Who Speaks About Childbirth? A Corpus-Driven Study of Parliamentary Discourse in Poland

Jolanta Bachan, Agnieszka Jeran

Adam Mickiewicz University, Poznań, Poland
jbachan@amu.edu.pl, jeran@amu.edu.pl

Abstract

This work presents a corpus-driven investigation into the discursive construction of childbirth in Polish parliamentary debates. The study utilizes two major corpora: the Polish Parliamentary Corpus (Ogrodniczuk, 2018) and ParlaMint 5.0 (Erjavec et al., 2024), which together offer both national and comparative European perspectives on political discourse. The research is grounded in corpus linguistics, focusing on quantitative dimension. Keywords and collocational patterns related to childbirth (e.g. "childbirth", "maternity", "midwife", "perinatal care") were extracted to detect recurring linguistic structures. The findings are relevant not only to scholars of discourse and corpus linguistics, but also to those interested in political communication, gender studies, and language in institutional settings. Furthermore, the study underscores the value of open-access parliamentary corpora in facilitating linguistically informed contributions to public policy debates.

Keywords: parliamentary debate, text corpus analysis, Polish Parliamentary Corpus, PCC, KDP, ParlaMint, childbirth, perinatal care

1. Introduction

The present paper focuses of the topic of childbird and the discussion about it in the Polish parliament. For the study two major corpora were used: the Polish Parliamentary Corpus (Ogrodniczuk, 2018) and ParlaMint 5.0 (Erjavec et al., 2024), which together offer both national and comparative European perspectives on political discourse.

The Polish Parliamentary Corpus (PPC) is a large, linguistically annotated collection of transcripts from sessions of the Polish Parliament (from lower and higher chambers, Polish Sejm and Senate), including plenary sittings, committee meetings, interpellations and inquiries. It spans from 1919 to the present and is continuously updated with new parliamentary material. The texts in the corpus are enriched with metadata and processed automatically — they are tokenised and annotated for morphosyntax, lemmatisation, syntactic structure and named entities. In May 2025 the corpus consisted of 879 million segments with detailed information about the origin and periods of creation[1].

The ParlaMint-PL 5.0 is the Polish component of the broader ParlaMint project – a multilingual, comparable corpus of parliamentary debates from 29 European countries and autonomous regions (Erjavec, 2025). The Polish parliament corpus consists of data from the period of time 2015-2022 and contains 43 million tokens (including 35 million words)[2].

2. Corpus search

The Polish Parliamentary Corpus has search engine which enables to look for keywords and phrases using a query language compatibile with Korpusomat.pl (Kieraś and Kobyliński, 2021). Table 1 shows the results of examplar queries for the keywords: childbirth, maternity, midwife, perinatal care. The choice of keywords was based on the belief that perinatal issues are important in public discourse, not only in parliament, especially in light of such disturbing phenomena as perinatal violence (cf. Nowakowska-Kutra and Jeran, 2024a).

Keywords	Query	Result
poród (childbirth)	poród	550
poród (childbirth)	[lemma = "poród"]	6408
macierzyństwo (maternity)	[lemma = "macierzyństwo"]	45774
położna (midwife)	[base = "położna"]	25127
opieka okołoporodowa (perinatal care)	[orth = "opieka"][orth = "okołoporodowa"]	24
okołoporodowy	[base="okołoporodowy"]	18
opieka (perinatal care)	[base="opieka"]	
adjective+poród (childbirth)	[pos="adj"][base="poród"]	1254

Table 1. Keyword search in PPC

Fig. 1 shows a concordance in PPC of the query *adjective+poród (childbirth)* [pos="adj"][base="poród"]. 1,254 results were found. The word "poród" appears in different declination forms in KWIC (key-word in context): "porodu" (Genitive), "porodzie" (Locative).

The result of PPC contains the following metadata: Left context, Result, Right context, Interpretation, Proper name, Dependency relation, Superordinate lemma, Chamber, Type of interaction, Author, Speaker, Speaker affiliation, Speaker role, Speaker gender, Committee, Date, Timestamp, Year, Political system, Number of the term of office, Label, Title, Session, Day, Number, Part. Some fields are empty, so for the following analysis only the selected characteristics were taken into account.

[1]The Polish Parliamentary Corpus / Korpus Dyskursu Parlamentarnego. Retrieved from: https://clip.ipipan.waw.pl/PPC (last access date: 2025-10-20)

[2]ParlaMint-PL 5.0 (Polish parliament). Retrieved from: https://www.clarin.si/ske/#dashboard?corpname=parlamint50_pl&corp_info=1 (last access date: 2025-10-20)

Jolanta Bachan, Agnieszka Jeran, *Who Speaks About Childbirth? A Corpus-Driven Study of Parliamentary Discourse in Poland*. In: Z. Vetulani, P. Paroubek, I. Kardava (Eds.), Human Language Technologies as a Challenge for Computer Science and Linguistics – 2025, pp. 7–11. Poznań: Adam Mickiewicz University Press, 2025.
© Authors, 2025. https://doi.org/10.14746/amup.9788323245414.1. Open Access chapter, distributed under the terms of the CC licence (BY-NC-ND, https://creativecommons.org/licenses/by-nc-nd/4.0/).

Znaleziono **1 254** wyniki.

Lp	Mówca	Lewy kontekst	Rezultat	Prawy kontekst	Data
1	Przedstawiciel Wydziału Polityki Społecznej Organizacji Współpracy Gospodarczej i Rozwoju	matki w szpitalu jest niesamowicie kosztowny. Dodatkowo w przypadku	**normalnego porodu**	nie ma dowodów, że jest to korzystne dla matki	2010-01-20
2	Przedstawiciel Wydziału Polityki Społecznej Organizacji Współpracy Gospodarczej i Rozwoju	medykalizacją systemu to liczba dni w szpitalu po standardowym,	**normalnym porodzie**	. Jeżeli chodzi o pobyt w szpitalu, Polska plasuje	2010-01-20
3	Jerzy Ziętek	by skin jest szalenie ważne do tworzenia więzów macierzyńskich od	**samego porodu**	. Są inne sposoby przeprowadzania porodu. Od wielu lat	2010-01-20
4	Eugenia Gienicka	zostały zagwarantowane w budżecie środki na dwa tygodnie – przy	**pierwszym porodzie**	z 16 do 18, a przy drugim z 18	2006-04-05

Fig. 1: Concordance for the query adjective+childbirth, 1254 results. The columns form the left: Number, Speaker, Left context, KWIC, Right context, Date.

Fig. 2 shows a concordance for the query *childbirth* in ParlaMint-PL. The KWIC "poród" is marked in bold and appears in different declination forms in the Polish text. 254 results were found.

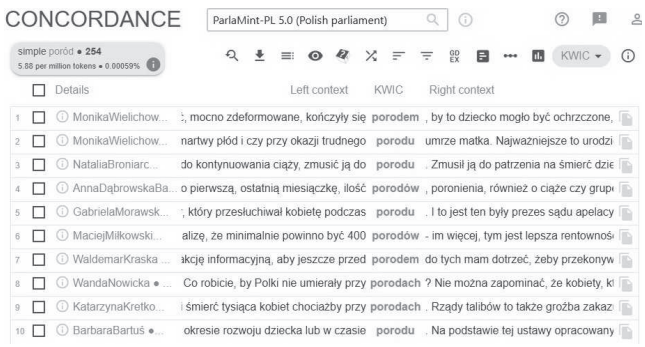

Fig. 2: Concordance for the query childbirth, 254 results.

3. Data characteristics

This section presents the search results of the words "childbirth" and "perinatal" in PPC and the word "perinatal" in ParlaMint-PL.

3.1. Selected variables and deficiences: PPC

Analysis of the data collected in the PPC reveals several important characteristics and deficiencies in the metadata that influence how the corpus can be interpreted. For the entry "childbirth", a total of 6,210 occurrences were identified, but as many as 2,474 of them (39.8%) lacked information on the date. Only after introducing analysis by parliamentary term was this gap reduced – within this framework, only 34 entries (0.5%) were missing, including 19 that required correction from term 11 to term 10. The oldest occurrence related to this entry dates back to 1919. A similar issue appeared for the speaker's gender: 3,209 entries (51.7%) did not include this information. After completing the data based on first names, the number of missing cases was reduced to just 123 (1.98%).

Comparable challenges emerge in the analysis of the term "perinatal". The PPC contains a total of 1,593 such occurrences. Initially, 510 of them (32%) lacked date information, but when grouped by parliamentary term, the number of missing cases fell to 34 (0.5%), with one entry requiring a correction from term 11 to term 10. The oldest record for this category dates back to 1961. Metadata deficiencies also concerned the speaker's gender—this information was missing for 999 entries (62.7%). After completing the data based on first names, the number of unclassified cases decreased to only 19 (1.19%).

These findings demonstrate that although the PPC is a valuable resource for studying the evolution of parliamentary discourse, its effective use requires substantial work on metadata completeness. Filling gaps in date information and speaker gender significantly enhances the quality of analyses and allows for more precise conclusions about the dynamics of parliamentary debate.

Table 2 shows the occurences of the words "childbirth" and "perinatal" in data characterised by political system (Second Republic of Poland – II RP, Third Republic of Poland – III RP, Polish People's Republic - PPR), parliamentaty chamber (lower – Sejm, higher – Senate), speaker gender and type of interaction in PPC.

	childbirth		perinatal	
Political system	Qt	%	Qt	%
II RP	78	1.3	0	0
III RP	5900	95.0	1581	99.2
PPR	211	3.4	11	0.7
PPR/III RP	21	0.3	1	0.1
Total	6210	100.0	1593	100.0

Parliamentary chamber	Qt	%	Qt	%
Sejm	5708	91.9	1534	96.3
Senate	502	8.1	59	3.7
Total	6210	100.0	1593	100.0

Gender	Qt	%	Qt	%
Female	1963	31.6	441	27.7
Male	4105	66.1	1126	70.7
Mixed*	19	0.3	7	0.4
Unknown	123	2.0	19	1.2
Total	6210	100.0	1593	100.0
*several people of different genders are listed as authors				

Type of interaction	Qt	%	Qt	%
Interpellation	613	9.9	194	12.2
Committee	1413	22.8	325	20.4
Response to interpellation	2223	35.8	750	47.1
Response to inquiry	250	4.0	38	2.4
Plenary session	1588	25.6	269	16.9
Inquiry	123	2.0	17	1.1
Total	6210	100.0	1593	100.0

Table 2. Occurences of "childbirth" and "perinatal" in PPC, Qt - quantity

3.2. Selected variables and deficiences: ParlaMint

In the ParlaMint corpus, the term "perinatal" appears 254 times in the years 2016–2022 (Table 3). The dataset contains only a limited set of metadata, as it includes

essentially the text of each speech along with information about the date and the speaker. Consequently, variables such as the year of occurrence and the gender of the speaker had to be inferred on the basis of the available metadata.

Year	Quantity	Percent
2016	71	28.0
2017	41	16.1
2018	26	10.2
2019	26	10.2
2020	37	14.6
2021	42	16.5
2022	11	4.3
Total	254	100.0

Gender	Quantity	Percent
Female	141	55.5
Male	113	44.5
Total	254	100.0

Table 3. Occurrences of "childbirth" in ParlaMin-PL

4. Analysis of data from the Third Republic of Poland period

This section focuses on the analysis of parliamentary material on perinatal care in the period of the Third Republic of Poland extracted from PPC, excluding all materials originating from the Polish People's Republic.

4.1. Distributions by term of office, speaker gender, type of interaction

Table 4 shows the distribution of references in PPC to the words "childbirth" and "perinatal" across parliamentary terms, speaker gender, and type of interaction, which reveals several noteworthy patterns. The frequency of occurrences of the word "childbirth" is highly uneven across terms of office, with the 6th, 7th, and especially the 8th term standing out. The marked increase in the 8th term aligns with a key policy event: in 2018, during this term, Minister of Health Ewa Kopacz signed the Regulation on Perinatal Care Standards, which may have intensified parliamentary attention to the issue. Gender distribution also shows relevant disparities. Although the 10th term of office features the highest proportion of female MPs in the history of the Polish higher chamber of the parliament (Sejm) – approximately 30% (rising from 29.35%, 135 female MPs, immediately after the elections, to about 31.3%, 144, in April 2025) – the speakers represented in the corpus are not limited to female MPs. Furthermore, although utterances containing "perinatal" are considerably fewer than those referring to "childbirth", they nevertheless exhibit a visible increase during the 8th term, suggesting growing institutional engagement with perinatal care issues.

4.2. Analysis by chamber of parliament based on gender, term of office, type: "childbirth"

Analysis of parliamentary discourse on childbirth reveals limited variation by speaker gender, with distributions in both chambers remaining comparable, even when restricting the analysis to women or men only (Table 5).

Term of office		childbirth		perinatal	
		Qt	%	Qt	%
I	1991-1993	38	0.6	4	0.3
II	1993-1997	173	2.9	27	1.7
III	1997-2001	540	9.2	54	3.4
IV	2001-2005	491	8.3	79	5.0
V	2005-2007	399	6.8	90	5.7
VI	2007-2011	1153	19.5	284	18.0
VII	2011-2015	1439	24.4	299	18.9
VIII	2015-2019	999	16.9	564	35.7
IX	2019-2023	318	5.4	115	7.3
X	2023-now	350	5.9	65	4.1
Total		5900	100.0	1581	100.0

Gender	Qt	%	Qt	%
Female	1852	31.4	437	27.6
Male	3925	66.5	1121	70.9
Mixed*	19	0.3	0	0
Unknown	104	1.8	23	1.5
Total	5900	100.0	1581	100.0
*several people of different genders are listed as authors				

Type of interaction	Qt	%	Qt	%
Interpellation	613	10.4	194	12.3
Committee	1305	22.1	317	20.1
Response to interpellation	2223	37.7	750	47.4
Response to inquiry	250	4.2	38	2.4
Plenary session	1386	23.5	265	16.8
Inquiry	123	2.1	17	1.1
Total	5900	100.0	1581	100.0

Table 4. Distribution of "childbirth" and "perinatal" in PPC

By contrast, substantial differences emerge across terms of office (Cramer's $V = 0.408$, $p < 0.001$). These disparities are driven primarily by the 4th (2001–2005; 22.2%), 9th (2019–2023; 38.1%), and 10th (current; 29.1%) terms, during which the proportion of childbirth-related statements in the Senate markedly exceeded the overall average, despite Senate speeches constituting only 7.6% of all statements (Fig. 3). Variation is also evident across types of parliamentary interaction (Cramer's $V = 0.35$, $p < 0.001$). Because several interaction types identified in the Sejm do not occur in the Senate, childbirth-related interventions in the higher chamber originate exclusively from committee meetings and plenary debates (Table 6).

Chamber	Qt	Gender of the speaker				Total
		Unknown	Female	Male	Mixed	
Sejm	Qt	103	1704	3625	19	5451
	%	1.9	31.3	66.5	0.3	100.0
Senate	Qt	1	148	300	0	449
	%	0.2	33.0	66.8	0.0	100.0

Table 5. Distribution of "childbirth" in PPC across genders

Type of interaction		Sejm	Senate	Total
Interpellation	Qt	613	0	613
	%	11.2	0.0	10.4
Committee	Qt	1169	136	1305
	%	21.4	30.3	22.1
Response to interpellation	Qt	2223	0	2223
	%	40.8	0.0	37.7
Response to inquiry	Qt	250	0	250
	%	4.6	0.0	4.2
Plenary session	Qt	1073	313	1386
	%	19.7	69.7	23.5
Inquiry	Qt	123	0	123
	%	2.3	0.0	2.1
Total	Qt	5451	449	5900
	%	100.0	100.0	100.0

Table 6. Distribution of "childbirth" in PPC across types of interaction

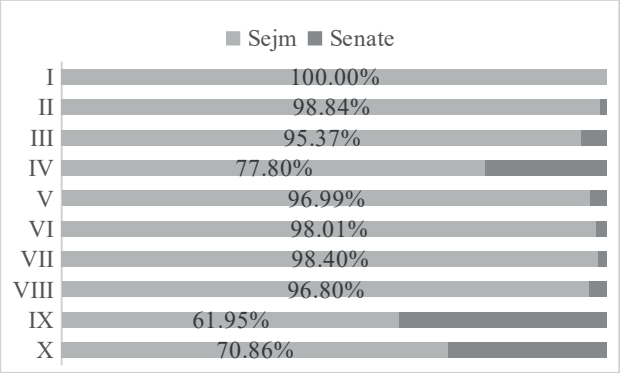

Fig. 3: Distribution of "childbirth" in PPC across terms of office

4.3. Analysis by term of office based on gender: "childbirth"

Across successive parliamentary terms, discourse on childbirth exhibited notable variation linked to both speaker gender and the type of parliamentary speech. Variation by gender is confirmed by the distributional differences observed across terms (Cramer's V = 0.149, $p < 0.001$). These differences stem primarily from the 8th (2015–2019; 35.08%), 9th (2019–2023; 46.54%), and 10th (current; 50.86%) terms, during which the proportion of women's statements substantially exceeded the overall average of 32.1% (and 28.1% between the 1st and 7th terms) (Fig. 4).

4.4. Analysis by speaker gender (within the female/male constraint) based on type on interaction: "childbirth"

The analysis indicates that the topic of childbirth, when raised by speakers of different genders, exhibits notable variation across parliamentary speech types (Cramer's V = 0.17, $p < 0.001$). Among male speakers, the distribution is clearly dominated by one genre: responses to interpellations constitute more than 43% of all childbirth-related speeches, followed at a considerable distance by responses delivered during plenary sessions and those presented in committee meetings. In contrast, women's contributions show a more even distribution across speech types. For female speakers, the three most frequent forms – responses to interpellations, speeches delivered during plenary sessions, and responses in committee – appear in comparable proportions. Moreover, women's speeches feature a higher overall share of interpellations compared with those of men (Table 7).

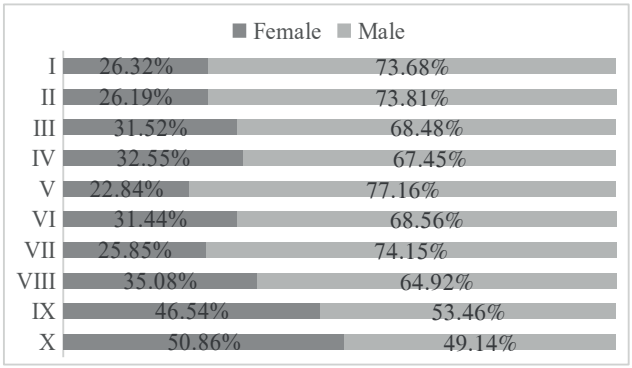

Fig. 4: Distribution of "childbirth" in PCC based on gender of speaker across terms of office

Type of interaction		Female	Male	Total
Interpellation	Qt	224	334	558
	%	12.1	8.5	9.7
Committee	Qt	495	800	1295
	%	26.7	20.4	22.4
Response to interpellation	Qt	502	1694	2196
	%	27.1	43.2	38.0
Response to inquiry	Qt	65	185	250
	%	3.5	4.7	4.3
Plenary session	Qt	545	839	1384
	%	29.4	21.4	24.0
Inquiry	Qt	21	73	94
	%	1.1	1.9	1.6
Total	Qt	1852	3925	5777
	%	100.0	100.0	100.0

Table 7. Distribution of "childbirth" in PPC based on speaker gender across types of interaction

4.5. Analysis by chamber of parliament based on gender, term of office, type of interaction: "perinatal"

The discussion of perinatal issues within parliamentary chambers reveals several dimensions of variation. With respect to speaker gender, the distribution of utterances is broadly comparable across both chambers, indicating no substantial gender-based differentiation in this regard (Table 8). In contrast, clear variation emerges across terms of office (Cramer's V = 0.348, $p < 0.001$). This pattern is primarily driven by the 3rd (1997–2001; 7.4%), 4th (2001–2005; 7.6%), 9th (2019–2023; 20%), and 10th (current; 23.1%) terms, during which the proportion of Senate statements was markedly higher than the overall average. Although Senate contributions constituted only 3.7% of all perinatal-related statements, in many terms these were limited to single interventions, making the elevated shares

in the aforementioned terms particularly notable (Fig 5). Variation is also evident across types of parliamentary statements (Cramer's V = 0.266, p < 0.001), largely due to specific interaction types that did not occur in Senate debates. As a result, Senate references to perinatal issues originate exclusively from committee meetings and plenary sessions (Table 9).

Chamber	Qt	Female	Male	Total
Sejm	Qt	420	1080	1500
	%	28.0	72.0	100.0
Senate	Qt	17	41	58
	%	29.3	70.7	100.0
Total	Qt	437	1121	1558
	%	28.0	72.0	100.0

Table 8. Distribution of "perinatal" in PPC across genders

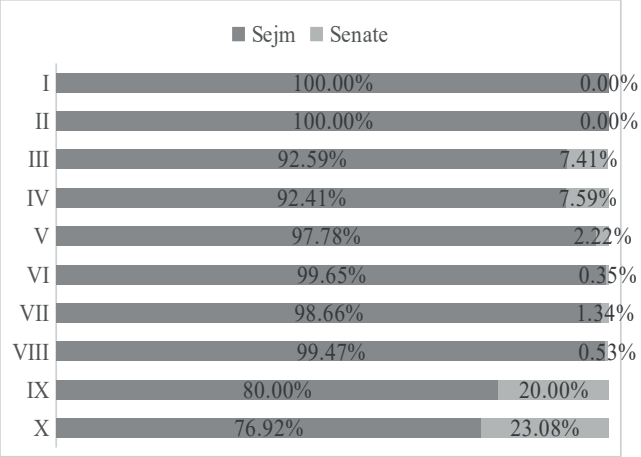

Fig. 5: Distribution of "perinatal" in PPC across terms of office

Type of interaction	Qt	Sejm	Senate	Total
Interpellation	Qt	194	0	194
	%	12.7	0.0	12.3
Committee	Qt	292	25	317
	%	19.2	43.1	20.1
Response to interpellation	Qt	750	0	750
	%	49.2	0.0	47.4
Response to inquiry	Qt	38	0	38
	%	2.5	0.0	2.4
Plenary session	Qt	232	33	265
	%	15.2	56.9	16.8
Inquiry	Qt	17	0	17
	%	1.1	0.0	1.1
Total	Qt	1523	58	1581
	%	100.0	100.0	100.0

Table 9. Distribution of "perinatal" in PPC across types of interaction

5. Conclusions

The analysis shows that women's use of the terms "childbirth" and "perinatal" constitutes approximately 30% of all relevant parliamentary data, and these terms appear across a wide range of interacction types. The findings indicate that, particularly with resources such as the PPC, further research should combine quantitative approaches (examining who speaks, when, and in what institutional role) with qualitative analyses focused on content and thematic patterns. Dynamic analyses would be especially valuable if the dataset were completed with precise dates or years of speeches, enabling more accurate correlations between parliamentary discourse and external sociopolitical events and with the results of other studies on perinatal situations (Nowakowska-Kutra and Jeran, 2024b). At the same time, the data reveal several limitations, including incomplete speaker metadata (such as party affiliation, functions, and often gender), missing temporal information, and occasional technical errors, which may constrain the depth and reliability of future studies.

References

Computational Linguistics in Poland. The Polish Parliamentary Corpus / Korpus Dyskursu Parlamentarnego. (2025) Retrieved from: https://clip.ipipan.waw.pl/PPC (last access date: 2025-10-20)

Erjavec, T., Kopp, M., Ljubešić, N., Kuzman T., Rayson P., Osenova P., Ogrodniczuk M., Çöltekin Ç., Koržinek D., Meden K., Skubic J., Rupnik P., Agnoloni T., Aires J., Barkarson S., Bartolini R., Bel N., Calzada Pérez M., Dargis R., Diwersy S., Gavriilidou M., van Heusden R., Iruskieta M., Kahusk N., Kryvenko A., Ligeti-Nagy N., Magariños C., Mölder M., Navarretta C., Simov K., Magne Tungland L., Tuominen J., Vidler J., Vladu A. I., Wissik T., Yrjänäinen V. and Fišer D. (2024). ParlaMint II: Advancing comparable parliamentary corpora across Europe. *Language Resources and Evaluation.* https://doi.org/10.1007/s10579-024-09798-w

Erjavec, T. (2025). ParlaMint: Comparable Parliamentary Corpora. Retrieved from: https://github.com/clarin-eric/ParlaMint (last access date: 2025-11-04)

Kieraś, W., and Kobyliński, Ł. (2021). Korpusomat – stan obecny i przyszłość projektu. *Język Polski*, (2), 49–58. https://doi.org/10.31286/JP.101.2.4

Nowakowska-Kutra, A. and Jeran, A. (2024a). Opowieści kobiet o przemocy lekarskiej i położniczej – wstępne ustalenia na podstawie analizy wypowiedzi respondentek badania „Głos Matek" z lat 2020-2021, *Przegląd Socjologiczny*, 2024, vol. 73, nr 3, s.99-120. DOI:10.26485/ps/2024/73.3/5

Nowakowska-Kutra, A. and Jeran, A. (2024b). Werbalna przemoc położnicza i lekarska wobec kobiet rodzących w Polsce, *Kultura i Społeczeństwo*, 2024, vol. 68, nr 4, s.185-203. DOI:10.35757/kis.2024.68.4.9

Ogrodniczuk, M. (2018). Polish Parliamentary Corpus. [in:] Darja Fišer, Maria Eskevich, Franciska de Jong (red.) *Proceedings of the LREC 2018 Workshop ParlaCLARIN: Creating and Using Parliamentary Corpora*, s. 15–19, Paris, European Language Resources Association (ELRA).