



**Universität
Zürich** ^{UZH}

DIS
INTEGRATION 
the mass politics of disintegration

International Organizations Corpus:

A corpus of news articles on International Organizations from major newspapers in seven German and English speaking countries between 1990 and 2022

Marco Martini*, Martin Juan José Bucher[^], & Stefanie Walter*

*University of Zurich

[^]ETH Zurich

Description of the data:

Our International Organizations corpus is a collection of close to 1 million news articles from major newspapers in 7 English and German speaking countries between 1990 and 2022, which are concerned with reporting on one or more of some 70 major international organizations. The corpus has been compiled using the LexisNexis news API. The compilation has taken place in the context of the DISINTEGRATION project,¹ headed by Stefanie Walter, at the University of Zurich (<https://www.disintegration.ch>).

Our country selection is designed to match the *IOParISpeech* data, a dataset on parliamentary speeches, collected in this project by Hunter (2023). Despite the targeted case selection strategy, our corpus achieves very broad coverage across the English speaking world. Our selection of IOs has been following *Measure of International Authority* (MIA) data compiled by Hooghe et al. (2017, 2019): That is, we focus on the set of IOs that are contained in the MIA to make our data comparable with theirs and allow researchers add the variables coded by the MIA project to our corpus (see 'IO coverage' below for a list of IOs and the 'Search terms' section for a discussion of occasional deviations of our data from the MIA selection).

Our newspaper selection is based on the largest (in terms of circulation) nation-wide appearing dailies and weeklies for the countries in our sample, that are available via LexisNexis. We generally aimed to achieve a rough balance in term of newspapers' political leaning within countries. In countries, this goal has been easier to achieve in countries with better data availability. In terms of time frame, we opted for coverage ranging back to the End of the Cold War.

The corpus contains not only the text of the articles (separated into heading, subheading, lead, and body), but also a range of meta data (see below for more details). All articles are in original language. The corpus might be of particular interest for researchers seeking to understand media discourse on International Organizations and/or international integration over the last decades in one or more of the countries in our data.

¹ This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement No 817582 (ERC Consolidator Grant DISINTEGRATION).

Temporal coverage:

01.01.1990 - 31.12.2022

Country coverage (ISO-3 codes):

AUS, AUT, CAN, DEU, GBR, NZL, USA

IO coverage:

African Union

Andean Community

Arab League

Arab Maghreb Union

Asia-Pacific Economic Cooperation

Association of Southeast Asian Nations

Bank for International Settlements

Caribbean Community

Central American Integration System

Central Commission for the Navigation of the Rhine

Centre for Agriculture and Bioscience International

Common Market for Eastern and Southern Africa

Commonwealth of Independent States

Commonwealth Secretariat

Council for Mutual Economic Assistance

Council of Europe

East African Community

Economic and Monetary Community of Central Africa

Economic Community of Central African States

Economic Community of West African States

European Economic Area

European Free Trade Association

European Organization for Nuclear Research

European Space Agency

European Union

Food and Agriculture Organization

General Agreement on Tariffs and Trade

Global Environment Facility

Gulf Cooperation Council

Intergovernmental Authority on Development

Intergovernmental Organization for International Carriage by Rail

International Atomic Energy Agency

International Civil Aviation Organization

International Coffee Organization

International Criminal Court

International Criminal Police Organization

International Labour Organization

International Maritime Organization

International Monetary Fund

International Organization for Migration

International Seabed Authority
International Telecommunication Union
International Whaling Commission
Latin American and the Caribbean Economic System
Latin American Integration Association
Mercosur
Nordic Council
North American Free Trade Agreement
North Atlantic Treaty Organization
Northwest Atlantic Fisheries Organization
Organisation internationale de la Francophonie
Organization for Economic Cooperation and Development
Organization for Security and Cooperation in Europe
Organization of American States
Organization of Arab Petroleum Exporting Countries
Organization of Eastern Caribbean States
Organization of the Islamic Conference
Organization of the Petroleum Exporting Countries
Pacific Community
Pacific Islands Forum
Permanent Court of Arbitration
Shanghai Cooperation Organisation
South Asian Association for Regional Cooperation
Southern African Customs Union
Southern African Development Community
United Nations Educational, Scientific, and Cultural Organization
United Nations Industrial Development Organization
Universal Postal Union
World Bank
World Customs Organization
World Health Organization
World Intellectual Property Organization
World Meteorological Organization
World Tourism Organization
World Trade Organization

News sources (by country):

AUS Herald Sun, Daily Telegraph, The Australian, The Sydney Morning Herald
AUT Der Standard, Die Presse
CAN Toronto Star, Toronto Sun, The Globe and Mail, National Post, The Gazette/ Montréal Gazette
DEU BILD, Die Welt, Die Zeit, Der Spiegel, Welt am Sonntag, Bild am Sonntag
NZL The New Zealand Herald, The Dominion Post, The Press, Otago Daily Times
GBR The Sun, Daily Mail, The Daily Telegraph, The Times, Financial Times, The Guardian, The Economist, The Spectator
USA USA Today, Wall Street Journal, New York Times, New York Post

Search terms (used to query the LexisUni API on the sources and specified above):

For a full list of all search terms used in the LexisNexis API query, see the separately available .txt files named "IO_search_terms_de.txt" and "IO_search_terms_en.txt." Note that the _de file contains search terms in German language, while the _en file contains search terms for the same IOs in English.

To avoid a large number of false positives stemming from ambiguous meanings of IO abbreviations, the query is based only on long forms of IO names (e.g., African Union rather than AU). We also excluded the "United Nations" as an Umbrella organization from the query because this is not distinguishable from other UN sub-organizations such as the United Nations Development Program.

Total number of articles (rows):

926'078

Temporal coverage (number of articles per year):

1990	10855
1991	11855
1992	17713
1993	16098
1994	19687
1995	20397
1996	17598
1997	22704
1998	25473
1999	26524
2000	23936
2001	21247
2002	21495
2003	29539
2004	29079
2005	30909
2006	29683
2007	28686
2008	30185
2009	31013
2010	29068
2011	34381
2012	32820
2013	30361
2014	31220
2015	32481
2016	43043
2017	35910
2018	45007
2019	46588
2020	58895
2021	31147
2022	10480

Country coverage (number of articles by country):

AUS 40348

AUT 52392
CAN 157006
DEU 90358
GBR 423875
NZL 37229
USA 124870

Available Text and Metadata (*columns*):

id	LexisNexis unique document ID
title_h1	Article title
title_h2	Article subtitle
author	Author name
published	Publication date
language	Publication language
text_lead	Text of the article lead
text_body	Main part of the article text
publication_name	Name of the news outlet
publication_type	Type of outlet (e.g. daily, weekly, tabloid)
publication_edition	Edition of the outlet (e.g. Sunday version)
publication_copyright	Copyright information
source_section	Section within the news source (e.g. politics, economy, sports)
wordcount	Total number of words in the article
country	Country code (ISO-3)
full_xml	Original XML data file as retrieved from LexisNexis

Available data formats:

The corpus is available in two formats: 1) It can be accessed as a PostgreSQL database on a UZH server. This way it can be effectively queried and subset via SQL commands prior to further analysis. This requires some SQL programming skills, however. 2) alternatively, the corpus can also be immediately accessed as a set of in simple .CSV files organized by country.

Terms of use:

We are happy to make this corpus available to researchers, who would like to work with the data. For legal reasons, we cannot however publish the data online or share it with researchers outside of UZH. This is due to the copyright and licensing restrictions enforced by news publishing houses on LexisNexis and their clients. As a consequence, the data need to remain located on UZH servers and interested researchers need to be physically present on UZH premises to work with the raw (text) data.

Researchers interested in working with the data are invited to get in touch with the project PI, Prof. Stefanie Walter, to arrange the data access (walter@ipz.uzh.ch). In the process, it will be necessary to sign a consent form, which specifies policies concerning data access, data handling, and non-disclosure.

References:

- Hooghe, Liesbet, Gary Marks, Tobias Lenz, Jeanine Bezuijen, Besir Ceka, Svet Derderyan. 2017. *Measuring International Authority*. Oxford: OUP.
- Hooghe, Liesbet, Tobias Lenz, Gary Marks. 2019. *A Theory of International Organization*. Oxford: OUP.
- Hunter, Tom and Stefanie Walter (2023). "The Politicization of International Organizations in Parliamentary Debates" *Working paper*.