

Bundesbank Speeches Data Report 2019-12

Data available from 2012 to 2017

Metadata ID: bbk_speeches_data_doc_v1.0

Deutsche Bundesbank, Research Data and Service Centre

Karim El-Ouaghli

Matthias Gomolka

Jens Orben

Abstract

We describe the research dataset “Bundesbank Speeches” using a structured metadata schema.¹⁾ This document consists of three sections. In the first section, we describe the general properties of the dataset as a whole, such as its scope, coverage, and the methods of data collection and data appraisal. The second section addresses the variable level. The last section provides information on the meaning of variables and terms widely used in the context of `bbk_speeches`.

Keywords: speeches, text mining, board members, central bank

Metadata ID: `bbk_speeches_data_doc_v1.0`

Citation: El-Ouaghli, K., Gomolka, M., and J. Orben (2019). Bundesbank Speeches, Data Report 2019-12 – Metadata ID `bbk_speeches_data_doc_v1.0`. Deutsche Bundesbank, Research Data and Service Centre.

¹ The metadata scheme is derived from the “Data Documentation Initiative” (DDI, <http://www.ddialliance.org>).

Contents

1 Dataset description	4
1.1 Dataset scope and coverage	4
1.2 Data collection	5
1.3 Data appraisal	5
1.4 Data accessibility	6
2 Description of variables	7
2.1 Overview of variables	7
2.2 Details of variables	7
References	14
A Codelists	15
A.1 spch_lang	15
A.2 spkr_bbk_role	15
A.3 spkr_nom_pol_body	15
B Validation rules	16

1 Dataset description

Bundesbank Speeches (bbk_speeches) is a text based dataset, entailing the speeches of current and past members of the executive board of the Deutsche Bundesbank beginning in January 2012 until December 2017. It comprises publicly available information from www.bundesbank.de/en/press/speeches, which is cleaned and enhanced by information on the respective speakers that is publicly available as well. The choice of variables is based on availability and gauged value. The latter stems from existing literature on the topic of central bank communication as in Malmendier, Nagel, & Yan (2017).. The dataset is ready for a variety of analyses, including textmining. Different from other Bundesbank's RDSC datasets only publically available information is entailed. It's value lies in experimental use of text based datasets with economic relevance, abstracting of webscraping techniques, which often require further knowledge of htm or xm languages.

The use of machine learning techniques, artificial intelligence, or what has become known as "advanced analytics" has spread across many disciplines. Textmining, as one of the methods in this field, is also present in economic analyses. Hence, Deutsche Bundesbank's Research Data and Service Centre (RDSC) wants to promote and foster this development and provides this text based and ready to analyse dataset.

1.1 Dataset scope and coverage

Origin

The speeches are published via the Bundesbank's webpage. The speeches are almost exclusively written by Bundesbank staff members and edited in collaboration with the respective speaker. The spoken word at the occasion itself may differ from what the script of the speech displays. Furthermore, the individual speakers most often and most likely use gesture, mimic and intonation, possibly delivering nuanced and unanticipated insights. This dataset does not include the spoken word, but the script of the speeches.

Unit of analysis

Speeches are separated into speech sections. Hence, speech section is the unit of analysis. Speeches were held at a specific point in time by a member of the executive board of the Deutsche Bundesbank and were published in a given language. Those speech attributes are used to identify speeches and to link the respective sections.²⁾

Time period covered

January 2012 until December 2017.

² Inacurate information on `spch_datetime` leads to duplicate identifiers.

Geographic coverage

World: The dataset covers all publicly held speeches of Bundesbank board members, no matter where the speech was held.

Universe

The dataset covers all publicly held speeches of Bundesbank board members, concerning a wide range of topics. It must be noted at this point, that speeches held by board members who resigned due to political pressure are not included.

The encompassed speakers in the dataset are assigned to the Deutsche Bundesbank's executive board according to the Bundesbank Act. Naturally the speakers themselves change in the course of time, as members of Deutsche Bundesbank's executive board are assigned for a maximum of eight years according to the Bundesbank Act.

1.2 Data collection

Data collection mode

Download from Bundesbank's homepage and manual enrichment with speaker attributes.

Collection frequency

Speeches by Bundesbank's executive board members are not held in a regular frequency or specific amount. The data is updated on an ad-hoc basis.

1.3 Data appraisal

Quality checks

Besides the institution's internal quality checks before holding the speech, the following quality checks are applied in the construction of the dataset.

- Manual check whether the composition of the executive board has changed.
- Check of the whole dataset against various validation rules (see the Appendix on validation rules)

Note however, that the spoken word at the occasion itself may differ from what the script of the speech displays. Thus, the content of this dataset may differ from the spoken word at the occasion.

Data editing

The pure information on the speeches is enhanced by manually collected speaker attributes.

1.4 Data accessibility

Public use file

The dataset is freely available as a public use file (PUF) and can be downloaded from

Contact

Deutsche Bundesbank, Research Data and Service Centre (RDSC)

E-mail: fdsz-data@bundesbank.de

Homepage: <https://www.bundesbank.de/rdsc>

Citation requirements

For any study or other document which is made available to the public and contains information derived from the provided data, the researcher is obliged to properly cite the data source as:

In addition to that, the DOI for the data employed in the research has to be cited as:

2 Description of variables

2.1 Overview of variables

Name	Label
spch_id	Unique speech identifier
spkr_surname	Speaker's surname
spkr_first_name	Speaker's first name(s)
spch_datetime	Speech's date, time and timezone
spch_title	Speech's title
spch_occasion	Speech's occasion
spch_loc	Speech's venue
spch_lang	Speech's language
spch_sec_num	Speech's section number
spch_sec_title	Speech's section title
spch_sec_body	Speech's section body
spkr_titles	Speaker's title(s)
spkr_gender	Speaker's gender
spkr_bbk_role	Speaker's role in Deutsche Bundesbank's executive board.
spkr_phd	Indication wheather the speaker has a PhD or a Dr
spkr_major	Field of the speaker's studies
spkr_apprenticeship	Indication wheather the speaker completed an apprenticeship successfully, and field of apprenticeship
spkr_phd_univ	City of the university
spkr_nom_pol_body	Nominating political body
spkr_nom_pol_geo	Nominating federal state
spkr_nom_pol_party	Nominating politcal party
spkr_pers_party	Speaker's personal party affiliation
spkr_birthplace	Speaker's place of birth
spkr_birth_year	Speaker's year of birth
spkr_birth_month	Speaker's month of birth
spkr_bbk_start	Date of speaker's first day in office as a member of the executive board
spkr_bbk_end	Date of speaker's last day in office as a member of the executive board
spkr_bbk_dismissed	Indication whether the term as a member of the executive board was ended under special circumstances

2.2 Details of variables

spch_id: Unique speech identifier

Notes

The **speech_id** is a unique identifier for each speech. Its structure is: date, time, speaker, language.

Inacurate information on **spch_datetime** leads to duplicate identifiers in few cases.

...

Available from – to 2012 – 2017
Variable type string

spkr_surname: Speaker's surname

Notes Speaker's sur- / last / family name
Available from – to 2012 – 2017
Source Bundesbank homepage
Variable type string

spkr_first_name: Speaker's first name(s)

Notes Speaker's first / given name. Possibly more than one.
Available from – to 2012 – 2017
Source Bundesbank homepage
Variable type string

spch_datetime: Speech's date, time and timezone

Notes Date and time when speech was held in the ISO format. The date and time show German local time, no matter where the speech was held. In most cases, where 00:00 is stated, the exact time is unknown.
Available from – to 2012 – 2017
Source Bundesbank homepage
Variable type datetime in ISO 8601 format (YYYY-MM-DDThh:mm:ssz)

spch_title: Speech's title

Available from – to 2012 – 2017
Source Bundesbank homepage
Variable type string

spch_occasion: Speech's occasion

Available from – to 2012 – 2017
Source Bundesbank homepage
Variable type string

spch_loc: Speech's venue

Notes Location where speech was held.
Available from – to 2012 – 2017
Source Bundesbank homepage
Variable type string

spch_lang: Speech's language

Notes	Language of the speech's textbody, possibly generated by the translation service, provided as ISO 639-1 code.
Available from – to	2012 – 2017
Source	Bundesbank homepage
Variable type	categorical

spch_sec_num: Speech's section number

Notes	Generally the speeches of members of the executive board of the Deutsche Bundesbank are structured in various sections. According to that structure, there are section numbers (spch_sec_num), section titles (spch_sec_title), and the according textbodies (spch_sec_body). <i>Note:</i> This means that a single speech is spread across as many rows as there are sections in the respective speech.
Available from – to	2012 – 2017
Source	Bundesbank homepage
Variable type	integer

spch_sec_title: Speech's section title

Notes	Generally the speeches of members of the executive board of the Deutsche Bundesbank are structured in various sections. According to that structure, there are section numbers (spch_sec_num), section titles (spch_sec_title), and the according textbodies (spch_sec_body). <i>Note:</i> This means that a single speech is spread across as many rows as there are sections in the respective speech.
Available from – to	2012 – 2017
Source	Bundesbank homepage
Variable type	string

spch_sec_body: Speech's section body

Notes	Generally the speeches of members of the executive board of the Deutsche Bundesbank are structured in various sections. According to that structure, there are section numbers (spch_sec_num), section titles (spch_sec_title), and the according textbodies (spch_sec_body). <i>Note:</i> This means that a single speech is spread across as many rows as there are sections in the respective speech.
Available from – to	2012 – 2017
Source	Bundesbank homepage
Variable type	string

...

spkr_titles: Speaker's title(s)

Notes	Speaker's title(s) accompanying the name, including Prof., Dr., Dr. h.c.
Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	string

spkr_gender: Speaker's gender

Notes	Female or male.
Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	categorical

spkr_bbk_role: Speaker's role in Deutsche Bundesbank's executive board.

Notes	President, vice-president or member or combinations of those. <i>Note:</i> Speakers may have worked in different functions in the Bundesbank's board. Then, spkr_bbk_role lists all functions separated by commas.
Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	categorical

spkr_phd: Indication wheather the speaker has a PhD or a Dr

Notes	Dummy variable; TRUE if the speaker has a PhD or doctorate, FALSE otherwise.
Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	boolean

spkr_major: Field of the speaker's studies

Notes	Categorical variable.
Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	categorical

spkr_apprenticeship: Indication wheather the speaker completed an apprenticeship successfully, and field of apprenticeship

Notes	Dummy variable; TRUE if the speaker has completed an apprenticeship successfully, FALSE otherwise.
-------	--

...

Available from – to 2012 – 2017
Source various, internet, publicly available
Variable type boolean

spkr_phd_univ: City of the university

Notes City of the university, where the speaker has done her/his PhD, if any.
Available from – to 2012 – 2017
Source various, internet, publicly available
Variable type string

spkr_nom_pol_body: Nominating political body

Notes The nominating political body is either the German federal government (Bundesregierung) – indicated as *Government* – or the German federal council (Bundesrat) – indicated as *Council*. Presidents, Vice-Presidents and one member (two members before 2007) are generally nominated by the federal government, whereas the remaining half is nominated by the federal council.
Available from – to 2012 – 2017
Source various, internet, publicly available
Variable type categorical

spkr_nom_pol_geo: Nominating federal state

Notes If the nominating political body (**spkr_nom_pol_body**) is the German federal council (*Council*), then the nomination rotates between groups of federal states.
Available from – to 2012 – 2017
Source various, internet, publicly available
Variable type string

spkr_nom_pol_party: Nominating political party

Notes Political party/parties in charge for nominating the respective speaker for the job at the Executive Board of the Bundesbank at the federal state or country level.
Available from – to 2012 – 2017
Source various, internet, publicly available
Variable type string

...

...

spkr_pers_party: Speaker's personal party affiliation

Notes	Categorical variable with speaker's personal party affiliation, if known. Else NA.
Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	categorical

spkr_birthplace: Speaker's place of birth

Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	string

spkr_birth_year: Speaker's year of birth

Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	integer (YYYY)

spkr_birth_month: Speaker's month of birth

Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	integer (MM)

spkr_bbk_start: Date of speaker's first day in office as a member of the executive board

Notes	Date of speaker's first day in office as a member of the executive board of the Deutsche Bundesbank.
Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	date in ISO 8601 format (YYYY-MM-DD)

spkr_bbk_end: Date of speaker's last day in office as a member of the executive board

Notes	Date of speaker's last day in office as a member of the executive board of the Deutsche Bundesbank.
Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	date in ISO 8601 format (YYYY-MM-DD)

spkr_bbk_dismissed: Indication whether the term as a member of the executive board was ended under special circumstances

Notes	Dummy variable; TRUE if speaker was dishonourably discharged from the executive board of the Deutsche Bundesbank, FALSE otherwise.
Available from – to	2012 – 2017
Source	various, internet, publicly available
Variable type	boolean

References

Malmendier, U., Nagel, S., & Yan, Z. (2017). *The making of hawks and doves: Inflation experiences on the fomc* (Working Paper No. 23228). Working paper series. National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w23228>

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

van der Loo, M., & de Jonge, E. (2018). *Validate: Data validation infrastructure*. Retrieved from <https://CRAN.R-project.org/package=validate>

A Codelists

A.1 spch_lang

Code	Description
de	German
en	Englisch
fr	French
it	Italian

A.2 spkr_bbk_role

Code	Description
member	Speaker is member of the Bundesbank's board of executives
vice-president	Speaker is vice-president of the Bundesbank's board of executives
president, vice-president	Speaker is either president or vice-president of the Bundesbank's board of executives
president	Speaker is president of the Bundesbank's board of executives

A.3 spkr_nom_pol_body

Code	Description
Council	Speaker was nominated by the German Federal Council
Government	Speaker was nominated by the German Federal Government

B Validation rules

The validation rules were formulated such that they can be applied using R (R Core Team, 2018) and the validate package (van der Loo & de Jonge, 2018).

No missing values

Description	The applicable variables must not contain missing values.
Formal definition	<code>!is.na(variable)</code>
Applicable variables	<code>spch_id, spch_sec_body, spkr_surname, spkr_first_name, spch_datetime, spch_title, spkr_bbk_start, spkr_bbk_role, spch_loc, spch_occasion, spkr_phd, spkr_apprenticeship, spkr_nom_pol_body, spkr_pers_party, spkr_birthplace, spkr_birth_year, spkr_birth_month</code>

No missing section title

Description	If the section number is greater than one, there must be a section title.
Formal definition	<code>if (spch_sec_num != 1) !is.na(spch_sec_title)</code>

Start prior to end

Description	Membership at the Bundesbank's executive board must start before it ends.
Formal definition	<code>if (!is.na(spkr_bbk_end)) spkr_bbk_end > spkr_bbk_start</code>

No missing values if speaker holds a PhD

Description	If the speaker holds a PhD, the applicable variables must not contain missing values.
Formal definition	<code>if (spkr_phd == TRUE) !is.na(variable)</code>
Applicable variables	<code>spkr_major, spkr_phd_univ, spkr_titles</code>

No missing values on state who nominated speaker in council

Description	If the speaker was nominated by the Federal Council, the information on the respective federal state must not contain missing values.
Formal definition	<code>if(spkr_nom_pol_body == "Council") !is.na(spkr_nom_pol_geo)</code>

BBk start date must be prior to date of speech

Description	The speech must be held after the speaker became a member of the Bundesbank's executive board.
-------------	--

Formal definition `spch_datetime > spkr_bbk_start`

Speech date must be prior to BBk end date

Description The speech should be held before the speaker left the Bundesbank's executive board. This is a weak rule since farewell speeches are included in the data.

Formal definition `if (!is.na(spkr_bbk_end)) spch_datetime < spkr_bbk_end`

Section numbers ascending

Description Section numbers must be ascending within `spch_id`.

Formal definition `if(spch_id == lag(spch_id, 1)) (spch_sec_num == lag(spch_sec_num, 1) + 1)`

First section number one

Description Section numbering has to start with one.

Formal definition `if(spch_id != lag(spch_id, 1)) spch_sec_num == 1`

Section body not allowed to start with number or [EMBEDDED]

Description This rules out some artefacts in the source data.

Formal definition `!grep1("^\\d | ^\\[EMBEDDED]", spch_sec_body)`

No variation within `spch_id`

Description The applicable variables must be constant within a given speech id.

Formal definition `spch_id variable`

Applicable variables `spkr_surname, spkr_first_name, spch_datetime, spch_title, spch_occasion, spch_loc, spkr_titles, spkr_gender, spkr_bbk_role, spkr_phd, spkr_major, spkr_phd_univ, spkr_apprenticeship, spkr_nom_pol_body, spkr_nom_pol_geo, spkr_nom_pol_party, spkr_pers_party, spkr_birthplace, spkr_birth_year, spkr_birth_month, spkr_bbk_start, spkr_bbk_end, spkr_bbk_dismissed, spch_lang`

Allowed languages

Description No other languages than German, English, French, and Italian are allowed.

Formal definition `spch_lang %in% c("de", "en", "fr", "it")`

No other genders than male & female

Description	Until now, speakers in the data must be either male or female.
Formal definition	<code>spkr_gender %in% c("male", "female")</code>
