IOParlspeech: Data Documentation

IOParlspeech is a dataset containing over 600,000 statements on IOs from legislative speeches held in national parliamentary debates in Austria, Canada, Germany, United Kingdom, United States, and New Zealand over the time period 1990 to 2018. This document provides documentation on the construction and contents of the dataset, available at https://doi.org/10.7910/DVN/MBLASU

Table of Contents:

1. IOParlspeech Description
2. IOParlspeech Construction
3. IOParlspeech Variables
4. Examples of IO statements
5. List of IOs in IOParlspeech
6. Number of statements per IO
7. Validation: Minimizing False Positives and False Negatives in IOParlspeech
8. References

Citation Details: When using IOParlSpeech, please cite the following article:

- *Hunter, T. & Walter, S (2025). International Organizations in National Parliamentary Debates. Review of International Organizations. https://doi.org/10.1007/s11558-024-09577-w*

Please also cite the relevant original parliamentary data source (see also Table 1):

- For Austria, Germany, the UK, and New Zealand:
  Rauh, C., Schwalbach, J. (2020). The ParlSpeech V2 Data Set: Full-Text Corpora of 6.3 Million Parliamentary Speeches in the Key Legislative Chambers of Nine Representative Democracies. Harvard Dataverse.

- For Canada:
  Beelen, K., Thijm, T. A., Cochrane, C., Halvemaan, K., Hirst, G., Kimmins, M., Lijbrink, S., Marx, M., Naderi, N., Rheault, L., Polyanovsky, R., & Whyte, T. (2017). Digitization of the Canadian Parliamentary Debates. Canadian Journal of Political Science, 50(3), 849–864.

- For the USA:
  Gentzkow, M., Shapiro, J.M., & Taddy, M. (2018). Congressional Record for the 43rd-114th Congresses: Parsed Speeches and Phrase Counts. Palo Alto, CA: Stanford Libraries

For questions regarding IOParlspeech please contact Tom Hunter (hunter@ipz.uzh.ch) .

Acknowledgments IOParlspeech is built thanks to scholars making excellent, machine readable datasets of parliamentary speeches openly available, most notably Parlspeech (Rauh and Schwalbach 2020). Sections 1 and 2 provide details on how the data in IOParlspeech were constructed from these sources. We are also grateful for excellent research assistance from Nicole Plotke-Scherly and Rabiya Abdullah and helpful comments from Mirko Heinzel, Farsan Ghassim, Christian Rauh, Ken Benoit, Stephanie Hoffmann, Svanhildur Thorvaldsdottir, and Sara Hobolt. Support for this data collection was provided by the European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme grant agreement No 817582 (ERC Consolidator Grant DISINTEGRATION).

1. IOParlspeech Description

IOParlspeech is a dataset on how international organizations (IOs) are debated in national parliaments. It consists of over 600,000 statements on IOs in national parliamentary debates between 1990 and 2018. The data covers six countries: Austria, Canada, Germany, United Kingdom, United States, and New Zealand.

IOParlspeech is generated from the Parlspeech V2 dataset (Rauh and Schwalbach 2020) and from parsed speeches in the US Congress and Canadian House of Commons (Gentzkow et al 2018, Beelen et al 2017), as shown in Table 1[1]. Keyword in context is applied to find mentions of IOs (full names and acronyms), as well as extensive validation to remove false positives and minimize false negatives (see section 7). These IO statements are three sentences in length, capturing both the sentence where the IO is mentioned, and the sentence before and after for additional context (Rauh et al., 2020).

Table 1 – Number of IO-related statements per legislature and data source

Legislature	**Number of IO-related statements**	**Data Source**	**Period**
Austria (Nationalrat)	42,450	Parlspeech V2 (Rauh and Schwalbach 2020)	1996 – 2018
Canada (House of Commons)	101,371	Digitization of the Canadian Parliamentary Debates (Beelen et al 2017)	1990 – 2018
Germany (Bundestag)	76,876	Parlspeech V2 (Rauh and Schwalbach 2020)	1991 – 2018
New Zealand (House of Representatives)	25,842	Parlspeech V2 (Rauh and Schwalbach 2020)	1990 – 2018
United Kingdom (House of Commons)	280,602	Parlspeech V2 (Rauh and Schwalbach 2020)	1990 – 2018
United States (Congress)	120,611	Congressional Record for the 43rd-114th Congresses (Gentzkow et al 2018)	1990 - 2016

IOParlspeech includes metadata such as information on date, speaker, party, the IO mentioned, and the relevant IO classifiers, amongst others. Hunter and Walter (2025) introduce the dataset and illustrate some uses of this data through sentiment and salience analyses.

The dataset, full documentation, and replication materials for the analyses in Hunter and Walter (2025) are available on Harvard Dataverse at *https://doi.org/10.7910/DVN/MBLASU*.

[1] For a review on comparative legislative research with large-scale computational text analysis (particularly on Europe), see Sebők et al (2025).

2. IOParlspeech Construction

To construct IOParlspeech, we draw on the growing range of machine-readable datasets of parliamentary speeches. Most notably, we use Parlspeech V2, which provides a corpus of speeches from Austria's Nationalrat, the British House of Commons, the German Bundestag, and New Zealand's House of Commons (Rauh and Schwalbach 2020). We complement the Parlspeech data with a dataset of speeches in the US Congress (Gentzkow et al 2018), and a dataset of speeches made in Canada's House of Commons (Beelen et al 2017). We applied the same pre-processing to all speeches, removing very short speeches of less than 10 tokens and removing full stops that do not denote the end of the sentence with the quanteda package (Benoit et al. 2018).

We then use keyword search for both the IO's name (e.g. United Nations) and acronym (e.g. UN) in the original language (English or German) to capture IO statements, with the sentence before and after also included for additional context. The list of IOs is based on all the IOs included in the *Measuring International Authority Database* (Hooghe and Marks 2017). Table 4 shows the IOs included and contains the list of search strings. It is important to note here that this method maximizes precision over recall. That is to say, it includes only terms for which an IO reference is evident to avoid false positives. For example, we exclude terms that reference specific IO policies without naming the IO. The choices made here means we capture IO statements rather conservatively but guard against false positives that might occur. Whilst one speech can contain several IO statements, we remove exact duplicates (for example if an IO is mentioned multiple times in the same sentence). Section 7 contains further details on validation.

Why settle on three-sentence windows as the length of these IO statements? One could for example extract the full speech in which an IO is mentioned, or identify debates that are explicitly about an IO and extract all the speeches from these debates, or even use tokens rather than sentences for the length of the window. We opt for three-sentence windows for a number of reasons. First, international cooperation is now rather integrated into the national politics of member states, so that references to IOs frequently appear in debates that are not explicitly about that IO. Only including IO-labelled debates would therefore run the risk of ignoring a significant number of statements about international cooperation. Furthermore, the labelling of parliamentary debates is sometimes incomplete (see Rauh and De Wilde 2018), whereas the speeches themselves are consistently well captured. Three sentence windows also replicate the methodology from other studies (Traber et al 2020; Rauh et al 2020) that experiment with one-sentence windows and windows based on number of tokens and opt for three-sentence windows because they are a natural context unit for human listeners, and because the resulting data windows are empirically most well-behaved in distributional terms.

3. IOParlspeech Variables

Table 2 – Variables in IOParlspeech Dataset

Variable	Description
parliament	Legislature where the IO statement is made (e.g. Nationalrat)
iso3country	Three-letter country code, as published by the International Organization for Standardization. (e.g. GBR, DEU)
date	Day the statement was made in a character vector (YYYY-MM-DD)
quarter	Quarter when the IO statement is made (e.g. 2003.3 for the third quarter of 2003)
year	Year when IO statement was made. Based on variable 'date' above
speaker	Character vector with the full name of the person making the IO statement.
party	Party or faction of the speaker as given in the plenary protocol (e.g. LibDem, Rep)
in_government	Binary variable ("in government" or "in opposition") that captures whether the speaker's party was in government at the time of the speech
party_facts_id	Numeric identifier for the party in Döring and Regel's (2019) Party Facts database. This in turn offers straightforward linkages with e.g. the Chapel Hill Expert Survey (Bakker et al 2015) Manifesto Data (Lehmann et al 2024) or the ParlGov database (Döring and Manow 2018)
party_family	Party family in a character vector (e.g. 'Liberal', 'Socialist')
io_statement	The key variable in IOParlspeech. Captures the three-sentence statement within the speech where an IO is mentioned. Note that one speech can contain multiple IO statements

io_statement_length	Overall length (number of words) of the IO statement.
keyword	The keyword used to identify the IO. This includes both the IO name (e.g International Monetary Fund) and the acronym (e.g. IMF)
ioname	Acronym of IO referenced in the IO statement (e.g. EEA, WTO)
iolongname	Full name of IO referenced in the IO statement (e.g. European Economic Area, World Trade Organization)
io_issue	Policy issue of the IO, drawn from COW IGO dataset. Categorical variable with three levels: 1 = social, 2 = economic, 3 = security)
io_authority	Sum of delegated and pooled authority from the *Measuring International Authority Database* (Hooghe et al 2017) used to create an overall measure of authority, as in Dingwerth et al (2020)
ioyear	Aggregation of ioname and year. Useful for aggregation (e.g. NATO 2007)
agenda	Name of the agenda item under which the IO statement was made (e.g. UK Human Trafficking Centre, Manufacturing Industry etc)
full_speech	Raw text of the full speech from which the IO statement is taken.
full_speech_length	Overall length (number of words) of the full speech from which the IO statement is taken..
positive_all	Number of positive terms in the IO statements, developed using sentiment dictionaries in English (Young and Soroka 2012) and German (Rauh 2018)
negative_all	Number of negative terms in the IO statements, developed using sentiment dictionaries in English (Young and Soroka 2012) and German (Rauh 2018)
pos_share	Share of the the IO statement dedicated to positive language. Calculated by dividing positive_all by io_statement_length

neg_share	Share of the the IO statement dedicated to negative language. Calculated by dividing negative_all by io_statement_length
netsent	Net sentiment of the IO statement. Calculated with (Pos.Share – Neg.Share). Higher values indicate more positive language.
member	Binary variable that captures whether the speaker's country was a member of the IO at the time of the speech. ('member' or 'not.member')
legelec	Binary variable that captures whether an election took place the year in which the IO statement was made (1 for election year, 0 otherwise). Important as research shows that electoral cycles play an important role in legislative speech (Schwalbach 2022, Rauh and De Wilde 2018)
unemployment	Yearly unemployment rate from the World Bank
IPI	Binary variable that captures whether the IO mentioned had an international parliamentary institution (IPI) at the time of the speech. IPI Data captured from Schimelfenning et al (2020)

4.Examples of IO statements

The default unit of analysis in IOParlspeech is the 'IO statement' – a three sentence passage where an IO is mentioned in a parliamentary speech. Table 3 below gives three examples of typical IO statements: an MP from New Zealand criticizing the allowances paid to peacekeeping forces, a British MP criticizing French dominance of the IMF, and a US Congressman lauding the benefits of NAFTA.

Table 3 – Exemplary Statements in IOParlspeech

Date	**Speaker**	**IO**	**IO Statement**
1994/09	G.Braybrooke (NZ - Labour)	UN	They found that compared with other United Nations peacekeeping forces they were grossly underpaid. The allowances paid to our peacekeeping forces were described as a pittance compared with those paid to other United Nations peacekeeping forces. They also discovered other things, which our troops will also discover when they get to Bosnia.
2014/02	J.Rees-Mogg (UK - Cons)	IMF	The IMF is not full of well-known leftists, but it does seem to be run, by and large, by the French, who have a very different understanding of economics, an absolutely rotten economy, and are the last people from whom I would take lessons. We will not in this Chamber go into the behaviour of the previous managing director it would shock the viewers of the Parliament channel if they were to consider how Monsieur Strauss-Kahn had behaved.
1993/11	C.Grassley (USA - Rep)	NAFTA	I would like to quote Glen Keppy, president elect for the National Pork Producers. How can any reasonable person in Iowa be against NAFTA? It will mean the opportunity for more rural development and jobs created by the sales of more of the agricultural products Iowa leads the nation in producing.

5. List of IOs in IOParlspeech

Table 4 - List of IOs in IOParlspeech

The table below includes the full list of IOs in IOParlspeech. These are drawn from the Measuring International Authority Database (Hooghe and Marks 2017). The search strings include both the IO's acronym and the IO's full name. The languages are English for the USA, UK, Canada, and New Zealand, and German for Germany and Austria.

Acronym in English	**Full name in English**	**COW issue area**
ALADI	Latin American Integration Association	economic
AMU	Arab Maghreb Union	economic
APEC	Asia-Pacific Economic Cooperation	economic
ASEAN	Association of Southeast Asian Nations	economic
AU	African Union	political
Benelux	Benelux	political
BIS	Bank for International Settlements	economic
CABI	Centre for Agriculture and Bioscience International	social
CAN	Andean Community	economic
Caricom	Caribbean Community	economic
CCNR	Central Commission for the Navigation of the Rhine	political
CEMAC	Economic and Monetary Community of Central Africa	economic
CERN	European Organization for Nuclear Research	economic
CIS	Commonwealth of Independent States	political
COE	Council of Europe	political
COMECON	Council for Mutual Economic Assistance	economic
COMESA	Common Market for Eastern and Southern Africa	economic
ComSec	Commonwealth Secretariat	political
EAC2	East African Community	economic
ECCAS	Economic Community of Central African States	economic
ECOWAS	Economic Community of West African States	economic
EEA	European Economic Area	economic

EFTA	European Free Trade Agreement	economic
ESA	European Space Agency	political
EU	European Union	economic
FAO	Food and Agriculture Organization	economic
GCC	Gulf Cooperation Council	economic
GEF	Global Environment Facility	social
IAEA	International Atomic Energy Agency	social
IBRD	International Bank for Reconstruction and Development	economic
ICAO	International Civil Aviation Organization	social
ICC	International Criminal Court	political
IGAD	Intergovernmental Authority on Development	economic
ILO	International Labour Organization	social
IMF	International Monetary Fund	economic
IMO	International Maritime Organization	economic
Interpol	International Criminal Police Organization	social
IOM	International Organization for Migration	political
ISA	International Seabed Authority	social
ITU	International Telecommunication Union	economic
IWhale	International Whaling Commission	economic
LOAS	Arab League/League of Arab States	political
Mercosur	Southern Common Market	economic
NAFO	Northwest Atlantic Fisheries Organization	social
NAFTA	North American Free Trade Agreement	economic
NATO	North Atlantic Treaty Organization	political
NordC	Nordic Council	political
OAPEC	Organization of Arab Petroleum Exporting Countries	economic
OAS	Organization of American States	political
OECD	Organization for Economic Cooperation and Development	economic
OECS	Organization of Eastern Caribbean States	economic

OIC	Organization of Islamic Cooperation	social
OIF	International Organisation of La Francophonie	social
OPEC	Organization of the Petroleum Exporting Countries	economic
OSCE	Organization for Security and Cooperation In Europe	political
OTIF	Intergovernmental Organization for International Carriage by Rail	economic
PCA	Permanent Court of Arbitration	political
PIF	Pacific Islands Forum	political
SAARC	South Asian Association for Regional Cooperation	political
SACU	Southern African Customs Union	economic
SADC	Southern African Development Community	economic
SCO	Shanghai Cooperation Organization	political
SELA	Latin American and the Caribbean Economic System	economic
SICA	Central American Integration System	economic
SPC	South Pacific Commission	political
UN	United Nations	political
UNESCO	United Nations Educational, Scientific and Cultural Organization	social
UNIDO	United Nations Industrial Development Organization	economic
UNWTO	World Tourism Organization	economic
UPU	Universal Postal Union	economic
WCO	World Customs Organization	economic
WHO	World Health Organization	social
WIPO	World Intellectual Property Organization	economic
WMO	World Meteorological Organization	social
WTO	World Trade Organization	economic

6. Number of Statements per IO

Table 5 – Number of statements per IO

	Austria		Germany		UK		Canada		USA		New Zealand		Full Sample	
IO name	Total	Share	Total	Share	Total	Share	Total	Share	Total	Share	Total	Share	Total	Share
EU	32,895	77%	44,661	58%	166,874	59%	7,355	7%	5,656	5%	1,420	5%	258,861	39%
UN	1,498	4%	3,349	4%	49,473	18%	32,611	32%	26,608	22%	8,083	31%	121,622	19%
NATO	2,476	6%	13,220	17%	20,692	7%	7,474	7%	20,030	17%	156	1%	64,048	10%
NAFTA	60	0%	142	0%	468	0%	24,792	24%	33,212	28%	36	0%	58,710	9%
WTO	462	1%	2,453	3%	7,810	3%	10,090	10%	13,120	11%	2,192	8%	36,127	6%
OECD	1,999	5%	2,492	3%	3,845	1%	4,684	5%	956	1%	6,796	26%	20,772	3%
IMF	401	1%	1,554	2%	5,067	2%	2,299	2%	3,853	3%	1,463	6%	14,637	2%
IBRD	169	0%	1,227	2%	3,418	1%	1,042	1%	2,031	2%	498	2%	8,385	1%
WHO	311	1%	890	1%	2,167	1%	0	0%	1,464	1%	461	2%	7,271	1%
COE	640	2%	719	1%	4,718	2%	339	0%	134	0%	19	0%	6,526	1%
OSCE	535	1%	2,250	3%	755	0%	401	0%	1,593	1%	0	0%	5,534	1%
ICC	27	0%	88	0%	2,049	1%	875	1%	779	1%	414	2%	4,232	1%
IAEA	26	0%	282	0%	1,184	0%	203	0%	2,254	2%	33	0%	3,982	1%
ILO	18	0%	94	0%	1,127	0%	431	0%	599	1%	1,069	4%	3,338	1%
UNESCO	236	1%	622	1%	930	0%	980	1%	436	0%	115	0%	3,319	1%
OPEC	47	0%	93	0%	147	0%	109	0%	2,756	2%	14	0%	3,166	0%
EEA	254	1%	68	0%	2,566	1%	0	0%	2	0%	1	0%	2,891	0%
APEC	1	0%	25	0%	7	0%	1,268	1%	56	0%	0	0%	2,849	0%
AU	3	0%	216	0%	1,105	0%	518	1%	388	0%	6	0%	2,236	0%
EFTA	27	0%	92	0%	1,243	0%	598	1%	8	0%	0	0%	1,968	0%
ASEAN	10	0%	192	0%	390	0%	95	0%	436	0%	572	2%	1,695	0%
NAFO	0	0%	2	0%	15	0%	1,572	2%	42	0%	3	0%	1,634	0%
LOAS	1	0%	51	0%	718	0%	142	0%	539	0%	12	0%	1,463	0%

ICAO	8	0%	53	0%	316	0%	339	0%	304	0%	101	0%	1,121	0%
IMO	0	0%	129	0%	500	0%	65	0%	249	0%	53	0%	991	0%
Interpol	59	0%	90	0%	439	0%	156	0%	186	0%	55	0%	985	0%
IWhale	6	0%	139	0%	330	0%	7	0%	324	0%	97	0%	903	0%
CIS	27	0%	489	1%	97	0%	13	0%	229	0%	3	0%	858	0%
OAS	0	0%	31	0%	20	0%	139	0%	630	1%	0	0%	820	0%
FAO	18	0%	265	0%	196	0%	138	0%	84	0%	35	0%	736	0%
ESA	45	0%	242	0%	179	0%	23	0%	81	0%	11	0%	581	0%
WIPO	5	0%	12	0%	15	0%	179	0%	297	0%	66	0%	574	0%
Caricom	0	0%	0	0%	118	0%	119	0%	278	0%	2	0%	517	0%
GEF	10	0%	58	0%	67	0%	15	0%	315	0%	1	0%	466	0%
SADC	4	0%	94	0%	329	0%	1	0%	18	0%	7	0%	453	0%
Mercosur	32	0%	127	0%	127	0%	75	0%	49	0%	6	0%	416	0%
ICFO	0	0%	1	0%	273	0%	1	0%	19	0%	0	0%	294	0%
ECOWAS2	0	0%	87	0%	104	0%	29	0%	73	0%	0	0%	293	0%
CERN	55	0%	9	0%	208	0%	6	0%	1	0%	2	0%	281	0%
SPC	0	0%	0	0%	0	0%	2	0%	48	0%	216	1%	266	0%
IOM	16	0%	41	0%	120	0%	65	0%	17	0%	4	0%	263	0%
PIF	0	0%	0	0%	28	0%	8	0%	25	0%	173	1%	234	0%
GCC	0	0%	2	0%	133	0%	1	0%	51	0%	42	0%	229	0%
Benelux	43	0%	70	0%	95	0%	12	0%	6	0%	0	0%	226	0%
OIC	0	0%	2	0%	8	0%	12	0%	181	0%	1	0%	204	0%
UNIDO	12	0%	69	0%	77	0%	1	0%	2	0%	1	0%	162	0%
ComSec	0	0%	1	0%	130	0%	10	0%	0	0%	13	0%	154	0%
COMECON	0	0%	86	0%	22	0%	0	0%	7	0%	1	0%	116	0%
WMO	1	0%	5	0%	14	0%	26	0%	50	0%	11	0%	107	0%
BIS	2	0%	11	0%	35	0%	17	0%	26	0%	11	0%	102	0%
UPU	4	0%	0	0%	14	0%	27	0%	8	0%	44	0%	97	0%

NordC	0	0%	0	0%	32	0%	3	0%	55	0%	42	0%	90	0%
IGAD	0	0%	10	0%	54	0%	19	0%	13	0%	0	0%	96	0%
WCO	6	0%	1	0%	6	0%	22	0%	31	0%	20	0%	86	0%
CAN	0	0%	24	0%	7	0%	4	0%	35	0%	0	0%	70	0%
ITU	0	0%	2	0%	15	0%	5	0%	18	0%	8	0%	48	0%
UNWTO	0	0%	22	0%	8	0%	11	0%	4	0%	1	0%	46	0%
ISA	0	0%	2	0%	35	0%	0	0%	0	0%	2	0%	39	0%
EAC2	0	0%	3	0%	19	0%	3	0%	0	0%	0	0%	28	0%
OIF	0	0%	0	0%	0	0%	20	0%	0	0%	0	0%	20	0%
OECS	0	0%	0	0%	2	0%	0	0%	14	0%	0	0%	16	0%
SAARC	0	0%	2	0%	8	0%	0	0%	6	0%	0	0%	16	0%
COMESA	0	0%	2	0%	12	0%	0	0%	1	0%	0	0%	15	0%
SCO	0	0%	0	0%	3	0%	4	0%	6	0%	0	0%	13	0%
SICA	0	0%	1	0%	0	0%	13	0%	0	0%	0	0%	14	0%
PCA	0	0%	0	0%	9	0%	0	0%	3	0%	0	0%	12	0%
SACU	0	0%	0	0%	1	0%	0	0%	6	0%	0	0%	8	0%
CABI	0	0%	0	0%	4	0%	0	0%	0	0%	0	0%	4	0%
ECCAS	0	0%	0	0%	1	0%	1	0%	2	0%	0	0%	4	0%
CCNR	1	0%	0	0%	0	0%	2	0%	0	0%	0	0%	3	0%
AMU	0	0%	0	0%	0	0%	0	0%	1	0%	0	0%	1	0%
CEMAC	0	0%	0	0%	1	0%	0	0%	0	0%	0	0%	1	0%
OAPEC	0	0%	0	0%	0	0%	0	0%	1	0%	0	0%	1	0%
ALADI	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
OTIF	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Total	42,450	100%	76,876	100%	280,602	100%	101,371	100%	120,611	100%	25,842	100%	647,752	100%

7 – Validation: Minimizing False Positives and False Negatives in IOParlspeech

Scholars using automated text analysis methods must validate their use (Grimmer and Stewart 2013). In particular, they should ensure that both false positives (in our case, capturing a statement which isn't in fact about an IO) and false negatives (excluding a statement that is in fact about an IO) are minimized. In our case, identifying statements about IOs requires using acronyms, but some of these could yield false positives. For instance, in the UK 'ISA' might refer to the Individual Savings Account rather than the International Settlements Authority. In Germany, 'WIPO' is not simply an acronym for the World Intellectual Property Organization, but also for 'Wirtschaft/Politik' (Economics/Politics), a subject taught in secondary schools. To guard against false positives, we therefore employ a number of steps. First, handcoders were given a random sample of 5 hits from each acronym search string for each parliament. If any of these random samples returned one or more false positives, hand coders were asked to hand code the full sample hits using the acronyms. Altogether, 17 per cent of acronyms included a false positive, and hand coders coded over 900 additional hits to ensure false positives were excluded from the dataset.

False negatives could also be an issue for IOParlspeech. Parliamentarians might reference an IO without mentioning the IO's name directly by alluding to its policies. For instance, an MP that references the 'Millennium Development Goals' is implicitly referencing the UN. If the majority of mentions of an IO come from mentions of its policies or internal institutions, false negatives are likely to be a significant problem in IOParlspeech. If the majority of mentions of an IO come from direct references of the IO's name or acronym we can be confident that our method does a decent job of capturing the majority of parliamentary communication on IOs.

To show that our method does not exclude the majority of communication on IOs, we take the case where false negatives seem most likely. The EU, widely considered the most authoritative IO (Hagemann et al 2016) has a considerable policy output and a maze of institutions and agencies that make up the Brussels 'bubble'. Additionally, it also has the advantage of a validated dictionary of EU-level terms in English and German (see De Wilde and Rauh 2018), that includes a range of policy, institutional, and polity-related EU terms. Altogether, this dictionary consists of 78 EU-level terms in English, and 145 EU-level terms in German. We identify all EU-related parliamentary discourse based on these dictionaries, and find that in all six of our legislatures, the majority of EU communication is drawn from simple mentions of either the 'European Union' or the ' EU'. The figure is higher in non-EU member states (65% in USA, 71% in Canada, 61% in New Zealand) and lower in EU member states, who naturally reference EU policy and institutions more frequently. Even in these case however here the majority of EU references come from these two EU terms (51% In Germany, 58% in Austria, and 60% in the UK). We conclude that the prospect of false negatives, whilst undoubtedly present in the dataset, does not fundamentally question the validity of the data collection process.

It is important to note here that false positives and negatives are unavoidable in quantitative models of language, and that all text as data are by nature 'wrong', in the sense that they cannot perfectly capture the variable of interest (in our case, the totality of IO communication in parliamentary debates). In their seminal article on automated text analysis, Grimmer and Stewart (2013) outline that the fact that 'all automated methods are based on incorrect models of language also implies that models should be evaluated based on their ability to perform some useful social scientific task'. We believe our data generation achieve this criterion: through our efforts to minimize false positives through hand coding and through our analysis of the extent of false negatives in the case where these are most likely, we conclude that IOParlspeech provides a useful approximation of IO communication in parliamentary debate

Table 6 – English Dictionaries of all EU-level terms for validation

EU polity	EU politics	EU policy
(eu|ec)[1-9]{1,2}	ecb	(cle)sdp
(europeanleurope'sleu's) constitutional treaty	ecj	(commonleuropean) foreign and security polic(ylies)
(romelmaastrichtlamsterdamlnicellisbon) treat(ylies)	ep	(commonleuropean) security and defen(slc)e polic(ylies)
ec('s){0,1}	european (official(s){0,1}|civil servant(s){0,1})	eurozoneleuro zoneleuro area
economic and monetary union	european (politicslpolicy)	cfsp
eec('s){0,1}	european central bank	european ([a-z]*){0,1}polic(ylies)
emu	european commission(erlers){0,1}	european ([a-z]*){0,1}(act(s){0,1}|bill(s){0,1}|law(s){0,1}|legislation(s){0,1}|statute(s){0,1}
eu('s){0,1}	european competenc(eleslies)	european ([a-z]*){0,1}(aim(s){0,1}|goal(s){0,1}|target(s){0,1}
euratom('s){0,1}	european council	european ([a-z]*){0,1}decision(s){0,1}
european ([a-z]*){0,1}(integrationlunificationlcooperation)	european court of justice	european ([a-z]*){0,1}directive(s){0,1}
european_communit(ylies)	european election(s){0,1}	european ([a-z]*){0,1}engagement(s){0,1}
european (economic latomic energy)communit(ylies)	european executive	european ([a-z]*){0,1}guideline(s){0,1}
european institutions	european level(s){0,1}	european ([a-z]*){0,1}(measure(s){0,1}|action(s){0,1})
european project(s){0,1}	european member state(s){0,1}	european ([a-z]*){0,1}(provision(s){0,1}|prescription(s){0,1}
european treat(ylies)	european parliament	european ([a-z]*){0,1}(requirement(s){0,1}|allowance(s){0,1}
european_union('s){0,1}	european procedure(s){0,1}	european ([a-z]*){0,1}(standard(s){0,1}|norm(s){0,1}
single european act	european summit(s){0,1}	european ([a-z]*){0,1}agenda(s){0,1}
treat(ylies) of (romelmaastrichtlamsterdamlnicellisbon)	mep(s){0,1}	european ([a-z]*){0,1}budget(s){0,1}
treaty establishing a constitution for europe	policy on europe	european ([a-z]*){0,1}f(ulo)nd(s){0,1}
treaty on (the functioning of the){0,1}european union		european ([a-z]*){0,1}programme(s){0,1}
		european ([a-z]*){0,1}regulation(s){0,1}
		european ([a-z]*){0,1}strateg(ylies)
		european (case-law|jurisprudencellegal)
		european (single linternal)market{0,1}
		european [a-z]* union
		european currenc(ylies)
		european mandate(s){0,1}
		police and judicial cooperation in criminal matters
		single currency
		stability and growth pact

Table 7: German dictionary of all EU level terms for validation

EU polity	EU politics	EU policy
europäische(n|r){0,1} union	(eu|eg)-kommission	(eu|eg)-[a-zäöüß]*politik(en){0,1}
europäische(n|r){0,1} (atom|wirtschafts){0,1}gemeinschaft(en){0,1}	europäische(n|r){0,1} kommission	europäische(n|r){0,1} [a-zäöüß]*politik(en){0,1}
eu	(eu|eg)-kommissar(e){0,1}	europäische(n|r){0,1} [a-zäöüß]+union
eg	(eu|eg)-kommissarin(nen){0,1}	europäische(n|r){0,1} mandat(e|s){0,1}
ewg	europäische(n|r){0,1} Kommissare(n){0,1}	europäische(n|r){0,1} binnenmarkt(s|es){0,1}
euratom	(eu|eg)-beamte(n|r){0,1}	einheitliche(n|r){1} binnemarkt(s|es){0,1}
(eu|eg)-vertr(ag|ages|ags|äge){1}	europäische(n|r){0,1} beamte(n|r){0,1}	europäische(n|r){0,1} [a-zäöüß]*integration
vertrag(s|es){0,1} von (maastricht|amsterdam|nizza|lissabon)	europäische(n){0,1} exekutive	gemeinsame(n|r){0,1} außen- und sicherheitspolitik
(maastricht|amsterdam|nizza|lissabon)-vertrag(s|es){0,1}	europäische(n|s){0,1} parlament(es|s){0,1}	europäische(n|r){0,1} außen- und sicherheitspolitik
(lissabonner|amsterdamer) vertrag(es|s){0,1}	europaparlament(es|s){0,1}	polizeiliche(n|r){0,1} und justizielle(n|r){0,1} zusammenarbeit
einheitliche(n|r){0,1} europäische(n|r){0,1} akte	(eu|eg)-parlament(es|s){0,1}	europäische(n|r){0,1} [a-zäöüß]*m(a|ä)rkt(e|s|es){0,1}
römische(n) verträge	ep	(eu|eg)-[a-zäöüß]*agenda
aeu-vertrag(es|s){0,1}	europawahl(en){0,1}	(eu|eg)-[a-zäöüß]*haushalt(s|es){0,1}
eu-verfassung(svertrag|svertrages){0,1}	europaabgeordnete(n|r){0,1}	(eu|eg)-[a-zäöüß]*programm(s|es|e){0,1}
europäische(n|r){0,1} verfassung(svertrag|svertrags|svertrages){0,1}	(eu|eg)-abgeordnete(n|r){0,1}	(eu|eg)-[a-zäöüß]*regulierung(en){0,1}
(eu|eg)-erweiterung(en){0,1}	(eu|eg)-ministerrat(s|es){0,1}	europäische(r|n){0,1} [a-zäöüß]*regulierung(en){0,1}
europäische(n|r){0,1} währungsunion	(eu|eg)-[a-zäöüß]*minister	(eu|eg)-[a-zäöüß]*vorschrift(en){0,1}
europa der [1-9]{1,2}	ratspräsidentschaft	europäische(r|n){0,1} [a-zäöüß]*vorschrift(en){0,1}
(eu|eg)-[1-9]{1,2}	(eu|eg)-ratspräsidentschaft	(eu|eg)-[a-zäöüß]*vorgabe(n){0,1}
europäische(n){0,1} projekt(es|s){0,1}	europäische(r|n){0,1} [a-zäöüß]*rat(s|es){0,1}	europäische(r|n){0,1} [a-zäöüß]*vorgabe(n){0,1}
europäische(n|r){0,1} einigung	(eu|eg)-gipfel(n){0,1}	(eu|eg)-[a-zäöüß]*ziel(e){0,1}
europäische(n|r){0,1} integration(sproze(ss|ß)|sproze(ss|ß)e|sproze(ss|ß)es){0,1}	europagipfel(n){0,1}	europäische(n|r){0,1} ziel(e){0,1}
(eu|eg)-institution(en){0,1}	europäische(n|r){0,1} gipfel(n){0,1}	(eu|eg)-[a-zäöüß]*maßnahmen
europäische(n|r){0,1} institution(en){0,1}	(eu|eg)-mitgliedstaat(en){0,1}	europäische(r|n){0,1} [a-zäöüß]*maßnahmen
wirtschafts- und währungsunion	europäische(n|r){0,1} mitgliedstaat(en){0,1}	(eu|eg)-instrumente(n){0,1}
ewu	(eu|eg)-mitgliedsl(and|änder){1}	europäische(n|r){0,1} instrumente(n){0,1}
wwu	europäische(n|r|s){0,1} mitgliedsl(and|änder)	(eu|eg)-[a-zäöüß]*standard(s){0,1}
ewwu	(eu|eg)-staat(en){0,1}	europäische(n|r){0,1} standard(s){0,1}
	(eu|eg)-l(and|änder)	(eu|eg)-[a-zäöüß]*norm(en){0,1}
	europäische(r|n){0,1} gerichtshof(s|es){0,1}	europäische(n|r){0,1} [a-zäöüß]*norm(en){0,1}
	eugh	(eu|eg)-[a-zäöüß]*zusammenarbeit
	(eu|eg)-gerichtshof(es|s){0,1}	europäische(n|r){0,1} zusammenarbeit
	(eu|eg)-gericht(s|e){0,1}	(eu|eg)-[a-zäöüß]*gesetzgebung
	europäische(n|r){0,1} zentralbank	europäische(r|n){0,1} [a-zäöüß]*gesetzgebung
	ezb	(eu|eg)-[a-zäöüß]*gesetz(e){0,1}
	ezb-direktorium	europäische(s|n|r){0,1} [a-zäöüß]*gesetz(e){0,1}
	ezb-rat	(eu|eg)-[a-zäöüß]*recht(es|s){0,1}
	hohe(r|n){0,1} vertreter(in|s){0,1} für außen- und sicherheitspolitik	europarecht(es|s){0,1}
	europapolitik	(eu|eg)-[a-zäöüß]*rechtsetzung
	europäische(r|n){0,1} ebene	(eu|eg)-[a-zäöüß]*richtlinie(n){0,1}
	(eu|eg)-ebene	europäische(n|r){0,1} [a-zäöüß]*richtlinie(n){0,1}
	europäische(n|r){0,1} verfahren	(eu|eg)-[a-zäöüß]*verordnung(en){0,1}
	europabühne	europäische(n|r){0,1} [a-zäöüß]*verordnung(en){0,1}
	(eu|eg)-kompetenz(en){0,1}	(eu|eg)-[a-zäöüß]*entscheidung(en){0,1}
	europäische(n|r){0,1} kompetenz(en){0,1}	europäische(n|r){0,1} [a-zäöüß]*entscheidung(en){0,1}
		(eu|eg)-[a-zäöüß]*leitlinie(n){0,1}
		europäische(n|r){0,1} [a-zäöüß]*leitlinie(n){0,1}
		(eu|eg)-[a-zäöüß]*reform(en){0,1}
		(eu|eg)-engagement(s){0,1}
		(eu|eg)-[a-zäöüß]*strategie(n){0,1}
		europäische(n|r){0,1} [a-zäöüß]*strategie(n){0,1}
		europäische(n|r){0,1} sicherheits- und verteidigungspolitik
		esvp
		europäische(n|r){0,1} sicherheits- und verteidigungsunion
		esvu
		gemeinsame(n|r){0,1} sicherheits- und verteidigungspolitik
		europäische(n|r){0,1} recht(sprechung|sordnung|setzung) {0,1}
		europäische(n|s){0,1} [a-zäöüß]*recht(es|s){0,1}
		vertragsverletzungsverfahren
		vorabentscheidungsverfahren
		aeuv
		europäische(n|r){0,1} währung(en){0,1}
		(eu|eg)-währung(en){0,1}
		gemeinschaftswährung
		eurozone
		euro-zone
		euroraum(s){0,1}
		euro-raum(s){0,1}
		europäische(n|r){0,1} [a-zäöüß]*fonds
		(eu|eg)-[a-zäöüß]*fonds
		stabilitäts- und wachstumspakt(s|es){0,1}

8. References

Beelen, K., Thijm, T. A., Cochrane, C., Halvemaan, K., Hirst, G., Kimmins, M., Lijbrink, S., Marx, M., Naderi, N., Rheault, L., Polyanovsky, R., & Whyte, T. (2017). Digitization of the Canadian Parliamentary Debates. *Canadian Journal of Political Science, 50(3),* 849–864.

Benoit et al., (2018). quanteda: An R package for the quantitative analysis of textual data. Journal of Open Source Software, 3(30), 774.

Dingwerth, K., Schmidtke, H., & Weise, T. (2020). The rise of democratic legitimation: Why International organizations speak the Language of Democracy. European Journal of International Relations, 26(3), 714–741.

Döring, H., & Manow, P. (2024). Parliaments and governments database (ParlGov): Release 2024. Harvard Dataverse, V1.

Gentzkow, M., Shapiro, J.M., & Taddy, M. (2018). Congressional Record for the 43rd-114th Congresses: Parsed Speeches and Phrase Counts. Palo Alto, CA: Stanford Libraries

Grimmer, J., & Stewart, B. M. (2013). Text as data: The Promise and pitfalls of Automatic Content Analysis methods for political texts. Political Analysis, 21(3), 267–297

Hooghe, L., Marks, G., Lenz, T., Bezuijen, J., Ceka, B., & Derderyan, S. (2017). Measuring international authority: A postfunctionalist theory of governance, Vol. III. Oxford University Press.

Jolly, S., Bakker, R., Hooghe, L., Marks, G., Polk, J., & Rovny, J., Steenbergen, M., Vachudova, M. A. (2022). Chapel Hill Expert Survey Trend file, 1999–2019. Electoral Studies, 75(February 1), 102420.

Lehmann, P., Franzmann, S., Al-Gaddooa, D., Burst, T., Ivanusch, C., Regel, S., Riethmüller, F., Volkens, A., Weßels, B., & Zehnter, L. (2024): The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2024a. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB)

Rauh, C., Schwalbach, J. (2020). The ParlSpeech V2 Data Set: Full-Text Corpora of 6.3 Million Parliamentary Speeches in the Key Legislative Chambers of Nine Representative Democracies. *Harvard Dataverse, 1*, 1

Rauh, C., Bes, B. J., & Schoonvelde, M. (2020). Undermining, defusing or defending European integration? Assessing Public Communication of European Executives in Times of EU Politicisation. European Journal of Political Research, 59(2), 397–423.

Rauh, C., & Wilde, P. D. (2018). The opposition deficit in EU accountability: Evidence from over 20 years of Plenary Debate in four Member States. European Journal of Political Research, 57(1), 194–216.

Rauh, C. (2018). Validating a sentiment Dictionary for German Political Language—a Workbench note. Journal of Information Technology & Politics, 15(4), 319–343.

Schimmelfennig, F., Winzen, T., Lenz, T., Rocabert, J., Crasnic, L., & Gherasimov, C., Lipps, J., Mumford, D. (2020).The rise of International parliaments: Strategic Legitimation in International Organizations. Oxford University Press.

Schwalbach, J. (2022) 'Going in circles? The influence of the electoral cycle on the party behaviour in parliament', European Political Science Review, 14(1), pp. 36–55. doi:10.1017/S1755773921000291.

Sebők, M., Proksch, S.-O., Rauh, C., Visnovitz, P., Balázs, G., & Schwalbach, J. (2025). Comparative European legislative research in the age of large-scale computational text analysis: A review article. *International Political Science Review*, *46*(1), 18-39

Traber, D., Schoonvelde, M., & Schumacher, G. (2020). Errors have been made, others will be blamed: Issue Engagement and blame shifting in Prime Minister speeches during the Economic Crisis in Europe. European Journal of Political Research, 59(1), 45–67.

Young, L., & Soroka, S. (2012). Affective News: The automated coding of sentiment in political texts. *Political Communication*, 29(2), 205–231.