Smitten by the Written: Corpus Analysis of Agency and Roles in Pornographic Video Titles

Tomáš Mach[*] iD

Charles University, Czech Republic

Abstract

Despite the magnitude of its consumption by the general public, online pornography has to date received comparatively very little attention within linguistics. The paper explores the language in video titles on the pornographic site Pornhub using a corpus-driven approach. A corpus of 17,000 titles compiled specifically for the purposes of the study yielded a word list, which, together with an interdisciplinary literature review, was used to formulate two research questions. With gender as a key variable, the first research question focuses on the linguistic representation of agency by looking at verb objects, while the second examines expressions (nouns and adjectives) used to ascribe roles to the participants. A quantitative evaluation of the data indicates an interplay between gender and the observed features in both questions. The lexicalisation of agency was found to be more complex than previously thought; nine categories were found, and the role of passive voice differed depending on gender. As for gender characterisation, descriptions of women's roles are not only more frequent but also more heterogenous and evaluative compared to those of men.

Keywords

Introduction

Although pornography has attracted some scholarly attention from many disciplines ranging from medicine and psychology to literary and media studies, only a handful of linguistic works have been dedicated to the subject. Even fewer have dealt specifically with online pornography. Apart from general reservations associated with the topic and possible fears of being ridiculed by the academic community (Attwood 2010; Irvine 2014), one of the main deterrents to linguists seems to be the peripheral role that language plays in pornography, which is by its very nature a multimodal phenomenon. In this vein, the users' motivations behind accessing the material are visual and (to a lesser extent) auditory stimulation – people *watch* porn(ography) – the role of language is thus relegated to a mere accompaniment of the first two. This is especially true of written language, consisting mainly of user commentaries and content information on the sites.

[*] **Corresponding Author:** Tomáš Mach, Institute of the Czech National Corpus, Faculty of Arts, Charles University, nám. Jana Palacha 2, 116 38 Praha 1, Czech Republic. E-mail: tomas.mach@mail.muni.cz.

The fascinating aspect of these is that however immaterial and counterintuitive their presence may seem given the context of the genre, they still exist and are indispensable. Like on other online platforms, users feel the need to share their experience (Pihlaja 2016) and like on other websites, information is presented in form of texts. Still, of how much value is textual information when thumbnails (visual) are utilised as icons pointing to videos (visual content)?

In this paper, I analyse a corpus of nearly 17,000 video titles from the pornography site Pornhub. The primary motivations behind embarking on this journey are twofold. First, in line with Dziadosz and Chandrasekar (2002), who suggest that combination of both thumbnails and text help internet users to navigate the web better, the titles must have a pragmatic function of their own. In this respect, especially the headline functions of framing the narrative and connecting it to previously known facts and events (Dor 2003: 714) are of importance. Second, based on how pervasive a role pornography plays in today's Western society, assumptions can be made about how it either shapes our beliefs about sex and/or (depending on the position taken) how these beliefs are reflected in the discourse and language.

The paper begins with an overview of research regarding pornography consumption and its effects to familiarise the reader with the current situation outside linguistics. Later, the overview is narrowed down to humanities and language research. The corpus compiled for the study is then described and the research questions are introduced. These are followed by the two respective analyses.

Background

Research on sexually explicit internet material (SEIM)

Decades ago, the choice between the perusal of a (possibly rare) nude magazine and actual sexual intercourse seemed a binary one. Due to the rapid development of information technology the situation nowadays has become more varied (Attwood, Smith 2014) and it is this factor, to which the proliferation of online pornography in recent decades can be chiefly attributed. In research, it is frequently referred to as SEIM (sexually explicit Internet material) and generally understood as "*professionally produced or user-generated (audio)visual material on or from the Internet that typically intends to arouse the viewer and depicts sexual activities and (aroused) genitals in unconcealed ways, usually with close-ups on oral, anal, and vaginal penetration*" (Peter, Valkenburg 2011: 751). Materials of this kind are nowadays being consumed by the masses – one even speaks of "pornification/pornographication" (see Attwood 2002; Smith 2010 for an elaboration). The exact extent of consumption is difficult to report as most studies almost always moderate the datum by an independent variable in question. One that can be regarded standard is gender; in their study Petersen and Hyde (2010) found out that consumption of pornography is one of albeit only a few gender differences in sexual behaviour between men and women. This finding is corroborated by data from both before the emergence of large pornographic sites such as Pornhub or XVideos, as well as after, as these have been established as a go-to source of pornographic imagery. A sizeable study (n = 3,387) by Træen, Nilsen and Stigun (2006) found that 95.7 % of men claim to have seen a pornographic film as opposed to 76.1 % of women. As far as SEIM goes, 33.9 % of men admit to having watched porn on the internet in contrast to only 13.6 % of women. In a smaller (n = 832) but more current paper, Blais-Lecours et al. (2016) found that 90.2 % men and 51 % women are reported to have watched SEIM over the last six months. Similar results for the past six months were observed by Daspe et al. (2018) – 98.1 % and 73.1 % for men and women respectively, out of which 80.3 % and 25.5 % watched SEIM at least once a week. A large-scale (n = 21,555) study of pornography in couple relationships by Carroll et al. (2016: 156) revealed "*a notable gap in reported use levels with women being about twice as*

likely to report that they never use pornography compared with men of similar relationship status." These differences are further moderated by religiosity, first sexual union and the number of sexual partners (see Poulsen, Busby, Galovan 2013: 80 for details). Lastly and Stella (2018) investigated preferences of men and women in watching online pornography. His study found that women only perceived three categories (bondage, BDSM, gay) to be more arousing than men, whereas males were aroused by fourteen (lesbian, milf, squirting, anal etc.) more than women.

The discrepancies between the sexes together with other factors under scrutiny have become the subject of research, particularly in sociology and clinical psychology, with the aim to ascertain and account for any (causal) relationships between the consumption of the omnipresent SEIM and people's sexual beliefs and behaviour. Consumption of SEIM has, for example, been linked to lower sexual satisfaction (Blais-Lecours et al. 2016; Wright et al. 2017) but causality has been difficult to establish. Lower sexual satisfaction in turn predicts the inverse relationship between the frequency of use and the control over pornography use (Daspe et al. 2018). Importantly, the consumption of SEIM among adolescents has sparked a lot of debate; in a research review Owens et al. (2012: 116–117) sum up that "youth who consume pornography may develop unrealistic sexual values and beliefs", which may result in "permissive sexual attitudes, sexual preoccupation, and earlier sexual experimentation." Similar points are mentioned in a comprehensive review by Peter and Valkenburg (2016) although they point out the non-homogeneity of the populations being sampled for studies, which allows for only a tacit admission that some effect *may* exist, which has yet to be fully demonstrated. Finally, in their meta-analysis, Mellor and Duff (2019: 45) focused on SEIM consumption in men and sexual offence, which do not appear to be interrelated: "[i]*t also consistently appears that men who sexually offend report less exposure to pornography and that exposure to pornography does not result in more harm being caused to the victim.*"

The studies above provide a valuable departure point in that there are clearly differences in consumption, preferences and effects between males and females. The overarching question that this paper seeks to address is if such differences are reproduced linguistically and if so, how? Moreover, it looks at possibilities how the surrounding discursive practices could contribute to the aforementioned sexual attitudes and beliefs.

Pornography in the social sciences

The attitude to pornography in the humanities is best summarised by Keilty (2016: 71) who argues that "*browsing online pornography is a social experience*". The perspectives on pornography in the humanities have been shaped and to various extents perhaps even monopolised by feminist scholars, which resulted in thematising power-asymmetries, equality, ideologies and representation of women in the discourse. As Hiramoto (2015: 189) puts it: "*there is also no denying that one's sexuality, like one's race, social-class, and gender, can affect one's access to* [. . .] *power, social status, and material resources in everyday life.*" Serious research into these was, however, hindered by the fact that it was "*seen as coming from either a pro-porn or an anti-porn position, and therefore always-already suspect*" (Gabriel 2016: 371) The scepticism is shared by Ciclitira (2004: 282) who notes that the middle ground was rarely actively promoted, which effectively rendered the research more argumentative at the unfortunate expense of objective empiricism. Correspondingly, in their recount of the study of pornography, Paveau and Perea (2014) even go as far as to distinguish between "the militant paradigm" and "the more academic approach" towards pornography within the humanities. Despite the reservations one may have about the first epithet, it seems particularly apposite when confronted with opinions such as those voiced in Dworkin (1981), for instance.

As for concrete examples of research, Cowan et al. (1988) analysed pornographic films and note that "*specific acts of physical aggression appeared in 73 % and rape in 51 % of the movies.*" They

conclude by stating that "*a significant level of hatred of women is now available for viewing in our living rooms and bedrooms*" (Cowan et al. 1988: 308–310). In some more recent research where objectification was operationalised differently, however, it did not significantly differ across genders (Klaassen, Peter 2014). McKee (2005) even found men to be objectified more than women on the whole. As the field became more intersectional and nuanced, scholars started to hypothesise that pornography may be instrumental in partially constructing identities (see e.g. Jones, Mills 2014: 241), influencing sexual behaviour (Brown 2009; Vannier, Currie, O'Sullivan 2014) and beliefs (Vandenbosch, van Oosten 2017) of the consumers. These presuppositions are in line with the widely accepted gender similarity hypothesis (Hyde 2014, 2016). The lack of differences between males and females would thus preclude the existence of distinct discourse representations of the two, which is not the case. This non-uniformity is felt to be biased against the discursive construal of women, which feminist researchers interpret as warranting uncovering, sensitising and perhaps indirectly changing the burden of the *status quo.*

The existence of different gender representations in pornography have led researchers to examine how these representations line up with traditional gender stereotypes, and what part pornography might play in their perpetuation. Although pornography abounds in gender stereotypical behaviour (Caputi 2014; Dines 2017; Carotte, Davis, Lim 2020), it remains unclear whether and how this affects the viewers. On the one hand, pornography consumption has been showed to be linked to objectifying, unequal gendered norms, and both hostile and benevolent sexism (Peter, Valkenburg 2007; Brown, L'Engle 2009; Hald, Malamuth, Lange 2013; Mikorski, Szymanski 2017; Willis, Bridges, Sun 2022, to mention just a few). In contrast to these findings, some studies link pornography consumption to liberalism and egalitarianism with respect to gender (Baron 1990; Jackson 2019) or report no relationship between pornography consumption and unequal gender norms (Vangeel, Eggermont, Vandenbosch 2020) and sexually objectifying behaviour (Yu et al. 2021). What further complicates the issue is that a lot of the research has been cross-sectional and demonstrating causality has proven difficult. Furthermore, a number of possible confounders have been identified, such as gender, agreeableness, and social context. Presently, the research seems inconclusive; more studies overall, however, tend to favour the hypothesis that pornography does affect the viewers' perceptions of women and men and their roles during sex. These might specifically include men being "in charge" of the sexual intercourse, thus having the power to decide when and how it happens. In most scenarios, men would be the active participants and would be the ones "on top". Moreover, violence (including spanking and gagging for instance) would be almost exclusively directed towards the female participant (Carotte, Davis, Lim 2020). Females would also be the more likely participant to be dehumanized and overall portrayed as submissive (Klaassen, Peter 2014).

Since pornography provides the unique (for most) opportunity to experience sexual intercourse of other individuals, it must shape the sexual life of the viewer somehow, be it on the level of attitudes or behaviour. Based on previous research and on the sexual script theory (Gagnon, Simon 1973), it is very likely that people seek out pornography for its edutainment value rather than for entertainment only. Escoffier (2007: 63) adds that "[t]*he effectiveness of video pornography depends upon the viewer's belief that sex is plausibly 'real' in some way; a pornographic film or video is both a 'documentary' and a fantasy*". This view is further corroborated by studies such as (Rothman et al. 2021), where online pornography was the number one source of information about sex among young adults (18–24 years).

Pornography in linguistics

A number of theoretical perspectives have been applied to the study of language and gender. The arguably most prevalent would be that of social constructionism where discourses structure,

perpetuate and construct realities. Accordingly, the linguistics choices one makes thus play a central role and changes to them can be expected to bring about changes in the real world. Other factors such as experience or biology, and their effect on linguistic choices are hardly ever considered. In this study, I do acknowledge that language may and does influence people's thought and beliefs, I am, however, sceptical of the oft-assumed extent, to which it does so, especially with regard to human sexuality. On the other hand, people use language to encode (among other things) the manifestations of human sexuality. Thus, when communicating about pornography and/or sex, people can (unwittingly) normalise certain attitudes/behaviours that have been shown to be unfavourable towards women or in general perpetuate gender stereotypes (see above).

Examples of the studies reflecting the social constructionist viewpoints include for example Koller (2015). Drawing on the notion of process types and conceptual metaphor theory, she analyses queer online written pornography to explore its potential "*to destabilise dominant discourses of gender and sexuality*" (Koller 2015: 268). McGowan's (2009) philosophical treatise makes use of Austin's Speech Act theory to substantiate the claims that pornography indeed contributes to gender oppression. Similarly, clichéd gender representations are a feature found in Virdis' (2015) empirical analysis of a pornographic novel, in which eroticism (e.g. allusions to body parts) is constructed by means of lexical choices, figurative language and metaphors from the source domain of natural environment. Morrish and Sauntson (2011) created a corpus of 40 texts of lesbian erotica to capture how the lesbian identity is constructed in these texts with notions like femininity, masculinity, power and submission in mind. Pornographic prose is also the focus of Lischinsky's (2017) quantitative corpus study, where he examines the agency of sexual acts in erotic fiction. Female characters are found to be more passive (occurring as an object) than their male counterparts. Finally, perhaps the closest study to this one is Pihlaja's (2016) paper in that it looks at online pornography (even the same website) using corpus methodology. It is a comprehensive analysis of the comment sections on Pornhub, revealing the solitary nature of such communication, functioning mainly to express arousal and pleasure without much interaction between the commenters.

Methods

It follows from the previous section that the underlying theoretical frameworks and therefore the methodologies of social sciences and humanities differ to some extent. McKee (2014) provides a list of methods used in porn studies listing four types of textual analyses, *ideological*, *post-structural*, *appreciation* and *exegesis*. Although I would argue that the boundaries between the first two are opaque at best (see Baker 2006: 5–10), it is still an immensely useful classification in contrast to the statistical approaches, which McKee also lists. He notes that the ideological and exegetic approaches are most popular in porn studies and literary analysis respectively and that perhaps adoption of quantitative methods would serve the field well. Apart from surveying population samples, there have been suggestions incorporating quantitative techniques (see Mazière et al. 2014), which, judging by the literature in the field, received a somewhat lukewarm response (see Motschenbacher 2018 though). Instead, manual qualitative processing of texts seems to be the preferred *modus operandi* of porn studies. This is where corpus assisted discourse studies (CADS) (Baker 2006; Partington, Duguid, Taylor 2017) methodology comes into play, allowing for querying the same sources much faster and producing quantifiable empirical results, not to mention making replicability a feasible option. An exemplary collection of such research on gender is Baker's (2014) work.

As opposed to other linguistic research published in the area so far, this study takes a corpus-driven approach due to the text type under investigation being somewhat uncharted territory.

Although one cannot resist making predictions, guesses and asking questions, the role of the corpus itself was deemed primary in the process of proposing the research questions. As Tognini-Bonelli (2001: 84) puts it:

> the commitment of the linguist is to the integrity of the data as a whole and descriptions aim to be comprehensive with respect to corpus evidence. The corpus, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well defined system. The theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus.

In this case, the statement consisted in compiling a wordlist, which would hopefully show tendencies or patterns that would be worthwhile to pursue in greater detail. The number of questions for this study was limited to a maximum of three prior to compiling the wordlist. Also, if there were a multitude of analyses worth conducting, priority would be given to those that would, at least in part, replicate the research available now.

Corpus

Description

To analyse pornographic video titles, a corpus had to be compiled and Pornhub was chosen as the most visited website (Alexa 2019). The videos were searched on 28 Nov 2018 based on the number of views using very liberal criteria; neither production (professional vs. amateur), nor category were taken into consideration. After some probing, the former was deemed too inaccurate (a lot of videos classified as amateur had a logo of a production company in them) to produce any meaningful results if encoded. In a similar vein, a lot of the videos were not ascribed any category, or this was done too haphazardly to be of any use. Since the thrust of the study is on what is viewed by people, the search was set to *World* and *All Time* as opposed to the default settings, where, based on user's IP, videos popular (in the week) in the respective country are shown. The contents of each page meeting the criteria were copied and this data then cleaned in Notepad and MS Excel. The only intervention into the text was deleting underscores where these clearly distorted the token count (cases such as *Cory_Chase_in_Mom_Has_a_Secret_HD 2*). To simplify the navigation, every title was enclosed in a structural tag (<t> *text* </t>) and the corpus was then uploaded into the Sketch Engine (Kilgarriff et al. 2014), during which it was lemmatised, PoS-tagged and duplicates were deleted. Since a lot of the video titles in the corpus had all words beginning with a capital letter, which in turn obscured the PoS-tagging, thereby diminishing the quality of word sketches (Kilgarriff et al. 2010), it was decided that another version of the corpus be created completely in lower case. After some initial testing, this version provided more reliable results, which was especially evident in the word sketches. The finished corpus comprises of 16,979 titles and totals 108,572 words (117,005 tokens). Mean title length was 6.89 (SD = 2.85) tokens with the minimum and maximum being 1 and 21 respectively (see Figure 1 for the distribution). These figures are in line with the evidence on headlines (Dor 2003), i.e. pornographic video titles tend to be relatively short.

Figure 1: Distribution of titles in the corpus by their token length
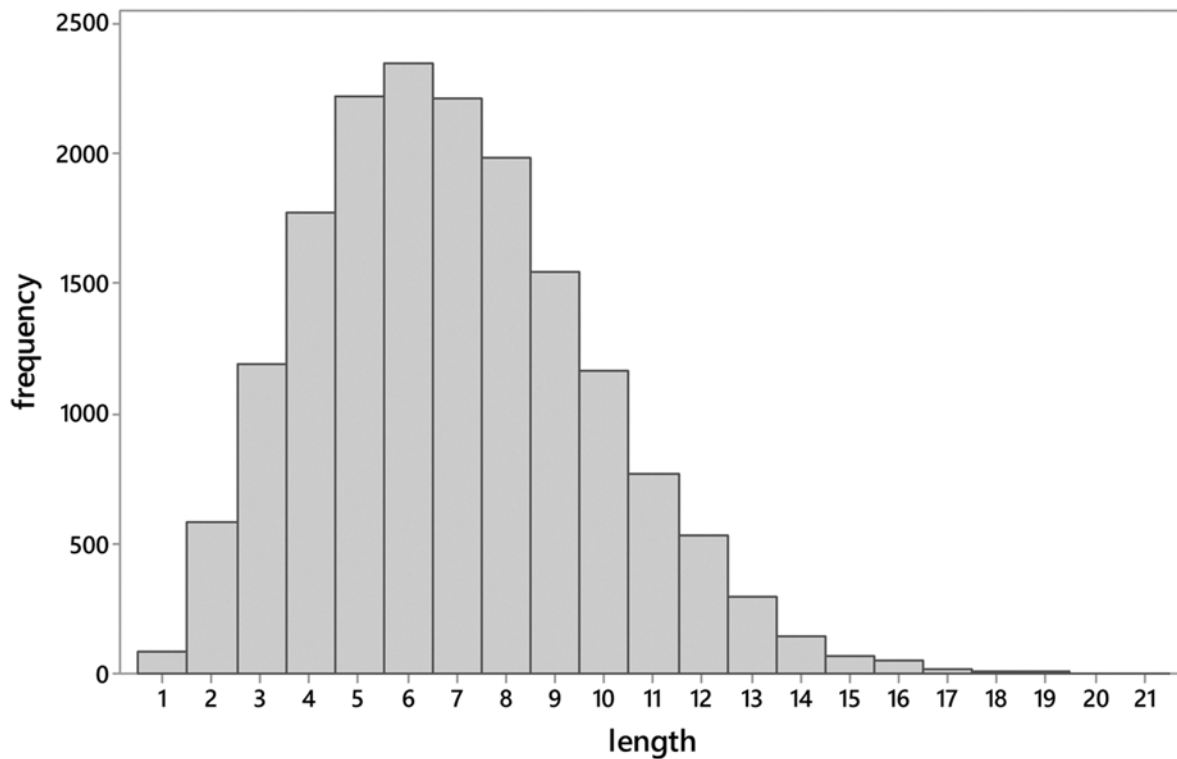
Frequent words

To obtain basic information about the corpus, the list of most frequent 100 lemmas was compiled. Absolute as well as normalised (to 1,000 tokens) frequencies are shown in the table below. Curiously enough, even grammar/function words are outnumbered by the lemma *fuck*. Also, there are, predictably, a lot of references to the domain of sex, 39 % of which also occur in Pihlaja's (2016) most frequent 100 words from the commentaries. The main difference are words related to categorisation within the genre (*amateur*, *orgy*, *massage*), accompanied by information about the participants (*blonde*, *brunette*, *busty*, *cute*). A lot of these characteristics have become categories of their own (*lesbian*, ma*ture/mom/MILF*, *teen*). In line with the data regarding the male-female differences in SEIM consumption, most of the words provide information about women as opposed to men. The only words that are explicitly male in nature are *cock*, *dick*, *guy* and *man*. Clearly, there is dissonance in role representation in the titles – this applies mostly to gender, but also to race and, to some extent, age.

Based on this list, the following two questions will be explored:

RQ1: What are the arguments of the verb *fuck*? Does their gender predict agency?

RQ2: How are male and female roles realised in the corpus?

Table 1: A list of the most frequent lemmas in the corpus with absolute and normalised frequencies

lemma	f_{abs}	f_{norm}	lemma	f_{abs}	f_{norm}	lemma	f_{abs}	f_{norm}	lemma	f_{abs}	f_{norm}
fuck	3533	30.20	*mom*	619	5.29	*suck*	343	2.93	*ebony*	231	1.97
and	2000	17.09	*sexy*	609	5.20	*have*	340	2.91	*party*	230	1.97
a	1723	14.73	*cum*	602	5.15	*horny*	332	2.84	*latina*	226	1.93
her	1711	14.62	*tit*	596	5.09	*good*	318	2.72	*cumshot*	225	1.92
in	1594	13.62	*lesbian*	582	4.97	*step*	298	2.55	*hd*	223	1.91
teen	1575	13.46	*my*	528	4.51	*ride*	293	2.50	*catch*	222	1.90
get	1522	13.01	*creampie*	497	4.25	*up*	289	2.47	*sister*	218	1.86
the	1377	11.77	*hard*	496	4.24	*Asian*	289	2.47	*couple*	215	1.84
cock	1243	10.62	*of*	496	4.24	*friend*	278	2.38	*me*	214	1.83
hot	1240	10.60	*love*	489	4.18	*orgasm*	269	2.30	*make*	214	1.83
big	1229	10.50	*dick*	479	4.09	*girlfriend*	267	2.28	*mature*	209	1.79
with	1203	10.28	*compilation*	475	4.06	*two*	265	2.26	*brazzer*	205	1.75
anal	902	7.71	*young*	457	3.91	*guy*	265	2.26	*man*	204	1.74
black	877	7.50	*take*	451	3.85	*scene*	261	2.23	*college*	201	1.72
sex	867	7.41	*busty*	446	3.81	*porn*	258	2.21	*video*	192	1.64
girl	867	7.41	*amateur*	431	3.68	*it*	254	2.17	*bang*	189	1.62
ass	829	7.09	*babe*	417	3.56	*she*	253	2.16	*orgy*	188	1.61
on	823	7.03	*wife*	416	3.56	*his*	253	2.16	*son*	186	1.59
be	796	6.80	*first*	399	3.41	*at*	246	2.10	*cute*	184	1.57
to	763	6.52	*brunette*	397	3.39	*hardcore*	244	2.09	*from*	182	1.56
pussy	762	6.51	*squirt*	394	3.37	*daughter*	242	2.07	*old*	182	1.56
blonde	739	6.32	*huge*	371	3.17	*i*	241	2.06	*bbc*	182	1.56
by	694	5.93	*threesome*	365	3.12	*white*	241	2.06	*time*	182	1.56
milf	670	5.73	*slut*	348	2.97	*massage*	239	2.04	*do*	180	1.54
for	644	5.50	*tight*	344	2.94	*blowjob*	232	1.98	*beautiful*	177	1.51

Agency analysis

Compared to Lischinsky's (2017) research, which dealt with fictional erotica, thus analysing more different verbs, the video titles feature predominantly the verb *fuck* (the second such verb being *bang* with only 117 hits), which is why the investigation is confined to the arguments of *fuck* only.

Procedure

The lemma *fuck* was searched in the corpus, separated into nouns ($n = 727$) and verbs ($n = 2806$). Both categories were manually checked for mis-tagged instances during the analysis, which yielded 2597 titles that were tested. Priority was given to recording the objects of the verbs rather than subjects. The former exhibit more varied and therefore more informative categorisation, whereas the latter were easier to reliably infer. The following table gives an overview of the categories.

Table 2. The categories used in the object analysis with examples

category	examples
male	*Alena Croft Fucks Her Step Son's Black Friend* *Phoenix Marie & Ava Addams (wait youre fucking erik) too? anal big tits*
female	*POVD Sexy Asian teen gets fucked wet after shower* *Shady Guys Fuck Little Shy Teen In Hotel*
male body part	*Pregnant teen fucks two cocks gingerly* *ExxxtraSmall Petite teen deepthroats fucks huge cock*
female body part	*Beauty Harley Dean gets pierced pussy fucked* *Gianna Michaels Gets Her Huge Natural Tits Fucked By Manuel Ferrera*
cooperative	*amateur couple fucking* *Pakistani Actress Fucking In Hotel Room With Director*
intransitive [SUBJ=M]	*His first threesome and he was fucking the same really.* *If You Won't Fuck, HE Will!*
intransitive [SUBJ=F]	*HD - POVD Hot teen is fucking by the pool in POV* *Girlfriend With Big Tits Sucks & Fucks*
lesbian	*Horny lesbian mistress anal strapon fucks her maid* *Christy Canyon and Ginger lynn scissor fucking*
reflexive	*Ami Emerson Fucks Machines* *ass fucking before i fucked my creamy pussy into a huge squirt*

It is worth noting that 104 titles fell into two or more categories; this was mostly the case in the lesbian category. *Bad cop fucks good prisoner* is thus both lesbian and female. Also, if the sex of the performers was difficult to infer from the title alone (as in the previous example), the video was searched, and the sex of the participants established. Besides the categorisation, diathesis (active/passive voice) of *fuck* was identified in each title, so that the titles featuring the participial word form *fucked* were labelled passive.

Results and discussion

The following table presents the counts of the titles in each category. The male:female (person) patient ratio (discounting the lesbian overlap) is 0.32:1, which is quite close to Lischinsky's (2017) 0.23:1, showing a clear preference towards female passivity across the two different text types. The differences between body parts as objects, on the other hand, indicates that the situation is more balanced, to say the least. Looking at the object's gender and where the agent directs their activity (body part or person) reveals a significant association ($\chi^2 = 123.08$; df = 1; p <.0001) with a moderate effect (Cramér's V = 0.24). A more fine-grained semantic analysis of the objects in the male (body part) patient category reveals that a surprisingly high number of the objects are marked for either a FAMILY MEMBER/(AGE), as in *Allysin chaynes fucks a dad and step-son* and *Fucking My Friends Son.* In comparison, only 4.1 % of all the sexual encounters were characterised by being cooperative. This tendency, albeit less prominent, was shared in the lesbian category – 72 instances (74 %) had a clear agent. The greatest discrepancy was between the intransitive descriptions of the sex acts. Women are visibly prioritised in instances where there is only participant mentioned in the title, which most likely reflects the viewer's (predominantly men) expectations.

Table 3: Categorisation counts and percentages

category	n	%
male (person)	457	16.8
female (person)	1495	55.1
male body part	142	5.2
female body part	108	4.0
cooperative	110	4.1
intransitive [SUBJ=M]	2	0.1
intransitive [SUBJ=F]	279	10.3
lesbian	97	3.6
reflexive	24	0.9

The second aspect of the analysis, diathesis, namely the use of the passive, was significantly associated with the gender of the patient ($\chi^2 = 644.92$; df = 1; p <.0001) showing a large effect (Cramér's V = 0.58). Out of 454 titles in the male category (3 discounted as they were also female), only 16 used the passive. In females the picture is vastly different – 1064 out of 1492 are described in the passive. Accordingly, there is a wholly new dimension to the crude "[m]*en do the fucking*[,] [w]*omen get fucked*" (Lischinsky 2017: 167). At least in the realm of online pornography, it could be made more precise by stating that women also sometimes fuck men, but men (almost) never get fucked.

I present two explanations of the phenomenon: There are still constructed imbalances between male and female representations in pornography. These are non-egalitarian in that women are portrayed as passive receivers and men as agents of the sexual intercourse. Even in cases where this is not supported by valency, it is nevertheless demonstrated by diathesis. Additionally, putting women in the foreground caters mostly to males and to some extent ignores women's experiences of sexual intercourse, which, among other things, may account for the differences in SEIM consumption (see above).

Alternatively, the language used in the titles is a mere representation of the logistics of sexual intercourse. It follows from this that the active-passive dichotomy operationalised linguistically can provide very limited information. The amount of physical movement during sex could be a potential variable that could explain more. The same goes for initiating sexual intercourse, where there are differences between pornography (see the studies cited in Klaassen, Peter 2014) and real life (Grøntvedt, Kennair, Mehmetoglu 2015; Kennair et al. 2009). As such, the linguistic evidence could be a reminder that people may take on more roles during a single sexual encounter. As the minute cooperative category suggests, there is always somebody "in charge". From this perspective, it seems wise not to try to conflate agency and pleasure.

Of Wives and Men: Performed Roles

Procedure

In this section, role representations of males and females in the titles are analysed and contrasted. Since roles are linguistically constructed mainly with the help of nouns and adjectives, one frequency (lemma) list was compiled for each. The minimum frequency was set to 20 occurrences in the corpus. Words that did not describe roles, i.e. did not refer to people or their qualities, were deleted. This process yielded 223 lemmas (171 nouns; 52 adjectives), 8 of which appeared in both lists due to mis-tagging. On closer inspection, the duplicates did not exhibit any differences

in meaning and were therefore aggregated under the label "ADJ+N", reducing the total to 215 lemmas. Further two-layered, gender categorisation ensued, dividing the lemmas first according to their typical denotative meaning to three groups: "male", "female", and "shared". Accordingly, *dude* and *bf* were assigned "male", *sister* and *busty* "female", and finally, *nerdy* and *doctor* to "shared". The last of these, which contained words that do not have a default-gender referent, became the subject of the second analytical layer; the words were observed in context and the gender of their real-world referents was determined. This was made possible using the word sketch function, context cues in the titles, and, in a small minority of cases, searches for the videos themselves. Despite this, 9.87 % (889) of the expressions remained unexplained. This was principally due to plurals and sometimes lack of contextual clues to establish, who the expressions refer to. Finally, semantic classification of expressions was carried out; when undecided on the class (henceforth in SMALL CAPITALS), WordNet (Princeton University 2010) was employed to supply ontological information and the decision was based on the hypernym.

To ensure the validity of the results, two further categories were introduced: "nationality" and "proprium". The first was devised to counter the skewness of the data towards females, thus maximising the differences but not providing any interpretative value. As we shall see, the representation of female referents in the corpus is much larger than male. The category of nationalities (words like *Czech*, *Indian* etc.) is as a whole symptomatic of this divide and therefore left out. Gendered national references, such as *latina*, were, however, preserved. The same motivations underlay the creation of "proprium". Female names (of the popular performers) account for as many as 37 % of the "female" category, but since they tell us virtually nothing about the roles, the category was also omitted from the analysis.

Results and Discussion

The table below provides frequencies for the gender categorisation in the corpus.

Table 4: Categories and gender differences

category		tokens	proportion	types	proportion
males (defaults)		1672	.17***	15	.25**
females (defaults)		8313		44	
shared	males	649	.07***	34	.38*
	females	8361		55	
total	males	2321	.12***	49	.33***
	females	16674		99	
unused	propria	3092		85	
	nationalities	1004		15	

binomial test (0.5) ***p <.0001; ** p <.001; * p <.05

As the table also suggests, the male and female proportions are far from being equally represented, which is true of both types and tokens, indicating that not only the number but also the variety of roles is larger for females. Nevertheless, it must be added that given the cut-off frequency (20), these are not entirely independent of each other.

Regarding shared roles, 56 were found and males were not represented in 22 of them at all (see Appendix). In order to guarantee a fair comparison, the absolute values of the female roles were weighted by the ratio of the gender-default expressions (1672[M]/8313[F] = 0.2011). This way, the overall gender representation was factored in, and the results were consequently more meaningful.

Two data matrices were created; a larger 56×2 one (see Appendix), which featured all the counts, and its smaller 35×2 alternative, which did not contain pairs where one value equalled zero, i.e. where the identifier was not gendered in the studied sample. Although it would have had little bearing on the results, the customary Fisher's exact test was replaced with Wilcoxon rank-sum test as a possible alternative proposed by Lijffijt et al. (2016). The test showed a significant difference ($Z = 2346$, $p < 0.0001$) between males and females. To sum up, the descriptive language indeed seems to be used differently in relation to gender.

If we look at the roles more closely, it is evident that only very few would exhibit equal distribution. Among these with frequencies higher than 10 are *old*, *young*, *white* (all $p > 0.1$; binomial test) and theoretically *teacher* (provided we are being more conservative; $p = 0.0169$). It follows from this that AGE (and partly RACE) seems to be the only role common to both groups. As for predominantly male (weighted counts) roles in the "shared" category, these include *friend*, *lover*, *boss*, *client*, *cop*, *neighbor*, *student* and *trainer*.

The complete profile of identifying devices for males in the corpus can be seen in Table 5.

Table 5: Semantic classes of identifiers referring to males (min. freq. = 5)

category	n	identifiers
RELATION	1223	*son* (186), *brother* (151), *friend* (129), *husband* (127), *boyfriend* (121), *stepson* (108), *dad* (104), *daddy* (85), *bf* (73), *step-dad* (65), *father* (30), *lover* (22), *neighbor* (22)
AGE	179	*boy* (85), *young* (55), *old* (39)
JOB	184	*student* (57), *teacher* (46), *boss* (46), *doctor* (32), *cop* (19), *driver* (19), *trainer* (11)
ATTITUDE	6	*horny* (6)
GENERIC	482	*guy* (265), *man* (190), *dude* (27)
RACE	72	*black* (51), *white* (21)
LAUDATIVE	55	*stud* (55)
MISCELLANEOUS	39	*client* (39)

If they are mentioned at all, men are characterised mostly based on their relationship to their sexual partner(s), which is suggestive of a narrative that strengthens the bond between the participants. This provides the video with a frame and facilitates a plot of some sort. A non-negligible proportion of these are references to fauxcest, highlighting the taboo aspect of consuming pornography. Most importantly, however, the roles of a male participant aids better identification of the woman. Considering that the second most searched term on Pornhub in 2018 was *milf* (Pornhub 2018), it is apparent that the definition of a concept like this would have to be reinforced by contextual means. Adding a male participant in the role of a relative is one way of achieving this. Furthermore, it is this kind of information that can only have language as its vehicle – it would be impossible to convey it via thumbnails, for example. Aside from RELATION, men are also identified generically. Relative to the other classes, the number of these references is comparatively high – this again points to the minor role of male identification – it is deemed sufficient that they are identified as male only. AGE and JOB are telling to some degree, but parallel to the previous categories, they complement the female roles in focus, e.g. *young/boy-MILF/mom*, *student-teacher*, *patient-doctor*. Regarding RACE and LAUDATIVE, the first is used almost exclusively (also as a woman identifier – see Table 6 below) in videos depicting interracial intercourse, whereas the second comprises only one word indicating that the only positive quality in males is their sexual prowess.

For women, the situation is markedly different. As Table 6 below outlines, there are more semantic classes with more identifiers in them, testifying to greater complexity of female characters.

Table 6: Semantic classes of identifiers referring to females (min. freq. = 5)

	category	n	identifiers
mental	RELATION	3424	*MILF* (653), *mom* (619), *wife* (416), *girlfriend* (267), *daughter* (242), *step-mom* (220), *sister* (218), *step-sister* (146), *friend* (129), *gf* (126), *mother* (112), *step-daughter* (111), *granny* (53), *mommy* (52), *stranger* (30), *neighbour* (18), *lover* (12)
	AGE	2320	*teen* (1549), *young* (369), *mature* (138), *old* (136), *nubile* (73), *cougar* (55)
	JOB	1104	*pornstar* (193), *babysitter* (155), *teacher* (125), *schoolgirl* (118), *nurse* (70), *student* (69), *maid* (65), *housewife* (54), *model* (51), *cheerleader* (47), *secretary* (43), *masseuse* (26), *stripper* (26), *doctor* (24), *boss* (11), *cop* (10), *trainer* (9), *driver* (8)
	ATTITUDE	1119	*horny* (324), *amateur* (292), *naughty* (107), *dirty* (93), *wild* (45), *kinky* (44), *shy* (38), *virgin* (32), *nasty* (24), *innocent* (20)
	CHARACTER	47	*crazy* (26), *nerdy* (21)
	GENERIC	1145	*girl* (867), *chick* (151), *woman* (92), *lady* (35)
	PEJORATIVE	615	*slut* (348), *whore* (103), *bitch* (91), *slutty* (73)
visual	RACE/ORIGIN	1059	*black* (421), *latina* (226), *ebony* (224), *white* (165), *filipina* (23)
	HAIR	1311	*blonde* (765), *brunette* (397), *redhead* (149)
	BODY	972	*busty* (446), *petite* (147), *skinny* (108), *tiny* (83), *little* (75), *curvy* (35), *chubby* (31), *fat* (29), *flexible* (18)
	LAUDATIVE	2883	*hot* (1113), *sexy* (595), *babe* (417), *cute* (180), *beautiful* (168), *beauty* (87), *gorgeous* (86), *stunning* (79), *sweet* (47), *hottie* (45), *pretty* (23), *lovely* (22), *bombshell* (21)
MISCELLANEOUS		775	*lesbian* (582), *drunk* (52), *pregnant* (42), *slave* (30), *queen* (30), *naked* (17), *swinger* (13), *client* (9)

A closer examination of the RELATIONS demonstrates the popularity of videos, in which the participants are related, more specifically, with an age gap of one generation between them. Much more than men, women are also identified by means of age, and their momentary attitude to sex. Both are primarily used to attract the viewer, the latter perhaps to encourage arousal by suggesting that the women might be less restrained in the expression of their sexuality (*horny, wild, naughty, dirty*), which justifies transitions in the direction of less conventional practices, such as BDSM and anal sex. Alternatively, descriptions like *shy, virgin, innocent* (and quite often *amateur*) invoke the fictitious narrative of submissiveness, which would appeal to certain groups of viewers. The submissiveness is partly mirrored in *job* identifiers selected to instil a sense of hierarchy. This is especially evident upon comparison with the previous table as this allows for establishing pairs like: *student ↔ teacher, doctor ↔ nurse, trainer ↔ client*, and *boss ↔ secretary*. Together with valency, this is another piece of evidence of a noticeable tendency to linguistically construe sex as an act of dominance of one participant over the other. This notion is connected to a class inherent to female descriptions, PEJORATIVE.

Defamatory labels for women are quite ubiquitous in the corpus and they seem to be used relatively haphazardly, although some loose patterns have been found: cases where women engage in an unconventional practice (*Turning my WIFE into a BI SLUT*; *Big booty latina deepthroat anal whore*), direct speech as an example of "dirty talk" (*Ride my cock hard old blonde slut!*; *I Am A Big Slut*), and means of distancing (*slut fucked in public*; *Slut Fucked On The Bus*; *Teen slut lovin' the cock*). Strangely enough, combinations of mental PEJORATIVES and visual

LAUDATIVES are not entirely uncommon (*Beautiful Ass Slut Gets Gangbanged*; *Incredibly Hot Arab Slut Shows Her Awesome Sucking And Fucking Sk[ills]*; *Awesome whore rides biggest cock ever*). The only identifiers that tend to repel PEJORATIVES are the "modest" attitudinal ones (*shy, virgin, innocent*). In this context, it is nearly impossible to escape the conclusion that whatever women's behaviour and attitude to sex is, they can always be labelled as *sluts*, especially if their enjoyment and/or experience are put on display or even implied (e.g. by their age). To that, one may object that dominance and submission are part of sexual life for some, and this is certainly the case for a number of the titles and the corresponding videos. In most of them, however, the use of the expletives is simply unwarranted, apart from, perhaps, being "inviting" to the consumer. In fairness, such extreme measures to attract attention were not taken in the vast majority of titles.

Conversely, the positively loaded LAUDATIVES were three times as likely to be used. The group is comprised of vague positive adjectives typically describing physical appearance, the most frequent of which being evocative of excitement and arousal. These elements are, on the other hand, completely absent from *cute*, *sweet*, and *lovely*, whose meanings arguably extend beyond the physical. Out of all the classes, LAUDATIVES presumably add the least information value; first, they are very subjective, hence unreliable, and second, their practical usefulness verges on redundant when one realises that the participants can be seen in the thumbnail. As such they would violate three points of Dor's (2003: 717) headline functions taxonomy. Ifantidou (2009) however takes an opposing view as she claims that "readers seem to value headlines for what they are, i.e. underinformative, creative, yet autonomous texts." Their utility also seems to lie in the effect of confirmation of the thumbnail. To put it bluntly, people might be more likely to click on *amateur hot latina extreme creamy pussy* as opposed to *amateur latina extreme creamy pussy* purely because the former primes the viewer for the visual beauty.

As for RACE/ORIGIN, the frequencies of the identifiers do not correlate with the distribution of races in the videos, although this could not be demonstrated empirically (inspecting every single video and crosschecking the roles was not a feasible option). This discrepancy suggests that white people are more often seen as unmarked, whereas other ethnicities as marked. The causes are likely the prevalence of white participants in the videos and intercourse between different ethnicities. RACE thus becomes accentuated only when the perceived ethnicities do not match. The most typical combination in this regard would be a black female and white male, that is where both roles are explicitly mentioned. The opposite combination tends to be verbalised slightly differently; e.g. *white college student gets **blacked***; *aching pussy rides longest **bbc** in the world – multiple angles*).

Similar to RACE, HAIR provides the viewer with very little when it comes to information not inferable directly from the thumbnail. Take for example the following two titles: *Brunette beauty fucked in the kitchen*; *Big-tit brunette MILF comes home to find college boy in her bed*. Since most women in the videos in the corpus have dark hair, I would argue that in cases like these, the only added value is the construction of artificial familiarity to the viewer. The more information is known about the (especially female) participants, the higher the chance that (especially male) viewers would be tempted to click on the video. The situation is slightly different with the identifier *blonde*, which could be interpreted as emphasising a desirable trait. The percentage of individuals with blonde hair is low in the population, and blonde women seem to be perceived as more attractive (Guéguen 2012), which would also explain the frequency of the identifier in contrast to the others in the class.

BODY offers a glimpse into what factors constitute "attractive", namely breast size, body size, and body weight. Note that opposites are represented, which makes the factor rather axial with individual preferences being spread on the axes. Breast size is the most frequent with 446 hits for

busty. It is also stressed when one observes the collocates of *tit* and *boob* in the corpus; the first two both are *big* and *huge* for both. Size is another dimension, of which only one side is mentioned – women, if identified based on their size, are mostly *small*, *tiny*, *little* and *petite*. Body weight, on the other hand, is distinguished from the previous two that both sides are featured as identifiers – *skinny* vs. *chubby, fat.* On the whole, the very existence of this class only as female identifiers is symptomatic of the fact that women, unlike men, are identified based on their looks. The only exception to this are identifiers of genitals, which are significantly more prominent with reference to male genitalia. This match between the focus of linguistic description and common, gender-specific anxieties regarding the body may, to some degree, be indicative of a contribution to the insecurities both sexes have about their own bodies and reinforce a distorted image of what an average person looks like. One may raise an objection that this is not a matter of language. I would argue that the combination of LAUDATIVE and BODY identifiers together with the images is certainly stronger than the effect of the images alone, since axiological statements are paired with the images. By that logic, if *hot/cute* is *busty* and the image is that of a woman with an above-average chest, this then pushes the notion of what hot/cute refers to in the real world, especially in sexual contexts.

To conclude, discernible differences on both quantitative and qualitative level were discovered with respect to how males and females are identified. Women are on the whole identified more than men and there are categories that appear to be exclusively female. These include the following classes: LAUDATIVE, HAIR, PEJORATIVE, ATTITUDE and CHARACTER. There are multiple reasons for this. Some gender stereotyping would certainly play a role (especially in PEJORATIVE, HAIR), some differences might be explained by being lexicalised in different ways in women and men (BODY), others simply by the generally heavier focus on women with the underlying assumption of a man as a typical viewer. Due to these factors, women are, nevertheless, described in a more complex fashion than men.

Conclusion

The study has empirically investigated the text type of pornographic video titles. It focused on agency and the properties of identifying expressions in relation to the gender of the participants in the videos. For agency, the central finding is that sex is almost always described asymmetrically, that is one participant being active and the other passive. As the analysis of the valency of the verb *fuck* revealed, females are much more likely to occur in the object position than males. If the subject of the verb was female, passive voice was also significantly more likely to be used than with men. This suggests that the linguistic framing of the porn video narrative mirrors the tendencies previously identified in the videos themselves. These include unequal gender norms and the hierarchical perspective on sexual intercourse, specifically the notions of female passivity and male dominance.

In the second question, I looked at the properties of identifying expressions used in the titles, again with gender in mind. A number of semantic classes were used to group the expressions and their meaning for the construal of male and female roles in pornography titles was discussed. Women's roles are linguistically realised much more frequently, and they are descriptively deeper; more categories are used to refer to women's roles. This suggests that for the creators/uploaders of the videos, women are the central part of the narrative rather than men. Not only does this reflect the expectations and desires of the predominantly male viewership, but it simultaneously excludes women who are therefore less likely to be interested in watching such pornography. As with the first question, a lot of phenomena observed in previous research on pornography have been demonstrated to also be true for the language surrounding it. A number of traditional

hierarchical scenarios were discovered (*teacher – student*; *cop – driver*; *boss – secretary*). A lot of attention was paid to the physical attractiveness of the female participant, as evidenced by the high number of labels in the "visual" category. Female participants were thus objectified, in that their identities were partly reduced to their bodies/appearance. Furthermore, the category PEJORATIVE, which included derogatory terms like *slut* or *bitch*, was only ascertained for females. This again showcases gender that the stark gender inequalities in pornography as reported in previous research are also frequently realised by linguistic means.

The study has shed light on the disparities in the treatment of male and female roles in online pornography. Due to the great pervasiveness of these materials, it can be concluded that the language used might contribute to the shift in views on the two sexes. Especially teenagers might start to construct their sexual roles and expectations around the language schemes that the videos are framed within. Men may expect unrealistic attitudes and develop unrealistic ideals of beauty, while at the same time being anxious about their performance, whereas women might fear being objectified, unsure of the expectations placed upon them, or having a distorted body image. Alternatively, young women might instead embrace the objectification and reproduce it in their sexual behaviour, which is, however, equally undesirable.

Acknowledgements

Bibliography

Alexa. 2019. "The top 500 sites on the web." [online] *Alexa.com* [retrieved 6. 24. 2019]. https://www.alexa.com/topsites.

Attwood, F. 2002. "Reading Porn: The Paradigm Shift in Pornography Research." *Sexualities* 5 (1): 91–105, https://doi.org/10.1177/1363460702005001005.

Attwood, F. 2010. "Dirty work: researching women and sexual representation." Pp 177–187 in R. Ryan-Flood, R. Gill (eds.). *Secrecy and Silence in the Research Process: Feminist reflections.* Oxon: Routledge.

Attwood, F., C. Smith. 2014. "Porn Studies: an introduction." *Porn Studies* 1 (1–2): 1–6, https://doi.org/10.1080/23268743.2014.887308.

Baker, P. 2006. *Using Corpora in Discourse Analysis.* London: Continuum, https://doi.org/10.5040/9781350933996.

Baker, P. 2014. *Using Corpora to Analyze Gender.* London: Bloomsbury.

Baron, L. (1990). "Pornography and Gender Equality: An Empirical Analysis." *The Journal of Sex Research* 27 (3): 363–380, https://doi.org/10.1080/00224499009551566.

Blais-Lecours, S., M.-P. Vaillancourt-Morel, S. Sabourin, N. Godbout. 2016. "Cyberpornography: Time Use, Perceived Addiction, Sexual Functioning, and Sexual Satisfaction." *Cyberpsychology, Behavior, and Social Networking* 19 (11): 649–655, https://doi.org/10.1089/cyber.2016.0364.

Carroll, J. S., D. M. Busby, B. J. Willoughby, C. C. Brown. 2016. "The Porn Gap: Differences in Men's and Women's Pornography Patterns in Couple Relationships." *Journal of Couple & Relationship Therapy* 16 (2): 146–163, https://doi.org/10.1080/15332691.2016.1238796.

Caputi, J. (2014). "The Pornography of Everyday Life." Pp 373–385 in G. Dines (ed.): *Gender, Race, and Class in Media: A Critical Reader.* Thousand Oaks: SAGE.

Carotte, E. R., A. C. Davis, M. S. Lim. 2020. "Sexual Behaviors and Violence in Pornography: Systematic Review and Narrative Synthesis of Video Content Analyses." Journal of Medical Internet Research 22 (5): 1–21, https://doi.org/10.2196/16702. Ciclitira, K. 2004. "Pornography, Women and Feminism: Between Pleasure and Politics." *Sexualities* 7 (3): 281–301, https://doi.org/10.1177/1363460704040143.

Cowan, G., C. Lee, D. Levy, D. Snyder. 1988. "Dominance and inequality in X-rated videocassettes." *Psychology of Women Quarterly* 12 (3): 299–311, https://doi.org/10.1111/j.1471-6402.1988.tb00945.x.

Daspe, M.-È., M.-P. Vaillancourt-Morel, Y. Lussier, S. Sabourin, A. Ferron. 2018. "When Pornography Use Feels Out of Control: The Moderation Effect of Relationship and Sexual Satisfaction." *Journal of Sex & Marital Therapy* 44 (4): 343–353, https://doi.org/10.1080/0092623X.2017.1405301.

Dines, G. 2017. "Growing Up With Porn: The Developmental and Societal Impact of Pornography on Children." *Dignity: A Journal on Sexual Exploitation and Violence* 2 (3): 1–9, https://doi.org/10.23860/dignity.2017.02.03.03.

Dor, D. 2003. "On newspaper headlines as relevance optimizers." *Journal of Pragmatics* 35 (5): 695–721, https://doi.org/10.1016/S0378-2166(02)00134-0.

Dworkin, A. 1981. *Pornography: Men possessing women.* London: Women's Press.

Dziadosz, S., R. Chandrasekar. 2002. "Do Thumbnail Previews Help Users Make Better Relevance Decisions about Web Search Results?" *SIGIR'02*: 365–366. Tampere, https://doi.org/10.1145/564376.564446.

Escoffier, J. 2007. "Scripting the Sex: Fantasy, Narrative, and Sexual Scripts in Pornographic Films." Pp 61–79 in M. Kimmel (ed.). *The Sexual Self: The Construction of Sexual Scripts.* Nashville: Vanderbilt University Press, https://doi.org/10.36019/9781978820180-006.

Gabriel, K. 2016. "The Subject of Porn Research: Inquiring Bodies and Lines of Resistance." Pp 311–327 in W. Harcourt (ed.). *Bodies in Resistance: Gender and Sexual Politics in the Age of Neoliberalism.* London: Palgrave Macmillan, https://doi.org/10.1057/978-1-137-47780-4_15.

Gagnon, J. H., W. Simon. 1973. *Sexual Conduct: The Social Sources of Human Sexuality.* Chicago: Aldine Publishing Company.

Grøntvedt, T. V., L. E. Kennair, M. Mehmetoglu. 2015. "Factors predicting the probability of initiating sexual intercourse by context and sex." *Scandinavian Journal of Psychology* 56 (5): 516–526, https://doi.org/10.1111/sjop.12215.

Guéguen, N. 2012. "Hair Color and Courtship: Blond Women Received More Courtship Solicitations and Redhead Men Received More Refusals." *Psychological Studies* 57 (4): 369–375, https://doi.org/10.1007/s12646-012-0158-6.

Hald, G. M., N. N. Malamuth, T. Lange. 2013. "Pornography and Sexist Attitudes Among Heterosexuals." *Journal of Communication* 63: 638–660, https://doi.org/10.1111/jcom.12037.

Hiramoto, M. 2015. "Who's really normal? Language and sexuality in public space." *Journal of Language and Sexuality* 4 (2): 183–192, https://doi.org/10.1075/jls.4.2.01hir.

Hyde, J. S. 2014. "Gender Similarities and Differences." *Annual Review of Psychology* 65 (1): 373–398, https://doi.org/10.1146/annurev-psych-010213-115057.

Hyde, J. S. 2016. "Sex and cognition: gender and cognitive functions." *Current Opinion in Neurobiology* 38: 53–56, https://doi.org/10.1016/j.conb.2016.02.007.

Ifantidou, E. 2009. "Newspaper headlines and relevance: Ad hoc concepts in ad hoc contexts." *Journal of Pragmatics* 41 (4): 699–720, https://doi.org/10.1016/j.pragma.2008.10.016.

Irvine, J. M. 2014. "Is sexuality research "dirty work"? Institutionalized stigma in the production of sexual knowledge." *Sexualities* 17 (5–6): 632–656, https://doi.org/10.1177/1363460713516338.

Jackson, C. A., A. Baldwin, B. G. Brents, P. J. Maginn,. 2019. "EXPOsing Men's Gender Role Attitudes as Porn Superfans." *Sociological Forum* 34 (2): 483–500, https://doi.org/10.1111/socf.12506.

Jones, L., S. Mills. 2014. "Analysing agency: reader responses to Fifty Shades of Grey." *Gender and Language* 8 (2): 226–244, https://doi.org/10.1558/genl.v8i2.225.

Keilty, P. 2016. "Embodied engagements with online pornography." *The Information Society* 32 (1): 64–73, https://doi.org/10.1080/01972243.2015.1107162.

Kennair, L. E., D. Schmitt, Y. L. Fjeldavli, S. K. Harlem. 2009. "Sex Differences in Sexual Desires and Attitudes in Norwegian Samples." *Interpersona: An International Journal on Personal Relationships* 3 (1): 1–32, https://doi.org/10.5964/ijpr.v3isupp1.67.

Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, V. Suchomel. 2014. "The Sketch Engine: ten years on." *Lexicography* 1: 7–36, https://doi.org/10.1007/s40607-014-0009-9.

Kilgarriff, A., V. Kovář, S. Krek, I. Srdanovič, C. Tiberius. 2010. "A quantitative evaluation of word sketches." *Proceedings of the 14th EURALEX International Congress*: 372–379.

Klaassen, M. J., J. Peter. 2014. "Gender (In)equality in Internet Pornography: A Content Analysis of Popular Pornographic Internet Videos." *The Journal of Sex Research* 52 (7): 721–735, https://doi.org/10.1080/00224499.2014.976781.

Kohut, T., J. L. Baer, B. Watts. 2016. "Is Pornography Really about 'Making Hate to Women'? Pornography Users Hold More Gender Egalitarian Attitudes Than Nonusers in a Representative American Sample." *Journal of Sex Research* 53 (1): 1–11, https://doi.org/10.1080/00224499.2015.1023427.

Koller, V. 2015. "The subversive potential of queer pornography: a systemic-functional analysis of a written online text." *Journal of Language and Sexuality* 4 (2): 254–271, https://doi.org/10.1075/jls.4.2.04kol.

Lijffijt, J., T. Nevalainen, T. Säily, P. Papapetrou, K. Puolamäki, H. Mannila. 2016. "Significance testing of word frequencies in corpora." *Literary and Linguistic Computing* 31 (2): 374–397, https://doi.org/10.1093/llc/fqu064.

Lischinsky, A. 2017. "Doing the naughty or having it done to you? Agent roles in erotic writing." *Porn Studies* 5 (2): 156–174, https://doi.org/10.1080/23268743.2017.1301215.

Mazières, A., M. Trachman, J.-P. Cointet, B. Coulmont, C. Prieur. 2014. "Deep tags: toward a quantitative analysis of online pornography." *Porn Studies* 1 (1–2): 80–95, https://doi.org/10.1080/23268743.2014.888214.

McGowan, M. K. 2009. "On Pragmatics, Exercitive Speech Acts and Pornography." *Lodz Papers in Pragmatics* 5 (1): 133–155, https://doi.org/10.2478/v10016-009-0002-1.

McKee, A. 2005. "The objectification of women in mainstream pornographic videos in Australia." *The Journal of Sex Research* 42 (4): 277–290, https://doi.org/10.1080/00224490509552283.

McKee, A. 2014. "Humanities and social scientific research methods in porn studies." *Porn Studies* 1 (1–2): 53–63, https://doi.org/10.1080/23268743.2013.859465.

Mellor, E., S. Duff. 2019. "The use of pornography and the relationship between pornography exposure and sexual offending in males: A systematic review." *Aggression and Violent Behavior* 46: 116–126, https://doi.org/10.1016/j.avb.2019.02.003.

Mikorski, R., D. M. Szymanski. 2017. "Masculine norms, peer group, pornography, Facebook, and men's sexual objectification of women." *Psychology of Men & Masculinity* 18 (4): 257–267, https://doi.org/10.1037/men0000058.

Morrish, L., H. Sauntson. 2011. "Discourse and identity in a corpus of lesbian erotica." *Journal of Lesbian Studies* 15 (1): 122–139, https://doi.org/10.1080/10894160.2010.490519.

Motschenbacher, H. 2018. "Corpus linguistics in language and sexuality studies." *Journal of Language and Sexuality* 7 (2): 145–174, https://doi.org/10.1075/jls.17019.mot.

Owens, E. W., R. Behun, J. C. Manning, R. C. Reid. 2012. "The Impact of Internet Pornography on Adolescents: A Review of the Research." *Sexual Addiction & Compulsivity* 19: 99–122, https://doi.org/10.1080/10720162.2012.660431.

Partington, A., A. Duguid, C. Taylor. 2017. *Patterns and Meanings: Theory and practice in corpus-assisted discourse studies (CADS).* Amsterdam: John Benjamins.

Paveau, M.-A., F. Perea. 2014. "An Object of Discourse for Studies of Pornography." *Questions de communication* 26: 1–9, https://doi.org/10.4000/questionsdecommunication.9495.

Peter, J., P. M. Valkenburg. 2007. "Adolescents' Exposure to a Sexualized Media Environment and Their Notions of Women as Sex Objects." *Sex Roles* 56: 381–395, https://doi.org/10.1007/s11199-006-9176-y.

Peter, J., P. M. Valkenburg. 2011. "The Influence of Sexually Explicit Internet Material on Sexual Risk Behavior: A Comparison of Adolescents and Adults." *Journal of Health Communication* 16: 750–765, https://doi.org/10.1080/10810730.2011.551996.

Peter, J., P. M. Valkenburg. 2016. "Adolescents and Pornography: A Review of 20 Years of Research." *The Journal of Sex Research* 53 (4–5): 509–531, https://doi.org/10.1080/00224499.2016.1143441.

Petersen, J. L., J. S. Hyde. 2010. "A Meta-Analytic Review of Research on Gender Differences in Sexuality, 1993–2007." *Psychological Bulletin* 136 (1): 21–38, https://doi.org/10.1037/a0017504.

Pihlaja, S. 2016. "Expressing pleasure and avoiding engagement in online adult video comment sections." *Journal of Language and Sexuality* 5 (1) 94–112, https://doi.org/0.1075/jls.5.1.04pih.

Pornhub. 2018. "2018 Year in Review." [online] *Pornhub INSIGHTS* [retrieved 4. 7. 2019]. https://www.pornhub.com/insights/2018-year-in-review.

Poulsen, F. O., D. M. Busby, A. M. Galovan. 2013. "Pornography Use: Who Uses It and How It Is Associated with Couple Outcomes." *Journal of Sex Research* 50 (1): 72–83, https://doi.org/10.1080/00224499.2011.648027.

Princeton University. 2010. "About WordNet." [online] *WordNet* [retrieved 3. 7. 2019]. https://wordnet.princeton.edu.

Rothman, E. F., J. B. Beckmeyer, D. Herbenick, T.-C. Fu, B. Dodge, J. D. Fortenberry. 2021. "The Prevalence of Using Pornography for Information About How to Have Sex: Findings from a Nationally Representative Survey of U.S. Adolescents and Young Adults." *Archives of Sexual Behavior* 50 (2): 629–646, https://doi.org/10.1007/s10508-020-01877-7.

Smith, C. 2010. "Pornographication: A discourse for all seasons." *International Journal of Media and Cultural Politics* 6 (1): 103–108, https://doi.org/10.1386/macp.6.1.103/3.

Stella, R. 2018. "Porn culture, embodied experiences and knowledge of sexual practices." *Sexualities* 23 (3): 1–17, https://doi.org/10.1177/1363460718791964.

Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work.* Amsterdam: John Benjamins, https://doi.org/10.1075/scl.6.

Træen, B., T. S. Nilsen, H. Stigum. 2006. "Use of pornography in traditional media and on the internet in Norway." *Journal of Sex Research* 43 (3): 245–254, https://doi.org/10.1080/00224490609552323.

Vandenbosch, L., J. M. van Oosten. 2017. "The Relationship Between Online Pornography and the Sexual Objectification of Women: The Attenuating Role of Porn Literacy Education." *Journal of Communication* 67 (6): 1015–1036, https://doi.org/10.1111/jcom.12341.

Vangeel, L., S. Eggermont, L. Vandenbosch. 2020. "Does Adolescent Media Use Predict Sexual Stereotypes in Adolescence and Emerging Adulthood? Associations with Music Television and Online Pornography Exposure." *Archives of Sexual Behavior* 49: 1147–1161, https://doi.org/10.1007/s10508-020-01677-z.

Vannier, S. A., A. B. Currie, L. F. O'Sullivan. 2014. "Schoolgirls and Soccer Moms: A Content Analysis of Free "Teen" and "MILF" Online Pornography." *The Journal of Sex Research* 51 (3): 253–264, https://doi.org/10.1080/00224499.2013.829795.

Virdis, D. F. 2015. "Eroticizing female and male bodies: a linguistic investigation of a pornographic novel from the Victorian magazine The Pearl." *Porn Studies* 2 (1): 19–34, https://doi.org/10.1080/23268743.2014.984924.

Willis, M., A. J. Bridges, C. Sun. 2022. "Pornography Use, Gender, and Sexual Objectification: A Multinational study." *Sexuality & Culture* 26: 1298–1313, https://doi.org/10.1007/s12119-022-09943-z.

Wright, P. J. 2013. "A Three-Wave Longitudinal Analysis of Preexisting Beliefs, Exposure to Pornography, and Attitude Change." *Communication Reports* 26 (1): 13–25, https://doi.org/10.1080/08934215.2013.773053.

Wright, P. J., R. S. Tokunaga, A. Kraus, E. Klann. 2017. "Pornography consumption and satisfaction: A meta-analysis." *Human Communication Research* 43: 315–343, https://doi.org/10.1111/hcre.12108.

Yu, C., A. E. Kågesten, S. De Meyer, C. Moreau, M. v. Reeuwijk, C. Lou. 2021. "Pornography Use and Perceived Gender Norms Among Young Adolescents in Urban Poor Environments: A Cross-site Study." *Journal of Adolescent Health* 69 (1): 31–38, https://doi.org/10.1016/j.jadohealth.2021.03.008.

About author

Tomáš Mach is currently a PhD student at the Institute of the Czech National Corpus at Charles University. He also works at the Department of English (Jan Evangelista Purkyně University). His main areas of interest include academic language, corpus linguistics and language teaching. ORCID: https://orcid.org/0000-0002-8178-717X.

Appendix: *Comparison of Male and Female Referents in the Category "Shared"*

lemma	M$_{abs}$	F$_{weighted}$	lemma	M$_{abs}$	F$_{weighted}$
amateur	1	59	*model*	1	10
beautiful	0	34	*naked*	0	3
black	51	85	*nasty*	0	5
blonde	0	154	*naughty*	0	22
boss	46	2	*neighbor*	22	4
client	39	2	*nerdy*	0	4
cop	19	2	*old*	39	27
crazy	1	5	*parent*	1	0
curvy	0	7	*pornstar*	0	39
cute	0	36	*pretty*	0	5
dirty	2	19	*sexy*	0	120
doctor	32	5	*shy*	1	8
driver	19	2	*skinny*	0	22
drunk	1	10	*slave*	0	6
ebony	0	45	*stranger*	0	6
fat	2	6	*stripper*	3	5
flexible	2	4	*student*	57	14
friend	129	26	*stunning*	0	16
gorgeous	0	17	*sweet*	0	9
horny	6	65	*swinger*	0	3
hot	5	224	*teacher*	46	25
chubby	2	6	*teen*	3	312
innocent	2	4	*tiny*	0	17
kinky	0	9	*trainer*	11	2
little	3	15	*virgin*	2	6
lovely	0	4	*white*	21	33
lover	22	2	*wild*	1	9
mature	2	28	*young*	55	74