lnvestigating gendered language through collocation: the case of mock politeness

Charlotte Taylor, University of Sussex

INTRODUCTION/DEFINITIONS

Overview

This chapter makes use of the notion of *collocation*, a key concept in corpus linguistic work. Collocation refers to the tendency of certain words or phrases to occur together with other words and phrases. It is an important aspect of language because it gives us a way of understanding the associations that particular words or phrases may carry for language users. While collocation comes from corpus linguistics work, and as such pertains to a more quantitative approach to language studies, it has been extensively used in the sub-field which combines corpus linguistics and (critical) discourse analysis and sits astride the traditional qualitative/quantitative divide.

Collocation analysis gives a way into understanding how words and phrases are used, and the associations they trigger, which is essential to identifying the discourses that surround the representation of groups - a topic which is often of interest in research on language, gender and sexuality. I would argue that collocation can also be an important preparatory stage in variation studies in this area. That is to say, before we can ask who does x (most frequently), we need to know the full semantic and pragmatic profile of x to be sure that we are measuring what we intend. In the case of mock politeness, which is the focus of the case-study in this chapter, previous research regarding gender has tended to report that men are more likely than women to perform sarcasm and more likely to perform patronising behaviours in mixed-sex interactions. In this chapter, I aim to step back from these binary comparisons to question whether the terms *sarcastic* and *patronising* are themselves gendered. For instance, is there a tendency to use *sarcastic* to describe behaviour by a male speaker when a different label would have been applied to same behaviour if it had been performed by a female speaker?

The chapter starts by defining what is meant by collocation and how it can be investigated in studies of language, gender and sexuality, including some guidance for good practice in the area. The case-study is then presented in which collocation is employed to investigate *sarcastic* and *patronising*. The chapter ends with indications of future directions and recommendations for further reading for scholars interested in using collocation analysis to investigate the relationship between language, gender and sexuality.

What is collocation?

Collocation is a fundamental notion within corpus linguistics, and is perhaps best summed up by Firth (1957: 11) who famously stated that 'you shall judge a word by the company it keeps'. The Firth quotation is important because it sums up not just *how* we understand collocation (the relationship between words), but *why* we are interested in collocation. That is how knowing which words tend to go together can tell us more about the contextual meanings (including evaluative potential) of the item we are particularly interested in. The role of corpus linguistics here is in allowing us to look at a greater number of instances than would be feasible by manual analysis, and in giving us information about the *significance* of collocation.

There are two different ways of thinking about the significance of collocation: the first relates to the identification of collocation and the second relates to why it is of interest to those of us studying language. Starting with the first, here we are trying to identify which collocates or pairings of words we should consider to be significant. To address this point we might want to go back to defining collocation. According to Stubbs (2001: 29), a collocate is 'a word-form or lemma which co-occurs with a node [the word the researcher is interested in] in a corpus. Usually it is frequent co-occurrences which are of interest, and corpus linguistics is based on the assumption that events which are frequent are significant'. However, frequency alone is insufficient because, as Biber, Conrad & Reppen (1998: 265) remind us, 'more

common words are more likely to occur in a collocate pair simply by chance'. Thus it would be overly simplistic to look at frequency alone, and potentially uninformative for our research purposes. For this reason, in the practical sections we will look at which measurements can be used to identify collocates.

To take the second aspect of significance in relation to collocation, we need to consider more fundamentally why word pairings should be of interest to a researcher interested in investigating the relationship between gender and sexuality and language. Here, the reason that collocation is seen as meaningful is because it gives us a way into understanding the evaluative meanings that are potentially bundled up with any lexical item.

In the most obvious cases, the evaluative meaning actually is the central meaning, as for instance in the term *bitch* where it is used as a term of abuse for a woman. In this context, we cannot separate out the negative evaluation from the thing to which the word refers; the evaluation is entirely intrinsic to the word. The unfavourable meanings, or negative connotations, are absolutely apparent to the fluent speaker and not in any sense peripheral or hidden. Similarly, to take a well-known gender pair, the connotations of *spinster* and *bachelor* are quite apparent to a fluent speaker; the difference between the two is not just that one is used for an unmarried man and one for an unmarried woman, but that *bachelor* evaluates that unmarried status more favourably than *spinster*.

Then, there are lexical items which are less obviously evaluative in function. For instance, Cameron (2003) discusses the term *openly gay*. At first glance, we might not see anything so obviously evaluative here but, as Cameron urges, we only have to think about what else we might describe someone as being *openly*. Using a corpus we can check those intuitions and, according to the general corpus EnTenTen13 (described below), the words which have the strongest collocation with *openly* are, in order of strength: *gay*, *hostile*, *homosexual*, *lesbian*,

racist, *contemptuous*, *bisexual*, *critical*, *defiant*, *anti-semitic*. Apart from the words that refer to sexuality and perhaps *defiant*, we might note that these are mainly attributes that are unlikely to be used to describe someone the speaker admires (*hostile, contemptuous, critical*) and describe highly offensive views (*racist, anti-semitic*). With this information to hand, we might re-consider whether *openly gay* is a neutral expression. This tendency for a lexical item to regularly occur with items that have favourable or unfavourable connotations has been discussed in corpus linguistics as 'semantic prosody' or 'evaluative prosody'. It is often referred to as the 'aura' that a word carries (following Louw 1993) and often we only realise that a word has a particular prosody when the expected pattern is broken. For instance, does *openly friendly* sound usual or familiar?

In addition to these evaluative meanings, there may be other connotations that are held as part of our conscious or unconscious knowledge of a lexical item, and these are the aspects now likely to be described as 'discourse prosodies'. To take an older example, Sinclair (2004: 38) discusses how the phrase *my place* has a prosody of 'informality' and 'invitation' as in 'Would you like come to back to my place for a while?'. The uptake of collocation and connotation in studies of language, gender and sexuality is discussed in more detail below.

Using collocation in studies of language, gender and sexuality

In a survey of work that employed corpus linguistics approaches in order to study the relationship between gender and sexuality and language, I identified 47 articles published in international journals between 2006 and 2016. Of these, 21 focussed on *variation* in language used by people classified as belonging to different gender or sexuality groups and 22 focussed on the *representation* of people classified as belonging to different gender or sexuality groups. Collocation was used as a tool only in the latter group of representation-based studies and was employed in 11 of those papers, so nearly one quarter of the studies overall. If we consider why collocation tends to be used most frequently in representation

studies, it is likely to be because of its close relationship to discourse and therefore to ideology.

As outlined in the previous section, the study of collocation provides us with an entry point to our data in terms of understanding what evaluations and other types of connotation may be packaged up with certain lexical items in given contexts. We may consider the contribution of collocation analysis as affecting two principal areas within discourse studies. In the first, as Bogetić (2013: 334) puts it, 'collocation analysis offers a productive means for understanding ideology, as lexical co-occurrence may shed new light on complex webs of identities, discourses and social representations in a community'. Indeed it is because of this notion of 'webs' that the concept of the collocational network is particularly useful. We may not be consciously aware of how particular ideas are persuasively grouped in discourses and so the analysis of collocates gives us a means for drawing out these connections and *non-obvious* meanings. The second important use of collocates is that they 'can be useful in revealing how meaning is acquired through repeated uses of language, as certain concepts become inextricably linked over time' (Baker 2014: 13). If we consider discourse to be cumulative, then looking at the accumulated associations around particular lexical items can help make this process more evident.

In the papers surveyed for the literature review, six of the eleven focussed explicitly on the gendered noun pairs MAN/WOMAN and GIRL/BOY and all but one focussed on gendered nouns. This focus on gender pairs may represent a strength of the approach because analysing the collocation patterns allows us to discover new information about how such seemingly straightforward pairs are used in different domains. However, it also shows where more research may be required in collocation studies of language, gender and sexuality because it reveals that the research to date tends to operate along binary gender lines. What this suggests

is that the knowledge which is valued is difference between two genders. More recent research into these gendered nouns, such as Baker & Levon (2016), digs deeper by examining the different language surrounding racialized and classed pre-modifiers of *man* and this intersectional approach reflects more accurately the wider movement in gender and sexuality studies.

Good practice in collocation analysis

In terms of good practice, those principles that apply to collocation analysis are the same as those that apply to most research and certainly most corpus linguistic work.

Transparency: When reporting collocation analysis, provide sufficient information for the reader to understand exactly how you manipulated your data. This would include details of which word forms you searched for, what software you used, the statistical measure of strength of collocation, the cut off points implemented, and the span examined for collocates (how many to the left and how many to the right of the node). Ideally, you should aim to produce enough information for a reader to repeat the analysis, thus fulfilling the goal of *replicability*. Although space is often an issue, it is usually possible to include this information at least as a footnote. The reason this is so important is that there are different ways of calculating collocation and these will produce different results (see discussion in McEnery & Hardie 2012; Baker 2014). To take an example which is relevant to the case-study we discuss in the following section, if we look at the collocates of SARCASTIC in EnTenTen13 (a very large corpus of texts gathered online), the Sketch Engine software (Kilgarriff et al. 2014) offers us four different measures for calculating collocates and the results are displayed in Table 1. The number given to the right-hand side of each column tells us how often these two words occur together.

Frequency	t-score	Mutual information	logDice

and	12,655	*And*	12,655	*insultsude*	14	*witty*	684
a	11,499	*a*	11,499	*quotesarcastic*	11	*snarky*	277
the	9,216	*I*	7,848	*jerksarcastic*	10	*Revive*	233
I	7,848	*the*	9,216	*sarkastisches*	5	*remark*	597
to	6,974	*to*	6,974	*bithcy*	14	*cynical*	469
of	6,236	*being*	5,070	*fringeheads*	13	*remarks*	928
being	5,070	*of*	6,236	*raucher*	3	*humor*	1,183
is	4,568	*was*	4,252	*fringehead*	8	*snide*	151
in	4,363	*is*	4,568	*Fringehead*	7	*wit*	430
was	4.252	*be*	3,835	*Marcot's*	8	*Dont*	323

Table 1. Comparison of collocate ranking according to different measures

In the first column we see that the ten most frequent collocates are mostly function words and this is not particularly surprising because these are the most frequent types of words in the corpus overall: in fact seven of the ten most frequent collocates are also among the ten most frequent words in the corpus. The following three columns use different statistical measures to calculate the likelihood of two items occurring together by chance and we can see great variation in the kinds of words that are identified. The t-score measure in the second column privileges high frequency of co-occurrence and indeed the ranking is very similar to the simple frequency ranking shown in the first column (nine of the top ten words are shared). By contrast, the words identified as significant collocates using the mutual information calculation (the third column) appear to be mainly usernames. They are very low frequency overall and are reported because the calculation foregrounds items that occur very infrequently and nearly always occur together with the node (SARCASTIC). In the fourth column we see that the logDice calculation provides items that are neither particularly high nor particularly low in overall frequency and so looks the most likely to be productive in better understanding the company kept by SARCASTIC. So we can see that the choice of statistic will depend on the purpose and 'the sort of words that the researcher is interested in obtaining' (Baker 2006: 102) and needs to be reported because the results can vary so greatly. The same would apply to other factors which will affect the words that are calculated to be

collocates such as the span and cut off points that are chosen by the researcher and/or set as default in the software.

Total accountability: When discussing collocates, it is good practice to show all the collocates that were gathered using the chosen settings and criteria. This is important in order to fulfil the requirements of total accountability, a term coined in Leech (1992), which encapsulates the principle that we account for all findings and do not simply select those which are favourable to our hypothesis. Restrictions on space often mean that we have to be selective about what we discuss, but it is usually possible to use the appendices to list all collocates so the reader can get a fuller picture.

Generalisations: Collocates are register specific and so we should be careful not to assume that the collocates found in one corpus will be found elsewhere. For instance, in a study of newspaper language I found that collocates of *girl* and *boy* which remained constant over a 12-year time period related to violence (for *girl* only: *abduct*, *burn*, *death*, *lure*; for *boy* only: *wound*). Within the context of newspaper discourse, these offered interesting routes for investigation but it is also the case that they are a result of the news values of our press which tends to report negative events.

Context: When we examine collocates we need to make sure we do not look at these lexical items in isolation. This is particularly the case when categorising collocates because that process of categorisation is a process of interpretation of meaning. Therefore, a key part of collocation analysis involves going back to the text (often via the concordance line) to see how the terms were used in that specific set of data. For instance, in my study of *girl* and *boy* in the press sometimes the two shared the same collocate but closer investigation showed that one was the object and the other was the subject of the same verb which requires a different interpretation.

Background: The last suggestion for good practice is to remember to look outside the corpus at the wider field of language, gender and sexuality studies. The rationale for filling a gap cannot simply be that no-one has done a collocation study of x previously; there needs to be a theoretically grounded reason as to why that is a worthy topic of investigation. More fundamentally, the researcher needs to be aware of how the theoretical framework in the field has developed and can support their research. As mentioned above, it is perhaps the case that collocation studies lag behind shifts in focus in the broader field at present.

CASE-STUDY: GENDERED LANGUAGE AND MOCK POLITENESS

Background

In the case-study I explore two issues relating to the topic of mock politeness and gender. Mock politeness is defined as occurring when there is an im/politeness mismatch leading to an implicature of impoliteness (see Taylor 2016 for a more detailed discussion). Thus, it encompasses utterances such as those in examples (1) and (2) which come from the corpus of forum interactions used in the case-study:

> (1) Lift came, doors opened, we stepped forward to get in and were almost knocked down by a couple with their own pushchair. I'm afraid I was rather sarcastic and exclaimed "Don't mind the queue".
>
> (2) people that carry on like you Alba, are often described as twattish, or a bit of a tit. hth. [hope that helps]

In the first, the impoliteness mismatch comes from the apparently polite move of *Don't mind the queue* and the context in which the targets had already entered the lift ahead of the speaker. The speaker intends that the incompatibility of what is said and the context lead to an implicature that what is meant is a reproach for non-observance of a perceived social norm (the impolite move). In the second, the mismatch is made more explicit as it occurs at the textual level where the polite move is given at the end of the utterance (*HTH*) following on

from an impolite move in which the speaker associates the target with unfavourable characteristics.

The term *mock politeness* is probably not one that you use in ordinary conversation, and as such we may consider it a 'second-order' label, which is to say it is a label for an academic concept (as summarised in the definition above). In previous work (Taylor 2016), I have found that the following are all 'first-order' or lay labels which were used to describe mock polite behaviours in a corpus of British English forum interactions: *patronising*, *biting*, *make fun*, *condescending*, *cutting*, *caustic*, *mock*, *bitchy*, *tease*, *passive aggressive*, *put down*, *overly polite*, *sarcastic*. This distinction between first and second order uses (see for instance Watts et al. 1992), between the lay and academic constructs and terms, is an important one, particularly when trying to elicit data because any researcher needs to be confident that they are employing terms with which the participants are familiar and which may be used to describe the full range of behaviours under study and all kinds of people who perform those behaviours. So, for instance, if a researcher wanted to collect accounts of mock politeness, they would not ask interviewees 'Can you tell me about a time someone was *bitchy* to you' unless they wanted to focus on stories of female mock politeness. In this case the gendering of the lexical item is obvious, in other cases, the researcher might need to study the collocates to check for such bias.

In the following sections I briefly describe the data used here and then introduce two areas that overlap significantly with mock politeness and which have consistently been associated with male behaviour: sarcasm and patronising behaviour.

Data

The dataset used in this study comes from an online forum which was selected because it allows access to 'everyday' or 'conversational' comments on mock politeness, while

retaining much of the context. The forum, mumsnet.com, is UK based and predominantly populated by people presenting as women (an imbalance which clearly has implications for any discussion of gender). By way of illustration of the size, as of January 2015, mumsnet claims to have over 14 million visits per month (Mumsnet 2015). The 61 million token corpus was compiled from the forum using the free software BootCat (Baroni and Bernardini 2004), which gathers text from entire webpages using seeds (search words).[i] The EnTenTen13 corpus, which is available through Sketch Engine, was also used.[ii] This is an English language corpus of online texts and contains approximately 19 billion words (Jakubíček et al. 2013).

Tools

The latest tool for visualising connections between words is Lancsbox (Brezina et al. 2015, applied in Baker & McEnery 2015) which shows the way that collocates link to the node and also to one another.[iii] The importance of visualising the collocational network is that it allows us to see the company that a word is keeping and, crucially, it places that company in context. As Brezina et al. (2015: 141) state, '[c]ollocates of words do not occur in isolation, but are part of a complex network of semantic relationships which ultimately reveals their meaning and the semantic structure of a text or corpus'. Furthermore, because the networks can be displayed simultaneously, it is also possible that we may be able to identify what is absent (Duguid & Partington 2018) by noting which items collocate with some nodes and not with others.

The Sketch Engine thesaurus (Rychlý and Kilgarriff 2007) allows us to see which words share similar collocates. It works by identifying collocates for a search word and then in the second stage identifies other words which share those collocates. So, for instance, in the previous study (Taylor 2013) of *boy* and *girl* in a corpus of newspaper texts, the Sketch Engine thesaurus identified the word with the most similar collocates to *boy* as *girl* and vice-

versa (more revealingly, the second word in the *girl* list was *woman*, while the second word in the *boy* list was *child*).

Sarcasm

Starting with research into sarcasm, to date attention from a language, gender and sexuality perspective has primarily focussed on *variation* in use as correlated with gender. The issue of frequency of use has received most attention and the consensus has been that men use sarcasm more than women. The most common measurement has involved self-assessment, for instance, Dress et al (2008: 83, my italics) asked participants the following questions:

1) What is the likelihood that you would use *sarcasm* with someone you just met?
2) How *sarcastic* do you think you are?
3) What is the likelihood that you would use *sarcasm* when insulting someone?
4) What is the likelihood that you would use *sarcasm* with your best friend?

The majority of studies using this method found men self-reported as being sarcastic more often than women (e.g. Rockwell & Theriot 2001; Dress et al. 2008; Bowes & Katz 2011; Milanowicz 2013). This method assumes that the participants are both self-aware and truthful and, perhaps not surprisingly, two of these studies (Bowes & Katz 2011; Dress et al. 2008) found that although the male participants reported using sarcasm more than the female participants, they did not do so in elicitation tests. Furthermore, the choice of metalanguage suggests a problematic blurring between first and second-order uses. The use of the word *sarcastic*/*sarcasm* when interacting with participants means that they will not answer with reference to the researcher's second order concept of sarcasm, that is the scientific construct, but the kinds of contexts in which they personally would describe a behaviour as *sarcastic*, that is the first order understanding. However, we know that lay and academic uses of

sarcasm are not the same (e.g. Creusere 1999) and lay uses will be influenced by sociolinguistic variables, including whether these terms are gendered.

In terms of *expectations* of gendered performance, previous research again points towards an association of sarcasm with male behaviour. In experimental conditions, Colston & Lee (2004) reported that speakers of sarcastic utterances were more likely to be assumed to be male. Furthermore, Katz, Piasecka, & Toplak (2001) found that the perceived gender of the producer of a sarcastic utterance affected processing, with reading times for texts featuring male producers of sarcasm being lower than for female producers. This was interpreted as occurring because 'sarcasm is more likely to be associated with males than females, comprehension of noncanonical usage is delayed as people attempt to integrate the text they are reading with their stored ''knowledge'' (stereotypes) of men and women' (Katz et al. 2004: 187). Indeed, what is not clear, and what this case-study aims to address, is the extent to which these gender effects are the results of stereotypes or actual gendered tendencies. For instance, Katz et al. (2004: 187, my italics) report that 'when the gender of the speaker is manipulated in a textoid [a short text], the *same* comment is rated as more sarcastic when made by a male than when made by a female'. This suggests that participants are drawing on stereotypes in associating *sarcasm* with male speakers.

To investigate the potential gender associations with *sarcastic* we start by visualising the collocates. Figure 1 visually displays the metapragmatic labels which I had previously found to indicate mock politeness and all their collocates which were gendered terms referring to people. In the LancsBox visualisation, the length of the line linking the search word and any given collocate reflects the strength of collocation between those items. The items in the central positions (*bitchy*, *patronising* etc.) are the nodes which are entered manually, and the items radiating out are the collocates.[iv]

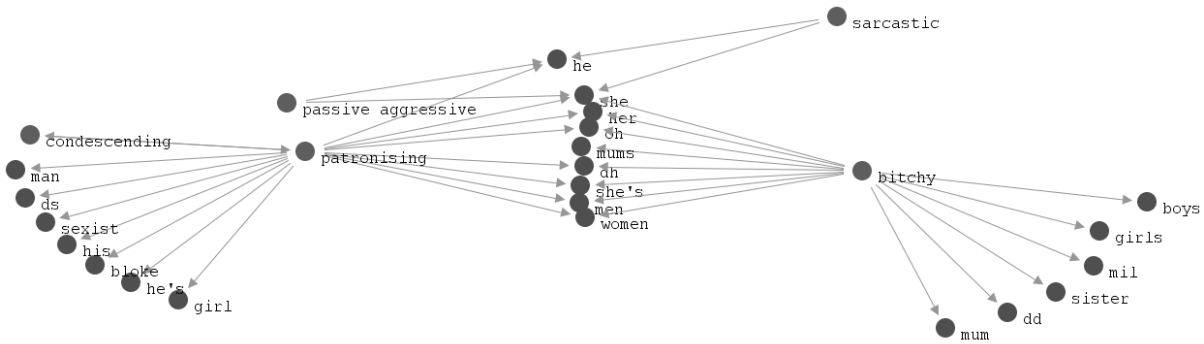

Figure 1. Gendered collocates of adjectival mock politeness labels

Two metapragmatic labels which emerge as gendered are *bitchy* on the right (the only one collocating with *mil* [mother-in-law], *girls*, *mum*, *dd* [dear daughter], *sister*) and *patronising* (the only one collocating with *man*, *ds* [dear son], *his*, *bloke*, *he's*). However, these two items also share a large number of collocates referring to both male and female participants (*she*, *her*, *mums*, *she's*, *women* and *dh* [dear husband], *men*). From this measure, *sarcastic* and *passive aggressive* collocate with just *he* and *she* which suggests a more neutral set in terms of gendering. Obviously at this point, what we do not know is how the gendered items relate to the node, that is – who is being described as mock polite and to whom they are being mock polite? This could be tackled by analysing collocates using a tool like Word Sketch which allows us to draw on grammatical (part of speech) information.

In this case-study, the behaviours described by the metapragmatic labels were all analysed to identify the gender of the performer, that is, the person who was described as being *sarcastic* etc. As Figure 2 shows, there were gender preferences for the different labels.[v]

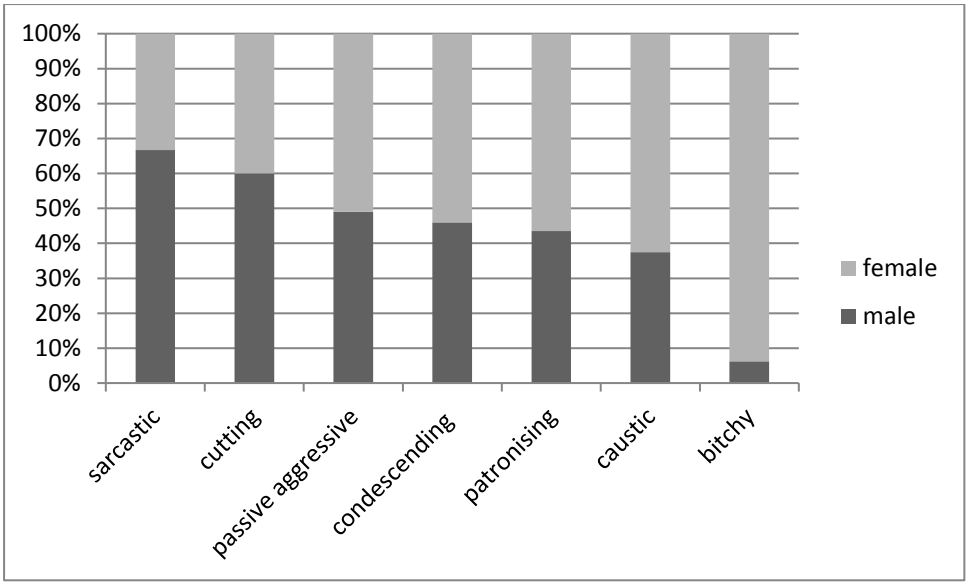

Figure 2. Distribution of male/female performance of behaviours

The most gendered metapragmatic label of those examined was *bitchy*, which showed a semantic preference for describing female behaviour while the item most strongly associated with male behaviour in terms of statistical significance (measured using log-likelihood) was *sarcastic*, these uses are illustrated in examples (3) and (4) with key terms in bold.

> (3) i ended up telling a couple of **bitchy** customers [I was pregnant], because I was lying down on the floor because I felt sick as shit, and this random **woman** came in and snottily said ‘oh! having a lie-down are we?’ ‘yes, I replied, I'm pregnant and feel sick’.
>
> (4) **DH** [dear husband] is happy for me to be at home BUT he moans at me if the house isn't tidy or I get behind. **He** is **sarcastic** and says things like ‘I know you're really busy’ or ‘if you could spare the time’....

From Figure 2 we see that both *bitchy* and *sarcastic* carry gendered associations in terms of who they tend to describe. So when studies of gender ask participants to self-report the extent to which they are *sarcastic*, they are asking participants to apply a gendered label to their

own behaviour. Understanding that *sarcastic* is actually gendered helps explain why men self-report as sarcastic more frequently than women even though there is no evidence of differences in actual practice: that is to say they self-report as *sarcastic* because they associate the term with a male behaviour (what remains unknown at this point is how a woman who performs the second-order construct of sarcasm would self-describe).

The collocation analysis also reveals differing patterns of evaluation for the labels investigated here, as shown in Figure 3. *Bitchy* and *sarcastic* and *patronising* share negative collocates (*rude*, *mean*). Additionally, *bitchy* collocates with a set of strongly unfavourable evaluations (e.g. *nasty*, *awful*) while those for *sarcastic* are less strong (e.g. *dry*, *odd*). Thus it appears that the two most gendered labels (*sarcastic* for male behaviour and *bitchy* for female behaviour) also carry very different evaluative prosodies, indicating the potential for male and female participants to be judged differently for similar mock polite behaviours.

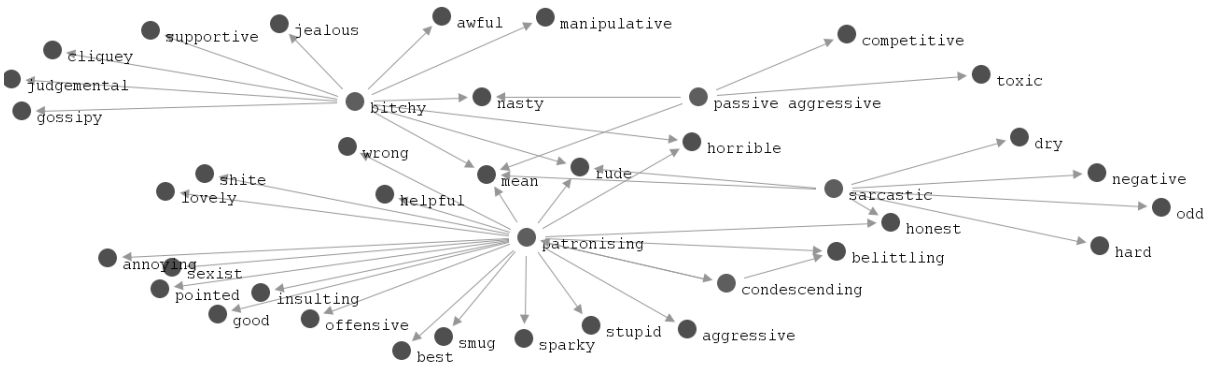

Figure 3. Evaluative collocates of adjectival mock politeness labels

To further explore the evaluative connotations, Sketch Engine Thesaurus was used to investigate which terms share similar lexical environments in a much larger web corpus (EntenTen13). The results are visualised in Figure 4 which presents those items with the most shared collocates in the biggest font (the colour variation is simply for ease of reading).

Figure 4. Sketch Engine Thesaurus output for *sarcastic*

From Figure 4 we can observe that terms with similar collocates include those previously identified as mock politeness labels (e.g. *condescending, ironic*) alongside a mixture of apparently favourable (e.g. *humorous, witty*) and unfavourable (e.g. *obnoxious, dismissive*) labels, in addition to some which would need context to disambiguate (e.g. *irreverent*). What emerges then is a picture in which the label *sarcastic* is more likely to be used to describe a man's behaviour and that the evaluation of the lexical item is mixed, with both favourable and unfavourable interpretations co-existing.

Patronising

As a second-order label, *patronising* is usually understood to encompass behaviours which might also be labelled as *condescending*, another under-investigated in terms of gender connotations. As Culpeper (2011) notes, '[b]eing patronized involves a kind of 'double whammy': your face is devalued in some way, but it is also devalued in a particular relational context that does not licence the "patroniser" to do so' (2011: 95). This latter point about 'licence' invokes another key aspect: power. Thus it is perhaps not altogether surprising that gender has often been a dominant focus of study in research into patronising behaviour from

social psychology (alongside intergenerational interactions), in particular so-called 'benevolent sexism' (Glick & Fiske 1997). In the conceptualisation of patronising and condescending behaviours in these areas, mismatch is given a central role because it is assumed that the patronising speaker is under-estimating the competence of the hearer. This potentially generates an im/politeness mismatch between the ostensibly helpful utterance and the simultaneous devaluing of the target by that under-estimation of competence. Another area where patronising behaviour has been researched in relation to gender is work on intimate relations. For instance, Buss (1989) identified 147 sources of upset (essentially impoliteness as it is conceptualised here) that men perform on women and vice-versa. One of these factors was labelled as 'condescending' and this was more frequently complained about by women with regard to men's behaviour than vice-versa.

To investigate the potential gendering of the term *patronising* itself, we can return to Figure 1 in which we saw that the collocates of *patronising* included more male participants than other mock politeness labels (*man*, *ds* [dear son], *his*, *bloke*, *he's*, *he*, *dh* [dear husband], *men*) but also collocated with female participants (*girl*, *she*, *her*, *mums*, *she's*, *women*), leaving the picture unclear. As previously, we may look to the larger corpus of EnTenTen13 to see what terms are used in similar environments through the Sketch Thesaurus function. The findings are reported in Figure 5 which shows there is a distinct negative evaluative prosody and the

presence of lexical items with clear gender associations (*paternalistic*, *misogynistic*, *sexist*).

Figure 5. Sketch Engine Thesaurus output for *sarcastic*

However, the data reported in Figure 2 did not show that this label was more likely to be used to discuss male behaviour in the context of the forum. One explanation for this which could be explored in further work is that *patronising* is not actually gendered in the same way as *sarcastic* but that it is so strongly associated with behaviours performed by those in positions of power that in many contexts it is biased towards male participants because they are more likely to hold such positions in societies.

Summary

In this short case-study, I hope to have shown how we can use collocation analysis as an entry point to our data and not only to start to unravel complex webs of connotations in relation to representation of gender but also as a preparatory stage for variation-focussed studies of language and gender.

What collocation analysis can offer language, gender and sexuality studies is an empirical basis for discussions of connotations and the two tools used here, *LancsBox* and the *Sketch Engine* Thesaurus, go further by offering visualisation techniques for displaying collocation patterns. The use of such techniques may help to make the integration of collocation analysis more accessible and something that can supplement language, gender and sexuality research coming from a broad range of methodological backgrounds (such as those discussed and illustrated in this volume), in addition to forming the centre point in research coming from corpus linguistics.

FURTHER DIRECTIONS

Given the adaptability of the method, the range of topics ripe for investigation using collocation research in language, gender and sexuality studies is almost as wide as the range of topics in the subject area itself. Labels can be interrogated for non-obvious meanings across a wide spectrum of questions. This may constitute the main scope of the research, as in some of the examples below which centre on the collocates of gendered nouns to understand how men and women are represented in public discourses. Alternatively, the collocation analysis may be integrated as way of offering another 'way in' to the data, as a form of triangulation alongside frequency analysis or non-corpus methods. Finally, the collocation analysis may be a preparatory stage to sociolinguistic variation investigations, as discussed above.

In terms of which future directions are likely to be particularly beneficial for the field, there is scope for research that takes a non-binary and/or intersectional approach in collocation analysis of language, gender and sexuality. From a methodological perspective, it would be highly interesting to see variation studies which look beyond frequency and integrate collocation analysis.

At a larger scale, important avenues for future research within language, gender and sexuality include studies of how lexical items accumulate and shed connotations over time, and the rise of diachronic corpora will facilitate this. Another significant avenue is to achieve better understanding of the extent to which people are aware of, or can be made aware of, the prosodies surrounding particular lexical items and this will be significant for creating impact from collocation studies. Finally, as a broad long term goal, understanding how the networks of evaluations and connotations relate and intertwine in representing and constructing gender and sexuality is an important area for future attention.

FURTHER READING

1. Baker, P. (2014) *Using corpora to analyze gender*. London & New York: Bloomsbury.
   This is a comprehensive overview of how corpus linguistics may be used in investigating the relationship between language and gender. Chapters 5 and 6 are particularly relevant for collocational analyses.
2. Baker, J.P. and Levon, E. (2016) 'That's what I call a man': representations of racialised and classed masculinities in the UK print media. *Gender and Language*, 10(1).
   This paper shows how collocation analysis may be integrated with other approaches to the study of language, gender and sexuality.
3. Bogetić, K. (2013) Normal straight gays: Lexical collocations and ideologies of masculinity in personal ads of Serbian gay teenagers. *Gender & Language*, 7(3): 333–367.
   This is an example of work in which collocation is absolutely central to the investigation. It is worth noting that collocation in this paper is calculated as simple frequency.
4. Brezina, V., McEnery, T. and Wattam, S. (2015) Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), pp.139-173.

This offers an overview of the importance of collocation networks as well as an introduction to the LancsBox tool.

5. Jaworska, S. and Hunt, S. (2017) 'Differentiations and intersections: a corpus-assisted discourse study of gender representations in the British press before, during and after the London Olympics 2012'. *Gender and Language* 11(3), pp.336–364.
   This is an example of how collocation analysis may be combined with frequency to investigate constructions of gender. It also addresses issues of intersectionality.

RELATED TOPICS

- Sexuality in South African news media
- Language of the law
- Representations of sexual orientation and national stereotypes

REFERENCES

Baker, P. (2006) *Using corpora in discourse analysis*. London: Continuum.

Baker, P. (2014) *Using corpora to analyze gender*. London & New York: Bloomsbury.

Baker, P. and McEnery, T. (2015) 'Who benefits when discourse gets democratised? Analysing a Twitter corpus around the British Benefits Street debate', in Baker, P. and McEnery, T. (eds) *Corpora and discourse studies.* Basingstoke: Palgrave Macmillan, pp. 244-265.

Baker, P. and Levon, E. (2016) ''That's what I call a man': representations of racialised and classed masculinities in the UK print media', *Gender and Language*, 10(1), pp. 106-139.

Baroni, M. and S. Bernardini. (2004) 'BootCaT: Bootstrapping corpora and terms from the web'. *Proceedings of LREC* (2004) Available from http://sslmit.unibo.it/~baroni/publications/lrec2004/bootcat_lrec_(2004)pdf

Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Bogetić, K. (2013) 'Normal straight gays: Lexical collocations and ideologies of masculinity in personal ads of Serbian gay teenagers'. *Gender & Language*, 7(3): 333–367.

Bowes, A., and Katz, A. (2011) 'When sarcasm stings'. *Discourse Processes,* 48(4), pp. 215-236.

Brezina, V. (2018) 'Statistical choices in corpus-based discourse analysis'. In Taylor, C. and Marchi, A. (eds) *Corpus approaches to discourse: a critical review*. Abingdon: Routledge, pp. 259-280.

Brezina, V., McEnery, T. and Wattam, S. (2015) 'Collocations in context: A new perspective on collocation networks'. *International Journal of Corpus Linguistics*, 20(2), pp.139-173.

Buss, D. M. (1989) 'Conflict between the sexes: strategic interference and the evocation of anger and upset'. *Journal of Personality and Social Psychology*, 56(5), pp. 735.

Cameron, D. (2003) 'Narrow church?'. *Critical Quarterly*, 45(4), pp. 109-112.

Colston, H. L. and Lee, S. Y. (2004) 'Gender differences in verbal irony use'. *Metaphor and Symbol*, 19(4), pp. 289-306.

Creusere, M. A. (1999) 'Theories of adults' understanding and use of irony and sarcasm: Applications to and evidence from research with children'. *Developmental Review*, 19(2), pp. 213-262.

Culpeper, J. (2011) *Impoliteness: using language to cause offence*. Cambridge: Cambridge University Press.

Dress, M.L., Kreuz, R.J., Link, K.E. and Caucci, G.M. (2008) 'Regional variation in the use of sarcasm'. *Journal of Language and Social Psychology*, 27(1): pp. 71.

Duguid, A. and A. Partington (2018) ‘Using corpus linguistics to investigate absence/s: You don’t know what you’re missing. Or do you?’. In Taylor, C. and Marchi, A. (eds) *Corpus approaches to discourse: a critical review*. Abingdon: Routledge, pp. 38-59.

Firth, J.R. (1957) *Papers in linguistics, 1934-1951*. Oxford: Oxford University Press.

Glick, P. and Fiske, S. T. (1997) ‘Hostile and benevolent sexism’. *Psychology of Women Quarterly,* 21(1), pp. 119–35.

Hummert, M. L., and Ryan, E. B. (2001). ‘Patronizing’, in Robinson, W.P. and Giles, H. (eds) *The new handbook of language and social psychology.* Chichester: Wiley, pp. 253-269.

Jakubíček, M., A. Kilgarriff, V. Kovář, P. Rychlý and V. Suchomel. (2013, July). ‘The TenTen corpus family’. Paper presented at *Corpus Linguistics Conference 2013*, Lancaster.

Katz, A., Piasecka, I. and Toplak, M. (2001) ‘Comprehending the sarcastic comments of males and females’. Poster presented at the *42nd annual meeting of the Psychonomic Society*, Orlando, FL.

Katz , A. N., Blasko, D.G. and Kazmerski, V.A. (2004) ‘Saying what you don't mean: Social influences on sarcastic language processing’. *Current Directions in Psychological Science* 13: 186-189.

Kilgarriff, A., Baisa, V. Bušta, J., Jakubícek, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014) ‘The Sketch Engine: ten years on’. *Lexicography*, 1(1): pp. 7–36.

Leech, G. (1992) ‘Corpora and theories of linguistic performance’, in Svartvik, J. (ed) *Directions in corpus linguistics: proceedings of Nobel symposium 82*. Berlin: Mouton de Gruyter, pp. 105-122.

Louw, B. (1993) ‘Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies’, in Baker, M, Thompson, G., Tognini-Bonelli, E. (eds) *Text and Technology. In Honour of John Sinclair*. Amsterdam: John Benjamins, pp. 240-251.

McEnery, T. and A. Hardie. (2012) *Corpus linguistics: method, theory and practice*. Cambridge: CUP.

Milanowicz, A. (2013) ‘Irony as a means of perception through communication channels. emotions, attitude and IQ related to irony across gender’. *Psychology of Language and Communication* 17: pp. 115-132.

Mumsnet, About us, http://www.mumsnet.com/info/aboutus (page last updated on 29/01/2015, retrieved on 12 May 2015)

Phillips, M. (1985) *Aspects of text structure: an investigation of the lexical organisation of text*. Oxford: Elsevier Science Publishers.

Rockwell, P. and E.M. Theriot (2001) ‘Culture, gender, and gender mix in encoders of sarcasm: A self-assessment analysis’. *Communication Research Reports*, 18(1), pp. 44-52.

Rychly, P. and Kilgarriff, A. (2007) ‘An efficient algorithm for building a distributional thesaurus’. *Proc ACL*. Prague, Czech Republic. Retrieved on 12 May 2015 from http://www.kilgarriff.co.uk/Publications/2007-RychlyKilg-ACL-thesauruses.pdf

Sinclair, J.M. (2004) *Trust the text: language, corpus and discourse*. London & New York: Routledge.

Stubbs, M. (2001) *Words and phrases. corpus studies of lexical semantics*. Oxford, Blackwell.

Taylor, C. (2016) *Mock Politeness in English and Italian: a corpus-assisted metalanguage analysis*. Amsterdam: John Benjamins.

Watts, R., Ide, S. and Ehrlich, K. (1992) 'Introduction'. In Watts, R., Ide, S. and Ehrlich, K. (eds) *Politeness in language: study in its history, theory and practice*. Berlin: Mouton de Gruyter, pp. 1-17.

[i] See Taylor (2016) for the full corpus description.
[ii] Sketch Engine is currently free to universities in EU member states
[iii] Freely available from http://corpora.lancs.ac.uk/lancsbox/
[iv] Calculated using a span of 5L/R, logdice, minimum collocation frequency of 5
[v] Only those instances where the behaviour of a third person was being described were counted in this stage because this mean that the gender was more likely to be specified and to avoid bias from the fact that the first and second person references would be disproportionately female in this particular corpus