

These women's verbs

A combined corpus and discourse analysis on
reporting verbs about women and men in
Czech media 1989–2015

Irene Elmerot

Institutionen för slaviska och baltiska språk, finska, nederländska och tyska

Examensarbete för magisterexamen 15 hp / Magister thesis 15 HE credits

Tjeckiska / Czech

Avancerad nivå / Advanced level

Höstterminen 2017 / Autumn 2017

Handledare/Supervisor: Tora Hedin

Examinator: Ludmila Pöppel



Stockholms
universitet

These women's verbs

A combined corpus and discourse analysis on reporting verbs about women and men in Czech media 1989–2015

Irene Elmerot

Abstract

This study aims to analyse how women and men in five different professions are portrayed and represented through reporting verbs in Czech media over a period of 25 years (end of 1989 to the beginning of 2015). The empirical data consist of entire newspapers and magazines in the source material, a subcorpus from the Czech National Corpus. The theoretical basis is Critical discourse analysis and the methodical basis is corpus-based statistical analysis. Binary categories from the Harvard Psychosociological Dictionary are used to classify the reporting verbs. After a quantitative study, the results are clear for some professions and less clear for others; these results are analysed.

This study could not (at least not without severe adjustments) have been performed in languages like English, where the distinction between the female and male professional concepts is less clear. In the chapter on previous research, special attention is given to the Czech context. That chapter also explains this study's contribution to previous research in language, power and corpus studies.

Nyckelord

Kritisk diskursanalys, korpuslingvistik, mediaspråk, tjeckiska, anföringsverb, könsdifferentiering

Keywords

Critical discourse analysis, corpus linguistics, media language, Czech, reporting verbs, gender differentiation

1. Introduction	1
2. Aim and focus	2
3. Theories	2
4. Previous research	3
4.1 Critical discourse analysis	3
4.1.1 Gender, language and power	4
4.2 Research on reporting verbs	4
4.3 Corpus-based discourse studies	5
4.3.1 Corpus-based discourse studies on reporting verbs	6
4.4. Gender discourse in the Czech Republic	7
5. Question and hypotheses	8
7. Material	9
8. Method	10
8.1 Analyses and work steps	11
8.1.1 The verbs	12
8.1.2 Professional denominations	14
9. Results	16
10 Conclusions	21
11 References	24

1. Introduction

Much previous research has been done on denominations, appellations and representations of women in society. For this study, a modern linguistic perspective on this issue is used. The present paper provides a case study of how five different professional denominations, in co-occurrence with reporting verbs used in media, may unveil social processes and, as Wodak mentions (1989, xiv), make otherwise unnoticed linguistic structures and systems visible. This is a theoretical framework called Critical Discourse Analysis, often abbreviated as CDA. The professions chosen are Members of Parliament, bosses, clerks, teachers and singers, and the reporting verbs are the 50 most common reporting verbs for the Members of Parliament.

Language is the common tool for everyone working with CDA, but this case is analysed through the linguistic structure of Czech, in which the nouns for different professions are gender-specific. To this is added the corpus linguistic analysis made from the empirical data. The combination of reporting verbs and professions thus form a linguistic structure that is analysed through the filter of corpus-based CDA.

This analysis can be seen as a part of a continued discourse work on linguistic othering conducted since 2015, in which a corpus-based method is used, and where absolute and relative figures from searches in the Czech National Corpus are calculated to give ratios. These figures then lay the basis for a more qualitative discourse analysis of the research question at issue. All parts of this work (Elmerot 2016; 2017 and the present study) are theoretically based on critical discourse analysis, as well as previous research on language and power, and are methodologically based on corpus linguistics.

This study would not (without severe problems or alterations) be feasible in English, nor in other languages where there is no morphological gender distinction for professional concepts

like “singer” or “teacher.” For other languages, like Arabic, French, Polish or Russian, such a study of the professions would have to be limited by the professions for which there is an accepted and widely used linguistic distinction. Czech, however, has a clear morphological distinction for most professional concepts in the standard written usage, and also is a language that has a corpus that is both large enough and available for a study like this.

2. Aim and focus

The aim of this study is to see if there is any visible gender differentiation in the Czech language, with a focus on the kind of reporting verbs that are used in co-occurrence with denominations for professional women and men in Czech media after 1989. To fulfil this aim, corpus linguistics will be used together with critical discourse analysis, enhancing reliability and returning a statistically significant and systematic result.

3. Theories

Apart from the methodological thought that large corpora may lead to a more reliable result, this study is based on critical discourse theories on gender, language and power. This leads to two main statements:

- Female politicians get more negative media coverage than their male counterparts (Gidengil & Everitt 2003, 209).
- Men and women are depicted in news media in proportions that are not representative of their numerical presence (Caldas-Coulthard 1995, 239).

In this study, these two main theories will be used in a corpus-based analysis for the case of the Czech Republic, by means of reporting verbs in Czech printed media about Members of

Parliament, bosses, clerks, teachers and singers. The study is also using the following CDA research as a theoretical framework.

4. Previous research

The previous research considers more theoretical CDA in general and the combination of gender, language and power in particular, as well as more method-based research on reporting verbs. Some corpus-based discourse analysis is also included in the previous research, but the combination of this methodology and gender studies is still considered rather new (Baker 2014, 13).

4.1 Critical discourse analysis

In this paper, a corpus analysis method is used with a critical discourse approach. CDA is based on theories explaining how certain language usage has come to be a matter of course (Stubbs 1997, 3), and especially how power is used and misused in discourse. One CDA aim is to reveal what Norman Fairclough calls “hidden” and Michelle Lazar calls “invisible” power (Fairclough 2015, 41; Lazar 2007, 148): some discourse is not always obvious when browsing, but may turn into a matter of course if it is repeated often enough. When a certain phrase, or a whole discourse for that matter, starts to get repeated, the receivers (listeners, readers etc.) import that phrase and keep it close at hand. One example in English is the phrase “illegal immigrant”, an alliterating, two-word phrase that originally consisted of two separate lexical items, but that we today see as a lexical unit, a matter of course, unless we think critically about it. CDA is also a good starting point for studies on gender in discourse (Sunderland 2004, 11), although the researcher must reflect on the results in light of what is known from other sources about the area in question. One extraordinary definition of CDA was coined by Teun van Dijk: “discourse analysis ‘with an attitude’” (van Dijk 2001, 96). The same author also states (van Dijk 2008, viii–ix) that it is important to study media as well as political, educational and scholarly

discourse in order to pin down the “socially shared” ideas and attitudes that lead to discrimination in society. In this study, the focus is on media language: that and educational material are probably the most widely spread of the types of material that van Dijk mentions.

4.1.1 Gender, language and power

For a linguist, the CDA approach fits very well when the aim is to see how gender is represented through language usage in society. Several gender scholars have concluded that gender studies should and could often overlap with CDA, since that method may give a stable theoretical basis for gender issues (e.g. Lazar 2007, 144; Wodak 2007). Wodak claims (2007, 93) that gender differentiation often is subtle, and we may expect that, when women hold the same positions as men, the differentiation is eradicated. A discourse analysis of a large source material is then a suitable way of making such subtlety clearly visible. However, according to Stubbs (1997, 1), there is no discourse analysis theory that clearly states how language usage might affect what reoccurs in its speakers’ minds. Stubbs also points out the necessity of large and relevant source data (idem 1997, 6). This is why corpus analysis based on a very large corpus has been used to test the theories in this study. This large corpus is the source material, and statistical corpus analysis is used as a method.

4.2 Research on reporting verbs

Reporting verbs have been the focus of studies in many languages. More than 20 years ago, in a study on reporting verbs in Swedish fiction, Martin Gellerstam concluded that when men talked, they were reported to do so “briefly” and “calmly”, whereas women spoke while “smiling” or with “trembling lips” (Gellerstam 1996, 23). A similarly detailed study for non-fiction would be welcome in the future.

The most frequent reporting verb in Indo-European languages is “say”; in Czech, this is the aspectual pair *řít/říkat*.¹ In newspaper text, however, verbs like “tell” are more common than

¹ In this paper, all translations into English are noted for the sake of understanding.

verbs like “ask” (Allén 1971, 146–147; Caldas-Coulthard 1995, 234). This is the case in Czech, in general: According to *A Frequency Dictionary of Contemporary Czech* (Čermák & Křen 2011), both *říct* and *říkat* (“say” or “tell”) come before *zeptat se* (the perfective form of “ask”), which, in turn, comes before the other reporting verbs like *tvrdit* (“claim”). This dictionary is based on spoken and written language, including fiction, non-fiction and newspaper texts. In the more recent *Encyclopaedic Dictionary of Czech*, the reporting verbs are explained as semantic variations of the verb “speak”, *mluvit* (Hirschová 2017). There, the reporting verbs are divided into different categories based on the character of the report:

- a means of communication like *telefonovat* (call on the phone) or the newer *mejlovat*, *chatovat*, *textovat* (emailing, chatting, texting)
- a sound character like *šeptat* (whisper), *volat* (call) or *křičet* (scream, shout)
- an explanation of the purpose of the communicative function like *říct* (say), *oznámit* (announce, report) or *zakázat* (forbid).

For this study, no such distinction has been made, since the purpose is to search for other differences.

4.3 Corpus-based discourse studies

The combination of CDA and corpus analysis has now been established enough as a research field (whether it is otherwise called corpus-based or corpus-assisted) to receive its own acronym, CADS (Corpus-Assisted Discourse Studies, Törnberg & Törnberg 2016, 404). To create a non-biased, methodical, systematic result out of something that could otherwise have been very vague (Franklin 2017), a corpus-based discourse analysis is used in this study. In his book on using corpora for gender studies, in particular, Baker states (2014, 13) that, despite the methodology being well established, few researchers seem to combine a CDA analysis on gender with quantitative corpus analysis. Such a method does, however, give a broader view of the subject and the research, making it easier to issue a more general scientific statement about the language in use in the media – in this case, the Czech media. Baker also notes (*idem* 2014, 90) that gender representation from a large corpus is a way around some issues with

interpreting the results – since the analysis gives a cumulative picture of views concerning gendered categories in the society at issue.

4.3.1 Corpus-based discourse studies on reporting verbs

There are only a few previous studies on the combination of gender, corpus analysis and reporting verbs. Caldas-Coulthard (1995) does an analysis of who is “given voice” and how this is reported in three newspapers from the United Kingdom. In that study, the English material consists of 200 news “narratives” from ten days in 1992, excluding such topics as sports, debates and interviews. That study is qualitative and its material is carefully chosen to be as gender-neutral as possible. Caldas-Coulthard concludes (1995, 230) that utterances are interpreted and re-interpreted until they eventually end up in a newspaper article, but what the readers see in print (on paper or digitally) is what reflects structures and systems and may be incorporated into their own discourse. Caldas-Coulthard’s study, although it is not quantitative, also concludes that men are quoted eight times more often than women (idem 1995, 235). It would have been interesting to see her conclusions about the differences in reporting verbs backed up by figures, since her results could possibly be verified using a larger corpus material and method.

A more recent combination of CDA and reporting verbs analysis is made in Gidengil & Everitt (2003). They conduct a qualitative study on the case of Canadian, female politicians depicted in TV and newspapers through 885 instances of reporting verbs (idem 2003, 217). The examples they discuss are all active and strong narratives, something that is considered typical for masculine politicians (idem 2003, 210). The authors differentiate between reported speech and reports on speech (idem 2003, 216), something that is not considered in this analysis, since, in Czech, the representation of the professional woman or man is gendered, either way. Their categorisation was studied manually, by letting 242 students evaluate their verbs according to a Positive/Negative scale of five. In addition, they measured how aggressive the students found the reported speaker without the students knowing whether the speaker was female or male.

Their tables also show how the students reacted to the reporting verbs categorised as aggressive. Gidengil & Everitt conclude *inter alia* that the female students took the female party leaders' speech as more aggressive than the male party leaders (idem 2003, 225). Their more general conclusions are, however, that the Canadian journalists interpreted female party leaders' speech more than they did male leaders' speech, using verbs other than the standard "say", "tell" and "talk about". In their study, several "aggressive" verbs were only used for women – and these were not in any way standard reporting verbs--*blast*, *bash*, *slam* and *rebuff*, to name a few (idem 2003, 227).

4.4. Gender discourse in the Czech Republic

Gender theories have not been developed solely in the Western part of the world, of course. Before the 20th century, several female, Czech writers (including Božena Němcová, Eliška Krásnohorská and Karolina Světlá) were the avant-garde of the emancipatory idea that women should take their well-deserved part in society. During the First Republic of 1918-1948, when the Czechoslovak nation was founded after the fall of the Habsburg Empire, both translated works of fiction and nationally produced magazines discussed the role of women in Czechoslovak society (Oates-Indruchová 2016, 923), putting forward the idea of capable women as a contrast to stereotypes about the two standard sexes. This is not to say that the period was extremely liberal; for example, women writing about homosexuality still mostly used pseudonyms (Lishaugen & Seidl 2011, 222; 234). This period was short-lived; the Nazi occupation of the Czechoslovak Republic in 1939 pushed women back out of politics and back into their homes. During the Communist era of 1948–1989, women were officially back in politics, but in practice, women's emancipation suffered many setbacks, and the word "feminism" even disappeared from the official, public discourse (Oates-Indruchová 2016, 924–925). Oates-Indruchová, therefore, aims to clarify the presumption that feminism as a concept and ideology was imported to the Czech(oslovak) Republic after 1990. It is here claimed (idem 2016, 938) that there was then (or is still) a hostile feeling against feminism and challenges to gender norms in the Czech media. Of interest here is Oates-Indruchová's conclusion that popular

books and media from the early 1990s quickly became notably sexist, with the examples (idem 2016, 939) of both re-published novels and one of the daily newspapers (*Blesk*) that is included in the source material for this study.² Rebecca Nash reports on three prominent Czech gender theorists from the first decade after the Velvet Revolution, and states (Nash 2002, 293) that the issue of having employment, something debated among gender scholars in the West at the time, was not an issue under discussion for these women, and that Czech women of the times were not supposed to aspire to any level of political involvement (idem 2002, 294). Havelková & Oates-Indruchová (2014) give a good overview of the general state of gender research in the Czech and Czechoslovak republics. None of the articles in that book, however, look more closely at the language usage as a whole. The authors also conclude that gender issues and history have not been sufficiently studied and that further discourse research needs to be done in order to complete the picture (idem 2014, 13).

5. Question and hypotheses

The question to be answered is:

- Are negative reporting verbs more frequent for women than for men in the Czech media after 1989?

From this question and the theories presented above, either of the following two hypotheses should be verified, or a null hypothesis should be verified:

- Hypothesis 1: Women get significantly more negative media coverage than their male counterparts.
- Hypothesis 2: Women get significantly more positive media coverage than their male counterparts.
- Null hypothesis: No significant gender differentiation is visible in the source material.

² Unfortunately, one of Oates-Indruchová's sources here is an unpublished article that seems impossible to obtain today.

These are mutually exclusive hypotheses. From them is derived the argument that, if the negative reporting verbs occur at a different frequency when used in statements referring to women and men, respectively, then there is visible gender differentiation in the material. The hypotheses can be tested and falsified either for individual professions or for a weighted average of all five chosen professions. In this study, both options will be tested.

7. Material

The source material is the latest (at the time of writing) version of SYN, version 5. This is empirical data collected in the Czech National Corpus, abbreviated ČNK (Křen et al. 2017). This version consists of 4 599 643 984 tokens, which makes 7 770 263 lemmata (words in their basic form, such as nominative or infinitive). The specific material used is the journalistic portion of the SYN version 5: 176 titles from the period ranging between 1989 and 2015, including several national daily newspapers (*Mladá fronta Dnes*, *Lidové noviny*, *Právo*, *Hospodářské noviny*, *Blesk*, and *Sport*), regional daily newspapers (mostly *Deníky Bohemia* and *Moravia*) and weekly or monthly magazines (*Reflex*, *Respekt*, and *Týden*), the latter from the years 1998–2014 (Křen, Richterová & Škrabal 2017). The journalistic portion is by far the largest in the SYN series version 5³. The corpus is, hence, not considered representative, since it does not cover all kinds of empirical data, but no document is to be found twice in it. It is traditionally tagged with standard metadata (Hnátková et al. 2014, 160). The searches were conducted mainly during September and October 2017. For all searches, only material with Czech as the source language is used. All the SYN series are monitor corpora (McEnery and Hardie 2012, 6), which means that they are well-made for searching large text volumes and applying a statistical method to get an overview of the everyday usage of expressions.

³ A table of the number of words from the respective areas is found here:
https://wiki.korpus.cz/lib/exe/detail.php/cnk:slozeni_syn_v5.png?id=en%3Acnk%3Asyn%3Averze5

8. Method

To methodically reach conclusions drawn from the material and make a systematic, quantitative study – in other words, in order to get a statistically significant overview – a large enough source material (text corpus) should be used (Baker 2014, 18; Törnberg & Törnberg 2016, 404; Stubbs 1997, 110). The material is the latest version of the SYN series of the Czech National Corpus, which consists of empirical data ranging from 1989 to the beginning of 2015, and is presented in more detail elsewhere. Baker (2014, 21) makes the methodological recommendation of putting the concordance hits into a table as both raw numbers and percentage frequencies. This is, therefore, the method undertaken in this study for ratio calculation.

The empirical data will be systematically researched through an analysis of reporting verbs found in the source material. To classify the reporting verbs, the Harvard Psychosociological Dictionary (Kelly & Stone 1975, 10; 12–13) is used for the classification of the reporting verbs. Since that dictionary is a work based on English words, two or three synonyms of each chosen verb will be studied to get a more complete meaning. This dictionary is still such a valid research work that its importance should not be overlooked. The focus is, then, on Charles Osgood's semantic differentials (Osgood, Suci & Tannenbaum 1971), noting whether or not the Czech words' English translations in the largest Lingea Czech–English dictionary (Lingea s.r.o. 2008) are categorised in the dictionary as Positive/Negative, Strong/Weak or Active/Passive (cf. Osgood, Suci & Tannenbaum 1971, 25, 66 & 120). Naturally, a future, thorough reading of the semantics of these 25 verbs should be made by studying their core meaning manually in Czech monolingual dictionaries.

In the calculations, only reporting verbs in a position of +/-3 from the keyword noun are included. This was chosen because Czech word order allows the predicative verb to be placed

both before and after its subject noun, which may, in turn, consist of more than one word. To choose a larger number, like 4 or 5, would create too much information noise. Two examples from the concordance for the most frequent of the chosen reporting verbs are shown in Figures 1 and 2.

Figure 1: Example of the positioning of verbs in Czech, from the search for female MP, *poslankyně* + *tvrdit* (“claim”, “assert”, “contend”).

<input type="checkbox"/>	Mladá fronta DNES	v úterý	Nejprve pořádek , pak možná platby pacientů	Západočeská	poslankyně	Milada Emmerová	tvrdí	že žádné nemocnice není třeba rušit
<input type="checkbox"/>	Mladá fronta DNES		Z tohoto pohledu tam pan Herold patří ,		poslankyně	ODS Kateřina Dostálová .	tvrdí	Vladimír Mlynář z Unie svobody se
<input type="checkbox"/>	Mladá fronta DNES		dostal geniální nápad vyrábět plenkové kalhotky ,	náchodská	poslankyně	Zdeňka Horníková .	tvrdí	Vydělal na plenkách
<input type="checkbox"/>	Lidové noviny		propagaci buržoazní kultury a západního stylu života " .	Slečna	poslankyně		tvrdí	že o zvlášti komunistického režimu nemá dostatek objektivní
<input type="checkbox"/>	Mladá fronta DNES		Myslím , že už není návratu ,		poslankyně		tvrdí	že chtěla kandidaturou za Zemanovce posílit levice v

Figure 2: An example of the combination of the noun for female MPs (*poslankyně*) and the verb *protestovat* (“protest”).

Návrh nebere v úvahu jedinečnost manželství ,	lidovecká	poslankyně	Michaela Šojdrová .	Svatba jen na matrice ?	Největší spory
Proti těmto závěrům České školní inspekce	pardubická	poslankyně	Ladislava Zelenková (ČSSD)	Česká školní inspekce ve své	
že byly zveřejněny části nahrávky osobního charakteru ,		poslankyně		Hodlá proto žalovat tvůrce divadelní hry .	„ Svá
mit i církevní školy , proti čemuž marně	komunistická	poslankyně	Marta Semelová .	Doma až později	Ze školních automatů by

The corpus lemma search does not differentiate between the participles of the verbs and their indicative forms, but manually browsing through the search hits revealed that this was evident in the case of only one verb, *přesvědčit* (in the sense “to convince”). Since this word, in context, meant that the person with the researched occupation was convinced, it was classified as falling within the positive category.

8.1 Analyses and work steps

With this combined method of CDA and corpus linguistics, it is sometimes difficult to differentiate between the qualitative and the quantitative analysis – they are intertwined in the progress.

The searches start with the nominative form of a term for a professional occupation. A so-called basic search is performed, as opposed to a lemma search, which includes all cases of the word in

question. First, the search is made within a context, with a so-called PoS (part of speech) filter, of any verb within one position to the right of the professional term. A frequency list is then created of the verb lemmata within one position to the right, to show which verbs are used at all with this keyword and how frequent they are (after the most frequent verbs *být* and *mít*, “be” and “have”). The reason for considering one position at a time is the current limitations of the corpus engine, which cannot create a frequency list like this for positions greater than one.

8.1.1 The verbs

The next step is to check manually which verbs in the frequency list are reporting verbs and chose which are the 50 most frequent (since the searches returned a few hundred reporting verbs). After that, a categorisation from the *Harvard Psychosocial Dictionary* is applied, and a qualitative analysis is made on that basis. *The Harvard Psychosocial Dictionary* was created to assist psychologists who wanted to assess meaning in text content – in other words to do content analysis (Kelly & Stone 1975, 1). The dictionary is currently published online, where the categories are also explained in more detail.⁴ It has been expanded over time to consider a wide range of binary values, but for this study, only three of the original value pairs have been used. Specifically, the semantic categories Positive/Negative, Strong/Weak and Active/Passive were used for categorisation, and the 50 most frequent reporting verbs for Members of Parliament were noted with the categories that their English equivalents have in the dictionary. Where there are more than one or two English synonyms for the Czech verb, the three most common are chosen. Here are some categorisation examples:

<i>obviňovat</i>	accuse, blame	(accuse & blame) Negative, Hostile.
<i>potvrdit</i>	confirm, affirm, verify	(affirm) Positive, Strong. (confirm) Strong. (verify) Positive, Active.
<i>mluvit</i>	speak, talk, say	(say) Active. (speak) Active. (talk) Active.

⁴ The Harvard Psychosociological Dictionary’s categories are explained briefly here: <http://www.wjh.harvard.edu/~inquirer/homecat.htm>

Of the three binary categories that were deemed relevant for the aim of this study, only one was chosen: 25 of the verbs were only either positive or negative. Both strong and weak were used for some English synonyms, but none of the categorised reporting verbs were classified as passive. A few verbs were not found in the dictionary, and these were then disregarded. That left 12 verbs that were negative and 13 that were positive, which made a comparable binary category for this study. The 25 Positive or Negative verbs were therefore divided into separate groups and analysed with the nouns (see Appendix 1).

Table 1: The final reporting verbs with their English translations and categorisations according to the Harvard Psychosociological Dictionary

Czech	English	Strong	Weak	Active	Passive	Positive	Negative
kritizovat	criticize, attack, denounce	X		X			X
obviňovat	accuse, blame						X
odmítnout	refuse, decline, pass		X				X
pohrozit	threaten	X					X
pomluvit	slander, defame, libel			X			X
přiznávat	confess		X				X
protestovat	protest, remonstrate, object			X			X
prozradit	reveal, disclose, leak						X
tvrdit	claim, assert, contend	X	X	X			X
vyhrožovat	threaten, menace, intimidate	X		X			X
vyplsnit	reproach, chastise, reprimand			X			X
zdůraznit	stress, emphasize, point out	X		X			X
hovořit	talk, discuss; address in a speech			X		X	
informovat	inform, notify, instruct	X		X		X	
poradit	advise, counsel, recommend	X				X	
potvrdit	confirm, affirm, verify	X		X		X	
považovat	consider, believe					X	
přesvědčit	convince, persuade, reason	X		X		X	
připouštět	admit, concede, acknowledge		X			X	
připustit	admit, concede, allow		X			X	
přislíbit	promise, vow, agree					X	
prohlásit	declare, state, affirm	X		X		X	
sdělit	communicate, inform, announce			X		X	
slíbit	promise, assure, vow to	X				X	
vysvětlit	explain, clarify, clear up			X		X	

Source for classification: Osgood, Suci & Tannenbaum (1971).

Now the corpus is searched again, this time for professional noun term + each of these 25 reporting verbs. As previously mentioned, the search is filtered to contain the respective verb within three steps from the keyword (+/-3), to get a more representative picture. The number of hits for each combination is noted (see Appendix 2). In the Appendix, there are formulas for calculating the following data for each verb and for each female and male profession:

- relative figures to all verbs
 - ratio for the figures that regard women
 - ratio for the figures that regard men
- relative figures to the searched reporting verbs
 - ratio for the figures that regard women
 - ratio for the figures that regard men
- total figures for each verb, for women and men

Using these tables, it will be possible to falsify either hypothesis (H), and answer the question.

If there are patterns to support H1, then H0 and H2 can be falsified, and in the result there will be a conclusion emerging from H1. Since the perfective and imperfective aspects of every verb are reported separately in the corpus, as well as in the most recent frequency dictionary (Čermák & Křen 2011), they are taken as separate verbs here, as well. In some cases, there are also semantic differences, as shown in Table 1 (*tvrdit* and *potvrdit*; *připouštět* and *připustit*).

Finally, for each reporting verb classified as negative or positive, a search was made for all five professions.

8.1.2 Professional denominations

To begin with, the study needed a professional noun that would be well represented in the media. Inspired by Gidengil and Everitt (2003), the starting point are the Czech nouns used to refer to is Members of Parliament, *poslankyně* and *poslanec* for female and male members, respectively. The other four professions were chosen partly based on previous research and partly on what professions are often cited in media in general and represent both public officials and popular culture. They were chosen because they come from diverset social groups with

different amounts of power in society, but they also had to have a strong enough representation in the source material to afford meaningful statistical analysis. In the news media, low-wage profession employees are rarely represented by themselves through reporting verbs (for a British study, see Owen 2012), which is why no cleaners (11 search hits for *uklízečka* + *tvrdit*, compared to 180 hits for *úřednice* + *tvrdit*, which was the lowest ranked of the chosen professions), bartenders (29 hits for *barmanka* + *tvrdit*) or similar professions could be chosen for this study. They have not been chosen based on statistical criteria for the most common professions, since the empirical data cover more than 20 years and also include foreign-based professionals, like the Members of the European Parliament.

The professions' key words were searched in their nominative, singular forms, i.e., with the search called "word form", with the reporting verb in question as a lemma in position 1–3 to both the right and the left of the word. Then the boxes for "journalism" and "source language: Czech" were ticked to get the relevant figures.

Finally, a few notes on what did not end up in the final result:

Reporting verbs that may be stereotypes for women, like *pomluvit* ("slander, libel") were not generally found for the male occupations (there was a single entry for a male boss who slandered someone), which is an example of the reinforcement of the idea that women do not belong in politics (Gidengil and Everitt 2003, 211). On the other hand, *rafat*, ("bark, yap") was actually found for both female and male occupations.

When choosing the word to use for a kind of leader, searches for the word *předačka* ("female political leader") returned two (!) hits, the form *šéfová* ("female boss", treated as a noun despite its adjectival form) returned 26 hits and *vůdkyně* (female leader) returned 75 hits. There is also a discussion in the Czech Republic (Anonymous 2006) concerning which form to use for a female leader using the loanword *lídr*; in this study, the form *lídrka* got zero hits with any verb at

position 1R, whereas the form *lidryně* got 23 hits, the first one from as late as in 2004. I, therefore, decided not to use any of these.

9. Results

In this section, the quantitative results are again intertwined with the qualitative, since the corpus-based (critical) discourse analysis, or CADS, method is a combination of the two.

The results show that *poslankyně* (female Member of Parliament), a female profession with a stable statistical significance and a high level of power in society, is negatively and non-representatively portrayed. On the other hand, the profession with the most unsteady income, *zpěvačka/zpěvák* (singers), are also the most negatively represented, but this might have more to do with the fact that in this source material, they are most frequently reported in tabloids “driven more by sleaze than by substance” (Ross 2017, 162) – surveyed here via such publications as *Aha!* and *Blesk* – that prefer to write about the more negative sides of society.

In Table 2, we see the 25 reporting verbs that were eventually chosen, because they were clearly either negative (Neg) or positive (Pos). The rows then show the figures for the first analysed noun, which is Member of Parliament (MP; *poslankyně* indicating female members and *poslanec* indicating male members). The first two columns show how many occurrences of each verb there were, as well as occurrences of all verbs and reporting verbs. The relative figures and fifth column (“Ratio female/male” for any verb) are then calculated based on the figures for any verb and any reporting verb, respectively. The relative figures for any verbs are important, giving a standardized view of the verbs’ relative importance, since there are so many more verbs that are associated with men than with women in the material. The ratio thus gives an easier comparison.

The most interesting column is probably the last one, “Ratio female to male” for the reporting verbs, where all numbers above one mean that this verb is used more often to portray women than men. There are more ratio figures above one in the top half of the table, which means more negative verbs are used to portray women here. There are two instances in which the verb was not used at all for the male MP, which creates a mathematical problem of division by zero, noted in the ratio columns. This is not important for the final result. When the Positive/Negative ratio is above one in the last row, it means that the positive verbs are more common than negative verbs.

The first 12 verbs are negative and the last 13 (from *sdělit*) are positive. In the first two columns, the absolute figures are presented, and the first row has the absolute figures for all verbs found with the noun in question, 19 387 verbs for female MP and 87 584 for male MP. The next row has the total number of reporting verbs in the table, 1 675 for women and 8 160 for men. The negative table includes the verb for “claim”, *tvrdit*, which is a very common reporting verb, in the corpus material used as well as in the frequency dictionary (Čermák and Křen 2011, 22). Since the ratios are calculated on relative numbers, it does not matter that this verb and a few other verbs have so many hits; the important thing is how often they are used about women as compared to men, which we can see in the rightmost ratio column.

Table 2: The numbers and performed calculations, here with the example calculations for *poslankyně/poslanec* (Member of parliament)

		Absolute figures		Relative figures to any verb		Ratio (female to male)	Relative figures to reporting verbs		Ratio female to male
		poslankyně	poslanec	poslankyně	poslanec		female	male	
	Any verb	19 387	87 584						
	Any reporting verb	1 675	8 160						
Neg	tvrdit	410	2049	2.11%	2.34%	0.90	17%	19%	0.97
Neg	odmítnout	85	325	0.44%	0.37%	1.18	4%	3%	1.27
Neg	přiznávat	38	141	0.20%	0.16%	1.22	4%	2%	1.31
Neg	prozradit	34	106	0.18%	0.12%	1.45	12%	4%	1.56
Neg	obviňovat	5	24	0.03%	0.03%	0.94	0%	0%	1.01

Neg	kritizovat	117	334	0.60%	0.38%	1.58	2%	2%	1.71
Neg	pohrozit	13	19	0.07%	0.02%	3.09	0%	0%	3.33
Neg	zdůraznit	57	330	0.29%	0.38%	0.78	2%	3%	0.84
Neg	protestovat	16	72	0.08%	0.08%	1.00	0%	0%	1.08
Neg	pomluvit	4	0	0.02%	0.00%	DIVI- SION/0	0%	0%	DIVI- SION/0
Neg	vyplísnit	3	2	0.02%	0.00%	6.78	0%	0%	7.31
Neg	vyhrožovat	8	11	0.04%	0.01%	3.28	0%	0%	3.54
Pos	sdělit	127	616	0.66%	0.70%	0.93	11%	10%	1.00
Pos	prohlásit	262	1499	1.35%	1.71%	0.79	8%	19%	0.85
Pos	slíbit	37	129	0.19%	0.15%	1.30	1%	1%	1.40
Pos	považovat	56	303	0.29%	0.35%	0.83	2%	2%	0.90
Pos	přislíbit	27	80	0.14%	0.09%	1.52	1%	0%	1.64
Pos	hovořit	39	213	0.20%	0.24%	0.83	1%	2%	0.89
Pos	vysvětlit	108	476	0.56%	0.54%	1.02	16%	10%	1.10
Pos	potvrdit	94	582	0.48%	0.66%	0.73	9%	12%	0.79
Pos	přesvědčit	46	220	0.24%	0.25%	0.94	2%	2%	1.02
Pos	informovat	36	254	0.19%	0.29%	0.64	5%	5%	0.69
Pos	poradit	3	21	0.02%	0.02%	0.64	1%	0%	0.69
Pos	připustit	30	226	0.15%	0.26%	0.60	1%	2%	0.65
Pos	připouštět	20	128	0.10%	0.15%	0.70	1%	1%	0.76
	Positive/ Negative ratio	1.12	1.39						

In these results, there are more positive than negative reporting verbs in the material used, but the positive ratio is higher for men, and thus negative verbs are more often used about women. The negative verbs with the highest ratio are *vyplísnit*, “to reprove or reproach”, with a ratio of 7,31 for female MPs, followed by *vyhrožovat*, “to threaten, intimidate or menace”, with a ratio of 3.54, and *pohrozit*, “to threaten”, which has a ratio of 3.33. None of these are among the most frequently used verbs, but it is still clear that Czech readers read that female Members of Parliament more often threaten others than male members do. A manual, contextual check of the search hits has been performed to ensure that it was not the women who were reproached or threatened, and it was not. This check also included the positive verb *přesvědčit*, “to convince or persuade”, for which many of the results consisted of the participle, meaning that the professional person was convinced or persuaded by someone, rather than convincing or persuading others. These verb forms were included in the calculations. More relevant for the

final result is that a verb like *kritizovat*, to criticize, has a relative ratio of 1.71, and is the second most frequent of these negative reporting verbs. Gidengil and Everitt (2003, 211) noticed that, in the Canadian TV debates they studied, the female party leaders were more often depicted as “attacking” than the men, something that is thus reflected here, where a verb that may include the concept of “to attack” is *kritizovat*.

The results of a search in the Czech National Corpus are displayed as a concordance, where the surrounding context is shown as well as the hits themselves. That makes it possible to perform manual controls of the context. This was also done here, especially in cases like that of “to threaten”, above. A few hits were removed before performing any calculations: *poradit si*, which with this dative reflexive particle included means “to handle” instead of “to advise” or “to recommend”, and in the case of negated occurrences of *protestovat*, “to protest” before the words for boss, since that would give a different meaning not relevant for this study.

In Table 2, only one occupation was chosen as an example of the calculations leading up to the result; we will now look at the results for all five. In Table 3, below, the figures in the ratio columns mean that, the higher the number, the more common positive reporting verbs were.

Table 3: The total result for each profession

	Number of occurrences		Positive/Negative verb ratio	
	female	male	female	male
<i>poslankyně/poslanec</i> (member of parliament)	1 675	8 160	1.12	1.39
<i>šéfka/šéf</i> (boss)	2 251	45 973	2.11	2.18
<i>úřednice/úředník</i> (clerk)	1 177	1 313	1.81	1.79
<i>učitelka/učitel</i> (teacher)	3 996	1 871	2.23	1.90
<i>zpěvačka/zpěvák</i> (singer)	3 080	3 331	0.65	0.85
All occupations (weighted average)	12 179	60 648	1.42	1.92

It seems clear from these results that there is a gender differentiation for these professions in the Czech printed media included in the National Corpus. The positive verbs are generally prevalent, but they are more prevalent for men (a total of 1.92) than for women (a total of 1.42). Since these ratios are relative to the number of reporting verb occurrences, it does not matter for the result that men are mentioned five times more often than women – but that is another clear result of gender differentiation in general.

The weighted average in the last row still points to a generally more negative picture of women for all five occupations in this study. A Chi2 test is performed, to test the probability (P) value of the figures, where the probability should be zero to show it to be statistically significant. The difference for all professions together is statistically significant at the conventional levels (P value = 0.000). However, we see a difference between the professions: The female MPs and singers are clearly more negatively depicted (P value = 0.000 for both occupations respectively). For the bosses (P value = 0.524) and clerks (P value = 0.887), however, the results show no statistically significant difference in language usage. The female teachers, finally, have a statistically significant overrepresentation of positive reporting, relative to their male colleagues (P value = 0.007).

None of the categorised reporting verbs were classified as passive (see Appendix 1), which makes it clear that female members of the Czech Parliament – since they were the starting point for the study – are at least not considered less active than their male counterparts. Only eight out of 50 reporting verbs were classified as weak. However, 14 of them were classified as negative, and 15 were classified as positive. Of these, two had synonyms that were the opposite, which leaves 12 negative and 13 positive in the basis of this study.

Men are in focus more often, with the extreme example of 45 973 male and 2 251 female bosses in Table 3, but that may be representative of the number of actual bosses, MPs etc., which would mean that the gender differentiation exists on a different level in society, not necessarily

in the language usage. Since this is a study spanning 25 years, it is rather difficult to find the statistics for how many people of each gender work in all of these professions. In the corpus searches, there are also results from other countries and the European Union Parliament, which makes such statistical figures for the Czech Republic rather biased. Singers may be particularly hard to pinpoint, since it is probable that many of them are not registered in any statistical records. It is equally striking that reporting verbs on singers are very negative (a Positive/Negative ratio below one for both genders) and reporting verbs for bosses are very positive (a ratio not only above one, but above two for both). Since the paper with the largest number of articles in the SYN series corpus is *Mladá Fronta Dnes*, a general newspaper not specializing in any business reports, the opposite scenario could perhaps have been expected. Female teachers also stand out, since they get so much positive reporting (2.23 versus 1.90 for male teachers), a result that points to possibilities for future, more in-depth, studies.

10 Conclusions

This study aimed at answering the question:

- Are negative reporting verbs more frequent for women than for men in the Czech media after 1989?

There were two hypotheses and a null hypothesis:

- Hypothesis 1: Women get a significantly more negative media coverage than their male counterparts.
- Hypothesis 2: Women get a significantly more positive media coverage than their male counterparts.
- Null hypothesis: No significant gender differentiation is visible in the source material.

The findings in this study indicate that, in general, women are more negatively portrayed than men, according to the binary distinction used. For two of the five occupations, Member of Parliament and singer, hypothesis 2 has been falsified. For another two of them, bosses and

clerks, the null hypothesis cannot be falsified, since the P value is too high. For the fifth, the teachers, hypothesis 1 has been falsified.

The statements behind the hypotheses, that women are negatively and very disproportionately pictured, have been analysed for the Czech case, and the resulting picture is ambiguous. When studying different forms of linguistic differentiation and othering, however, it is important to remember that gender is not equivalent to categories like class or ethnicity (Eckert 1989, 253; Lazar 2007, 143). Minority groups, like the ones that have been in focus in the previous studies (Elmerot 2016; 2017), are more obviously “others”, whereas gender differentiation often is more complex, a dynamic that is also visible in the results this study, where the numbers differ from occupation to occupation. There is a clear gender bias in discussing certain professions, without a doubt, but not for all. The number of articles on men and women seems to be in the men’s favour, but this is due to the fact that men are also more often bosses and women are more often teachers in the Czech Republic. The exact figures for percentages by sex in these professions are not included in the present study. It is perhaps not surprising that men are reported about in the media five times more often than women, but that is still a clear indication that women are under-represented in a society where there is no great difference between the total number of women and the total number of men (Czech Statistical Office 2016). The professions in this study (Members of Parliament, singers, bosses, clerks and teachers) are represented by hundreds of thousands of sentences in the source material, and are by no means either obscure or irrelevant to society as a whole.

There are several facts in this study that may inspire future studies, quantitative as well as qualitative. First of all, other verbs than reporting verbs could be analysed with the same qualitative method to verify the results of this study. Far-fetched verbs have not been considered in the current study, but that could be done in the future. The study could also be widened in terms of time. Rebecca Nash (2002, 296–299), for example, looks back on gender research conducted during the First Czechoslovak Republic, making it even more interesting to conduct a

combined critical discourse and corpus analysis study on material from that time, if and when the Czech National Corpus contains useable material for the years 1918–1938. From the current study, however, it seems that female Members of Parliament were hardly portrayed at all in the Czech news media before the mid-1990s, since many of the keywords for them do not even appear until about 1994, when their male counterparts return search hits from 1990 and 1991 (with a few from 1989, as well). A larger study, and perhaps one exploring a wider range of time, could tell us whether it is a general fact that men are five times more reported about than women in the Czech printed news media, as the results here suggest. Future corpus-based research may also examine changes over time, since the ČNK is projected to eventually cover similar material from 1850 onwards.

11 References

- Allén, Sture. (1971). *Nusvensk frekvensordbok baserad på tidningstext. 2, Lemman = Lemmas*. Stockholm: Almqvist and Wiksell international.
- Anonymous. (2006). “Z dopisů jazykové poradně”. *Naše řeč* 89(3). Prague: Academia, 167–168.
- Baker, Paul. (2014). *Using corpora to analyze gender*. London: Bloomsbury Academic.
- Caldas-Coulthard, Carmen Rosa. (1995). “Man in the news: The misrepresentation of women speaking in news-as-narrative-discourse.” In Sara Mills (ed.), *Language and gender: Interdisciplinary perspectives*. London: Longman, 226–275.
- Čermák, František and Křen, Michal. (2011). *A frequency dictionary of contemporary Czech: core vocabulary for learners*. Routledge Frequency Dictionaries. New York: Taylor and Francis.
- Czech Statistical Office. (2016). “Gender: Demography – data”.
https://www.czso.cz/csu/gender/2-gender_obyvatelstvo [read 10 November 2017].
- Eckert, Penelope. (1989). “The whole woman: Sex and gender differences in variation.” *Language Variation and Change* 1, 245–267.
- Elmerot, Irene. (2016). “Är en zigenare mer oanpassningsbar än en rom? En pilotstudie om kollokationer för orden Cikán och Rom i modern, tjeckisk tidningstext” *Slovo: journal of Slavic languages and literatures* 57, 9–23.
- Elmerot, Irene. (2017). “Language and Power in Czech Corpora”. *Computational and Corpus-based Phraseology – Recent Advances and Interdisciplinary Approaches*. Proceedings of the Conference. Volume II. Geneva: Editions Tradulex, 174–177.
- Fairclough, Norman. (2015). *Language and power*. 3., [updated] ed. London: Routledge.

- Franklin, Emma. (2017). "Towards a Corpus-lexicographical Discourse Analysis", *Computational and Corpus-based Phraseology – Recent Advances and Interdisciplinary Approaches*. Proceedings of the Conference. Volume II. Geneva: Editions Tradulex, 190–196.
- Gellerstam, Martin. (1996). "Anföringens estetik: om dialogformler i tvärspråkligt perspektiv." In Josephson, Olle (ed). *Stilstudier*. Uppsala: Hallgren and Fallgren, 12–29.
- Gidengil, Elisabeth and Joanna Everitt (2003). "Talking Tough: Gender and Reported Speech in Campaign News Coverage". *Political Communication* 20(3), 209–232.
- Havelková, Hana and Libora Oates-Indruchová. (2014). *The Politics of Gender Culture under State Socialism: An Expropriated Voice*. Abingdon, Oxon: Routledge.
- Hirschová, Milada. (2017). "Verbum Dicendi." I *CzechEncy - Nový encyklopedický slovník češtiny*, edited by Petr Karlík, Marek Nekula and Jana Pleskalová.
<https://www.czechency.org/slovník/VERBUM DICENDI> [read 5 October 2017].
- Hnátková, M., Křen, M., Procházka, P., Skoumalová, H. (2014). "The SYN-series corpora of written Czech." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavík: ELRA, 160–164.
- Jones, Owen. (2012). *Chavs: the demonization of the working class*. Updated ed. London: Verso.
- Kelly, Edward F. and Stone, Philip J. (1975). *Computer recognition of English word senses*. Amsterdam: North-Holland.
- Křen, Michal, Olga Richterová and Michal Škrabal (2017). *Corpus SYN version 5*.
<http://wiki.korpus.cz/doku.php/en:cnk:syn:verze5> [read 1 November 2017]
- Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kovářiková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondříčka, P., Zasina, A. (2017). *Corpus SYN, version 5 from 24. 4. 2017*. Prague: Ústav Českého národního korpusu FF UK. [Available online: <http://www.korpus.cz>].

- Lazar, Michelle. (2007). "Feminist Critical Discourse Analysis: Articulating a Feminist Discourse Praxis", *Critical Discourse Studies*, 4(2), 141–164.
- Lingea s.r.o. (2008). *Lexicon 5 Anglický slovník Platinum*, version 5.1.0.6.
- Lishaugen, Roar and Jan Seidl (2011). "Generace Hlasu: Česká Meziválečná homoerotická literatura a její tvůrci," in Martin C. Putna, ed., *Homosexualita v dějinách české kultury*, 209–280. Prague: Academia.
- McEnery, Tony and Andrew Hardie. 2012. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- Nash, Rebecca. (2002). "Exhaustion from explanation – Reading Czech gender studies in the 1990s", *European Journal of Womens Studies*, 9(3), 291–309.
- Oates-Indruchová, Libora. (2016). "Unraveling a Tradition, or Spinning a Myth? Gender Critique in Czech Society and Culture." *Slavic Review*, 75(4), 919–943.
- Osgood, Charles E., George J. Suci and Percy H. Tannenbaum. (1971). *The measurement of meaning*. Urbana, Ill.: University of Illinois Press.
- Ross, Karen (2017). *Gender, politics, news: a game of three sides*. Chichester: Wiley Blackwell.
- Stubbs, Michael. (1997). "Whorf's children: critical comments on critical discourse analysis (CDA)" in Ann Ryan and Alison Wray (eds): *Evolving Models of Language*. Clevedon: Multilingual Matters, 100–116.
- Sunderland, Jane. (2004). *Gendered discourses*. Basingstoke: Palgrave Macmillan.
- Törnberg, Anton and Petter Törnberg. (2016). "Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum". *Discourse and Society* 27(4), 401–422.
- van Dijk, Teun Adrianus. (2001) "Multidisciplinary CDA: a plea for diversity". In Wodak, Ruth and Michael Meyer (eds). *Methods of critical discourse analysis [E-book]*. London: SAGE.
- van Dijk, Teun Adrianus. (2008). *Discourse and power*. Basingstoke: Palgrave Macmillan.

Wodak, Ruth. (1989). *Language, Power and Ideology: Studies in political discourse*, edited by Ruth Wodak, John Benjamins Publishing Company.

Wodak, Ruth. (2007). "Gender Mainstreaming and the European Union: Interdisciplinarity, Gender Studies and CDA". In Lazar, Michelle M. (red.) *Feminist critical discourse analysis – Gender, power and ideology in discourse*. Basingstoke: Palgrave Macmillan.