-as-data is useful for the study of PTAs
  - Reduces the trade-off between detail and coverage
  - Generates new insights on PTA design and its determinants
  - Increases precision of estimates of the impact of PTAs on trade flows, welfare...
- Beyond PTAs: Text-as-data may become more and more important (for example, "Panama papers")
Existing PTA Databases

• Past decade: a push for a more fine-grained understanding of PTAs
  • Horn, Mavroidis and Sapir (2010)
    • 52 subject areas, "WTO-plus" versus "WTO-extra", legal enforceability, 28 PTAs
  • Extended by WTO (2011) (World Trade Report) and
  • Hofman, Osnago and Ruta (2017)
    • WB database "Content of Deep Trade Agreements" (279 PTAs)
  • Baccini, Dür and Elsig (2014)
    • WTI database "Design of Trade Agreements (DESTA)"
    • More than 100 content variables; over 600 PTAs

• In common: hand-coding of data
  • Labor intensive, time consuming; costly to add variables ex post

• The value added of text-as-data analysis
  • Reduces trade-off between details and coverage
  • Different methods allow to gain different insights than from traditional coding

• Used in other fields of the social sciences; most closely related work is by Alschner and Skougarevskiy on bilateral investment treaties (BITs) and Allee, Lugg, Elsig and Wuethrich on PTAs
The Dataset

• Corpus of full texts based on WTO regional trade agreements database
  • 447 treaties notified to the WTO, signed by 202 countries, 1948-2015
  • 413 in English, 33 in Spanish, 1 in French
  • 4 types (depending on which article in GATT/GATS they relate to): Customs unions, goods FTAs, services FTAs, partial scope agreements
  • 60% currently in force

• Plus: TPP agreement (not in force yet)

• Transformed all texts (excluding annexes and schedules) into a unified marked-up text format (XML)
Measure: Textual similarity

• Text-as-data: a lot of different methods (dictionary methods, supervised machine learning, unsupervised machine learning...)
• We use textual similarity (like plagiarism detection software)
• Textual similarity can aid legal analysis of PTAs:
  • Legal language is relatively uniform
  • In practice, legal drafters build on existing formulations even if they innovate
  • PTAs often replicate reference language, such as WTO agreements
  • Domestic political context privileges textual uniformity and continuity (e.g., Trade Promotion Authority bill, passed by US Congress in 2002)
Measure: Textual similarity

• Split each document into 5-character components
• Compute the overlap between each pair of treaties
• Jaccard similarity: \( s_{ij} = \frac{|A_i \cap A_j|}{|A_i \cup A_j|} \)
• \( A_i, A_j \): texts of treaties \( i \) and \( j \) (sets of 5-character components and)
• Note: In this analysis, we use only English-language treaties (414, or 87% of our database)
• We compute similarity at the treaty- and chapter-level (article-level to be added)
Chapter-level analysis

• Existing work using text-as-data analysis to study PTAs has used the full text of agreements (Allee, Lugg, Elsig and Wuethrich)

• We disaggregate to the chapter level (later: article level):
  • 71% of our English-language PTAs are structured in chapters
  • We structure the treaties into chapters, and use chapter headers to investigate the scope (subject matters)
  • Classify headers into 57 categories ("Investment", "Competition", etc.)

• This allows us to conduct more detailed analysis: for example, comparing the similarity of the investment chapters of different PTAs
Global level: Clusters, not country models

Heat map: Matrix of PTA similarity (English)
(yellow = low, red=high similarity; ordered by similarity)

3 large clusters

1. "Eastern European cluster"
   • (upper left)
   • Goods FTAs from the 1990s

2. "Global cluster"
   • (center)
   • Goods & services FTAs from the 2000s

3. "Central Asian cluster"
   • (lower right)
   • Goods FTAs from the 1990s
   • Less chapters and subject matters covered than 1)
National level: Customization and Standardization

Heat map: European PTAs

- The EU/EC adapts its PTAs according to
  - Political relationship
  - Type of economic cooperation envisaged
- Example: same time frame, but systematic difference between
  - Agreements with EU accession candidates
  - Neighborhood policy agreements with non-accession countries
Chapter level: Variation between subject areas

- Government procurement and competition chapters are often included in the same treaty ("Modern Global Cluster")
- Government procurement chapters follow a common design (due to existing WTO language(?))
- Competition chapters are very dissimilar
- In the same set of agreements, treaty practice can converge in one area and diverge in another
Ex. 1: Economics - Impact on trade flows (TPP-11)

- What is the impact of (deep) PTAs on trade flows?
- Heterogeneity of PTAs implies that the impact is not uniform
- Existing literature recognizes this; uses mostly unidimensional approaches, which do not capture "depth" well in all cases:
  - Baier, Bergstrand and Clance (2015): Agreement type (Stages of economic integration: FTA, customs union, common market, economic union)
    → NAFTA < Eurasian Economic Community
  - DESTA depth index: Agreement scope
  - Presence of 7 key features: reduction of tariffs to 0, provisions on services, intellectual property, investment, standards, government procurement, competition
    → Maastricht treaty < EU-CARIFORUM EPA
- Text-as-data methods can yield indicators that capture more than one dimension of the agreement
- For example, textual similarity is expected to captures both scope and type of agreement
Ex. 1: Economics - Impact on trade flows (TPP-11)

- Goal: Predict the impact of the TPP-11 (without US) on trade between members
- In a gravity framework (PPML, IM-year, EX-year and IM-EX fixed effects, s.e. clustered by IM and EX), we successively include different types of variables to capture trade policy
  1) PTA dummy
  2) Dummies for PTA type (as in Baier, Bergstrand and co-authors)
  3) DESTA depth indicator (0-7)
  4) Textual similarity to the TPP, in quintiles

\[ X_{ijt} = \beta \times TradePolicy_{ijt} + \gamma_{it} + \delta_{jt} + \zeta_{ij} + \varepsilon_{ijt} \]

- We use the coefficients from the regressions to predict the expected impact of the TPP-11 from the four different approaches
- Intuition: PTAs that resemble each other are expected to have a similar impact on trade

Text-as-data analysis of PTAs
Ex. 1: Economics - Impact on trade flows (TPP-11)

- Predicted impact of TPP on members' trade flows varies with approach used.
- Similarity quintiles yield a significantly higher prediction than unidimensional trade policy indicators.
- More disaggregate analysis is necessary to uncover the exact drivers of these differences.
- Text-as-data methods allow to construct more disaggregate variables (chapter-level, article-level similarity...)

**Predictions: Impact of TPP-11 on members' trade**
(incl. 95% confidence interval)
Ex. 2: Politics - Design diffusion (NAFTA)

- Diffusion of trade-related norms, policies and institutions is a core research area for international political economy scholars
- Here, treaty design is not the independent, but the dependent variable
- Example of NAFTA: most far-reaching trade agreement when signed in 1992
  - Covers wide range of issues (trade in goods and services, investment, labor...)
  - NAFTA model had a strong influence on treaty design
- Literature has suggested different drivers of diffusion of NAFTA design
  - US power and hegemony
    - US want to implant neoliberal rules abroad, especially in Latin America
  - Rational design
    - States choose NAFTA regional integration model because it fits their interest
    - Based on precise, enforceable rules (as opposed to EU institutions-based approach)
  - Bureaucratic inertia
    - Negotiators find it convenient to "specialize" in a single treaty model
Ex. 2: Politics - Design diffusion (NAFTA)

Diffusion in 4 waves
1. MEX, CAN with Latin-American partners
2. MEX, CAN with new partners; partners from 1) sign South-South FTAs with each other and new partners
3. PAN, CHL, USA export NAFTA design to Asia; return of USA
4. East Asian states among themselves, new treaties between North and South America, between America and South East Asia
   - NAFTA = dominant model in the Pacific Rim region
   - Culmination = TPP (58% overlap with NAFTA)
Ex. 2: Politics - Design diffusion (NAFTA)

• US power and coercion probably not the reason for NAFTA design diffusion

• Interest-based considerations unlikely in many cases
  • Investment, intellectual property rights chapters in NAFTA were motivated by asymmetric North-South considerations
  • Not optimal design for South-South economic integration

• Path dependency, emulation, etc. are more likely to be drivers of NAFTA design diffusion
Ex. 3: Law - Convergence (BITs and PTA investment chapters)

• Important debates among legal scholars concern convergence and divergence of legal norms
• PTAs, by definition, pose a risk of fragmenting trade law, as they diverge from the multilateral WTO law
• But they may also be a source of convergence between legal fields through connecting trade law with its neighboring fields (investment, competition, environment, labor...)
• Example: Investment regulations in bilateral investment treaties (BITs) and the investment chapters of PTAs
• Investment and trade are increasingly intertwined through global value chains
• Are trade and investment rules converging over time or growing apart?
Ex. 3: Law - Convergence (BITs and PTA investment chapters)

Heat maps: BITs and PTA investment chapters (ordered chronologically)

USA: Before 2003, only BITs, not PTAs with investment chapters except NAFTA

Post-2003: PTAs and BITs look alike, modelled after NAFTA

AUS: BITs and PTAs don’t resemble each other

BIT network = consistent, PTA network = more inconsistent

Power asymmetries: AUS may be rule-maker in BITs with developing countries, rule-taker in PTAs with developed countries
Conclusion

• Highlights from the PTA database (textual similarity indicator)
  • PTA clusters go beyond country models
  • EU PTAs customized to the political and economic relationship with partner countries, but standardized within groups
  • In the same set of treaties, texts can converge in one area and diverge in another

• Text-as-data represents a valuable complement to build on existing research on PTAs, for example:
  • Economics: obtain more fine-grained estimates and predictions of treaty impact
  • Political science: explain the drivers of PTA design diffusion
  • Law: understand where international law converges and diverges

• Future projects
  • Expand database (chapter-level similarity; eventually more agreements)
  • Deepen research on questions presented here and others
*** THANK YOU FOR YOUR ATTENTION ***
Most similar agreements

<table>
<thead>
<tr>
<th>RTA 1</th>
<th>In force</th>
<th>RTA 2</th>
<th>In force</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC - Algeria</td>
<td>1976</td>
<td>EU - Algeria</td>
<td>2005</td>
<td>.99333191</td>
</tr>
<tr>
<td>Slovak Republic - Israel</td>
<td>1997</td>
<td>Czech Republic - Israel</td>
<td>1997</td>
<td>.99290115</td>
</tr>
<tr>
<td>EC - Czech Republic Europe Agreement</td>
<td>1995</td>
<td>EC - Slovak Republic Europe Agreement</td>
<td>1995</td>
<td>.98343811</td>
</tr>
</tbody>
</table>
Similarity over time

- Mean similarity relatively stable over time
Data

• Textual similarity
• Trade flows: UN COMTRADE
• Production (for "internal trade"): UNIDO, FAO, WB, CEPII
• Alternative trade policy indicators: Baier/Bergstrand database, Design of Trade Agreements (DESTA) database
• Dataset: 90 countries, 1994-2012
## Results

<table>
<thead>
<tr>
<th>Trade</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTA</td>
<td>0.135*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-way PTA</td>
<td>0.0672</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-way PTA</td>
<td>0.151</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTA</td>
<td>0.181*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customs Union</td>
<td>0.820***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Market</td>
<td>0.804***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic Union</td>
<td>0.998***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTA not in BB data</td>
<td>0.0448</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth indicator</td>
<td></td>
<td>0.0337*</td>
<td>(0.0131)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarity to TPP</td>
<td></td>
<td>0.631***</td>
<td>(0.142)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTA not in our data</td>
<td></td>
<td>0.117**</td>
<td>(0.0442)</td>
<td>0.0775</td>
<td>(0.0496)</td>
</tr>
<tr>
<td>Similarity Quint1</td>
<td></td>
<td></td>
<td></td>
<td>-0.302**</td>
<td>(0.113)</td>
</tr>
<tr>
<td>Similarity Quint2</td>
<td></td>
<td></td>
<td></td>
<td>0.0680</td>
<td>(0.108)</td>
</tr>
<tr>
<td>Similarity Quint3</td>
<td></td>
<td></td>
<td></td>
<td>0.0513</td>
<td>(0.0765)</td>
</tr>
<tr>
<td>Similarity Quint4</td>
<td></td>
<td></td>
<td></td>
<td>0.0967</td>
<td>(0.0580)</td>
</tr>
<tr>
<td>Similarity Quint5</td>
<td></td>
<td></td>
<td></td>
<td>0.330***</td>
<td>(0.0676)</td>
</tr>
<tr>
<td>N</td>
<td>55930</td>
<td>55930</td>
<td>55930</td>
<td>55930</td>
<td>55930</td>
</tr>
</tbody>
</table>