1. Introduction

Normativity is a central mechanism affecting the discursive construction of sexual identities, desires and practices. It stipulates which sexual aspects are deemed preferable or normal vs. stigmatised or abnormal, and this has an influence on which sexual aspects can be made explicit in a given context, the way people communicate about sexual matters, and how we conceptualise sexuality. Accordingly, normativity has played an important role as an explanatory tool in language and sexuality studies (see Baker 2013, Koller 2013, Motschenbacher 2014a, forthcoming). The focus on normativity was more of an implicit kind during the emergence of the field in the late 1990s and 2000s, as most researchers concentrated on the investigation of language and linguistic practices in relation to gay male, lesbian, and transgender subjectivities, which are traditionally viewed as non-normative at the societal macro-level (e.g. Besnier 2003; Leap 1996; Queen 1997). Research on non-heterosexualities continues to be an important driving force for the field, but recently many researchers have turned to more epistemological questions of how normativity affects the way we use language to talk or write about sexuality in all its manifestations (e.g. Koller 2013 on lesbian sexualities; Milani 2013 on gay male sexualities; Motschenbacher 2014a, 2014b on objectophilia; or Motschenbacher 2011, 2018a, forthcoming on normativity in language and sexuality studies more generally). While earlier language-centred inquiry tended to focus on the experiences of sexual subjects traditionally deemed non-normative, newer work concentrates more on normativity as a discursive system and its relations to power structures. The central focus has therefore shifted from the documentation of language phenomena connected to gay male, lesbian and transgender experiences to investigations of
how sexual normativity surfaces in language use and thus shapes the discursive formation of sexuality, including heterosexualities, same-sex sexualities, and other sexuality-related phenomena. This development has been fostered by previous theoretical debates on identity vs. desire in language and sexuality research (Bucholtz & Hall 2004; Cameron & Kulick 2003), which have advanced our understanding of sexuality as (multiply) discursively shaped. The focus on normativity in relation to language and sexuality highlights an additional, and perhaps alternative, perspective on this issue, as it cuts across the influential identity-desire divide and may therefore be used to describe sexuality in a more comprehensive manner. Both sexual identities and sexual desires, as well as other sexuality-related aspects, can be described in terms of their (non-)normativity values, and communication about sexuality frequently orients to such values.

Despite the broader perspective on sexuality that a normativity-based approach permits, research on language and heterosexuality (e.g. Cameron 1997; Coates 2013; Motschenbacher 2012; Schneider 2013) is still less common within the field of language and sexuality. This is remarkable, since heterosexuality is similarly shaped by normative mechanisms as non-heterosexualities, and such normative pressures may cause heterosexually identified people great distress, as we will see in the empirical part of this article. The present study contributes to the new strand of language and sexuality research that views normativity as a central factor in the discursive construction of sexuality and aims to address a research gap by providing a normativity-oriented analysis of the discursive regimes governing male heterosexuality. We concentrate on how heteronormativity may adversely affect the lives of heterosexual men – more specifically, of men who pathologically doubt their heterosexual identities. To this end, we study a corpus of forum posts by
heterosexual men who suffer from sexual orientation obsessive-compulsive disorder (henceforth SO-OCD), using corpus-assisted discourse analysis.

We start with a theoretical section in which we introduce central concepts and previous research on the topic of language and sexual normativity (section 2). After a brief outline of the phenomenon of SO-OCD (section 3), we discuss the methodological basics of our study (section 4). Section 5 presents three types of corpus-assisted analysis carried out on the forum post data, namely keyword, n-gram and concordance analyses. The concluding section (section 6) recapitulates key findings and reflects on how the present study helps to shed light on normative mechanisms.

2. Language, sexuality and normativity

Normativity has been an important concept for critical inquiry in queer studies – a field that has tended to conduct research from an anti-normative position, finding fault with the negative effects that norms have for non-conforming people (see Jagose 2015; Wiegman & Wilson 2015). In language and sexuality studies more specifically, the concept has been used as a central, though under-theorised, explanatory tool as well. Researchers in this field have traditionally placed a strong focus on the documentation of non-normative (mainly non-heterosexual) phenomena, and there is evidence that scholars have become increasingly aware of the ways in which normativity shapes the discursive construction of sexuality (see Baker 2013).

The theoretical discussion of normativity has been advanced fairly recently in language and sexuality studies (Motschenbacher 2014a, 2018a, forthcoming), with hetero-, homo-, and cisnormativity being central targets of queer linguistic inquiry and critique. While
the term heteronormativity relates to discursive phenomena that stipulate (certain forms of) heterosexuality as normal, preferable or natural, and that privilege essentialist gender binarism discourses (e.g. Motschenbacher 2011; Schneider 2013), the more recent concept of homonormativity describes practices through which (certain forms of) same-sex sexualities are treated as preferable.\footnote{Note that \textit{homonormativity} as used here places greater emphasis on sexuality and thus departs from Duggan’s (2002) use of the term in connection with neoliberalism and gay consumption patterns.} Examples of homonormative phenomena include certain gay men’s preference for ‘straight acting’ sexual partners (Milani 2016), the privileging of politicised lesbian identities in certain lesbian communities (Koller 2013), or the homonationalist showcasing of sexual minorities for political reasons (Milani & Levon 2016). Cisnormativity refers to the privileging of identities that show a neat correspondence between biological sex and gender, to the detriment of trans identities (Borba & Milani 2017; Ericsson 2018). As can be deduced from this discussion, “normativity” is a cover term that unites various more specific normative discourses (“normativities”). The latter are context-dependent, that is, what counts as normative or non-normative differs across particular communities of practice.

The theorisation of normativity more generally has brought forward a distinction between descriptive and prescriptive forms of normativity (e.g. Bicchieri 2006; Hall & LaFrance 2012; Hogg & Reid 2006). ‘Descriptive norms’ are quantitatively based and relate to what many people do – behaviour that is widely perceived as ‘normal’. ‘Prescriptive norms’, by contrast, stipulate how people should (not) behave and are a corollary of ‘normative’ stances, which are in principle independent of majority behaviours.

These two types of normativity also surface linguistically. Earlier work on the discursive construction of sexual normativity showed that certain linguistic features play a role in this, among them the use of adjectival descriptors to sketch out non-normative
aspects of sexual behaviour as ‘crazy’, ‘weird’ or ‘sick’ (Motschenbacher 2014a), the use of auxiliaries in deontic modal functions to stipulate what should or should not be done (prescriptive normativity), and the use of quantifiers and frequency-denoting adverbs to describe how common or uncommon a certain behaviour is (descriptive normativity; Motschenbacher 2018a). The present study seeks to contribute to the exploration of normativity-related linguistic features through corpus linguistic evidence.

3. Sexual orientation obsessive-compulsive disorder (SO-OCD)

Obsessive-compulsive disorder (OCD) affects up to 2.3% of the general population (Ruscio, Stein, Chiu & Kessler 2010) and is characterised by recurrent unwanted thoughts or images that intrude persistently upon an individual’s mind (American Psychiatric Association 2013; see also Friedrich 2015 on the discursive construction of OCD). Some cognitive theories (e.g. Rachman 1997) claim that if an ego-dystonic thought – i.e. one whose content does not align with one’s sense of self – is experienced negatively, its frequency, duration and intensity may increase, resulting in an obsessive doubt that generates distress. Compulsions (rituals) are repetitively performed to control the doubt and anxiety, but offer only temporary relief until a new doubt launches another obsessive-compulsive cycle. People tend to stereotypically associate OCD with obsessive urges for cleanliness or perfectionism (McCarty, Guzick, Swan & McNamara 2017). However, beside obsessions about contamination, symmetry, hoarding, or aggression, intrusive sexual thoughts are commonly experienced by up to 25% of OCD sufferers (Grant, Pinto, Gunnip, Mancebo, Eisen & Rasmussen 2006). These obsessions may be associated with any form of sexual taboo (e.g. rape, incest, paedophilia, zoophilia, etc.) or sexual identity. Unlike sexual fantasies, which are
viewed as pleasant, harmless, and desired, sexual obsessions are experienced as unpleasant and unwanted, and rarely cause sexual arousal (Gordon 2002).

SO-OCD is a sub-type of OCD in which sufferers fear that their sexual identification is threatened. It is estimated that 11.9% of OCD sufferers experience obsessions about homosexuality, and that 65.3% of these are male (Williams & Farris 2011). The most common form of SO-OCD thus involves heterosexual men fearing that they might be gay (e.g. Gordon 2002; Williams 2008). For male heterosexual SO-OCD sufferers, obsessions revolve around various kinds of fears that can be interpreted as an outcome of a pathological orientation to heteronormative pressures and notions of idealised heterosexual masculinity (see Coates 2005, 2007): fears of becoming or being seen as gay, fears that heterosexual relationship issues or sexual problems are signs of ‘turning’ gay, fears of being in denial or losing one’s heterosexual identity. Consequently, affected people may compulsively avoid contexts that could trigger intrusive thoughts (locker rooms, movies with same-sex themes, or places frequented by gay people). Other commonly attested reassuring compulsions include the persistent monitoring of one’s groinal reactions while meeting or observing people, and excessively watching (gay and/or straight) pornography to check one’s sexual desires (Williams, Slimowicz, Tellawi & Wetterneck 2014). Some of these obsessive ruminations are reflected in the following prototypical excerpt from a post written by a male heterosexual SO-OCD sufferer on the website Psychforums, the source our data for the present study:

(1) now every time I kiss my girlfriend some gay thought happens, but before I had my girlfriend I was hooking up with girls […] and the thought never came into my head, [...] i cant even hang out with my guy mates with out gay thoughts happening and mens faces popping up in my head… [...] i have nothing against gays or anything i have had gays friends, its just my orientation is under attack and well frankly its driving me [insane.]
Since the advent of the Internet, OCD sufferers have started using online forums as places where they can receive support and advice (Stein 1997). Around 2004, the term Homosexual OCD (HOCD) emerged in these online environments, and its use became established among some therapists and a growing community of affected people (Gross 2011). Although many sufferers continue to use the acronym HOCD in online forums, we employ the term SO-OCD here to avoid any potential homophobic undertones and to make the point that it is not a phenomenon that affects heterosexual people exclusively (e.g. gay sufferers can obsessively fear becoming straight). Since the term is not recognised by the American Psychiatric Association – although the Diagnostic and Statistical Manual for Mental Disorders does mention ‘intrusive sexual ideas’ about homosexuality (APA 2013: 700) – the LGBTQ+ community and people without expert knowledge on the phenomenon have interpreted the recognition of SO-OCD as homophobic, because it may be seen as legitimating highly controversial forms of reparative therapy (Strudwick 2015). These understandable concerns are based on reservations regarding the former pathologisation of same-sex sexualities. However, SO-OCD is treated with LGBTQ+ friendly cognitive-behavioural therapy, where sufferers are exposed to anxiety triggers and taught how to accept the possibility of being gay without performing compulsions (Gordon 2002).

Linguistic traces of normative discourses are usually more difficult to observe in contexts where people adhere to social norms, because norm-conforming behaviour is generally deemed unremarkable (Raymond, this volume). Breaking social norms regularly induces social actors to orient to such norms in their language use. Therefore, forum posts by heterosexual men who pathologically doubt their identity can be expected to show a substantial amount of normativity-indexing language use, as the authors frequently express
ideas that in their eyes clash with normative notions of heterosexual masculinity. In accordance with a queer linguistic framework (see Motschenbacher 2010, 2011), we thus approach the discursive construction of heterosexual masculinity from the perspective of the sexually marginalised (SO-OCD sufferers) in order to question dominant discourses of heteronormativity.

4. Data and methodology

The strength of using corpus linguistics in critical discourse studies lies in its quantitative foundations, which makes it a useful tool for identifying macro-level discourses through frequency-based evidence (e.g. in disorder-related online forums; see Hunt & Harvey 2015). Corpus methods are therefore well-suited to uncovering quantitatively based, descriptive normativities, that is, communicative practices that are treated as common or ‘normal’ in a given context. Prescriptive normativities, by contrast, can often be better analysed via qualitative forms of corpus linguistic analysis which look at forms in their syntactic context, providing insights on usage patterns. Our study, therefore, draws on both types of analysis. Although there has been substantial research on the discursive construction of sexuality using corpus linguistic methods (see Motschenbacher 2018b for an overview), little of this research has focused on sexual normativity as a discursive system or on heteronormative pressures on heterosexual men more specifically.

In this study, we carry out a corpus-assisted discourse analysis (for comparable work, see the contributions in Baker & McEnery 2015) in order to shed light both on common topics and communicative practices drawn on by male heterosexual SO-OCD sufferers. The data has been collected from the publicly accessible online platform Psychforums, which
offers mental health support to its users. The forum, whose first posts date from 2003, hosts more than 180,000 anonymous members from around the world, who communicate in English about their problems. The platform is divided into several sub-forums dedicated to a wide range of psychological conditions, including OCD. As online communication has been shown to foster disinhibition effects (Suler 2004), such forums facilitate self-disclosure when discussing potentially embarrassing or socially sanctioned topics. The chosen online platform is thus an ideal site to investigate communication about non-normative identities and practices. Since the forum is freely accessible, the posts can be considered public language use, minimising any ethical implications. Moreover, as it is unrealistic to obtain informed consent for the collection and subsequent analysis of users’ posts (Beisswenger & Storrer 2008), we deemed it sufficient to anonymise any information in the data which might reveal users’ identities (such as names and e-mail addresses).

In the quantitative parts of this study, we compare two corpora compiled from Psychforums: a target corpus of thread-initial posts written by male heterosexual SO-OCD sufferers (OCD Corpus, henceforth: OC), and a reference corpus of thread-initial posts by heterosexual men who turn to the forum to seek relationship- or sexuality-related advice but do not question their heterosexuality as such (Non-OCD Corpus, henceforth: NOC). The qualitative parts of the analysis focus solely on material from the OC, highlighting typical usage patterns.

Thread-initial posts proved particularly suitable for our research purposes, since they contain descriptions of symptoms or problems that users perceive as non-normative (Harvey & Koteyko 2013: 169). We drew on different sub-forums in order to create the two corpora. First, we used the search term HOCD to retrieve introductory posts written by male heterosexual SO-OCD sufferers for our target corpus. This implies that users have already
been exposed to a variety of institutionalised and folk discourses about the disorder and employ them to make sense of their mental health. Second, since heterosexuality is often merely implied in the posts, it was impossible to use a specific search term to identify heterosexually themed messages. Hence, we looked more closely into the various sub-forums to compile a reference corpus of posts by other advice-seeking heterosexual men: we skimmed the sub-forums on ‘Sexuality’ and ‘Marriage & Divorce’ up to December 2016, the first 25 pages of the ‘Relationship’ sub-forum, and the first 20 pages of the ‘Sexual Addiction’ sub-forum.

In total, both corpora contain 671 thread-initial posts (455,497 word tokens), dating from 2004 to 2016. They are similar in size, allowing for a direct comparison of absolute frequencies: OC contains 299 messages (227,482 word tokens); NOC contains 372 messages (228,015 word tokens). Age and gender of the users were identified on the basis of their messages: all users specified in their posts that they are male; 396 users mentioned their age (age range: 16 to 68 years). The average age of users is somewhat higher in NOC (28.1 years) than in OC (20.7 years), which suggests that younger men are more likely to belong to the group of SO-OCD sufferers (see Ruscio et al. 2010, who specify an average OCD onset age of 19.5 years). The ethnicity of these anonymous men is mostly unknown, since their narratives are the sole demographic data source. We estimate that the majority of users in OC are from Anglophone countries, because 26 users apologised in their posts for their non-native use of English and/or mentioned a non-Anglophone home country (e.g. France, Germany, Poland, Sweden, and Philippines, South America, Turkey). The countries specified suggest a predominance of users from moderately or highly developed societies. However, race and ethnicity did not emerge as relevant intersectional categories during the analysis.
We used the corpus software AntConc to generate and analyse keywords, n-grams and concordance lines of particular constructions (Anthony 2018). These procedures represent a combination of quantitative and qualitative corpus linguistic methods that provides a detailed picture of the discursive mechanisms at hand. Our analysis yields insights into how SO-OCD sufferers experience and conceptualise their obsessions. Keywords (section 5.1) provide quantitative evidence of ‘normal’ topics and discourses frequently drawn on in the posts of SO-OCD sufferers. N-grams (section 5.2) allow for a better understanding of how obsessions as reactions to heteronormative pressures are verbalised in the shape of recurring word combinations. Finally, the analysis of the syntactic contexts of particular constructions that play a role in the orientation to normativity via concordance lines (section 5.3) captures the inner thought processes of forum users that appear to be influenced by normative ideas (I feel/felt like), aspects which have high prestige for them because they take them to be in accordance with the norms of heterosexual masculinity (I want), and aspects that they believe to clash with these norms and therefore are associated with low prestige among this group of men (I don’t want).

Through these procedures, we seek to explore how normativity surfaces in the ritualised linguistic practices of male heterosexual SO-OCD sufferers, that is, in a virtual community of practice (Meyerhoff & Strycharz 2013) in which an orientation to heteronormativity shows pathological traits. Two normative aspects play a role here: (a) the local communicative norms that have become ritualised in the community of practice under study, which often derive from institutionalised discourses of SO-OCD, and (b) an orientation to heteronormativity as a dominant discourse in society at large.

5. Corpus-assisted discourse analysis of SO-OCD sufferers’ forum posts
5.1 Keyword analysis

Keywords are words that occur significantly more often in a target corpus when compared to a reference corpus (Baker 2004). A keyword analysis highlights linguistic differences between two corpora, yielding evidence on the ‘aboutness’ of the texts and prominent discourses in the data (Bondi 2010). When texts associated with a particular community of language users are studied, keyword analysis provides insights on descriptive local normativities, i.e. the ‘normal’ communicative practices of that community.

Table 1 presents the top 50 keywords in OC – forms that occur unusually frequently in this corpus when compared to NOC. Keywords are ranked by log likelihood (LL) values, a measure of the confidence that a certain word is key. LL values do not indicate the actual size of differences between corpora (Gabrielatos 2018). This is why such a ranking list tends to cover a wider range of keyword types, including not just lexical items but also grammatical function words. Although the latter are often neglected in corpus-assisted discourse studies, they play a role in the formation of n-grams (section 5.2). We have selected the top 50 keywords for closer analysis, because they cover all major semantic areas and discourses in the data (an analysis of a larger set of keywords would not have yielded additional insights).

<table>
<thead>
<tr>
<th>Rank</th>
<th>Keyword</th>
<th>Freq.</th>
<th>Keyness (LL)</th>
<th>Rank</th>
<th>Keyword</th>
<th>Freq.</th>
<th>Keyness (LL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gay</td>
<td>2178</td>
<td>2481.286</td>
<td>26</td>
<td>intrusive</td>
<td>97</td>
<td>106.563</td>
</tr>
<tr>
<td>2</td>
<td>hocd</td>
<td>736</td>
<td>997.166</td>
<td>27</td>
<td>anxious</td>
<td>123</td>
<td>106.421</td>
</tr>
<tr>
<td>3</td>
<td>thoughts</td>
<td>906</td>
<td>795.718</td>
<td>28</td>
<td>groinal</td>
<td>73</td>
<td>101.37</td>
</tr>
<tr>
<td>4</td>
<td>i</td>
<td>18981</td>
<td>766.191</td>
<td>29</td>
<td>denial</td>
<td>98</td>
<td>98.511</td>
</tr>
<tr>
<td>5</td>
<td>ocd</td>
<td>388</td>
<td>459.206</td>
<td>30</td>
<td>crush</td>
<td>112</td>
<td>93.286</td>
</tr>
<tr>
<td>6</td>
<td>girls</td>
<td>806</td>
<td>353.785</td>
<td>31</td>
<td>myself</td>
<td>784</td>
<td>90.435</td>
</tr>
</tbody>
</table>
At the most general descriptive level, the keyword list contains few function words or grammatical items, which is probably an outcome of the fact that the two comparative corpora contain material of the same text type. At the same time, those function words that show up in the list tell us something about discursive practices in OC more specifically. Besides four first person pronoun forms (\textit{i, myself, my, im [= I’m]}), there is one more function word, namely \textit{if}. These forms play a role within common n-grams in OC, as will be shown below (section 5.2). The predominance of first person singular references in OC is
remarkable, since both corpora consist of personal narrative texts. It reflects SO-OCD sufferers’ focus on their internal mental struggles to maintain a heterosexual subject position. This becomes particularly evident in syntactic constructions in which the subject pronoun I, which refers to the user as a heterosexual agent or experiencer, occurs in connection with possessive my or reflexive myself:

(2) I started forcing myself to watch gay porn to prove to myself that I didn’t like it, and once I was sufficiently nauseated by it, I would switch back to regular porn […]

Extract (2) illustrates the performance of a checking compulsion. Notice here both the writer’s urge to watch gay porn to find proof of his distaste, and his orientation towards ‘regular’ (i.e. straight) porn, to establish his subject position as heterosexual. While gay porn is here constructed in negative terms, as something that the writer ‘forces himself’ to watch and that ‘nauseates’ him, his experience of straight porn is described in a more neutral or even normative fashion (switch back; regular).

One prominent set of keywords is connected to OCD as a medical condition (hocd, ocd) and its symptoms, which revolve around notions such as anxiety (anxiety, fear, panic, anxious, scared), attack (intrusive, attack) or insecurity-triggered compulsions (checking, denial, check, questioning). The fact that OCD symptoms are primarily of a mental kind is further illustrated by nouns and verbs denoting mental states or feelings (thoughts, thought, remember, think, thinking, felt), and the nouns mind and head. There are three progressive verb forms (checking, thinking, questioning) among the keywords. Unlike in NOC, such forms
are here used to construct excessive, repetitive processes that are typical of obsessive-compulsive behaviours:

(3) I’m constantly [...] checking my thoughts. I was thinking of accepting the fact that I’m gay but like I said - I can’t, I know it’s not me, yet I feel gay then I don’t. WHAT A PARADOX! Whenever I feel straight for a moment my mind convinces me I’m not.

122.OCD-corpus.txt

Another lexical field that is represented in the keyword list is sexual identity labels (gay, straight, homosexual, bi), which demonstrates that sexual identification is a salient, and probably the most central issue in OC. The item gay unsurprisingly ranks first among the keywords (2178 tokens), since a potential gay identification is the major trouble source for SO-OCD sufferers. In fact, because any semantic connection to same-sex sexualities produces anxiety (e.g. the imagined possibility of feeling attracted to, or having sex with, the same-sex), it is equally unsurprising to find bi (rank 49; 95 tokens) among the top 50 keywords. The label straight also figures prominently (rank 7; 420 tokens), and is regularly used as the binary antonym of gay in the OC posts:

(4) I constantly question myself whether I am gay or straight.

32.OCD-corpus.txt

(5) I masturbate to women and straight porn (gay is just weird and is a no go [...] )

35.OCD-corpus.txt
Due to its pathological connotations, *homosexual* (rank 33; 92 tokens) could be interpreted as producing a more distanced attitude towards homosexuality. Although it seems that it is mostly used interchangeably with *gay*, the cluster *a homosexual* (15 tokens) connotes more hostile attitudes towards homosexuality:

(6) *I was unable to perform in bed [...] and it somehow confirming my suspicions I am in fact a homosexual.*  
62.OCD-corpus.txt

(7) *I truly believe I brainwashed my self into being a homosexual.*  
149.OCD-corpus.txt

Gendered identities surface among the keywords in the shape of lexically gendered personal nouns (*girls, guys, guy, men, man*) and the form *male*. Interestingly, the plural forms of gendered personal nouns occur in general more frequently than the singular forms, which indicates that the posts in OC talk more frequently about female or male people in general than about individual women or men. As far as references to desired objects are concerned, this means that SO-OCD sufferers sketch out their experiences in a totalising, binary fashion, as a matter of being (not) attracted to women (men) as social macro-groups:

(8) *I have always been crushing on girls (NEVER ON GUYS! NEVER) until that very day when HOCD struck.*  
255.OCD-corpus.txt
Remarkably, while women are merely represented as an undifferentiated group with the term *girls*, the semantic field of male person specification is represented by five terms (*guys, guy, men, man, male*). This higher degree of male in-group differentiation (versus female out-group homogenisation) in OC indicates that same-sex attraction is treated as a more salient issue or a greater problem than heterosexual attraction when compared to NOC.

A final semantic area that is overrepresented in OC is the domain of desire, which surfaces in keywords such as *aroused, attracted, attraction, crushes, groinal, crush, arousal* and *porn*. Except for *crush(es)*, all of these terms relate to the construction of (physically mediated) sexual attraction, which turns out to be a central concern for SO-OCD sufferers. To reassure themselves that they are straight, sufferers repeatedly check their (lack of) sexual attraction towards (men) women by monitoring their genitals for signs of sexual arousal (Gordon 2002). Accordingly, 60 out of 73 occurrences (82.1%) of *groinal* appear in the collocation *groinal response(s)* – a term coined by OCD specialists which sufferers use as part of their ‘normal’ communicative practices:

(9) *I am continuously getting groinal responses whenever these thoughts occur, or whenever im looking at a dude. [...] my mind keeps reinforcing the thoughts, and tells me that is what I want, when IT’S NOT, but I still get a groinal response, so I keep wondering if it is real attraction.*

263.OCD-corpus.txt

Taken together, the keyword analysis highlights, on the one hand, which topics and discourses SO-OCD sufferers frequently draw on, i.e. themes that surface as ‘normal’ in this
particular community of practice: sexual/gender identities, sexual desire, and SO-OCD as a pathology. On the other hand, it also tells us something about how these themes are written about, because the data document a clear focus on sufferers’ internal mental struggles (first person references) and a medical or pathological discussion of their experiences intertwined with institutionalised discourses of SO-OCD (medical terminology, description of obsessive-compulsive behaviours).

5.2 N-gram analysis

N-grams are word clusters in a corpus – alternatively called ‘lexical bundles’ (Biber & Conrad 1999) or ‘recurrent word combinations’ (Altenberg 1998) – with the ‘n’ standing for the number of orthographic words in a cluster. An n-gram analysis provides additional insights on the communicative practices prevalent in a corpus (in fact, many of the frequent n-grams do not contain content keywords), through a syntactically more contextualised view on the linguistic data. N-grams in general do not coincide with syntactic categories such as phrases, clauses or sentences. Still, they document how forms are used within larger grammatical constructions and thus add a layer of analysis to the merely lexical focus of the keyword analysis. Like the keyword analysis, an n-gram analysis yields evidence for descriptive normativities, that is, communicative practices that are ‘normal’ for the virtual community of SO-OCD sufferers. On top of this, it can provide insights into prescriptive normativities that forum users orient to. While individual keywords do not normally possess the illocutionary force of making a normative statement, larger constructions may be used to convey such normative messages.
The identification of the most frequent n-grams in a corpus is corpus-driven in the sense that it does not rely on pre-defined linguistic constituents or forms. We used AntConc to generate lists of the 30 most frequent 2-, 3-, 4- and 5-grams in OC. From these lists, we selected 5-grams for closer inspection, as they showed fewer identical types when compared to the top 30 lists of NOC. The top 30 5-grams in OC are listed in Table 2. Note that nine of these 5-grams (marked with an asterisk) also occur among the top 30 5-grams in NOC. These clusters, therefore, are common in sex- and relationship-oriented problem posts by heterosexual men more generally speaking, while the remaining 21 5-grams point to aspects that are more typical of SO-OCD sufferers’ posts.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Freq.</th>
<th>5-Gram</th>
<th>Rank</th>
<th>Freq.</th>
<th>5-Gram</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36</td>
<td>i am a year old *</td>
<td>16</td>
<td>15</td>
<td>have always been attracted to</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
<td>i don't want to be *</td>
<td>17</td>
<td>15</td>
<td>porn to see if i</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
<td>don't want to be gay</td>
<td>18</td>
<td>15</td>
<td>want to be gay i</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>don't know what to do *</td>
<td>19</td>
<td>14</td>
<td>gay porn to see if</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>am a year old male *</td>
<td>20</td>
<td>14</td>
<td>had a crush on a</td>
</tr>
<tr>
<td>6</td>
<td>23</td>
<td>i don't know what to *</td>
<td>21</td>
<td>14</td>
<td>i feel like i am</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
<td>i'm a year old male *</td>
<td>22</td>
<td>13</td>
<td>i felt like i was</td>
</tr>
<tr>
<td>8</td>
<td>21</td>
<td>to the point where I *</td>
<td>23</td>
<td>13</td>
<td>in the back of my</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>that i might be gay</td>
<td>24</td>
<td>12</td>
<td>i don't know if I *</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>to see if i was</td>
<td>25</td>
<td>12</td>
<td>i have always been attracted</td>
</tr>
<tr>
<td>11</td>
<td>19</td>
<td>as long as i can</td>
<td>26</td>
<td>12</td>
<td>i want to be with</td>
</tr>
<tr>
<td>12</td>
<td>17</td>
<td>always been attracted to girls</td>
<td>27</td>
<td>12</td>
<td>if i was gay i</td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>i dont want to be</td>
<td>28</td>
<td>12</td>
<td>the rest of my life *</td>
</tr>
<tr>
<td>14</td>
<td>17</td>
<td>long as i can remember</td>
<td>29</td>
<td>11</td>
<td>dont want to be gay</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>a year old male and</td>
<td>30</td>
<td>11</td>
<td>for as long as i</td>
</tr>
</tbody>
</table>

Table 2: Top 30 5-grams in OC

The shared 5-grams can be seen as indexical of the text genre at hand, as they epitomise aspects that could be described as typical of ‘problem posts’. For example, there is a cluster of 5-grams that revolve around the notions of ‘not knowing’ (don’t know what to do; i don’t know what to; i don’t know if i) and ‘not wanting’ (i don’t want to be), thus
indicating the users’ personal struggles and need for advice. Connected to the notion of knowing are question words (what and if), which occur in three of the nine shared top 30 5-grams. Another strongly represented aspect that seems to be typical of problem posts is the mentioning of certain life stages, be it in terms of a specification of the writer’s age (i am a year old; am a year old male; i’m a year old male), his future (the rest of my life) or a decisive point in his life (to the point where i).

Besides these similarities between the two corpora, the remaining 21 5-grams point to aspects that are more typical of OC. The function word if occurs in four top 30 5-grams in OC. It is not used as a conditional conjunction in these clusters but introduces indirect ontological questions (synonymous with whether), which indicate identificational struggles centring on hypothetical gayness (to see if i was; porn to see if i; gay porn to see if; if i was gay i; compare also the modal construction that i might be gay).

Furthermore, the ‘not wanting’ discourse is in OC specifically connected to gay identification (don’t want to be gay; want to be gay i; dont want to be gay), and there are no further traces of the ‘not knowing’ discourse among the top 30 5-grams. As ‘not wanting’ is connected to a more strongly normative position than ‘not knowing’, SO-OCD sufferers can be said to adopt stronger stances on their problems (note that want occurs in five of the top 30 5-grams in OC):

(10) I really hope what I’m dealing with is HOCD, because I don’t want to be gay.

151.OCD-corpus.txt

Even if sufferers want to get rid of HOCD symptoms, note how in (10) HOCD serves paradoxically as a heteronormative ground to hold writers’ heterosexual subject position.
Often (especially during therapy) the anxiety wanes in the presence of intrusive thoughts, which sufferers interpret as a sign of them being gay (Gordon 2002). Consequently, they wish to feel anxious, because the anxiety ‘proves’ that they are ‘heterosexual with HOCD’.

With regards to the discursive construction of sexuality, the top 30 5-grams in OC pay witness to two central conceptualisations: one in terms of sexual identity (with six 5-grams containing the sexual identity label gay: don’t want to be gay; that i might be gay; want to be gay i; gay porn to see if; if I was gay i; dont want to be gay), and one in terms of sexual desire. The latter conceptualisation surfaces in references to a longer history of sexual attraction (always been attracted to girls; have always been attracted to; i have always been attracted), to the compulsive scrutinising of one’s physical reactions to porn consumption (porn to see if i; gay porn to see if), and to romantic feelings (had a crush on a; i want to be with) within the 5-grams. Extract (11) illustrates the use of three such overlapping 5-grams (in the string I have always been attracted to girls) and of several other constructions that draw on sexual desire discourses:

(11) I have ALWAYS been attracted to girls, never EVER questioned my sexuality, [...] I ALWAYS masturbated to straight porn [...] Never ever watched gay porn [...] I was always calling myself straight and never doubted it, was always attracted by girls [...] I never ever had these thoughts, I wasnt attracted by guys,

258.OCD-corpus.txt

It is remarkable that the 5-grams that conceptualise sexual attraction and romantic feelings are reinforced with the frequency-denoting adverb always, and that constructions with always and never figure prominently in the forum posts (see extract 11). In fact, the
temporal references in the top 30 5-grams of OC indicate a discourse of recapitulating one’s
entire life rather than individual past events (*always been attracted to girls; long as i can
remember; have always been attracted to; i have always been attracted; for as long as i*),
which reflects OCD sufferers’ inclination to ponder their sexuality-related history.

The description of mental processes also plays a prominent role in the top 30 5-
grams of OC. These processes represent once more SO-OCD sufferers’ obsessive mental
struggles (*long as i can remember; i feel like i am; i felt like i was; in the back of my
[mind/head]; plus references to attraction illustrated above):

(12) *I was so scared and believed everything was so real, I felt like I was legit gay all of
sudden despite knowing deep down I have loved women my whole life*

275.OCD-corpus.txt

In summary, the analysis of the most frequently used 5-grams has yielded additional
quantitative evidence of topics and discourses that SO-OCD sufferers commonly draw on.
Their discussion of sexuality shows a leaning towards a discursive construction in terms of
sexual identity and desire, with sufferers generally perceiving their identities and their
desire-related intrusive thoughts to be in contrast. An orientation to prescriptive
normativities surfaces, on the one hand, in the discursive construction of obsessions as
reactions to normative pressures (prevalence of mental processes, sexual life history
recapitulation), and, on the other hand, in the expression of strongly normative stances
(wanting and not wanting discourses). These latter forms of orientation to qualitatively
based normativities are studied in the following section in greater detail through a
concordance analysis of particular constructions in OC.
5.3 Concordance analysis

The final step of our analysis is a qualitative investigation of the collocational behaviour of constructions which are central to the expression of prescriptive normativities in OC. The construction *I feel/felt like*, which occurs several times within the top 30 5-grams in OC, can in principle be used to indicate that the speaker is in the mood to do something or would like to do something (*I felt like going out*). However, in OC, this structure is hardly ever used in this way; rather, it often expresses an evaluation of the sufferer’s internal thought processes as non-normative (see also Bellés-Fortuño & Campoy-Cubillo 2010). Performing an AntConc regex search with the search query *(i|I) (feel|felt) like* yields 169 hits, which allows us to identify this construction in its various forms in OC. Figure 1 presents an excerpt from the concordance line list, which displays instances in which the construction introduces clauses with the subject pronoun *I*:

```
Every time *I feel like* I am attracted to guys I would open porn with women to check if I would get aroused
I feel like I'm ejaculating to a man, which starts the vicious cycle over again.
I was so angry and upset *I feel like* I got made a prison bitch, felt like I was sexually intimidated.
my face get bright red and *I feel like* I was having an anxiety attack, I had to leave and go home.
then id lose it because *I feel like* I couldn't perform. It was terrible I feel like even if I know im not gay
this is a new problem and *I feel like* I am facing a losing battle.
I walk in a masculine way if *I feel like* I am walking like a woman. I used to just stare at girls and girls only
Sometimes *I feel like* I should just "come out" and these worries will all be gone, but I cant
*I feel like* I am losing attraction to women as well.
I enjoyed her greatly, *I feel like* "had to do it" to prove I'm not a homosexual.
but sometimes *I feel like* I m not as aroused as I should be.
If I am gay, then so be it. *I feel like* I would have known that from day one, WITHOUT ANY true attraction to women.
```

Figure 1: Concordance lines of *I feel/felt like* in OC

These concordance lines show that behaviours that are widely deemed non-normative for heterosexual men are specified to the right side of this construction. These anxiety-
provoking aspects revolve mainly around having sex with or being attracted to men (*I am attracted to guys; I’m ejaculating to a man; I got made a prison bitch*), and not behaving in a masculine way or failing to perform in ways deemed typical of ‘real’ men (*I couldn’t perform; I am walking like a woman; I am losing attraction to women; I “had to do it” to prove I’m not a homosexual; I’m not as aroused as I should be*). They indicate SO-OCD sufferers’ fear of a potential gay self destroying their ‘true’ heterosexual self. Other tokens represent discursive constructions of anxiety-provoking mental struggles (*I was having an anxiety attack; I am facing a losing battle; I should just “Come out”; I would have known that from day one*).

Similar messages that reflect sufferers’ orientation to the norms of heterosexual masculinity are often conveyed through the construction *I don’t want*, which also figures prominently within the top 30 5-grams in OC. A search for this construction in OC yields 93 hits. As Figure 2 illustrates, the construction is commonly followed by a non-finite to-clause:

How can I accept something *I don’t want to* be and never were? I am 100% sure I never was gay but in reality I know that *I don’t want to* have gay sex. I’m frightened.

The problem is that: *I don’t want to* be gay or bi but I feel like it could be enjoyable it could be enjoyable now(* I don’t want to* but it seems that after 3 month of tears and hate

This is freaking me out, *I don’t want to* be gay, it’s more personal based

But *I don’t want to* be bisexual. He was also saying its good to try out trying out gay sex, but *I don’t want to* want to do that either. I don’t know what to do.

*I don’t want to* be gay. I don’t want any of this.

If I am bisexual, or if I am gay, *I don’t want to* lose my heterosexual and and *I don’t want to* do something I know isn’t me.

Figure 2: Concordance lines of *I don’t want* in OC

The data show that the construction *I don’t want* is predominantly used in OC to describe the writer’s unwillingness to embrace a gay identification or to engage in gay sexual practices (*I don’t want to have gay sex; to be gay or bi; to be gay; to be bisexual; to lose my*
heterosexuality). Thus it plays a crucial role in the expression of an identification with a ‘true’ heterosexual identity that is regulated by strict heteronormative ideals and threatened by gay intrusive thoughts.

While what heterosexual SO-OCD sufferers don’t want is usually something that is perceived as non-normative, what they do want tends to be perceived by them as desired, normative or normal. A search for I want in OC produces 203 hits. Figure 3 presents an excerpt from the concordance line list:

Figure 3: Concordance lines of I want in OC

As can be seen, SO-OCD sufferers express a longing for the most traditional aspects normatively associated with heterosexual masculinity, namely a female partner, marriage, a family household, career success, confidence, or sex with women (i want a wife, children, a good career and a house; a wife and kids someday; confidence; have sex with women and girl; her to be my girlfriend). The combination with a boyfriend in the first concordance line, by contrast, reflects the ego-dystonic thought processes of SO-OCD sufferers (my mind wants to tell me I want a boyfriend).
6. Conclusion

A combination of various quantitative and qualitative types of corpus linguistic analysis in the present study has shed light on how male heterosexual SO-OCD sufferers orient to normativity in their thread-initial forum posts. More specifically, we gained insights on the discursive construction of quantitatively based, descriptive normativities, that is, communicative practices that can be deemed ‘normal’ for the virtual community of practice under study, and of qualitatively based, prescriptive normativities, surfacing in the expression of normative stances on what forum users deem to be in accordance with heteronormative imperatives. Note that the way they narrate their experiences as prescriptively non-normative forms an integral part of ‘normal’ communication in this community. This suggests that an orientation to normativity can be a multi-layered process and that ‘normative’ and ‘normal’ behaviour need not coincide. The triangulation of the various corpus linguistic methods resulted in a multidimensional picture. The findings provide insights on the thought processes of SO-OCD sufferers (which in most private and public contexts are unlikely to be linguistically explicated) and on how these are shaped (and maybe even caused) by normativities. The analyses have shown that SO-OCD sufferers orient to heteronormativity in their posts in the sense that they feel they fall short of embodying the imperatives of male heterosexuality as idealised by them (such as an unquestioned attraction to female people and no attraction whatsoever to male people). They construct hypothetical non-heterosexuality (a potential gay identity; potential desires for other men) as their central problem. It needs to be viewed critically that they orient to the most traditional forms of heteronormativity in pathological ways. That is, it may even be argued that they privilege the claiming of SO-OCD symptoms (and thus of a mental disorder) over having (healthy) non-heterosexual desires, because it
enables them to leave their heterosexual identity temporarily intact and anxiety-free. Moreover, the analysis revealed how sufferers pick up certain institutionalised discourses of SO-OCD that serve as normative guidelines to make sense of their obsessive-compulsive behaviours and sexual identity.

A limitation of our study is that it has exclusively analysed data representing the views of people who think they suffer from SO-OCD but might not have been diagnosed with it, while alternative perspectives (the medical perspective, the perspective of relatives or acquaintances of SO-OCD sufferers, the perspective of complete outsiders to this topic) have not been investigated. Future research may therefore wish to complement the findings of this study by incorporating these other perspectives, which may well yield a different picture of the phenomenon.

References


Hunt, Daniel; Harvey, Kevin (2015). Health communication and corpus linguistics: Using corpus tools to analyse eating disorder discourse online. In Paul Baker & Tony


