

# The Language of Institutional Design: Text Similarity in Preferential Trade Agreements

Soo Yeon Kim\*

National University of Singapore

sooyeon.kim@nus.edu.sg

September 2015

## Abstract

This paper analyzes the degree of text similarity across preferential trade agreements (PTAs). The analytical framework takes the texts as templates for trade liberalization, and investigates the degree to which the text content is replicated from one agreement to the next. As PTAs continue to rise in their numbers, an interesting question to raise is how they reflect the development of different templates of trade liberalization and whether they are subsequently adopted in other agreements. The analysis compares pairs constituted from 416 PTA texts to generate similarity values that captures the degree of text commonality. Variation in similarity measures is examined for their longitudinal and regional patterns and differences across regional and trans-regional agreements. This paper finds that the extent of text commonalities in PTAs is actually very low. A comparison of common word sequences of 4 or more words across a pair of PTAs averages less than 4 percent, with a median value of less than 1 percent. This is somewhat contrary to expectations, as the rise in the numbers of PTAs, and especially the signing of multiple agreements by the same country, would suggest that countries are likely to employ much of the same text content across these agreements.

---

\*The author thanks Guanfeng Wang and especially Matin Ling for excellent research assistance.

# 1 Introduction

”There is a tendency to replicate trade-opening rules in PTAs because template approaches are often used for PTAs.” (World Trade Report 2011, 171)

Preferential trade agreements (PTAs) continue to grow in number and influence as rule-making institutions for trade. According to the World Trade Organization (WTO), 7 April 2015, some 612 PTAs have been notified to the WTO, of which 406 are in force.<sup>1</sup> PTAs liberalize trade between agreements partners through preferential market access for members, and they promote economic regionalism as states cooperate in institution-building to coordinate trade policies ((Mansfield and Milner, 1999, 591); Fishlow and Haggard (1992)).

The existing scholarship on the politics of PTAs has focused on three major questions. First, the long-standing Vinerian (Viner (1950)) debate on the trade-creation versus trade-diversion effects of PTAs on trade was perhaps the first question to animate a substantial but largely inconclusive body of literature. The second question has focused on the domestic and international political factors affecting the formation and expansion of PTAs. Third, the most recent scholarship has examined PTAs from the perspective of institutional design, shifting the analytical focus from whether states commit to trade liberalization to how they do so. This literature provides insights on the political economy of design, the effects of particular design features, and the politics of implementation of PTAs.

This paper focuses on PTA texts themselves, as a progression of scholarship that has engaged in the mapping or coding of PTA provisions. The analytical framework is premised in on the assumption that texts are templates, and for PTAs, their texts represent templates for trade liberalization. As PTAs continue to rise in their numbers, an interesting question to raise is the extent to which they reflect the development of different templates of trade liberalization and to what extent they are subsequently adopted in other agreements.

The variation of interest is the degree to which the text content of PTAs is replicated—’copied and pasted’—from one agreement to the next. The analysis compares pairs constituted from 416 PTA texts to generate similarity values that captures the degree of

---

<sup>1</sup>The WTO refers to reciprocal trade agreements, but the terms PTA and reciprocal trade agreement are treated as equivalent in this paper. The former is the widely used, generic term to refer to trade agreements of all types. The latter is WTO nomenclature, and it refers specifically to agreements in which partners agree to mutually liberalize trade through the exchange of concessions. The WTO also refers to preferential trade agreements, but these refer to agreements in which only one agreement partner provides concessions, such as in the Generalized System of Preferences (GSP) offered by individual WTO members that grant preferential access to certain trade partners such as least-developed countries. The figures from the WTO count notifications for goods, services, and accessions separately. When these are considered as part of the same agreement, the WTO reports 449 RTAs of which 262 are currently in force.

text commonality. Variation in similarity measures is examined for their longitudinal and regional patterns and differences across regional and trans-regional agreements.

This paper finds that the extent of text commonalities in PTAs is actually very low. A comparison of common word sequences of 4 or more words across a pair of PTAs averages less than 4 percent, with a median value of less than 1 percent. This is somewhat contrary to expectations, as the rise in the numbers of PTAs, and especially the signing of multiple agreements by the same country, would suggest that countries are likely to employ much of the same text content across these agreements.

## 2 The Language of Institutional Design

This paper builds on scholarship that has examined the evolution of the international trade system through the observation of trade agreements that co-exist with the multi-lateral trade regime. In doing so, scholarship has evolved in addressing questions concerning trade-creation versus trade-diversion, the formation and expansion of PTAs, and the causes and consequences of institutional design.

The literature most immediately relevant to this study concerns the institutional design of PTAs. These studies have focused on *how* states make commitments as observed in specific provisions of trade agreements. Institutional design refers to features such as membership conditions, the scope of issue areas covered by legal commitments, the centralization of institutional activities, enforcement and flexibility mechanisms, and voting rules (Koremenos et al. (2001)). In examining such institutional features, studies have also analyzed the extent to which specific PTA provisions go beyond current levels of obligation under the World Trade Organization (WTO).

A number of 'mapping' projects have provided classifications of PTA provisions to illustrate and investigate the sources and consequences of variation in institutional design. The Design of Trade Agreements (DESTA) project (Dür et al. (2014)) is perhaps the the largest mapping project in the current scholarship. Covering 591 PTAs, the project classifies provisions in 10 issue areas that have produced about 100 data points per agreement: market access in industrial goods, services, investments, intellectual property rights, competition, public procurement, standards, trade remedies, non-trade issues, and dispute settlement. Estevadeordal, Suominen, and Teh's (Estevadeordal et al. (2009)) study is a more specialized mapping project that focuses on a sample of 74 PTAs chosen for the diversity of agreement partner characteristics such as economic development, trade, and geography. This project covers in great detail six issue areas, including the traditional areas of market access and trade remedies, and relatively newer areas such as technical barriers to trade, services, investment, and competition. Chauffour and Maur (Chauffour et al. (2011)) focus on provisions that are particularly relevant and challenging to developing countries: trade facilitation, labor mobility (GATS Mode 4), government

procurement, intellectual property rights, environment, labor rights, and human rights.

A common theme that runs through these mapping projects is the distinction between 'shallow' and 'deep PTAs,' which rely on the quality of commitments made especially in regulatory areas related to trade. Cited as an important new development in the PTA design (WTO (2011)), deep PTAs have strong commitments toward 'deep integration' that involve the adoption of domestic trade-related regulations that are WTO-consistent. Deep integration has three main properties: (a) liberalization of behind the border trade rules; (b) protection of foreign firms' interests; and (c) harmonization of domestic regulatory systems for managing international production and trade (Kim (2015)).

Other studies have delved into specific institutional design features such as enforcement through a dispute settlement mechanism and flexibility provisions. McCall Smith (Smith (2000)) finds that more legalized dispute settlement mechanisms in PTAs involving large economies, inequality between partners, and high levels of economic integration. Flexibility mechanisms include trade remedies and other provisions that allow a time-delimited suspension of trade liberalization commitments. In a study of the political economy of flexibility provisions, Kucik (Kucik (2012)) finds that import-competing industries benefit from flexibility provisions in PTAs while export-dependent industries bear the costs.

## 2.1 Texts as Templates

This paper contributes to scholarship on institutional design by directing attention to the texts of the PTAs, focusing on the role of texts as templates. The texts of PTAs represent templates for trade liberalization and the adoption of particular text content from existing PTAs reflects the acceptance and support of these templates. Countries utilize the text materials provided in existing PTAs to indicate their acceptance of and intention to carry out certain trade liberalization commitments. PTAs as templates for trade liberalization is most evident in the 'deep integration' PTAs that have become increasingly visible in the global network of trade agreements. They not only widen the scope of issues covered under the PTA but also establish trade rules that may go beyond the current levels of obligation (WTO-plus) or are not currently covered under any agreement under the WTO (WTO-x).

The considerable variation in PTA texts is also markedly different from the template-based approach of bilateral investment treaties (BITs). It is widely recognized that countries rely on templates in negotiating and signing BITs. The same cannot be said, however, for PTAs, which appear to exhibit wide variation in their scope, depth, and other features of institutional design.

In PTAs, one source of variation can be found in the degree of WTO-plus or WTO-x provisions of the PTA text. For example, What is 'legally enforceable' ((Horn et al.,

2010)) or not cannot be sufficiently captured by the inclusion of particular issue area but rather in the language itself of the text that will indicate the degree to which the legal obligations stipulated for, for example, gender equality, is symbolic or legally binding. Moreover, even commitments in areas such as competition policy, there exists variation across agreements signed by the United States and by the European Union. US agreements are considered to be much more stringent in competition commitments than EU agreements, and this variation can be measured through the analysis of PTA texts.

Text analysis of PTAs is complementary and corroborative of the various mapping projects that have been reported in the existing scholarship. Text analysis is complementary in that it goes beyond identifying the scope of commitments by the issue areas that are included to provide a measure of the degree of legal obligation that is evidenced in the text of the agreement. Indeed, text analysis is geared toward capturing the underlying latent dimension of the strength of trade liberalization commitments. It is also corroborative in that it enhances and strengthens any measure of the quality of a PTA by providing detailed and nuanced text evidence of the level of legal obligation, and it can also be considered as confirmatory evidence to boost the results of manual coding.

## 2.2 Examples: Government Procurement Provisions

The World Trade Organization's *World Trade Report 2011*, which is devoted to the role of PTAs in the world trade system, suggests strongly that "There is a tendency to replicate trade-opening rules in PTAs because template approaches are often used for PTAs" (WTO (2011), 171). For example, NAFTA's telecommunications provision has been adopted by a large number of countries in their PTAs, to the point that this provision is increasingly becoming a norm (Baldwin et al. 2009). Baldwin et al. argue that this replication of templates is equivalent to regulatory harmonization, in which states apply common rules to firms irrespective of national origin. As such, replication of templates is not preferential and may well be effective in promoting competition and trade.

Replication of text can also be found in PTA provisions for liberalization of government procurement.<sup>2</sup> For example, the Article 27 of Turkey-Albania FTA (2006) and Article 28 of the Serbia-Turkey FTA (2009) on Public Procurement contain exactly the same text with the only difference being the insertion of 'of this Article' (See box text).

---

<sup>2</sup>The choice of government procurement provisions is somewhat arbitrary and accidental. These commonalities in text were discovered while coding competition-related provisions in PTAs. This section provides the results of further investigation.

Turkey-Albania FTA (2006) Article 27 and Serbia-Turkey FTA (2009) Article 28  
Public Procurement

1. The Contracting Parties consider the liberalization of their respective public procurement markets as an objective of this Agreement. The parties aim at opening up of the award of public contracts on the basis of non-discrimination and reciprocity.
2. The parties will progressively develop their respective rules, conditions and practices on public procurement with a view to granting suppliers of the other Party access to contract award procedures on their respective public procurement markets not less favourable than that accorded to companies of any country or territory.
3. The Joint Committee shall examine developments related to the achievement of the objectives of this Article and may recommend practical modalities of implementing the provisions of paragraph 2 of this Article so as to ensure free access, transparency and mutual opening of their respective public procurement markets.
4. During the examination referred to in this paragraph 3 of this Article, the Joint Committee may consider, especially in the light of international developments and regulations in this area, the possibility of extending the coverage and/or degree of the market opening provided for in paragraph 1 (of this Article).<sup>3</sup>
5. The parties shall endeavor to accede to the relevant Agreements negotiated under the auspices of the GATT 1994 and the Marrakesh Agreement, establishing the WTO.

The above agreements have Turkey as a common signatory in the two PTAs, which suggests that Turkey's public procurement template was employed for these two agreements. Moreover, even trade agreements concluded by different country pairs can contain virtually the same text with minimal difference. This can be seen, for example, in Article 29 of the Ukraine-FYROM FTA (2001) and Article 23 of the Albania-Moldova FTA (2003), both entitled Public Procurement (See box text). Signed two years apart, these two texts were signed by two different pairs of countries but show remarkable similarity in both the language and substance of their public procurement commitments.

The public procurement provisions in these two PTAs is remarkably similar in content, with minimal differences in the actual text and virtually no differences in the substantive content. The text box comparing the two provisions shows that the differences are not substantive but rather grammatical, with the later Albania-Moldova FTA making only small and cosmetic changes to the earlier Ukraine-FYROM FTA. In substance, the public procurement provisions in the two PTAs are identical. They include non-discrimination and reciprocity as the basis for awarding of public procurement. In addition, both agreements also commit to "free access, transparency, and full balance of rights and obligations" in implementation.

Ukraine-FYROM FTA (2001) Article 29 and Albania-Moldova FTA (2003) Article 23  
Public Procurement<sup>4</sup>

1. The Contracting Parties consider liberalization of their public procurement markets (as) an objective of this Agreement. The parties shall seek to (aim at) open(ing) up (of the award)ing of public contracts on the basis of non-discrimination and reciprocity.
2. The Contracting Parties shall(will) progressively develop their respective rules and practices of(on) public procurement and shall grant suppliers of the other contracting Party access to contract award procedures on their respective public procurement Markets(, which will) not (be) less favorable than that accorded to companies of any third country.
3. The Joint Committee shall review a list of tasks specified in (examine developments related to the achievement of the objectives of) this Article and may offer(recommend) practical recommendations concerning (modalities of) implementation of (implementing the) provisions in (of) Paragraph 2 of this Article (so as) to ensure free access, transparency and full balance of rights and obligations. During the examination of this situation (referred to this paragraph from this article), the Joint Committee may consider, especially in the light of international regulations in this area, the possibility of extending the coverage and/or degree of openness of the market provided for in paragraph 2 of this Article.
5. The parties shall endeavor to accede to the relevant Agreements negotiated under the auspices of the GATT 1994 and the (Marrakesh) Agreement(,) establishing the WTO.

Another pattern of replication in PTA texts is the replacing of partner country names, as appears to be the practice in two FTAs concluded by the European Free Trade Association (EFTA) with the Slovak Republic and Bulgaria (See box text). Comparison of paragraph 2 of Article 16, the same in both the EFTA-Slovak Republic FTA (1992) and the EFTA-Bulgaria FTA (1993) shows that in the later agreement, trade agreement partner the Slovak Republic is replaced with Bulgaria, but otherwise the text is exactly the same.

EFTA-Slovak Republic FTA (1992)

2. As of the entry into force of this Agreement, the EFTA States shall grant companies from **the Slovak Republic** access to contract award procedures on their respective procurement markets according to the Agreement on Government Procurement of 12 April 1979, as amended by a Protocol of Amendments of 2 February 1987 negotiated under the auspices of the General Agreement on Tariffs and Trade. **The Slovak Republic** shall, taking into account the restructuring and development process of its economy, gradually ensure that companies from the EFTA states have access on the same principles to contract award procedures on its public procurement market.

EFTA-Bulgaria FTA (1993)

2. As of the entry into force of this Agreement, the EFTA States shall grant companies from **Bulgaria** access to contract award procedures on their respective procurement markets according to the Agreement on Government Procurement of 12 April 1979, as amended by a Protocol of Amendments of 2 February 1987 negotiated under the auspices of the General Agreement on Tariffs and Trade. **Bulgaria** shall, taking into account the restructuring and development process of its economy, gradually ensure that companies from the EFTA states have access on the same principles to contract award procedures on its public procurement market.

### 3 Analyzing PTA Texts

What is the extent of text replication in PTAs? The examples above of public procurement provisions suggest that countries do adopt text from existing PTAs. However, public procurement provisions comprise only a small part of a trade agreement, and the PTA is likely to include numerous other provisions covering a wide range of trade rules. This paper extends the comparison of PTA texts to the entire document, including the main documents and the accompanying appendices and additional protocols.

The goal of this paper is to identify the extent of and patterns of text replication in PTAs. These features of text overlap provide informative and important insights into the diffusion of institutional design features insofar as they are embedded in the texts of PTAs. The analysis proceeds within the framework of descriptive inference (King and Verba (1994); Brady and Collier (2004)), in which descriptive insights speak to questions concerning the choice of templates in institutional design.

The analysis addresses two main questions: first, what is the degree of commonality across PTA texts? This paper develops a measure of text similarity that is based on common  $n$ -word groups found in a comparison of a given pair of agreement texts. The analysis also examines patterns across time and space. It investigates text commonalities between agreements in different years, between agreements from the same region and

trans-national agreements that involve signatories from different regions, and regional variation that shows whether agreements signed by countries in particular regions have more (or less) text commonality with other PTAs.

The second question of this paper concerns the content of the text commonalities. That is, what kind of content do agreements have in common? As this study is premised on the view that the texts of PTAs serve as templates for trade liberalization, the frequency with which certain text content suggests the adoption of the model that that text represents. In addressing this question, the analysis employs existing tools of text analysis to identify key words concerning trade liberalization that occur most frequently in the common content of PTA texts.

This paper analyzes the texts of 317 PTAs. Each agreement text is compared with every other agreement in the sample.<sup>5</sup> The unit of analysis is a pair of PTA texts, and the sample of analysis includes trade agreements inclusive of the years 1960 - 2013. The sample of analysis includes only English language texts, and excludes PTA-pairs in which the first agreement is signed earlier than the second, as only texts in later agreements can replicate materials from previous PTAs. The text analysis also utilizes all the documents that comprise the trade agreement. This includes not only the main document but also the annexes that specify reservations and exceptions or provide supplementary materials. These annexes often comprise the additional protocols (Moravcsik (2000)) that are important sources of variation states' commitments and adherence to international treaties. This paper thus includes these supplementary materials to examine the extent to which replication of PTA texts and templates also apply to them.

The sample of analysis, and the parts of the agreement used for the analysis differ from the study by Allee and Elsig (Allee and Elsig (2015)). This paper analyzes trade agreements of varying sizes, comparing agreements signed during the years 1960-2013 and including the entire corpus of text available for the agreement. These differences in the design of study suggest important avenues of investigation into the sources of variation in PTA commitments.

### **3.1 Text Similarity in Preferential Trade Agreements**

Text similarity captures the degree to which a given pair PTA texts shares common content. In this paper, the degree of text similarity is the number of words in  $n$ -word sequences that two PTAs have in common, expressed as a proportion of the total number of words in each agreement. For every pair of agreements paired, there are two similarity measures, which capture the extent to which agreement A takes language from B and vice versa.

This paper employs routines from the Natural Language Toolkit (NLTK, Bird et al.

---

<sup>5</sup>The list of PTAs is provided in the Appendix.

(2009)) to construct measures of the two text similarity measures: word similarity and semantic similarity.

The implementation of this measure of text similarity proceeded in two steps. In the first step, a given pair PTA texts was compared to identify the groups of sequential words that are common to both agreements. The minimum number of words in the group of sequential words set at 4 at the start. This procedure thus identified how many groups of 4 or more sequential words the two agreements have in common. The minimum number of words to compare was then increased one word at a time until the minimum number of sequential words for comparison reached 20. This procedure yielded 17 common word groups from the comparison. These are labeled N4 to N20 and used to generate distributions as shown in 1.

The second step involved calculating the actual measure of text similarity: the number of common single words found in the common word sequences as a proportion of the total number of words in each of the agreements being compared. For the case in which a minimum of 4 consecutive words is required to be identified as common text, construction of the measure involves identifying the number of single words that are in these common word groups and expressing them as a proportion of the total number of words in each agreement in the paired PTA. similarity measures were calculated for the 17 common word groups.

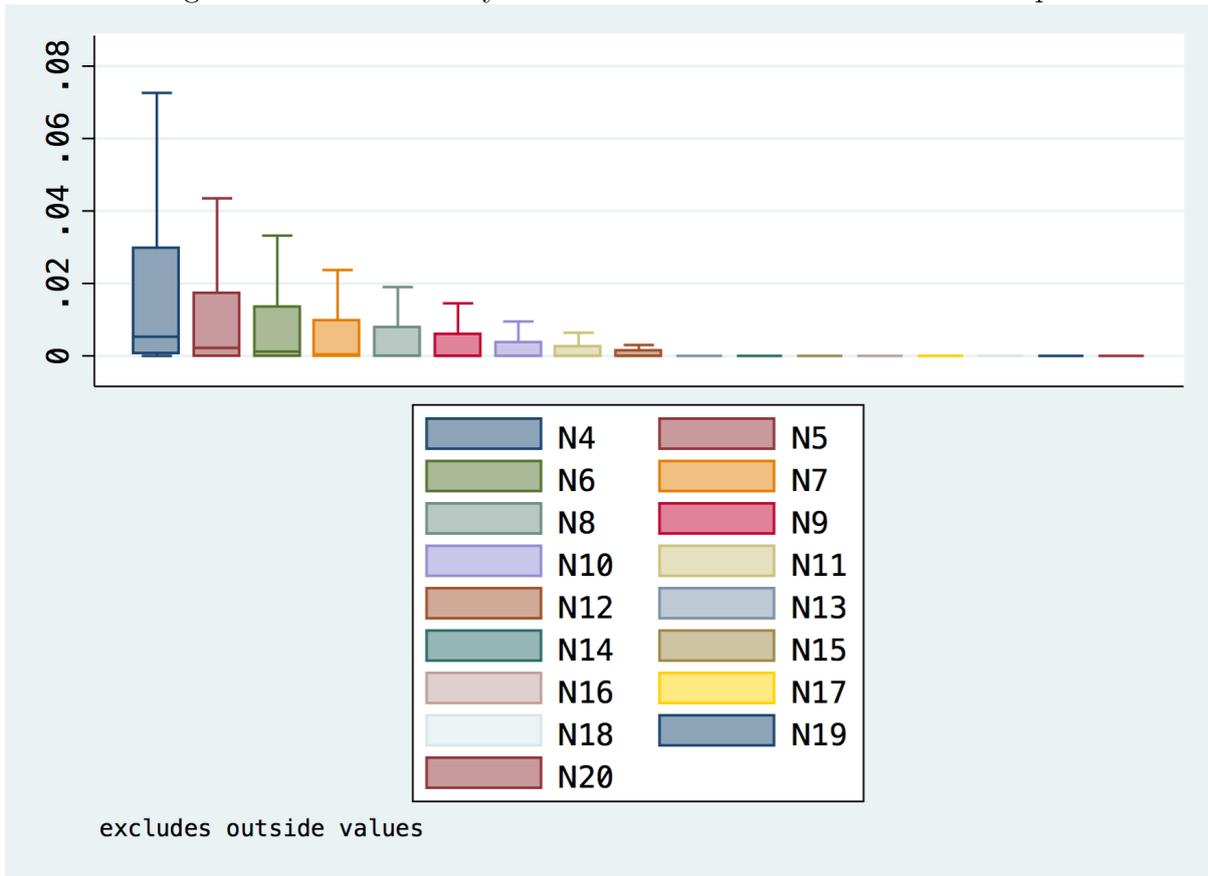
## 4 How Much do PTA Texts Have in Common?

A first analysis of the similarity data shows that PTA texts do not actually have that much in common. That is, countries do not appear to be adopting the text content of other agreements to any significant extent. Figure 1 shows boxplots that track the distribution of text similarities given the minimum number of words in a common word group across a pair of PTAs. Thus N4, which is the first boxplot, shows the distribution for word groups of 4 or more common words. N5 shows the distribution for word groups of 5 or more, and so on, to N20, which shows the distribution of word groups of 20 or more. Table 1 of the Appendix provides the corresponding descriptive statistics.<sup>6</sup>

---

<sup>6</sup>The boxplots exclude outside values, or outliers, using the *nooutsides* option in Stata. Outside values are those that skew the y-axis range of the box plot, defined conventionally as those lying outside 1.5 times the interquartile range of a variable, in this case, the similarity values.

Figure 1: Text Similarity across Minimum Common Word Groups



The distributions in Figure 1 show that overall, a given pair of PTA texts has very little common text content. Text similarity across agreements is very low, with the median for the 4-word minimum common word group is approximately 0.25% of the total number of words. The mean for this group is approximately 3.2%, which also indicates that most of the values are also concentrated at very low values. As expected, as the minimum number of words in a common word group is increased, the values become lower. The median similarity measure reaches zero when the minimum number of words is set to 8 or higher.

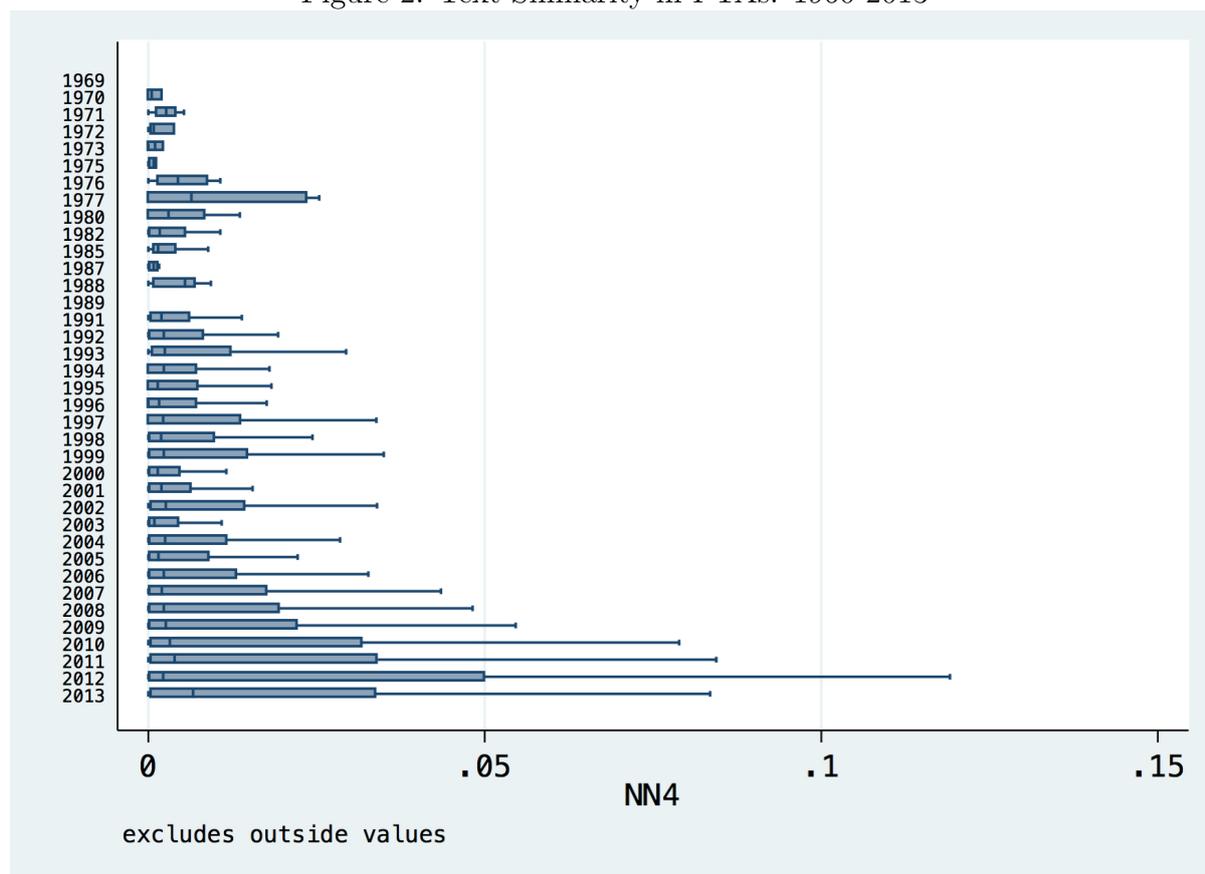
#### 4.1 Patterns of Variation

This section examines patterns of variation in the text similarity of PTAs. Specifically, I examine whether there are significant differences in text similarity across time, across regional and trans-regional agreements, and between agreements concluded by countries in particular regions.

#### 4.1.1 Longitudinal Variation in the similarity of PTA Texts

Figure 2 shows stacked boxplots for the years in which PTAs were signed.<sup>7</sup> The data include PTAs signed from 1960 to 2013. In tracking the longitudinal variation in text similarity of PTAs, the expectation is that the later PTAs may have higher levels of similarity with PTAs signed in previous years, for the simple reason that PTAs will emulate existing agreements signed in previous time periods.

Figure 2: Text Similarity in PTAs: 1960-2013



The longitudinal data show, however, that in terms of median values, there is not strong pattern of longitudinal variation in text similarity across PTAs. The values are overall low across the years, indicating that the low levels of text similarity are a consistent pattern rather than driven by time to any significant extent. The variation that does occur time is in the range of values, which shows that in more recent years, there are more PTAs that have text in common with other agreements. The data do not indicate, however, indicate whether these commonalities are with past or contemporary PTAs.

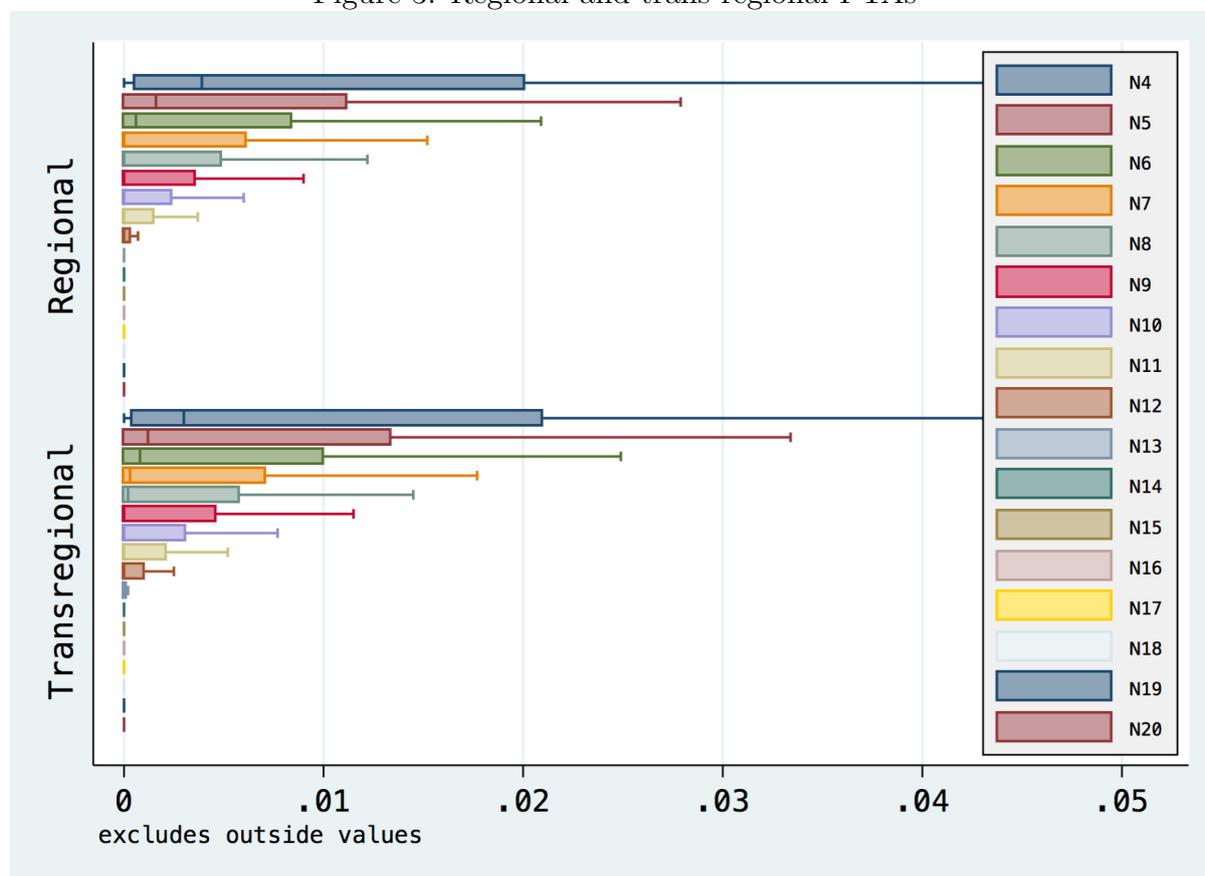
<sup>7</sup>These plots are for 4-word minimum common word groups.

### 4.1.2 A Comparison of Regional and Trans-regional PTAs

The analysis also distinguishes between regional and trans-regional PTAs. Regional PTAs are agreements between countries in the same region, while trans-regional PTAs are those signed by countries of different regions. Figure 3 shows the distribution of text similarity values for the range of minimum common word groups for regional and trans-regional PTAs. The expectation is that trans-regional agreements are more likely to refer to more international 'templates' as they are signed by countries that do not share regional characteristics. They are therefore likely to exhibit higher levels of similarity with other PTAs. Regional PTAs, on the other hand, are more likely to share text commonalities with other PTAs in the region, but such text similarity values are on a smaller scale than trans-regional agreements.

Figure 3 shows the distribution of text similarity values for the range of minimum common word groups for regional and trans-regional PTAs. The modal values of the text similarity values are not significantly different between the two types of PTAs. However, trans-regional PTAs exhibit a wider interquartile range in the similarity measures, and also include much higher values than those of regional PTAs, which suggest that there is some interesting variation in the similarity values of these agreements.

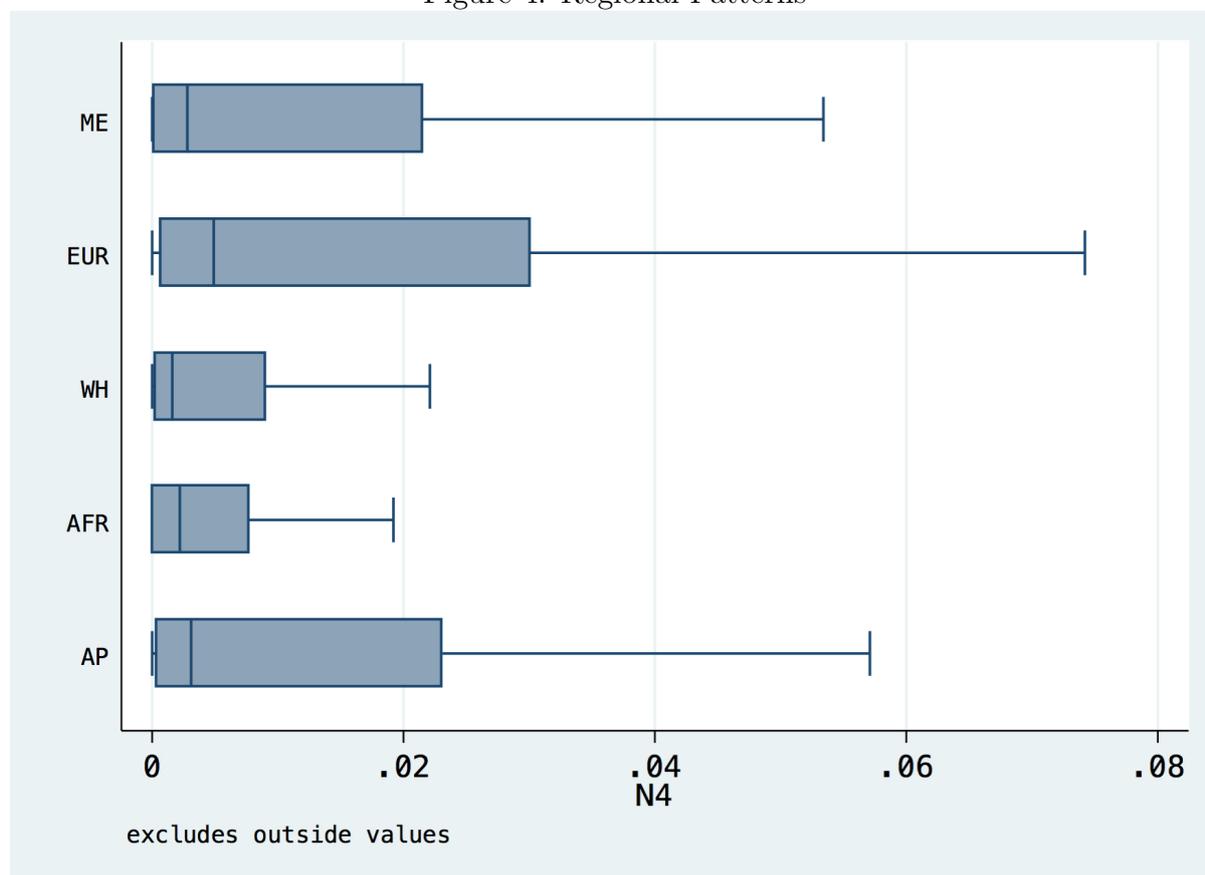
Figure 3: Regional and trans-regional PTAs



### 4.1.3 PTA Text Similarity by Region

This study also takes a closer look at the pattern of text similarity values by regions and subregions.<sup>8</sup> Figure 4 shows variation across the regions divided into Middle East (*ME*), Europe (*EUR*), Western Hemisphere (*WH*), Africa (*AFR*), and Asia-Pacific (*AP*). Figure 5 shows variation across the subregions of North Africa (*N Afr*), Sub Saharan Africa (*SS Afr*), Australia and New Zealand (*A & NZ*), Central Asia (*Ctl Asia*), *East Asia*, Pacific Islands (*Pac Is.*), *South Asia*, Southeast Asia (*SEA*), Eastern Europe (*E Eur*), Western Europe (*W Eur*), Middle East (*ME*), Central America (*Ctl Amer*), North America (*N Amer*), South America (*S Amer*), and Europe overall (*EUR*). Figures 4 and 5 show similarity patterns for 4-word minimum common word groups, the lowest value for generating text similarity values.

Figure 4: Regional Patterns



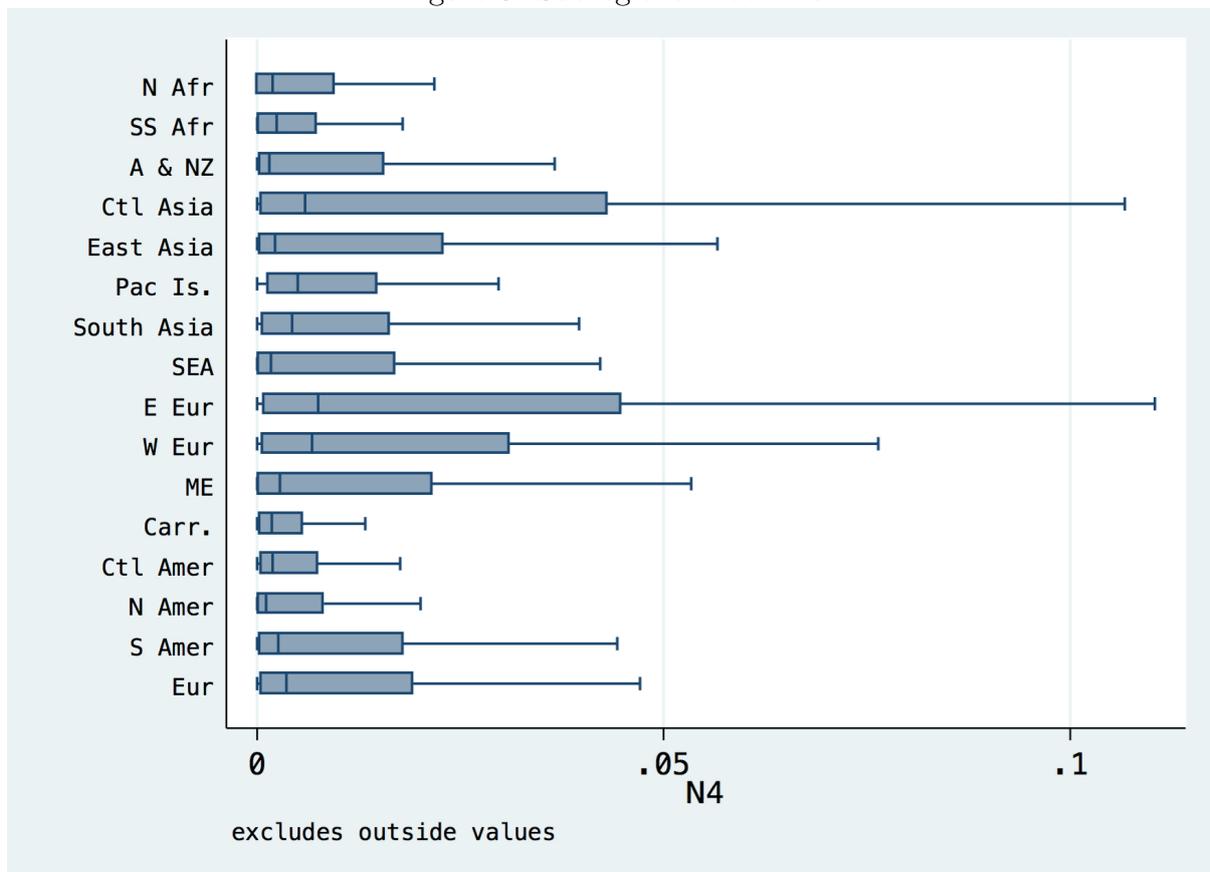
The broad regional patterns in Figure 4 show that PTAs signed by European countries have higher text commonalities with other PTAs. These are followed by agreements signed by countries in the Asia-pacific and in the Middle East. PTAs signed by countries

<sup>8</sup>This paper adopts the classification of countries by region and subregion as defined by the IMF. <http://www.imf.org/external/datamapper/region.htm>

in the Western Hemisphere, encompasses the Americas, show the lowest levels of similarity with other agreements.

Figure 5 provides greater detail to the regional patterns by showing the similarity distributions for specific sub-regions. These figures must be considered, of course, Within the context of generally low levels of replication across PTA texts. For the relatively higher figures for Europe, the higher text similarity values are more evident in PTAs signed by countries in eastern Europe rather than western Europe. These agreements also have a wider range of similarity values. For the Asia-Pacific, East Asian PTAs have the widest range of similarity values, relative to agreements signed by countries in Central, South, and Southeast Asia, and also by Australia, New Zealand, and the Pacific Islands. However, in terms of median similarity values, PTAs from South Asia and the Pacific Islands appear to be marginally higher. As for the Western Hemisphere, PTAs signed by Caribbean countries higher range of text commonalities with other agreements than those signed by countries in North, South, and central America. For the African region, North Africa and Sub-Saharan Africa have about the same median similarity values, though North Africa's PTAs appear to have a slightly wider range of values.

Figure 5: Subregional Patterns

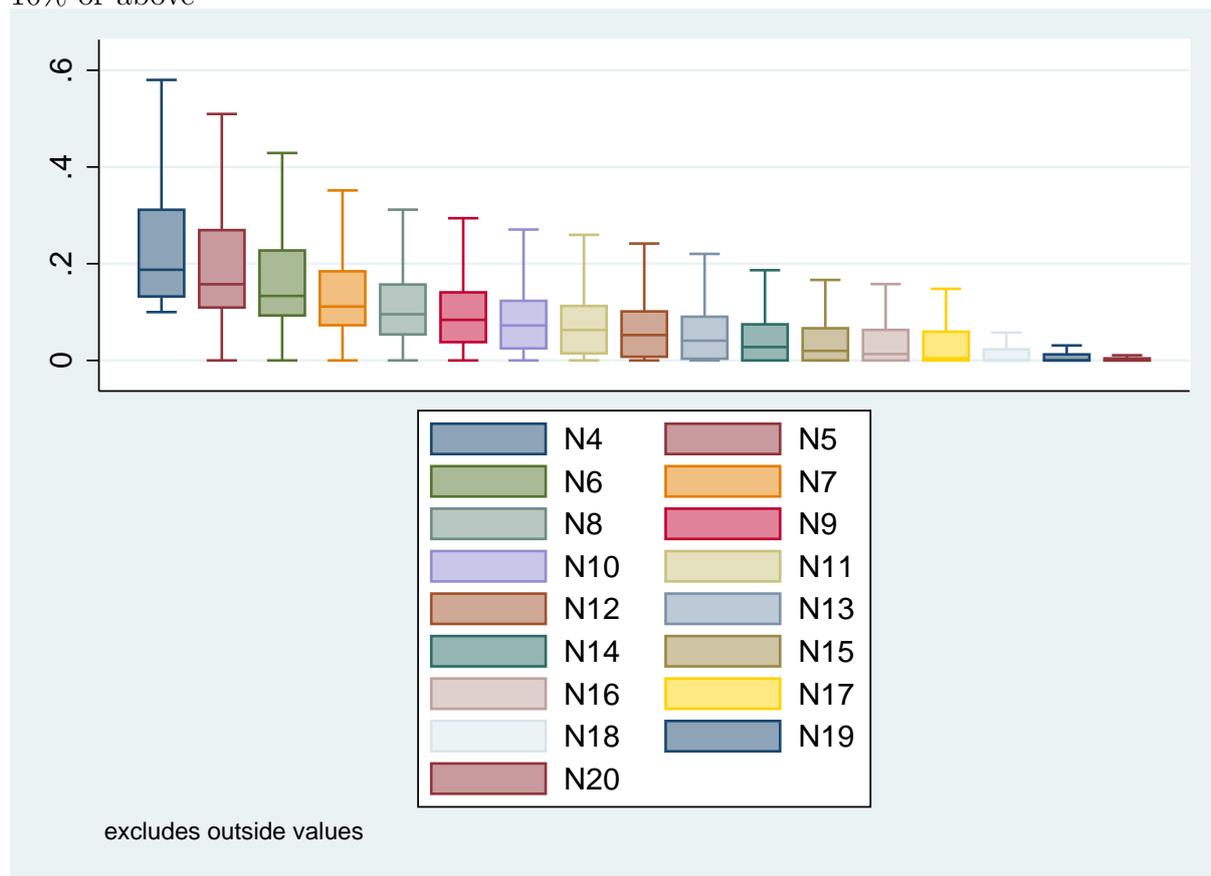


## 4.2 High similarity PTA Pairs

This section presents patterns of variation for pairs of PTAs that have similarity values of 10 percent or more. These comprise approximately one-tenth of the sample of analysis. Though 90 percent of the PTA pairs analyzed have very low text commonality, examining more closely the patterns of variation for agreement pairs that do appear to have text overlap provides further insights into their sources.

Figure 6 shows the distribution of similarity values. The patterns of variation evident in the full sample can also be seen for these cases that have similarity values of 10 percent or more. The median value is higher, as expected given this slice of the sample. It is not 'high,' registering at less than 20 percent and declining as the minimum number of sequential words in a common word group is set at higher levels. However, this median value is 10 times greater than that of the full sample.

Figure 6: Text Similarity Values across Minimum Common Word Groups: Similarity of 10% or above



In terms of longitudinal variation, Figure 7 shows that there is more fluctuation in the median and interquartile range. The median value appears to fluctuate more for these high similarity case. The range of interquartile values also appears to fluctuate more, with higher ranges appearing in the earlier years and lower ranges also appearing in the

more recent period.

Figure 7: Longitudinal Patterns in PTA Texts: 1960-2013, similarity of 10% or above

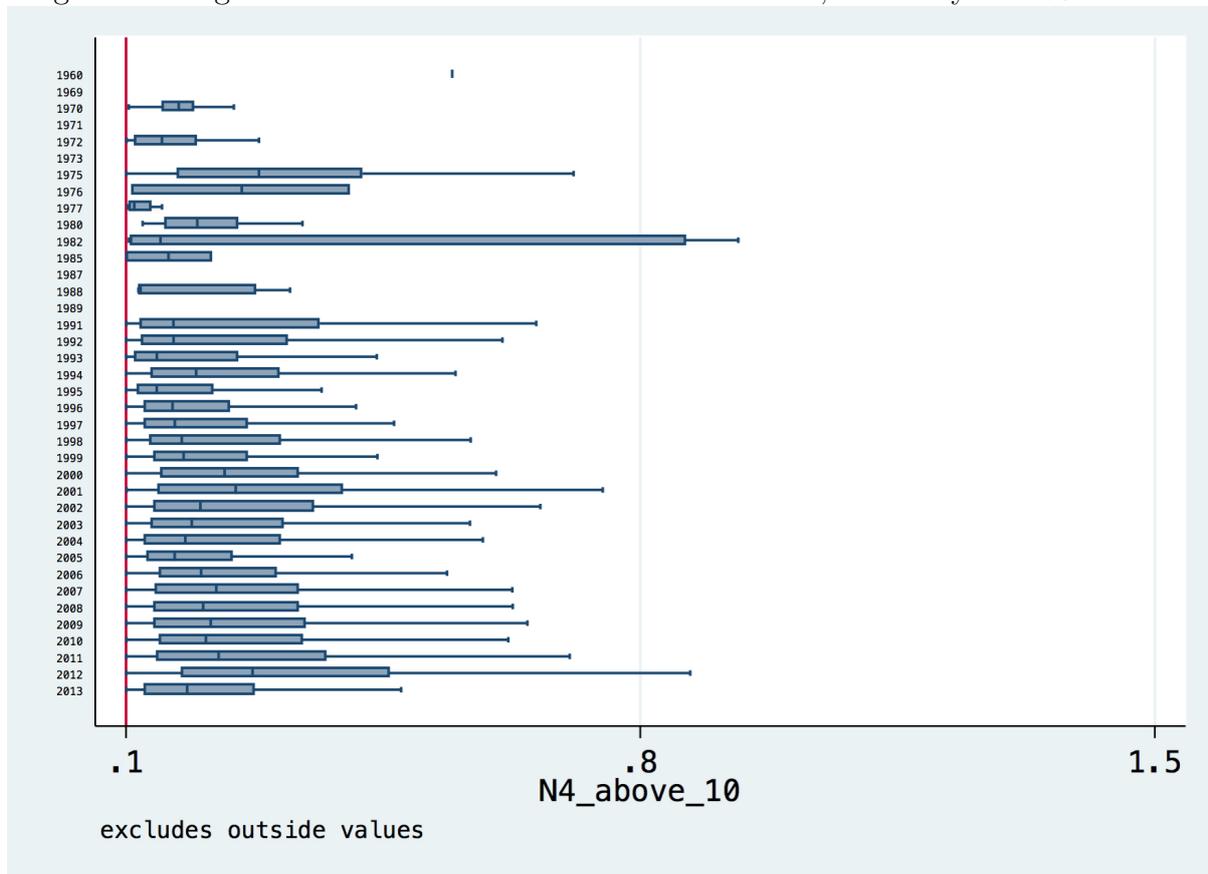
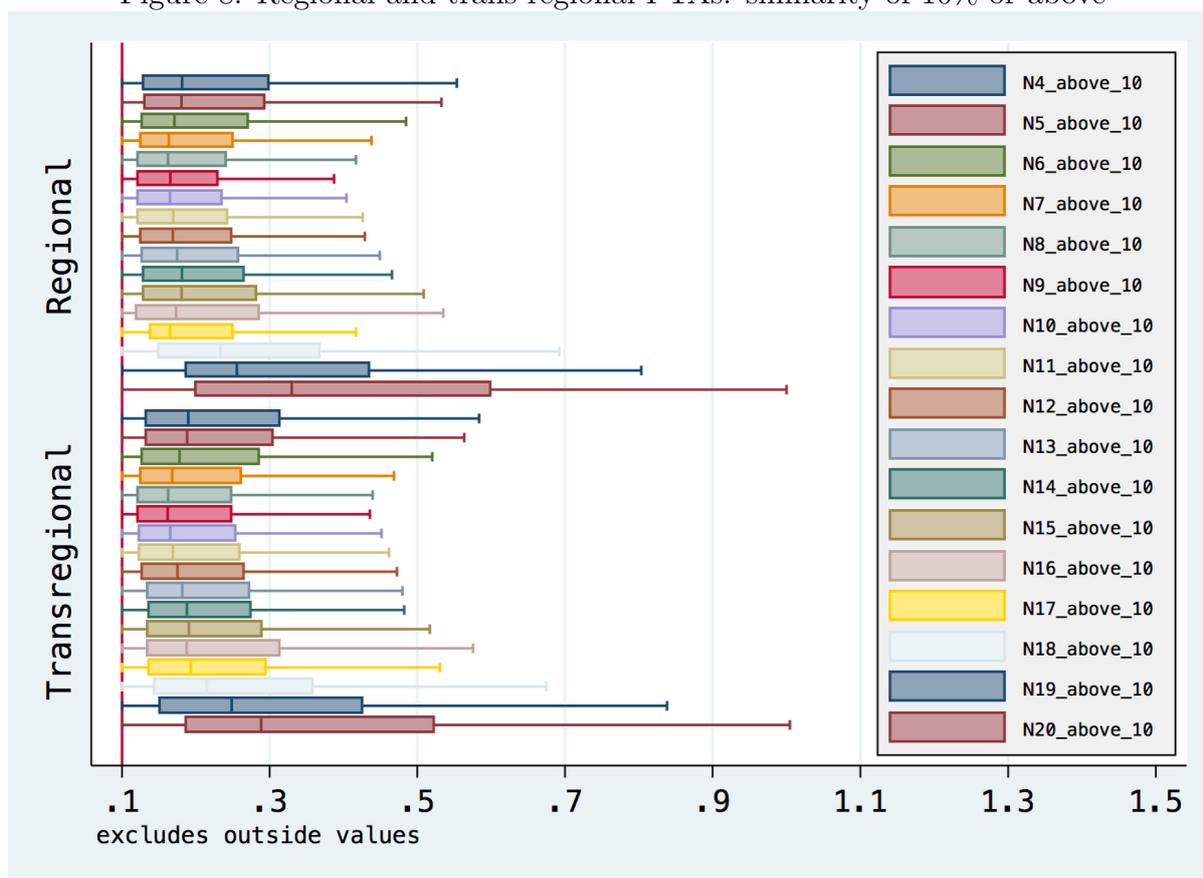


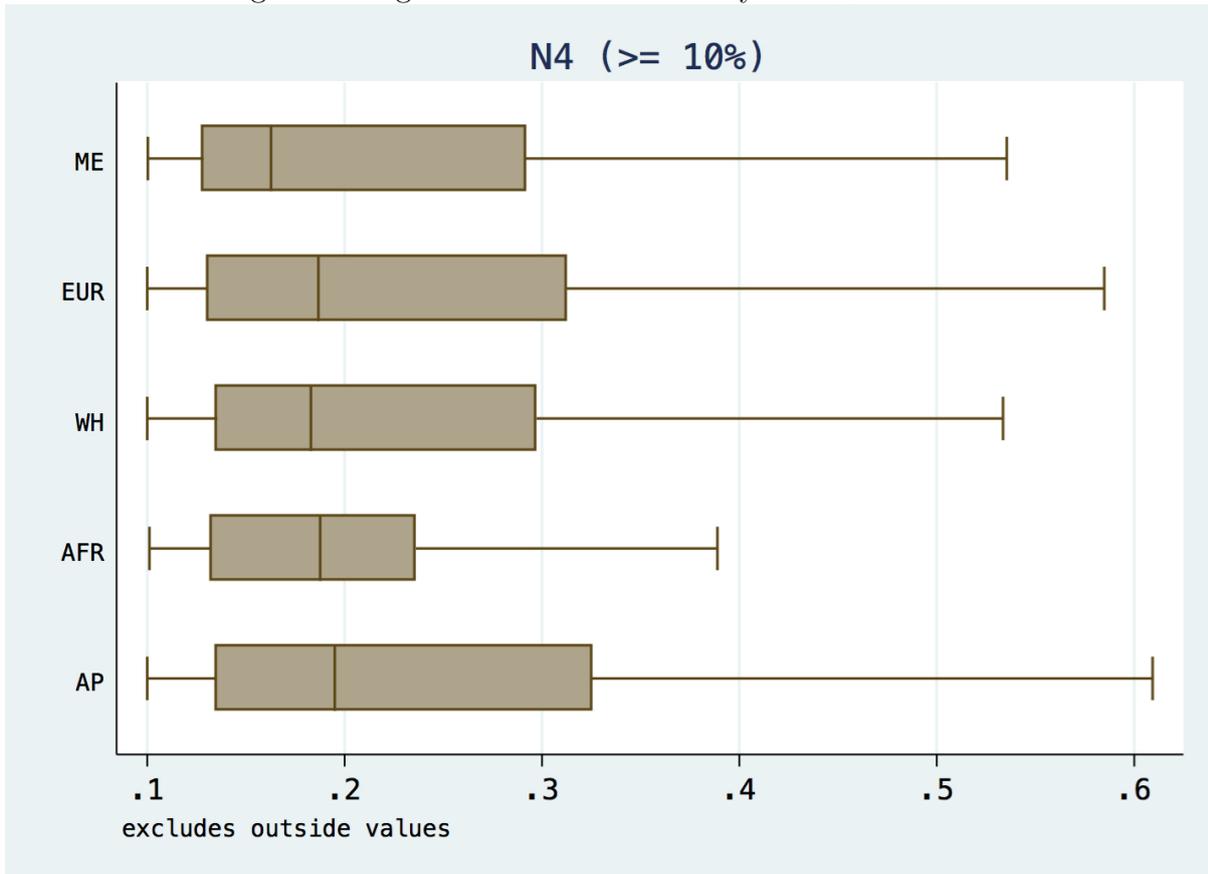
Figure 8 shows the distribution of similarity rates for regional and trans-regional PTAs. As was the case for the full sample, there appears to be no significant difference in the median similarity rates between PTAs signed by countries in these same region and those signed by countries of different regions. However, what is different from the full sample is that there also appears to be no notable difference in the range of similarity values found across these two agreements. Where there relatively higher text commonality between two agreements, both regional and trans-regional agreements do not differ in their degrees of similarity with other PTAs.

Figure 8: Regional and trans-regional PTAs: similarity of 10% or above



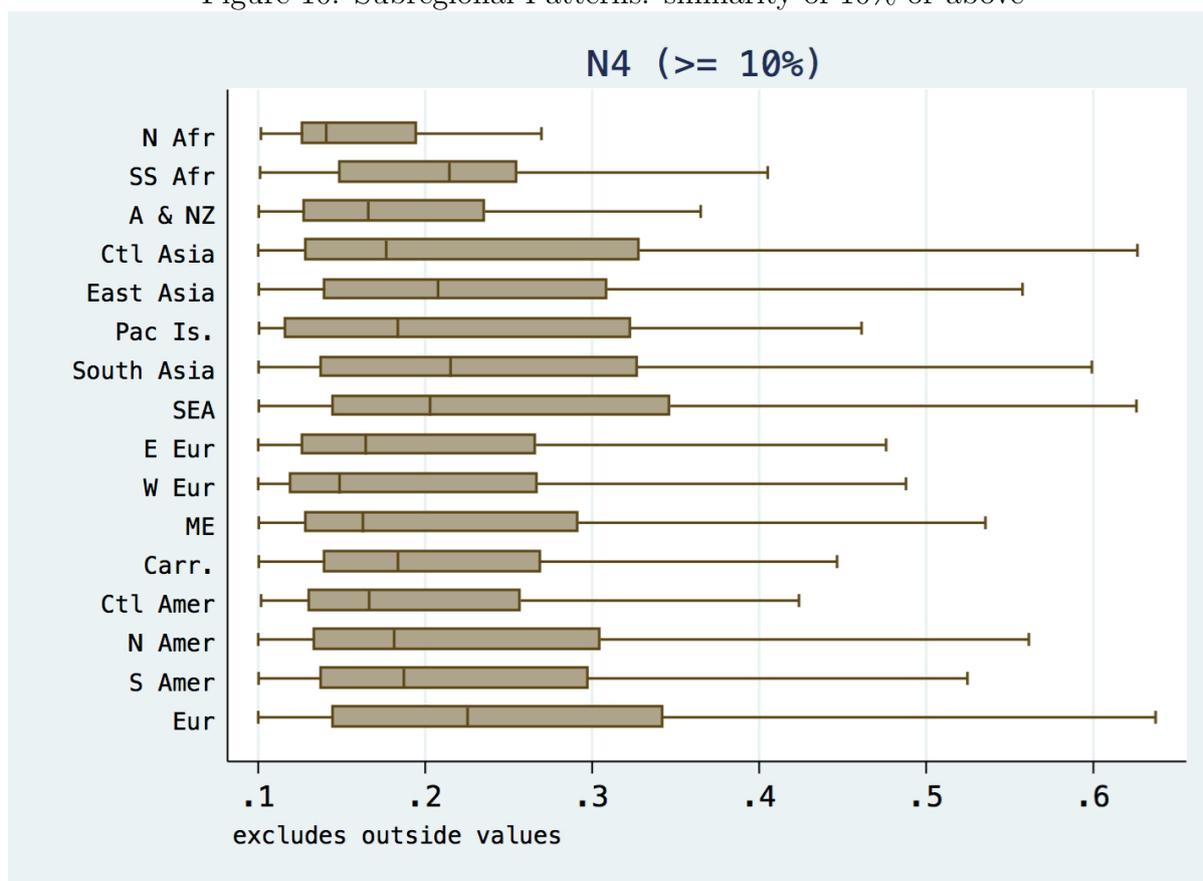
The most interesting patterns of variation that are distinct from those of the full sample are evident in regional and suregional distributions, as shown in Figures 9 and 10. First, there is much less variation across the regions. Comparisons of PTAs signed by countries of the Middle East with others have the lowest levels of similarity. Comparisons for PTAs from other regions are higher, but there are no strong differences between them as were evident in the full sample. PTAs from Europe, the Western Hemisphere, Africa, and Asia-Pacific are all have median similarity values of approximately 20 percent.

Figure 9: Regional Patterns: similarity of 10% or above



Second, an examination of subregional patterns provides more information about the major drivers of text commonalities in PTAs. For Africa, PTAs originating in Sub-Saharan countries have higher similarity rates than those of countries from North Africa. In the Asia-Pacific, PTAs from East Asia and South Asia have the highest levels of similarity, followed by Southeast Asia, the Pacific Islands, and Central Asia. In Europe, PTAs from Eastern European countries have distinctly higher similarity rates than those of Western Europe. Among the Western Hemisphere countries, North and South America and the Caribbean countries' PTAs have relatively higher similarity values than those of countries from Central America.

Figure 10: Subregional Patterns: similarity of 10% or above



## 5 What do PTA Texts Have in Common?

In this second section of the paper, I analyze the common text found in PTAs. For this purpose, I consider only the pairings of PTA texts that have similarity rates of 10 percent or more, which comprises approximately one-tenth of the cases. This approach is reasonable given that for the majority of cases, text similarity is low and thus an analysis of common text in these cases is not likely to yield any useful information.

I employed Wordfish (Slapin and Proksch (2008); Lo et al. (2015)), a scaling technique to extract political positions based on the frequencies of words found in text documents.<sup>9</sup> From the first stage analysis, in which pairs of PTAs texts were compared, I extracted the common text from comparisons of 4-word groups for those pairs of PTAs that have 10 percent or more in common text content. Using this smaller sample to focus on what countries replicate in their PTAs, I analyzed the common text to identify what is most frequently copied and which agreements look the most similar.

<sup>9</sup><http://www.wordfish.org/>.

## 5.1 Common Words in High similarity PTA Pairs

To identify the words that occur most frequently in the common texts of PTA, I generated a term document matrix, which identified 2420 word stems.<sup>10</sup> As the term document matrix indicates both the word stem and the documents–PTA pairs–in which it appears, the frequency of a particular word stem indicates how often it appears among document pairs in the analysis.<sup>11</sup>

1 presents two sets of information from the term document matrix. The top half of the table identifies the words that are among the top 100 most frequently appearing words in pairs of PTA texts. What is evident from this list is that the words refer to goods in the manufacturing industry, and suggest that they comprise lists of goods that are for some reason singled out in PTA texts. Information from the word count matrix, which provides only the frequencies of these words, does not indicate whether these goods are identified for liberalization, exclusion, or something else such as inclusion in rules of origin provisions. This task would involve going back to the texts themselves. Nevertheless, the list suggests that goods from the manufacturing industry still figure prominently in the provisions of PTAs.

The bottom half of 1 illustrates the importance of a specific class of provisions: words associated with behind-the-border commitments. Words indicating national treatment, provisions on phytosanitary issues, harmonization, standards, technical (regulations, part of technical barriers to trade), competition, and dispute settlement are some of the key principles and issue areas associated with depth (Dür and Elsig (2014)) and deep integration (Kim (2015)). If their importance is measured on the basis of word frequencies, the table shows that they are not as important as specific goods. National (197) treatment (291) appears outside the top 100 words in frequency rank, while technic(al regulations or barriers to trade, 486), investment (707), dispute(s) (862), and phytosanitary (1169) appear in the top half of words that most frequently appear in texts. Terms related to deep integration, such as harmon(ization) (1417), competition (1306), and standard(s) (1319), rank in the bottom half of the most frequently words.

---

<sup>10</sup>The text-mining package TM in *R* was employed to generate the word count/term document matrix. The document processing phase also removed a standard set of stopwords such as articles, conjunctions, and other frequently occurring words that do not have a substantive meaning in this analysis.

<sup>11</sup>Word stems capture similar words as one—for example, 'machinery' and 'machines'—thus minimizing the words that comprise a word count matrix. Generally, the stemming process removes endings from words and returns the word stems as single entries.

Table 1: Common Words in PTAs

---

Word Frequencies\*

---

From the top 100:

textil-	metal
machin-	yarn
oil	fabric
paper	fibr-
hair	iron
steel	wool
acid	

Terms of interest (rank in frequency)

treatment (197)	nation- (291)
technic- (486)	invest- (707)
disput- (862)	settlement (1695)
phytosanitary (1169)	competition (1306)
standard- (1319)	harmon- (1417)

---

\*Based on 2420 Word Stems

## 5.2 Which Agreements are Most Alike in their Content?

This section discusses the results of the Wordfish estimation, which places PTA pairs on a single dimension based on the frequencies of common words that appear in the texts.<sup>12</sup> 2 identifies the pairs of PTAs that comprise opposite ends of this dimension, which may indicate the extent to which these PTAs commit to trade liberalization. Given that the documents chosen for identification purposes include a pair of older agreements and another from among the most recently signed pair of PTAs, this dimension may also be indicative of generational differences in PTA templates.

2 shows two groups of PTAs that are position on opposite sides of the policy dimension. Group 1 consists of PTAs at the 'low' end of the estimates, which indicate that they differ most from Group 2, which includes PTAs from the 'high' end of the estimates. Though they may indicate as well the PTAs' positions on trade liberalization, the main finding of this analysis is that these groups are the most distinguishable based on their texts.

---

<sup>12</sup>For identification purposes, the oldest and most recent pairs of PTAs were used to indicate the different extremes of the policy spectrum.

Table 2: PTAs with Common Content\*

---

Groups of PTAs\*

---

Group 1

Faroe Islands/Denmark-Norway FTA

EFTA-Estonia FTA

EFTA-Slovenia

Slovenia-Turkey FTA

Egypt-Turkey

Poland-Turkey

Israel-Slovenia FTA

Israel-Slovak Republic FTA

Slovak-Republic-Turkey FTA

Hungary-Israel FTA

Hungary-Lithuania FTA

.....

Group 2

EFTA-Chile FTA

EFTA-FYROM FTA

Hong Kong-China-New Zealand FTA

New Zealand-Malaysia FTA

Peru-Singapore FTA

Singapore-Costa Rica FTA

Peru-Malaysia FTA

US-Australia FTA

Dominican Republic-Central America FTA

CAFTA-DR FTA

---

\*Based on 2420 Word Stems

Of the two groups, Group 1 appears to be a largely European group that often includes PTA signed by members of the European Free Trade Association (EFTA), several Eastern European countries, Turkey, and Israel. Group 2, on the other hand, has a mixed set of PTAs that includes EFTA's PTAs as well as PTAs signed by countries from Asia and the Americas. The predominance of Asia and the Americas in this group also suggests that this may be a PTA-grouping based on members of the Asia-Pacific Economic Cooperation (APEC) forum.

The identification of two groups suggests that PTA templates may follow a regional pattern. PTAs signed by Eastern European and EFT countries appear to be the most different from those signed by Asian countries and countries from the Americas. However, the EFTA agreements also appear to be widespread, as they appear in both groups identified by this analysis.

## 6 Conclusion

This paper has analyzed the degree of text similarity across PTAs. The analytical framework takes the texts as templates for trade liberalization, and investigates the degree to which the text content is replicated from one agreement to the next. The analysis compared pairs constituted from 416 PTA texts to generate similarity values that captures the degree of text commonality. Variation in similarity measures was examined for their longitudinal and regional patterns and differences across regional and trans-regional agreements. This paper has found that text commonalities across PTAs is lower than would be expected given the dramatic increase in the numbers of PTAs and the fact that countries negotiate and sign multiple trade agreements.

The results of the analysis so far suggest several interpretations. First, the commonalities across PTA texts may well be substantive rather than text. That is, the models of trade liberalization that countries are not necessarily couched in the same language but rather in the quality of the commitments themselves. The lack of text replication may well be attributed to different drafters of PTAs. The implication for scholarship is that mapping projects that apply coding templates to PTA texts may be a more effective way to gauge the strength and quality of liberalization commitments.

Second, what text analysis of PTA texts does contribute, however, is the insight that trade agreements may often be tailor-made for negotiating partners. Text analysis of the PTAs shows, moreover, where these important variations may be found. Comparing the results of this study with those of Allee and Elsig (2015), for example, which finds high levels of text replication using the main documents of PTAs, indicates that the annexes and supplementary documents may be the source of individual variation. The main document may contain the major commitments of the agreement partners, but the supplementary documents often contain reservations and exceptions. The results of this study, which included those supplementary documents and also found low levels of text similarity, indicates that reservations and exceptions may contribute significantly to variations in PTA commitments, and they differ markedly across agreements.

The next stage of this project is to undertake in greater depth and with the use of more sophisticated methodological tools the analysis of common text that can be found across trade agreements. This may involve examining specific issue areas covered in PTAs or adopting clustering routines to identify the agreements that are the closest in their

text content. In doing so, the objective is to identify the drivers of text commonalities across the ever increasing number of PTAs.

## Appendix 1. PTAs Included in the Analysis

- |   |   |
|---|---|
| 1. African Economic Community                                   | 22. Australia-New Zealand (ANZCERTA)      |
| 2. ALADI (Latin American Integration Association)               | 23. Azerbaijan-Russian Federation         |
| 3. ANZTEC (New Zealand and Taiwan <sup>13</sup> )               | 24. BIMST-EC                              |
| 4. Asia-Pacific Trade Agreement                                 | 25. Bahrain-Jordan                        |
| 5. ASEAN - Australia - New Zealand                              | 26. Bangladesh-India                      |
| 6. ASEAN - China  | 27. Bolivia-Chile                         |
| 7. ASEAN - India  | 28. Brunei Darussalam - Japan             |
| 8. ASEAN - Japan  | 29. CARICOM                               |
| 9. ASEAN - Korea, Republic of                                   | 30. CARICOM-Colombia                      |
| 10. ASEAN Free Trade Area (AFTA)                                | 31. CARICOM-Costa Rica                    |
| 11. Agadir (Free Trade Area among Arab Mediterranean Countries) | 32. CARICOM-Cuba                          |
| 12. Albania-Moldova   | 33. CARICOM-Dominican Republic            |
| 13. Albania-UNMIK (Kosovo)                                      | 34. CEFTA-Croatia                         |
| 14. Andean Community  | 35. Australia-New Zealand                 |
| 15. Armenia - Kazakhstan  | 36. Canada Colombia                       |
| 16. Armenia - Moldova   | 37. Canada - Costa Rica                   |
| 17. Armenia - Russian Federation                                | 38. Canada - Israel                       |
| 18. Armenia - Turkmenistan                                      | 39. Canada - Peru                         |
| 19. Armenia - Ukraine   | 40. Canada Chile                          |
| 20. Asia Pacific Trade Agreement (APTA)–Accession of China      | 41. Central European Free Trade Agreement |
| 21. Australia - Chile   | 42. Chile - China                         |
|   | 43. Chile - India                         |
|   | 44. Chile - Japan                         |

<sup>13</sup>Taiwan is referred to PTAs as the Separate Customs Territory of Taiwan, Penghu, Kinmen, and Matsu

- |   |                                  |
|---|----------------------------------|
| 45. Chile - Mexico  | 68. EC-Latvia                    |
| 46. Chile-Venezuela   | 69. EC-Lithuania                 |
| 47. China - Hong Kong, China                                  | 70. EC-Poland                    |
| 48. China - Macao, China                                      | 71. EC-Romania                   |
| 49. China - New Zealand                                       | 72. EC-Slovak Republic           |
| 50. China - Singapore   | 73. EC-Slovenia                  |
| 51. China-Iceland   | 74. EFTA - Albania               |
| 52. China-Switzerland   | 75. EFTA - Canada                |
| 53. Common Economic Zone (CEZ)                                | 76. EFTA - Chile                 |
| 54. Common Market for Eastern and<br>Souther                  | 77. EFTA - Egypt                 |
| 55. Commonwealth of Independent<br>States (CIS)               | 78. EFTA FYR Macedonia           |
| 56. Croatia-Lithuania   | 79. EFTA - Israel                |
| 57. Croatia-Moldova   | 80. EFTA - Jordan                |
| 58. Croatia-Slovenia  | 81. EFTA - Korea, Republic of    |
| 59. Dominican Republic - Central Amer-<br>ica                 | 82. EFTA - Lebanon               |
| 60. Dominican Republic - Central Amer-<br>ica - US (CAFTA-DR) | 83. EFTA - Mexico                |
| 61. EC (15) Enlargement                                       | 84. EFTA - Morocco               |
| 62. EC (25) Enlargement                                       | 85. EFTA - Palestinian Authority |
| 63. EC (27) Enlargement                                       | 86. EFTA - Peru                  |
| 64. EC-Bulgaria   | 87. EFTA - SACU                  |
| 65. EC-Czech Republic   | 88. EFTA - Serbia                |
| 66. EC-Estonia  | 89. EFTA - Singapore             |
| 67. EC-Hungary  | 90. EFTA - Tunisia               |
|   | 91. EFTA - Turkey                |
|   | 92. EFTA Bulgaria                |
|   | 93. EFTA Colombia                |
|   | 94. EFTA Croatia                 |

- |                                  |   |
|----------------------------------|---|
| 95. EFTA-Bulgaria                | 120. EU - Lebanon                                       |
| 96. EFTA-Czech Republic          | 121. EU - Mexico  |
| 97. EFTA-Estonia                 | 122. EU - Montenegro                                    |
| 98. EFTA-Hungary                 | 123. EU - Morocco                                       |
| 99. EFTA-Latvia                  | 124. EU - Palestinian Authority                         |
| 100. EFTA-Lithuania              | 125. EU - Papua New Guinea / Fiji                       |
| 101. EFTA-Poland                 | 126. EU - San Marino                                    |
| 102. EFTA-Romania                | 127. EU - Serbia  |
| 103. EFTA-Slovenia               | 128. EU - South Africa                                  |
| 104. EFTA-Slovak Republic        | 129. EU - Tunisia                                       |
| 105. EU - Albania                | 130. EU - Turkey  |
| 106. EU - Algeria                | 131. EU-Bulgaria  |
| 107. EU - Andorra                | 132. EU-Moldova   |
| 108. EU - Bosnia and Herzegovina | 133. EU-OCT   |
| 109. EU - CARIFORUM States EPA   | 134. EU-Romania   |
| 110. EU - Cameroon               | 135. EU-Switzerland-Liechtenstein                       |
| 111. EU Chile                    | 136. EU-Syria   |
| 112. EU - Croatia                | 137. East African Community (EAC)                       |
| 113. EU - Côte d'Ivoire          | 138. Economic Community of West African States (ECOWAS) |
| 114. EU - Egypt                  | 139. Economic Cooperation Organization (ECO)            |
| 115. EU - Faroe Islands          | 140. Egypt - Turkey                                     |
| 116. EU FYR Macedonia            | 141. Egypt-Jordan                                       |
| 117. EU - Israel                 | 142. Eurasian Economic Community (EAEC)                 |
| 118. EU - Jordan                 | 143. European Economic Area (EEA)                       |
| 119. EU - Korea, Republic of     | 144. FYROM-Moldova                                      |

- |  |                             |
|--|-----------------------------|
| 145. Faroe Islands - Switzerland                     | 169. India - Japan          |
| 146. Faroe Islands - Norway                          | 170. India - Malaysia       |
| 147. GSTP  | 171. India - Singapore      |
| 148. Georgia - Armenia                               | 172. India - Sri Lanka      |
| 149. Georgia - Azerbaijan                            | 173. India - Nepal          |
| 150. Georgia - Kazakhstan                            | 174. India-GCC              |
| 151. Georgia - Russian Federation                    | 175. India-Mongolia         |
| 152. Georgia - Turkmenistan                          | 176. India-Thailand         |
| 153. Georgia - Ukraine                               | 177. Iran-Pakistan          |
| 154. Georgia-EU                                      | 178. Israel - Mexico        |
| 155. Gulf Cooperation Council (GCC)                  | 179. Israel-Jordan          |
| 156. Gulf Cooperation Council-Singapore<br>FTA       | 180. Israel-Poland          |
| 157. Honduras - El Salvador and Taiwan               | 181. Israel-Slovak Republic |
| 158. Hong Kong, China - New Zealand                  | 182. Israel-Slovenia        |
| 159. Hong Kong, China-Chile                          | 183. Japan - Indonesia      |
| 160. Hong Kong, China-European Free<br>Trade (EFTA?) | 184. Japan - Mexico         |
| 161. Hungary-Israel                                  | 185. Japan - Philippines    |
| 162. Hungary-Latvia                                  | 186. Japan - Singapore      |
| 163. Hungary-Lithuania                               | 187. Japan - Switzerland    |
| 164. Hungary-Turkey                                  | 188. Japan - Thailand       |
| 165. IGAD  | 189. Japan - Vietnam        |
| 166. Iceland - Faroe Islands                         | 190. Japan - Malaysia       |
| 167. India - Afghanistan                             | 191. Japan-Vietnam          |
| 168. India - Bhutan                                  | 192. Jordan - Singapore     |
|  | 193. Jordan-Morocco         |
|  | 194. Jordan-Syria           |
|  | 195. Jordan-Tunisia         |

- |  |   |
|--|---|
| 196. Jordan-United Arab Emirate                  | 219. Malaysia-Australia                           |
| 197. Korea, Republic of - Chile                  | 220. Malaysia-Chile Free Trade Agreement          |
| 198. Korea, Republic of - India                  | 221. Malaysia-US                                  |
| 199. Korea, Republic of - Singapore              | 222. Mauritius-Pakistan                           |
| 200. Korea-United States Free Trade Agreement    | 223. Melanesian Spearhead Group (MSG)             |
| 201. Kyrgyz Republic - Armenia                   | 224. Moldova-Bosnia and Herzegovina               |
| 202. Kyrgyz Republic - Kazakhstan                | 225. Moldova-Monternegro-Serbia                   |
| 203. Kyrgyz Republic - Moldova                   | 226. Moldova-Romania                              |
| 204. Kyrgyz Republic - Russian Federation        | 227. Namibia-Zimbabwe                             |
| 205. Kyrgyz Republic - Ukraine                   | 228. New Zealand - Malaysia                       |
| 206. Kyrgyz Republic - Uzbekistan                | 229. New Zealand - Singapore                      |
| 207. Lao People's Democratic Republic - Thailand | 230. Nicaragua and Taiwan                         |
| 208. Latvia-Poland                               | 231. North American Free Trade Agreement (NAFTA)  |
| 209. Latvia-Slovak                               | 232. PTN  |
| 210. Latvia-Slovenia                             | 233. Pacific Island Countries Trade Agreement     |
| 211. Latvia-Turkey                               | 234. Pakistan - China                             |
| 212. Lithuania-Poland                            | 235. Pakistan - Malaysia                          |
| 213. Lithuania-Slovakia                          | 236. Pakistan - Sri Lanka                         |
| 214. Lithuania-Slovenia                          | 237. Pakistan-Mauritius                           |
| 215. Lithuania-Turkey                            | 238. Pan-Arab Free Trade Area (PAFTA)             |
| 216. MERCOSUR - India                            | 239. Panama - Costa Rica (Panama-Central America) |
| 217. MERCOSUR-Andean Community                   | 240. Panama - Singapore                           |
| 218. MERCOSUR-Bolivia                            | 241. Panama - Taiwan                              |

- |  |   |
|--|---|
| 242. Papua New Guinea-Australia                                      | 264. Southern Common Market (MERCOSUR)            |
| 243. People's Republic of China-Costa Rica                           | 265. TPS-OIC                                      |
| 244. People's Republic of China-Taipei                               | 266. Thailand - Australia                         |
| 245. Peru - China  | 267. Thailand - New Zealand                       |
| 246. Peru - Korea, Republic of                                       | 268. Thailand-Bahrain                             |
| 247. Peru - Singapore  | 269. Thailand-Peru Free Trade Agreement           |
| 248. Poland-Turkey   | 270. Trans-Pacific Strategic Economic Partnership |
| 249. Preferential Tariff Arrangement-Group of 8 Developing Countries | 271. Turkey Albania                               |
| 250. Romania-Turkey  | 272. Turkey - Bosnia and Herzegovina              |
| 251. Russia-Kazakhstan-Belarus                                       | 273. Turkey - Chile                               |
| 252. Russian Federation-Tajikistan                                   | 274. Turkey - Croatia                             |
| 253. SPARTECA  | 275. Turkey - FYR Macedonia                       |
| 254. Serbia-Montenegro-Romania                                       | 276. Turkey Georgia                               |
| 255. Singapore - Australia   | 277. Turkey - Israel                              |
| 256. Singapore-Costa Rica FTA  | 278. Turkey - Jordan                              |
| 257. Slovak Republic-Turkey  | 279. Turkey - Montenegro                          |
| 258. Slovenia-FYR Macedonia  | 280. Turkey - Morocco                             |
| 259. Slovenia-Turkey   | 281. Turkey - Palestinian Authority               |
| 260. South Asian Free Trade Agreement (SAFTA)                        | 282. Turkey - Serbia                              |
| 261. South Asian Preferential Trade Arrangement (SAPTA)              | 283. Turkey - Tunisia                             |
| 262. Southern African Customs Union (SACU)                           | 284. Turkey-Albania                               |
| 263. Southern African Development Community (SADC)                   | 285. Turkey-Korea                                 |
|  | 286. Turkey-Lebanon                               |
|  | 287. Turkey-Mauritius                             |
|  | 288. US - Australia                               |

289. US - Bahrain
290. US - Chile
291. US - Jordan
292. US - Morocco
293. US - Oman
294. US - Peru
295. US - Singapore
296. US-Afghanistan
297. US-Albania
298. US-Central Asia
299. US-Colombia
300. US-Israel
301. US-Lao
302. US-Pakistan
303. US-Panama
304. US-Vietnam
305. Ukraine - Azerbaijan
306. Ukraine - Belarus
307. Ukraine - FYR Macedonia
308. Ukraine - Kazakhstan
309. Ukraine - Moldova
310. Ukraine - Russian Federation
311. Ukraine - Tajikistan
312. Ukraine - Uzbekistan
313. Ukraine -Turkmenistan
314. United States-Marshall Islands Compact of Free Association
315. United States-Micronesia Compact of Free Association
316. United States-Palau Compact of Free Association
317. Uzbekistan-Russian Federation

Table 3: text similarity values across PTA Texts

Minimum number of words in common word group	Mean	Standard Deviation	Median	N*
4	0.033	0.096	0.003	53,014
5	0.026	0.084	0.001	53,014
6	0.021	0.073	0.0004	53,014
7	0.017	0.055	0.0001	53,014
8	0.015	0.048	0	53,014
9	0.013	0.044	0	53,014
10	0.012	0.041	0	53,014
11	0.011	0.041	0	53,014
12	0.010	0.041	0	53,014
13	0.009	0.042	0	53,014
14	0.008	0.040	0	53,014
15	0.007	0.038	0	53,014
16	0.007	0.039	0	53,014
17	0.007	0.040	0	50,782
18	0.005	0.040	0	49,514
19	0.005	0.043	0	44,385
20	0.005	0.050	0	39,940

\*Number of PTA pairs

## References

- Allee, T. and M. Elsig (2015). Are the contents of international treaties copied-and-pasted? evidence from preferential trade agreements. Paper presented at the annual meeting of the Political Economy of International Organizations.
- Bird, S., E. Klein, and E. Loper (2009). *Natural Language Processing with Python* (1st ed.). O'Reilly Media, Inc.
- Brady, H. and D. Collier (2004). *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. G - Reference, Information and Interdisciplinary Subjects Series. Rowman & Littlefield.
- Chauffour, J.-P., J.-C. Maur, et al. (2011). *Preferential Trade Agreement Policies for Development: A Handbook*. World Bank Publications.
- Dür, A., L. Baccini, and M. Elsig (2014). The design of international trade agreements: Introducing a new dataset. *The Review of International Organizations* 9(3), 353–375.
- Dür, A. and M. Elsig (2014). *Trade Cooperation: The Purpose, Design and Effects of Preferential Trade Agreements*. Cambridge: Cambridge University Press.
- Estevadeordal, A., K. Suominen, and R. Teh (2009). *Regional Rules in the Global Trading System*. Cambridge: Cambridge University Press.
- Fishlow, A. and S. Haggard (1992). *The United States and the Regionalisation of the World Economy*. Paris: OECD Development Centre Research Project on Globalization and Regionalization.
- Horn, H., P. C. Mavroidis, and A. Sapir (2010). Beyond the wto? an anatomy of eu and us preferential trade agreements. *The World Economy* 33(11), 1565–1588.
- Kim, S. Y. (2015). *The Oxford Handbook of the Political Economy of International Trade*, Chapter Deep Integration and RTAs. Oxford: Oxford University Press.
- King, Gary, R. O. K. and S. Verba (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Koremenos, B., C. Lipson, and D. Snidal (2001). The rational design of international institutions. *International Organization* 55(04), 761–799.
- Kucik, J. (2012). The domestic politics of institutional design: Producer preferences over trade agreement rules. *Economics & Politics* 24(2), 95–118.

- Lo, J., S.-O. Proksch, and J. B. Slapin (2015, 8). Ideological clarity in multiparty competition: A new measure and test using election manifestos. *British Journal of Political Science FirstView*, 1–20.
- Mansfield, E. D. and H. V. Milner (1999). The new wave of regionalism. *International Organization* 53(03), 589–627.
- Moravcsik, A. (2000, 3). The origins of human rights regimes: Democratic delegation in postwar europe. *International Organization* 54, 217–252.
- Slapin, J. B. and S.-O. Proksch (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science* 52(3), 705–722.
- Smith, J. M. (2000). The politics of dispute settlement design: Explaining legalism in regional trade pacts. *International Organization* 54(1), pp. 137–180.
- Viner, J. (1950). *The Customs Union Issue*. New York: Carnegie Endowment for International Peace.
- WTO (2011). The wto and preferential trade agreements: From co-existence to coherence. Technical report, World Trade organization.