

Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine¹

Michael Pearce²

Abstract

In this paper, I examine the representation of men and women in the *British National Corpus* (BNC) by focussing on the collocational and grammatical behaviour of the noun lemmas MAN and WOMAN (i.e., the nouns *man/men* and *woman/women*). Using Sketch Engine (a powerful corpus query tool, which is described) I explore the functional distribution of the target lemmas, and reveal the structured and systematic nature of the differences in the way these terms for adult male and female human beings pattern with other word forms in different grammatical relations.

1. Introduction

This article is a contribution to a growing body of work in which corpus-analytic techniques are used to derive social and cultural information from electronic corpora (see, for example, Stubbs, 1996; Baker, 2005; Baker and McEnery, 2005; Piper, 2000; Johnson and Ensslin, 2006). My concern here is the representation of men and women in a general corpus of British English, derived from an analysis of the collocational patterns associated with the lemmas MAN and WOMAN. Given the numerous social, economic and political inequalities between men and women in most societies throughout human history, it would be surprising if a large corpus of English speech and writing did not reveal significant contrasts, and this paper offers a new perspective on these differences. Using Sketch Engine, I explore the way the basic terms for adult male and female human beings pattern with other word forms in different grammatical relations. At various points, I supplement my corpus analysis with information drawn from, for instance, research on the psychology of gender and official government statistics, in order to offer possible explanations for the patterns of difference uncovered.³

¹ My thanks go to Adam Kilgarriff and Sebastian Hoffmann for fielding my questions about Sketch Engine and BNCweb respectively, and to the two anonymous reviewers whose helpful and supportive comments were much appreciated.

² Department of English, School of Arts, Design, Media and Culture, University of Sunderland, Sunderland, SR1 3PZ, United Kingdom.

Correspondence to: Michael Pearce, *e-mail:* mike.pearce@sunderland.ac.uk

³ Although these differences should not blind us to the fact that patterns of similarity are also evident – a point I return to in Section 4.

<i>man</i>	<i>men</i>	<i>man + men</i>	<i>woman</i>	<i>women</i>	<i>woman + women</i>
57,699 (60.88% of all occurrences of MAN)	37,078 (39.12% of all occurrences of MAN)	94,777	21,999 (36.52% of all occurrences of WOMAN)	38,238 (63.48% of all occurrences of WOMAN)	60,237

Table 1: *Man/men* and *woman/women* in the BNC (derived using BNCweb)

A number of studies have looked at gendered items in corpora, particularly in relation to asymmetry and sexism. Kjellmer (1986) examined the frequency and distribution of masculine and feminine pronouns, together with the words *man/men* and *woman/women* in the 1961 Brown and London–Oslo–Bergen (LOB) corpora. He found that, overall, there were more ‘masculine’ items than ‘feminine’ ones in both corpora, but noted a masculine bias in the North-American Brown corpus, compared with the British LOB corpus. Later frequency studies have shown a similar bias towards the masculine, but with interesting diachronic variation. For example, Sigley and Holmes (2002) also studied the comparative frequencies of *man/men* and *woman/women* in the Brown and LOB corpora, together with the *Wellington Corpus of Written New Zealand English* (1986–90), the *Freiburg–Brown Corpus of American English* (1991–2) and the *Freiburg–LOB Corpus of British English* (1990–1). They found that the frequency of *women* in writing doubled between the early 1960s and the early 1990s, while references to *man/men* significantly decreased. But the frequency of references to women as individuals remained below references to individual men, though the ratio of *woman:man* went up from 1:5 in the earlier corpora to 1:2 in the later corpora (Sigley and Holmes, 2002: 141). Table 1 shows the figures for the BNC. Overall, they confirm the masculine bias observed in other corpora, with MAN occurring over one and a half times more frequently than WOMAN.⁴ But when we consider the proportion of singular to plural, some interesting contrasts emerge: *man* is over one and a half times more frequent than the plural *men*, but the plural *women* is nearly twice as frequent as *woman*. It seems that adult males are more commonly referred to in the singular, while adult females are more commonly referred to collectively.

Such approaches, though useful in giving a sense of how often men and women are talked and written about, are ‘broad-brush’ and limited. But

⁴ A bias is also present in the frequency of gendered pronouns: *he* ($n = 640,614$), *she* ($n = 352,844$), *him* ($n = 153,651$), *her* ($n = 100,352$), *his* ($n = 4,616$), and *hers* ($n = 2,367$). Masculine pronouns are 1.75 times more frequent than feminine ones. Comparison of the proportion of subject pronouns (*he/she*) to object pronouns (*him/her*) reveals that 76.02 percent of masculine pronouns are the subject, compared to 71.56 percent of feminine pronouns (derived using BNCweb).

	<i>woman</i>	<i>girl</i>	<i>man</i>	<i>boy</i>
blonde	25	28	1	1
frigid	2	0	0	0
honest	11	2	68	1
hysterical	14	1	0	0
intelligent	17	9	44	3
loose	3	2	1	1
neurotic	2	2	2	0
silly	16	35	0	10
ugly	6	4	0	0

Table 2: Collocates with *man/woman* and *boy/girl* in a three million word sub-corpus of the BNC (from Romaine, 2000: 110)

corpus studies in language and gender have not been confined simply to counting lexical items. A central concern of analysts has been *collocation* – the phenomenon of certain words frequently occurring in close proximity (Baker, 2006: 96). Romaine (2000) shows how sexism in language can be demonstrated with collocational evidence. In English there are several well-known pairings of gendered items which display various kinds of semantic and discursive asymmetry. These include *master* and *mistress*, *god* and *goddess*, *governor* and *governess*, *wizard* and *witch*, and *bachelor* and *spinster*. Romaine examined the collocates of *bachelor* and *spinster* in the BNC, focussing on a single grammatical relation: the adjectives modifying *spinster*. These are usually negative or pejorative. Amongst Romaine's examples are *gossipy*, *nervy*, *ineffective*, *jealous*, *eccentric*, *frustrated*, *repressed*, *lonely*, *prim*, *cold-hearted* and *despised*. Romaine claims that such asymmetries extend to basic terms for male and female human beings. She looked at *man/woman* and *boy/girl* and found that words with negative overtones are used more frequently with *woman/girl* than with *man/boy* (see Table 2). Similar claims, based on corpus evidence, have been made by others. For example, Caldas-Coulthard and Moon (1999, cited by Hunston, 2002: 121) examined the adjectives collocating with the words *man* and *woman* in a corpus of UK newspaper articles, and found that only *woman* is modified significantly by adjectives referring to physical appearance (e.g., *beautiful*, *pretty* and *lovely*), and only *man* is modified significantly by adjectives indicating importance (e.g., *key*, *big* and *main*). Collocational patterns such as these can reveal the associations and connotations of words and, therefore, the assumptions they embody (Stubbs, 1996: 172).

This study builds on the earlier work on collocation and gender, outlined above, by sorting the collocates of MAN and WOMAN in the BNC into grammatical categories, making possible a fuller account of their collocational behaviour, and rendering the cultural meanings they embody, and so transmit, more immediately accessible. The powerful and innovative software I use in my analysis is described in the next section.

	<i>G1</i> subject		<i>G2</i> object			<i>G3</i> adjective modifier		
	19,174	4.0		15,847	1.7		28,802	2.4
die	275	31.84	arrest	225	38.42	young	3,719	65.69
stand	331	29.95	kill	318	34.73	old	2,431	51.26
sit	274	29.26	accuse	132	28.86	gay	205	43.08
walk	197	28.05	convict	64	28.54	tall	355	42.95
wear	193	27.57	marry	108	28.52	middle-aged	138	41.56
live	212	24.86	age	65	28.47	older	352	39.03
come	619	25.53	jail	55	28.12	wise	160	37.78
work	267	22.46	charge	123	26.29	homosexual	93	37.62
look	341	21.65	meet	266	26.27	younger	224	37.05
nod	56	21.51	name	102	25.32	married	209	39.91

Table 3: Part of a word sketch showing three grammatical relations for MAN⁵

2. Sketch Engine and the British National Corpus

Sketch Engine was developed by Adam Kilgarriff, Pavel Smrz and David Tugwell. The software was originally designed as a tool for dictionary makers, and is currently used by lexicographers at Oxford University Press, ChambersHarrap and Macmillan Publishers. Sketch Engine is available as a web-based Corpus Query System (CQS), through which users have access to a number of corpora, including the BNC.⁶ The software produces a ‘word sketch’ for the target lemma. This is an automated summary showing how a word combines with other words, with the various combinations grouped into grammatical relations (Kilgarriff, 2002; Kilgarriff and Tugwell, 2002; Kilgarriff *et al.*, 2004). The usefulness of this procedure is revealed when the output of a word sketch for the noun lemma MAN is compared with its ‘traditionally’ derived collocates. The twenty words which collocate most strongly with MAN in the BNC are *odd-job*, *o’war*, *repo*, *Denard*, *bald-headed*, *inhumanity*, *thin-faced*, *measureless*, *sandy-haired*, *84-year-old*, *bobsleigh*, *Lechner*, *self-made*, *thick-set*, *distinguished-looking*, *tallish*, *tattle*, *youngish*, *best-dressed* and *Piltown*.⁷ Admittedly, different statistical operations result in slightly different lists of collocates, but whatever statistic is applied, once function words have been set aside, adjectives tend to predominate, and unusual words seem over-emphasised (see Baker, 2006: 100–4). If we compare this list to the word sketch for MAN, we can immediately see the benefits: a much fuller picture of a word’s behaviour can be built up. Table 3 shows the top ten items collocating with MAN in

⁵ Top 10 collocates ordered according to saliency.

⁶ The research presented in this article is based on an earlier version of Sketch Engine. The current version gives access to additional corpora in several languages, see: <http://www.sketchengine.co.uk/>

⁷ Calculated with a Mutual Information statistic using BNCweb, based on words occurring in a +5 to –5 span of the node. Collocates occurring in a single corpus text are excluded. MI scores are in the range 6.59–5.56. See: <http://www.bncweb.info/>

three grammatical relations. G1 is the relation between a verb and its subject (e.g., *The man died*). G2 is the relation between a verb and its object (e.g., *The officers arrested the man*). G3 is the relationship between an attributive adjective and the noun it modifies (e.g., *The mourners were young men*). What do the figures in the table mean? In the BNC, the noun lemma MAN appears as the subject of a verb 19,174 times; it is the object of a verb 15,847 times, and is modified by a preceding adjective 8,234 times. Word Sketch calculates the likelihood of MAN occurring in these relationships compared with nouns in general (that is, all words tagged in the BNC as a noun). MAN is 4.0 times more likely to occur as the subject of a verb than nouns in general, 1.7 times more likely to occur as the object of a verb, and 2.4 times more likely to be modified by a preceding adjective.⁸ Word Sketch also provides a list of the lemmas which occur with statistical significance in each grammatical relation with the target lemma. Table 3 shows that the most significant collocates are *die* (31.84) when MAN is subject, *arrest* (38.42) when MAN is object, and *young* (65.69) when MAN is premodified by an adjective. These figures are known as the *saliency* of a particular relation.⁹

Word Sketch reveals patterns which can be difficult to uncover using an ordinary concordancer. It shows which grammatical roles a lemma prefers or avoids, and also displays its collocates in dozens of grammatical relations.¹⁰ Sketch Engine also allows the behaviour of two target lemmas to be compared, using a tool called Word Sketch Difference. The software's creators describe this function as 'a neat way of comparing . . . words: it shows those patterns and combinations that the two items have in common, and also those patterns and combinations that are more typical of (or even unique to) one word rather than the other' (Sketch Engine user guide¹¹). For example, a word sketch difference of MAN and WOMAN reveals that both occur as subject of the verb *scream* in the BNC. However, *scream* has a higher saliency with WOMAN (18.6) than it does with MAN (8.6). Conversely, the verb *climb* has a higher saliency with MAN (15.9) than it does with WOMAN (1.4). It is possible to infer from such comparisons that, in the BNC, climbing is more likely to be given prominence in representations of men, while screaming is more strongly associated with women.

As well as providing information about patterns and combinations which two words have in common (hereafter 'common' patterns), Word

⁸ Each of these figures is a ratio of ratios. There are 93,269 occurrences of MAN, of which 19,174 are the subject of a verb, giving a ratio of $93,269/19,174 = 4.86$. If the whole corpus contains x nouns, of which y are subject of a verb, the second ratio is $x/y = z$. The ratio of ratios is calculated by dividing 4.86 by z . In the case of MAN as subject, this gives 4.0 (Kilgarriff, personal communication).

⁹ The statistic is $MI \times \log \text{frequency}$. MI is Mutual Information, which is $\log(\text{base } 2) (N \times f_{xy}) / (f_x \times f_y)$, where f_{xy} is the frequency of the words (x and y) occurring together, f_x is the frequency of x , f_y is the frequency of y and N is the number of instances of the relationship in the corpus as a whole, (Kilgarriff, personal communication).

¹⁰ In fact, Sketch Engine sorts the collocates of high frequency items such as MAN and WOMAN into over 100 categories.

¹¹ See: <http://www.sketchengine.co.uk/Sketch-Engine-User-Guide.htm>.

Sketch Difference also identifies combinations which occur exclusively with one lemma in the target pair ('exclusive' patterns). For example, WOMAN (but not MAN) is modified by the adjectives *pretty*, *dumpy* and *scarlet*; and MAN (but not WOMAN) is modified by *burly*, *balding* and *dirty*. We might describe exclusive patterns such as these as 'gendered', since they signal a somewhat sharper contrast than those reflected in the common patterns.

The corpus used in this study is the *British National Corpus* (BNC) – a 100-million-words general corpus designed to represent contemporary British English by incorporating as many text types as possible (McEnery *et al.*, 2006: 15). The written component of the BNC (which accounts for 90 percent of the corpus) contains 'extracts from regional and national newspapers, specialist periodicals and journals ... academic books and popular fiction, published and unpublished letters and memoranda, school and university essays'. The spoken component (which accounts for 10 percent of the corpus) includes 'unscripted informal conversation, recorded by volunteers selected from different age, region and social classes ... together with spoken language collected in ... different contexts, ranging from formal business or government meetings to radio shows and phone-ins' (BNC website¹²). Its 'heterogeneous' nature (Partington, 2003: 4), in terms of the kinds of texts and speech events it contains, means that the BNC can be seen as, 'a repository of cultural information about [British] society as a whole' (Hunston, 2002: 117). Therefore, findings about how MAN and WOMAN are being used here can be extended to 'British English in general' with at least some degree of confidence. But before I outline these findings, two notes of caution need to be sounded. First, it is important to remember that in the fast-paced world of corpus linguistics, the BNC is now somewhat out of date. Of the texts in the written component of the BNC, 89.2 percent were produced in the period 1975–93 (2.3 percent were produced 1960–1974 and 8.5 percent are undated), and the spoken part consists of transcripts of speech produced in the period 1991–4. This means that the most *recent* material is from fourteen years ago. Odd though it may sound, the BNC is becoming an historical corpus, which means that my findings about MAN and WOMAN are inevitably 'dated'. An analysis of a more contemporary corpus might yield different results. The second note of caution relates to the fact that Sketch Engine analyses the BNC *as a whole*. Inevitably, there will be contrasts in the collocational behaviour of the target lemmas across text-types – an issue I return to in Section 4.

3. The collocational behaviour of MAN and WOMAN

In this paper I concentrate mainly on the three grammatical relations mentioned in the previous section: MAN/WOMAN as subject, MAN/WOMAN

¹² See: <http://www.natcorp.ox.ac.uk/corpus/>

as object, and attributive adjectives associated with MAN/WOMAN. This allows me to say something about what men and women are represented as doing and experiencing as agents, and undergoing as patients and beneficiaries, and also how they are described and categorised.¹³ Where relevant, I shall also refer to data from other grammatical relations. What emerges from the data are patterns of collocation which reflect persistent gender differences in the representation of men and women in the domains of power and deviance, social categorisation, personality and mental capacity, appearance and sexuality.¹⁴

3.1 Power and deviance

Power is not distributed equally in society: some groups of social actors possess and exercise more power than others. The distribution of power between genders is asymmetrical: men are, in general, physically stronger than women (or are at least *expected* to be so), and the bulk of financial resources and economic and political power is held by men (Jutting *et al.*, 2006). Such asymmetries are reflected in the data. A number of verbs referring to actions requiring physical strength and endurance are more strongly associated with MAN as subject than with WOMAN. These include *chase, climb, jump, leap* and *march* (Table 6). And only MAN (but not WOMAN) collocates with these verbs of physically demanding activity: *dig, hammer, haul, heave, lunge, plough, pounce, race, saw, stomp* and *struggle* (Table 7). Such patterns conform to gender role expectations about male behaviour, which is ‘expected’ to be active, aggressive, strong and dominant. Verbs such as *dig, saw, heave, hammer, plough* and *haul* also suggest that, in western cultures, men are more likely to be represented working with

¹³ In order to control the amount of data returned for each query, Sketch Engine allows for the setting of search parameters. Collocates in each grammatical relation, and the parameter settings used, can be seen under Appendix A. I should also point out that when expressions such as ‘only WOMAN’ or ‘no collocates’ are used in the analysis, they refer only to the BNC.

¹⁴ An objection to the patterns discussed in what follows might be raised in relation to the so-called ‘generic’ masculine, or ‘false’ generic (Hellinger and Bußmann, 2001: 9). This is where certain masculine forms are intended by the speaker or writer to ‘include’ male and female referents. Within this category, Holmes (1994: 36) makes a distinction between instances of MAN referring to ‘generic man’ or ‘humankind’ (as in *the ascent of man*), and ‘pseudo-generics’ – forms which ‘claim to be generic while in fact suggesting male’ (e.g., phrases such as *the man in the street, the tax man*, and so on). If the writers or speakers whose words have been incorporated into the BNC had ‘intended’ their use of the lemma MAN to include females as well, then claims about asymmetries in the representation of men and women will need to be adjusted, depending on the frequency of such occurrences. But in fact, only a minority of instances of MAN in the BNC are either ‘generic’ or ‘pseudo-generic’. For example, from a random sample (derived using BNCweb) of 100 occurrences of the noun *man* in the BNC, seven were unambiguously ‘generic’ and five were ‘pseudo-generic’. Such statistics would suggest that most instances of MAN in the BNC refer to a specific male person or persons.

tools and engaging in heavy lifting, hence this representation in the BNC. (Of course, women have toiled throughout history, but since much of this labour has been ‘domestic’ and unwaged, it has often been unacknowledged by men.)

Adjectives referring to physical size and potency also pattern more strongly with MAN: *able-bodied, big, broad-shouldered, fastest, fit, stocky, strongest, tall* and *well-built* (Table 10). Often, when women are described with these adjectives, a figurative or more general sense is being used. For example, whereas all the instances of *broad-shouldered* and *strongest* modifying MAN refer to physique and strength, this is not the case with WOMAN, for example, ‘[She] was a *broad-shouldered woman*, though not physically’(AP0), and ‘She is the *strongest woman* er heroine that we’ve read’ (K60).¹⁵ Men (but never women) are described as *barrel-chested, burly, muscular, strong-armed* and *thick-set* (Table 11). Interestingly, a number of animal terms suggesting potency are also used exclusively to describe men: *beefy, bull-necked, hawk-faced* and *ram-headed*. The absence in the BNC of noun phrases such as ‘burly woman’ and ‘bull-necked women’ would indicate that when noun phrases attributing these physical characteristics to women *are* used, they are ‘marked’. For example, of the first ten hits returned by Google for the search string ‘burly woman’, five occurred in contexts where the woman was presented as ‘deviant’ in some way (transsexual, ‘mannish’ or committing a violent act on a man).¹⁶

Some of the verbs patterning strongly with MAN as subject have core meanings associated with the exercise (or ownership) of power more generally. For example, *dominate* and *lead* associate more strongly with MAN than WOMAN, and only MAN collocates with *build, captain, conquer, hunt, mastermind, outrank* and *raid* (Tables 6 and 7). MAN as subject also patterns more strongly with the verbs *possess* and *own*. Ownership and power are often closely associated in western society and there are contrasts in what men and women are represented as owning. Women, like men, own property, money and livestock; but, unlike women, men are also represented as owning businesses (e.g., shops, restaurants, hotels), shares, machinery, land, teams, estates, clothes, vehicles, educational establishments, boats and farms. These patterns reflect contemporary and historical disparities of wealth, power and resources – a contrast that is also apparent in relation to attributive adjectives. MAN patterns more strongly with *distinguished, eminent, grand, great, influential, leading, mighty, outstanding, powerful, rich, self-made, senior, top* and *wealthy*. And only men are *monied* (Tables 10 and 11). There are contrasts here in the contexts in which these adjectives occur. For example, when a woman is *self-made*, the attribution is often accompanied by the suggestion that this is a rare phenomenon or an odd expression: ‘She was one of the few *self-made women* in Britain’ (FPB), and ‘“*Self-made women*”

¹⁵ BNC document identification codes.

¹⁶ Search carried out in August 2007.

you seem to hear less of' (HH3). And of twenty-one occurrences of *rich* pre-modifying WOMAN, over half occur in the context of their relationship to men, clothes or consumption: 'Hilbert . . . took care of himself by marrying a *rich woman*' (CDB), '*Rich women* with closets full of clothes they'll never get around to wearing' (BMW), and '*Rich women* with Chanel bags rested their weary shopping feet and met their friends for a drink' (JYD). What seems to get represented about wealthy women is a propensity to spend money, rather than make it.

One way in which human beings exert power over others is to take part in criminal and deviant activities. In the BNC, men are more strongly associated with crime, violence and the criminal justice system than women. This is unsurprising, perhaps, since crime is overwhelmingly a male activity. In the UK, 85–95 percent of burglaries, robberies, drugs offences, and crimes of violence against the person are committed by males (Ayres and Murray, 2005). This fact is reflected in the data. MAN patterns more strongly with adjectives associated with deviancy, for example, *condemned, cruel, dangerous, drunk, drunken, evil, guilty, nasty, sinful* and *violent* (Table 10). And only MAN is modified by *accused, armed, arrested, convicted, evil-looking, gun-toting, hanged, jailed, masked, strong-arm, vengeful* and *wanted* (Table 11). MAN as subject patterns more strongly with *abduct, abuse, assault, attack, beat, fight, kidnap, kill, murder* and *shoot* (Table 6). Men (but not women) *abscond, bludgeon, burgle, con, conquer, fiddle, libel, mistreat, muscle, mutilate, oppress, pounce, raid, ransack, rape* and *strangle* (Table 7). Even when the subject verbs in question might appear to have little connection with crime or violence, an examination of the concordance lines reveals a criminal or violent context. For example, MAN collocates with *jump* thirty-five times – fourteen instances relate to crime/violence, for example, 'Two armed *men* had *jumped* from a car' (ARK), and 'When a police patrol approached the *men* *jumped* into the courier's car' (HJ3). WOMAN collocates five times with *jump* – and three of these are 'jump' as an involuntary response rather than a willed action, for example, 'Both *women* *jumped* in their seats' (CR6). Also, seventeen of the twenty-two occurrences of *burst* collocating with MAN are *burst + into*, in or through a door, room or building – a sudden, often violent entrance. And fourteen of these are related to a criminal act, for example, 'The family's ordeal began when the *men* *burst* into their Colchester bungalow' (CEN). WOMAN collocates three times with *burst* – and two of these occurrences involve 'bursting' into laughter or words. Only once does a woman burst through some doors. Men also *brandish* a variety of objects (mainly weapons): a sword, a Samurai sword, a hand gun, sub-machine guns, a Bowie knife, baguettes and a garden trowel. The only things brandished by women are white feathers (in a newspaper article about the First World War). Similarly, men *wield* aluminium clubs, a gun, a blade, a knife, iron bars, a large knife, discipline, power and (by contrast) feather dusters. Women are limited to wielding pints.

There is also a contrast between men and women in the kind of power which is exercised over them by others. Many of the verbs which strongly associate with MAN as object (Table 8) position men as undergoing the powerful actions of the legal system (e.g., *accuse, arrest, catch, charge, convict, fine, hang, jail, question* and *sentence*). Verbs exclusively occurring with MAN in this category include *apprehend* and *censure* (Table 9). MAN is also presented as the victim of violence, patterning particularly strongly as object with *kill, drown, shoot* and *wound*. Some of the verbs occurring exclusively with MAN as object also position him as the victim of violent acts, while others represent him as undergoing incarceration and physical restraint. For example, only men are *apprehended, beheaded, bitten, blindfolded, clouted, devoured, handcuffed, incarcerated, knifed, restrained, shackled, slain* and *slaughtered*. Again, this is a reflection of social context: just as men commit more crimes than women, men (particularly young men) are also more likely than women to be the victims of crime, and especially violent crime (though an exception to this is rape, see below). An additional social factor influencing this pattern is that more men than women are involved directly in armed combat, and are, therefore, more likely to be killed or injured. Furthermore, MAN is also the exclusive recipient of non-violent powerful actions by others, such as verbal reprimand and abuse (e.g., *antagonize, bait, censure, curse, deflate, ridicule* and *taunt*) and seduction (e.g., *bewitch, captivate, charm, enthrall, entice* and *flatter*).

As with some of the subject verbs, even when a verb does not have a core meaning associated with the domain, a scrutiny of the concordance lines reveals that it is often being used in the context of criminality. For example, seventeen out of twenty-five occurrences of *spot* with MAN as object are in relation to a crime and/or police action, for example, 'A postman *spotted* two *men* breaking into Laurie's Chemist' (K47). By contrast, of the six occurrences of *spot* with WOMAN, only one occurs in the context of illegal activity. A similar pattern occurs with *catch*, where twenty-seven out of fifty-eight occurrences with MAN occur in a criminal context, in contrast with only one out of the sixteen occurrences of *catch* with WOMAN.

Women are also the victims of violence. WOMAN patterns particularly strongly as the object of *rape* (Table 8). This can be ascribed to the fact that rape is a crime of which females (rather than males) are overwhelmingly the victims.¹⁷ Other verbs patterning strongly with WOMAN as object, position females as the recipients of negative actions (sometimes as the victims of male sexual aggression), for example, *abduct, abuse, assault, degrade, oppress, procure* and *segregate*. Even when women are positioned as the beneficiaries of positive actions, many of the verbs favouring WOMAN in this category seem to imply some sort of weakness, lack or shortcoming on the part of the beneficiary (e.g., *advise, encourage,*

¹⁷ According to the UK Department of Health (2005) in 2003/4, '52,070 sexual offences were recorded by the police. Of this number, 13,247 were offences of rape, of which 93 percent were rape of a female and 7 percent were rape of a male'.

help, protect, screen and *treat*). What is exclusively undergone by WOMAN is often also related to the exercise of power by others. Violent acts which WOMAN (but not MAN) collocate with as object are *gag, suffocate, terrorise* and *violate*. All these verbs are frequently used in descriptions of sexual violence. Frequently, the power being exercised is ideological, though sometimes the verb implies both ideological and physical coercion, for example, *coerce, discriminate, disempower, dislodge, downgrade, dump, hoodwink, interrogate, limit, marginalize, mistreat, objectify, omit, penalise, prescribe, restrict, shame, trivialize, use* and *violate* (Table 9). As far as attributive adjectives are concerned, women are more likely than men to be *vulnerable* and *disadvantaged*, and only women are *abused, sickly* and *tired-looking* (Tables 10 and 11). Furthermore, only women ‘possess’ (in such constructions as ‘woman’s ordeal’) *dependency, inferiority, invisibility, ordeal, powerlessness, shackle(s), softness, subordination, tear(s)* and *unhappiness* – all attributes associated with lack of status and power.

WOMAN also collocates exclusively as object with verbs of observation, categorisation, analysis and intervention – processes involving the exercise of power by others, for example, *assist, categorise, compensate, conceptualise, construct, cushion, define, direct, equate, exhibit, highlight, immunise, impregnate, integrate, interpret, monitor, nurse, organize, provide, regulate, section* and *sterilise* (Table 9). It is worth noting that many of these verbs position women as undergoing medical or psychiatric interventions and procedures (e.g., *section, sterilise, immunise, impregnate, nurse* and *monitor*).

The findings outlined in this section suggest important asymmetries in the way men and women are represented in relation to power and deviance. In the representation of men, emphasis is placed on strength and potency. Men are stronger than women, and exercise other forms of power, such as ownership, more readily. They commit more crime and are more violent than women. Women, on the other hand, are more likely to be represented as the recipients of the exercise of power by others, especially in relation to sexual violence (e.g., *rape, gag* and *violate*), limitation (e.g., *disempower, limit* and *marginalize*), categorisation, analysis and interpretation (e.g., *conceptualise, define* and *categorize*). Arguably, this is due to the fact that there is a tendency for women in the corpus to be represented as objects of sociological enquiry and discussion, which involves their marginalisation and oppression being written and talked about. This tendency extends to the way women are categorised.

3.2 Social categories

Particularly prominent amongst the adjectives patterning with WOMAN are those which represent women as belonging to social categories. For example, by comparison with men, they are more often characterised by adjectives signalling marital/reproductive status and sexual orientation. Thus,

WOMAN patterns more strongly with *celibate, childless, fertile, heterosexual, lone, married, non-married, separated, single, unmarried* and *widowed* (Table 10), and only women are, for instance, *barren, childbearing, ex-married, lesbian, menopausal, menstruating, motherly, motherly-looking, multiparous, postmenopausal, pregnant, remarried* and *subfertile* (Table 11). In this semantic domain, MAN patterns more strongly only with *gay* and *homosexual*. Arguably, some of these adjectives reflect biological facts: only women menstruate and bear children. However, ‘simple’ biological differences cannot account for the fact that, in the BNC, there are nearly three times as many instances of WOMAN modified by *married* ($n = 657$) as MAN ($n = 289$), and fifteen times as many instances of *childless* modifying WOMAN ($n = 15$) than MAN ($n = 1$).¹⁸ It seems that these aspects of a woman’s identity are of greater interest and concern than a man’s, and are given more prominence in the corpus. The same might be said of aspects of national, religious, ethnic and class identity. For example, WOMAN is saliently or exclusively modified by adjectives of nationality (*American, Bangladeshi, Bengali, British, Filipino, French, Indian, Iranian, Irish, Pakistani, Palestinian, Salvadorean, Saudi* and *Scottish*); religion (*Catholic, Hindu, Muslim* and *Sikh*); ethnicity (*African, African-American, Afro-Caribbean, Arab, Asian, Coptic, Euro-American, gentile* and *gipsy*); and class (*high-caste, lower-class, middle-class, upper-class* and *working-class*). No markers from these domains occur more strongly or exclusively in association with MAN.

Two factors seem to be at work here. First, as was the case with adjectives signalling marital/reproductive status and sexual orientation, nationality, ethnicity and so on are important in discussions of women in sociological discourse. Second, the prominence of these adjectives might also be connected to the fact that women are marked in this area of meaning. Some nationality/religion/ethnicity adjectives can be used as nouns and applied to people, but when the adjective is not marked for gender on the surface (as in *Arab, American, British, French, Muslim* and so on) the gender of the individual referred to in this way is nearly always ‘understood’ to be male.¹⁹ Once again, asymmetries in the representation of men and women are revealed in patterns of collocation.

¹⁸ These are raw frequencies derived from BNCweb; they have not been normed to take into account the fact that MAN occurs more frequently in the corpus than WOMAN, which makes these figures all the more startling.

¹⁹ BNCweb identifies 113 instances of *American* + WOMAN in the BNC, compared with seventeen instances of *American* + MAN. Issues of markedness are also involved in collocates where MAN/WOMAN modifies a noun, as in ‘woman doctor’. Examples of nouns referring to professional roles strongly or exclusively associated with WOMAN include: *artist, constable, deacon, detective, doctor, film-maker, jockey, journalist, lawyer, MP, newsreader, novelist, photographer, poet, priest, pro-vice-chancellor, psychologist, reporter, solicitor* and *writer*. In these instances the noteworthiness and perhaps rarity of a female occupying these roles is indicated by the use of WOMAN as a noun modifier. In other words, the ‘default’ doctor, lawyer or MP is male.

3.3 Personality and mental capacity

The collocational evidence reveals contrasts in how men and women are categorised and which aspects of their social identity are given prominence in representations, but it also points to differences in the mental and behavioural characteristics of men and women. A standard and widely used taxonomy of human personality is based on the ‘lexical hypothesis’, which states that ways for describing how people differ have become encoded in language, and these accumulations of person descriptors, ‘can serve as signposts that guide personality psychologists toward particularly important individual difference dimensions’ (Schmitt and Buss, 2000: 142). Five dimensions of description have been derived from the lexical hypothesis, representing personality at the broadest level of abstraction: these are extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience (or intellect). The so-called ‘Big Five’ (Goldberg, 1981) is a useful point of departure for an examination of the collocational behaviour of MAN and WOMAN in relation to the representation of personality. Table 4 shows those adjectives used attributively (e.g., ‘she is a mad woman’) and predicatively (e.g., ‘the man is mad’) which collocate with MAN and WOMAN. In this section I will also refer to other grammatical relations to build up a picture of how ‘male’ and ‘female’ personality is represented in the corpus.

Extraversion is associated with activity, friendliness, sociability, assertiveness and talkativeness. MAN seems to be more strongly associated than WOMAN with words that convey activity and assertiveness. We have already seen evidence of this in the verbs patterning more strongly (or exclusively) with MAN, some of which can be associated with extraversion, including the verbs of energetic action discussed under Section 3.1 (e.g., *climb*, *jump*, *leap* and *pounce*). Furthermore, some of the attributive adjectives patterning more strongly with MAN which are associated with power are also associated with extraversion (e.g., *powerful*, *eminent* and *influential*). Fewer extrovert behaviours and characteristics are associated with WOMAN, and it is noteworthy that several of these have strong sexual connotations (see Section 3.5). For example, only women *flaunt*, and WOMAN patterns more strongly with the adjectives *spirited* and *promiscuous*, and exclusively with *vivacious*.

Talkativeness is also a marker of extraversion. Figure 1 shows a contrast in the evaluative aspect of some of the subject verbs referring to speech or other forms of vocal expression. Men’s voices are often used to convey intensity and passion: they use ‘bad’ language (*swear*, *curse*), they are noisy (*shout*, *yell*) and verbally aggressive (*snarl*, *growl*). Women’s voices are represented as more emotionally intemperate: verbs referring to heightened emotional states such as *weep*, *cry* and *sob* pattern more strongly with WOMAN (see the discussion of neuroticism below). Women (but not men) also indulge in verbal pestering and fussing. They *berate*, *nag* and *cluck*, and are exclusively characterised as *bossy*, *chattering* and *gossiping*. These negative evaluations correspond with widely-held folklinguistic beliefs about

	<i>MAN</i>	<i>WOMAN</i>
+ extraversion	eminent, garrulous, gregarious , influential, powerful	bossy, chattering, gossiping , promiscuous, spirited, vivacious
– extraversion	ascetic, cautious , humble, quiet, reserved, sensitive, shy, unassuming,	submissive
+ agreeableness	affable , amiable, amiable-looking , avuncular , charming, considerate, content, contented, courteous, funniest , funny, generous, good-natured , happier, happiest, happy, happy, jolly, jovial , kind, kindest, <u>kindly</u> , likeable , merry, mild-mannered , nice, nicest, personable , polite	<u>glad</u> , <u>grateful</u>
– agreeableness	arrogant, cruel , cruel, <u>dangerous</u> , dour , embittered, evil, hateful, <u>impossible</u> , indifferent , insensitive , insufferable , nasty, <u>proud</u> , sinful, <u>unwilling</u> , violent	bitchy
+ conscientiousness	braver , conscientious, earnest, faithful, generous, good, humane , law-worthy , loyal, patient , prudent, reasonable, sincere , thoughtful, tolerant , trusted , trustworthy , truthful , upright, upstanding	
– conscientiousness		
+ neuroticism	anxious , insane , mad, <u>scared</u> , <u>sensitive</u> , <u>upset</u>	<u>dissatisfied</u> , <u>distraught</u> , hysterical , mad, neurotic, silly, weeping
– neuroticism	sane	<u>satisfied</u>
+ openness to experience	astute , brilliant, clever, gifted, learned, rational, reasonable, scholarly , self-educated , shrewd, thoughtful, wise, wiser	resourceful, strong-minded
– openness to experience	ignorant , retarded	daft, <u>dependent</u> , dumb

Table 4: Adjectives of personality (shared adjectives are in regular font and divided into columns according to strength of association; shared predicative adjectives are underlined; adjectives occurring exclusively with the target lemma are in boldface)

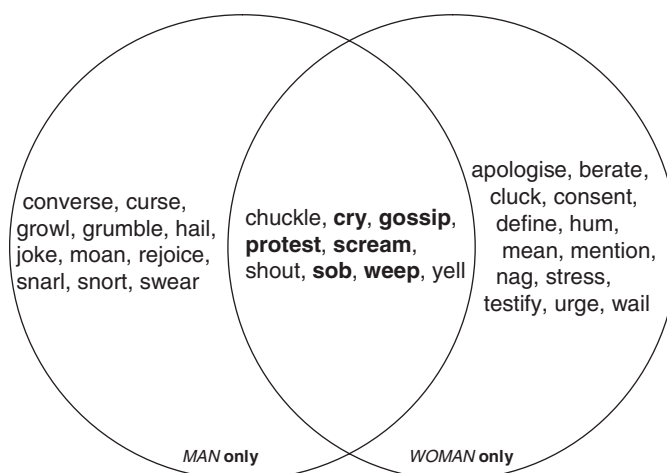


Figure 1: Speech and vocal expression (common items that pattern more strongly with WOMAN are shown in boldface)

male and female speech. Men are associated with taboo language and verbal aggression; women are nagging chatterboxes (Talbot, 2003).

Personality traits are dimensional. At the opposite end of the extroversion scale is introversion (indicated by ‘– extraversion’ in the table). Introvert people tend to be reserved, quiet, passive and sober. Some adjectives patterning more strongly with MAN potentially signal introversion, for example, *quiet*, *reserved*, *shy* and *unassuming*. There is only one marker of introversion associated with WOMAN: *submissive* (as a predicative adjective).

Agreeableness is associated with traits such as trustfulness, good-naturedness, kindness and affection. MAN patterns more strongly with adjectives such as *amiable*, *generous*, *kindly*, *likeable* and *nice*. And only MAN is modified by *affable*, *amiable-looking*, *avuncular*, *good-natured*, *gregarious* and *personable*. MAN also patterns more strongly with adjectives associated with humour and happiness, such as *contented*, *funny*, *happy*, *jolly* and *merry*, and exclusively with *funniest* and *jovial*. Other ‘agreeable’ characteristics associated with MAN include *considerate*, *courteous*, *mild-mannered* and *polite*. At the other end of the scale, disagreeableness is associated with ruthlessness, suspicion, uncooperativeness and selfishness. MAN patterns more strongly with a number of ‘disagreeable’ states and characteristics, including *arrogant*, *cruel*, *embittered*, *evil*, *hateful*, *nasty*, *sinful* and *violent*. One might also consider some of the subject verbs of violence and aggression, discussed in Section 3.1, as indicative of general disagreeableness (e.g., *attack*, *abuse* and *mistreat*). Items patterning more strongly or exclusively with WOMAN in this category are limited to *glad* and *grateful* at the positive end of the scale, and *bitchy* at the negative end.

Conscientiousness is associated with traits such as efficiency, thoroughness and discipline. MAN patterns more strongly with adjectives such as *conscientious, earnest, faithful, good* and *prudent*, and exclusively with, for example, *braver, patient, trusted* and *upstanding*. At the other end of this dimension are traits such as laziness, aimlessness and negligence, and MAN does not collocate significantly with any adjectives from this area of meaning. No adjectives from anywhere on this personality dimension associate significantly with WOMAN.

However, this is not the case with **neuroticism**, a dimension involving traits such as worry, insecurity, anxiety and depression. Here WOMAN patterns more strongly with *distraught, dissatisfied, mad, neurotic* and *silly*, and only WOMAN collocates with *hysterical* and *weeping*. MAN patterns more strongly with *mad, scared* and *upset*, and exclusively with *anxious* and *insane*. Interestingly, all the adjectives referring to negative emotional states patterning with MAN are predicative, while six of the seven adjectives patterning with WOMAN are attributive. An important difference between attributive and predicative use of adjectives is that, 'attributive adjectives tend to characterise a thing in terms of a stable, inherent property, whereas predicative adjectives tend to denote more temporary, circumstantial properties' (Taylor, 2002: 455). This would suggest that in a phrase such as 'mad woman' the condition of madness is more closely tied to the noun than it is in a phrase like 'the man is mad', which suggests there are possible temporal constraints to the condition, or that the 'madness' is limited to a particular entity often referred to in a complement (e.g., *the man is mad on sport*). This evidence suggests that women, more than men, are presented as suffering from permanent (or at least more intractable) negative mental states.

Openness to experience is associated with traits such as imagination, independence, creativity and intellectual curiosity. Some of these relate to intellect and/or wisdom. For example, MAN patterns more strongly with *brilliant, clever, gifted, learned, rational, reasonable, shrewd, thoughtful* and *wise*. And only MAN is modified by *astute, scholarly* and *self-educated* (but also *retarded*). WOMAN is not strongly associated with any adjectives referring to intellect and wisdom. However, WOMAN does pattern more strongly than MAN with several adjectives of mental fortitude and flexibility: *resourceful, spirited* and *strong-minded*. It is also interesting to note that, with the exception of *retarded* that patterns exclusively with MAN, adjectives of mental weakness and incapacity tend to be more associated with WOMAN: *daft* and *dumb*.

How far do the patterns of collocation revealed by Word Sketch Difference confirm findings in the psychological research into gender and personality? Costa *et al.* (2001), in their review of research on gender and the 'big five' model, found that men generally test higher than women in agentive facets of extraversion, such as assertiveness and excitement-seeking. To some extent, this is supported by the collocational data presented here, as are the claims made by Costa *et al.* (2001) that women test consistently higher in

neuroticism, and men are higher in openness to ideas. However, other claims appear to run counter to the corpus data. For example, Costa *et al.* (2001) found that women rank consistently higher in agreeableness, but this is not reflected in the patterns of collocation. And as far as conscientiousness is concerned, men in the corpus are represented as more dutiful than women; yet Costa *et al.* (2001) show that in most cultures, women are in fact more dutiful than men.

3.4 Appearance

As well as gender differences in the representation of personality in the BNC, there are also contrasts in the physical appearance of men and women. As we saw in Section 3.1, MAN is more strongly associated with attributive adjectives referring to physical strength and prowess (e.g., *able-bodied*, *fastest*, *fit* and *strongest*). It is not surprising that MAN also patterns more strongly or exclusively with adjectives referring to physical size, weight and bulk: *barrel-chested*, *beefy*, *big*, *bigger*, *broad-shouldered*, *bulky*, *bull-necked*, *burly*, *heavy-set*, *portly*, *solid*, *squat*, *stocky*, *stout*, *stoutish*, *tall*, *taller*, *tallest*, *tallish*, *thick-set*, *tubby* and *well-built*. The adjectives patterning more strongly or exclusively with WOMAN refer to a more limited range of bodily types, shapes and elements. Some refer to weight and size (*dummy*, *obese*, *pear-shaped*, *plump*, *plumpish* and *slender*), and others to breasts (*big-bosomed*, *buxom* and *large-breasted*). Certain cultural preoccupations are revealed here: an emphasis on strength and physique for males, and weight and breast-size for females (see also Section 3.5).

Men's facial appearance is more variously described than women's (Figure 2). Men's faces are compared with a wider range of animals (e.g., *falcon*, *hawk*, *ram*, *rat* and *weasel*) than women's (*ferret*). Adjectives referring to hair colour and style pattern more strongly or exclusively with MAN, including *blond*, *bald*, *bald-headed*, *balding*, *clean-cut*, *curly-haired*, *fair-haired*, *ginger* and *sandy-haired*. In this domain, WOMAN is associated with *blonde* and *blonde-haired*. And men's facial expressions are also more varied than women's: only men *beam*, *leer*, *scowl* and *squint*, and only women *arch* (their eyebrows). In relation to evaluative terms, WOMAN patterns more strongly with *attractive*, *beautiful*, *pleasant-looking* and exclusively with *pretty*, while male attractiveness is captured in adjectives such as *best-looking*, *devastating-looking* and *fantastic-looking*.

3.5 Sexuality

The final area I will consider is gender differences in the representation of sexual/intimate behaviour. As subject, MAN patterns more strongly with *kiss*, and occurs exclusively with *rape*; while WOMAN favours *undress* and occurs

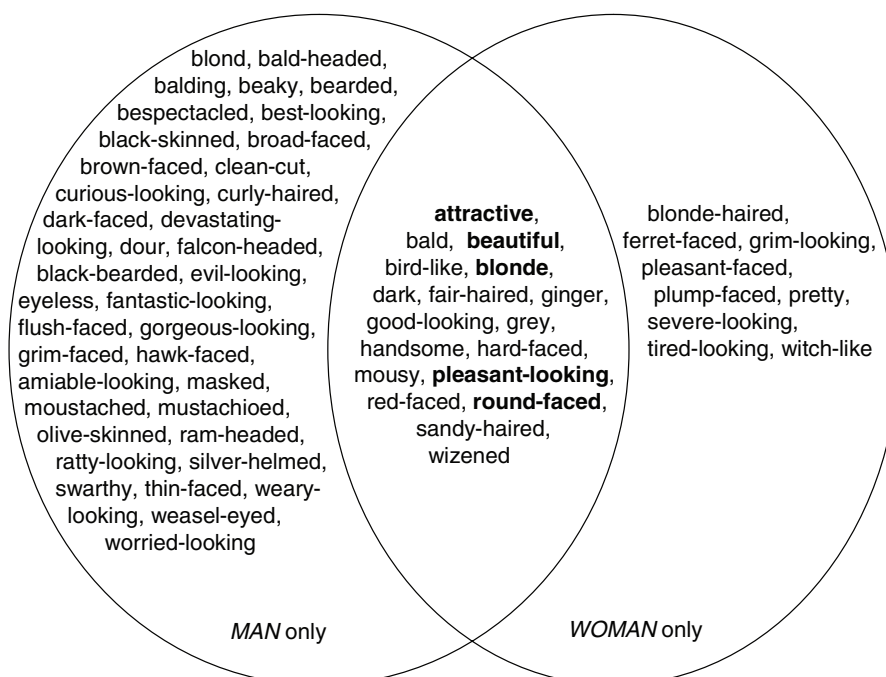


Figure 2: Facial appearance: attributive adjectives (shared terms that pattern more strongly with WOMAN are shown in boldface)

exclusively with *cuddle*, *flaunt* and *hug*. As object WOMAN patterns more strongly with *rape* and exclusively with *bed*, *date*, *dump*, *impregnate*, *ravish*, *sexualize*, *shag* and *violate*; while MAN occurs exclusively with *bewitch*, *captive*, *charm*, *enthral*, *entice* and *flatter*. The contrasts here are clear. We have already seen in Section 3.1 how women are more likely than men to be represented as the victims of male sexual violence (e.g., *rape* and *violate*) while men are more likely to be seduced by women (e.g., *bewitch* and *entice*). Other inferences which might be drawn from the collocational evidence are that women are sexually provocative (e.g., *flaunt* and *entice*), and at the same time capable of a gentle sort of intimacy (*cuddle* and *hug*); while men do not provoke women in this way, nor are they intimate. Women are also represented as ‘recipients’ of sexual activity (*bed*, *ravish* and *shag*) – a role that is not played by men.

Sexuality and sexual attractiveness is also captured in attributive adjectives. MAN patterns more strongly than WOMAN or exclusively with adjectives of sexual orientation (*homosexual*, *homosexual/bisexual*), ‘maleness’ (*masculine*, *virile*), general attractiveness (*best-looking*, *devastating-looking*, *fantastic-looking*, *good-looking*, *gorgeous-looking*, *handsome* and *sexiest*). Only two adjectives have clearly negative connotations: *lecherous* and *macho*. Adjectives with negative connotations are more strongly associated with WOMAN, including some associated

with sexual promiscuity (*blowsy, fallen, promiscuous* and *scarlet*), sexual inhibition (*frigid*) and 'deviance' (*butch*). As far as aspects of appearance relating to sexuality are concerned, adjectives in this category collocate only with WOMAN: *bare-breasted, big-bosomed, buxom* and *large-breasted*. Adjectives referring to aspects of reproduction are also exclusively associated with WOMAN, including *barren* and *childbearing*.

4. Discussion and conclusion

What conclusions about the representation of gender in the BNC can be drawn from the evidence presented here? Across the five domains, the collocates of MAN and WOMAN seem often to represent gender in stereotypical ways. Gender stereotypes, as ordered and culturally shared sets of beliefs about the characteristics of men and women, include information about physical appearance, attitudes and interests, psychological traits, social relations and occupations (Golombok, 1994). Stereotypical representations of the 'masculine' personality, emphasise traits such as competitiveness, adventurousness, independence, rationality and aggression. Physically, the stereotypical male is strong, rugged and muscular. The 'feminine' personality is co-operative, gentle, dependent, emotional and sympathetic, while the stereotypical female is physically weak (Golombok, 1994; Diekman and Eagly, 2000). Certainly, there is a marked contrast in what seems to be emphasised in relation to physical size and strength. MAN patterns more strongly than WOMAN as subject with verbs of action requiring strength and endurance, as well as with adjectives of physical size and potency (Table 5). Behaviours and traits associated with competitiveness, rationality, aggression and dominance are also evident in the collocates of MAN (e.g., *lead, conquer, wise* and *assault*). WOMAN, on the other hand, often patterns with items that are associated with the stereotypical female. For example, emotional intemperance is captured in subject verbs such as *weep, cry* and *wail*, and adjectives such as *distraught, neurotic* and *hysterical*. Physical weakness and subordination are evident in the extent to which women are represented as the victims of violence (in object verbs such as *rape* and *assault*) and the recipients of powerful actions by others (e.g., *coerce* and *marginalize*). Further evidence pointing to the presence of stereotypical representations of gender in the BNC include certain adjectives of appearance (e.g., men are *barrel-chested, broad-shouldered* and *stout*, while women are *buxom, plump* and *slender*), and verbs positioning women as the 'recipients' of the sexual actions of others (e.g., *bed, ravish* and *shag*).

The fact that the BNC contains stereotypical representations of men and women should come as no surprise. Gender is a social construct, established and reproduced in discourse (Bradley, 2007). Because texts are a product of discourse, the examination of gendered items in a large corpus is bound to reveal culturally-prominent patterns of representation. But stereotyping is not the whole story. The collocational evidence for WOMAN

MAN as subject	WOMAN as subject
Actions requiring physical strength and endurance (<i>dig, climb, jump, leap</i>)	
Exercise/ownership of power (<i>conquer, dominate, lead</i>)	
Criminal and/or violent acts (<i>assault, attack, rape, strangle</i>)	
Intense and passionate verbal/vocal expression (<i>shout, snarl, swear, yell</i>)	Emotionally intemperate verbal/vocal expression (<i>cry, scream, wail, weep</i>)
MAN as object	WOMAN as object
Undergoing actions of legal system (<i>apprehend, arrest, convict, sentence</i>)	
Victims of violence (<i>kill, knife, shoot, wound</i>)	Victims of violence (<i>assault, gag, rape, violate</i>)
Incarceration and physical restraint (<i>catch, handcuff, jail, restrain</i>)	
	Ideological and physical coercion (<i>coerce, disempower, downgrade, marginalize</i>)
	Observation, categorisation, analysis, intervention (<i>define, exhibit, interpret, monitor</i>)
	'Recipients' of sexual activity (<i>bed, ravish, shag</i>)
MAN modified by attributive adjectives	WOMAN modified by attributive adjectives
Physical size and potency (<i>big, burly, fit, tall</i>)	
Power, wealth, influence (<i>great, powerful, rich, self-made</i>)	
Deviancy (<i>cruel, evil, violent, wanted</i>)	
	Marital/reproductive status and sexual orientation (<i>childless, heterosexual, married, remarried</i>)
	Nationality, religion, ethnicity, class (<i>American, Arab, Catholic, French</i>)
Personality traits (<i>affable, arrogant, clever, evil, faithful, generous, good, gregarious, hateful, kind, nice, quiet, reserved, scholarly, shy, wise</i>)	Personality traits (<i>bossy, distraught, hysterical, neurotic, resourceful, spirited, strong-minded, vivacious</i>)
Appearance (<i>barrel-chested, broad-shouldered, stout</i>)	Appearance (<i>buxom, plump, slender</i>)

Table 5: Summary of example collocates in three grammatical relations (items that occur exclusively with the target lemma are in boldface)

points also to a ‘sociological’ discourse; in other words, women are presented as the objects of sociological enquiry within a discourse which acknowledges their subordinate status and attempts to redress it (see Section 3.2). Again, this is not surprising, especially given the date of the BNC’s constituent texts (mainly 1975–94). This period overlaps with what is sometimes referred to as the ‘second-wave’ of feminism, a time of extensive discussion about gender and the nature of women’s oppression, in academic circles and beyond (see James, 2003). If the stereotypes are depressing, then perhaps this ‘strand’ is more encouraging.

A further factor that needs to be taken into account when considering the collocational patterns revealed here is related to the composition of the BNC.²⁰ A full analysis of these patterns in the different constituent text-types of the corpus is beyond the scope of this article. However, by using BNCweb to compare the collocates of MAN and WOMAN in different parts of the BNC, it is possible to find evidence which at least suggests that some of the patterns discussed so far are not distributed evenly, but are limited mainly to particular text domains. Three examples will suffice, but others could equally well be cited. First, as we might expect, the aberrant male is a staple of news reporting, and the distribution of the sequence *adjective of deviancy* + MAN (e.g., *dangerous*, *armed* and *convicted*) seems to confirm this. These constructions occur over four times more frequently (per million words) across all news genres than they do across the entire BNC.²¹ The second example concerns the adjectives of ‘neuroticism’ patterning with WOMAN. 61.45 percent of instances of the sequence *adjective of neuroticism* + WOMAN (e.g., *distraught*, *hysterical* and *silly*) occur in prose fiction texts, even though prose fiction makes up only 16 percent of the word count of the BNC. Third, markers of social class (e.g., *working-class* and *middle-class*) premodify WOMAN over nineteen times more frequently per million words in the genre of social science than they do in the BNC as a whole, which suggests that the ‘sociological’ discourse referred to earlier might be particularly associated with this domain.²²

The final area of concern in relation to the patterns shown here is the privileging of difference over similarity. Inevitably, with a tool (Word Sketch Difference) that is designed, as its name suggests, to reveal contrasts, the analyst is in danger of exaggerating the differences and overlooking similarities. For example, both target lemmas associate at similar strengths (as subject and/or object) with common verbs like *get*, *go*, *know* and *come*. However, it seems that even when the strength of association as calculated

²⁰ For a description of the structure of the BNC, see Burnard, 2000.

²¹ A search using BNCweb for *condemned/dangerous/drunk/drunken/guilty/violent/accused/armed/arrested/convicted/gun-toting/hanged/jailed/masked/strong-arm/wanted* + MAN resulted in 242 hits (2.99 per million words). These strings occur 12.75 times per million words in the news genres of the BNC.

²² A search with BNCweb for *working-class/upper-class/middle-class/lower-class/high-caste* + WOMAN resulted in 207 hits (2.12 per million words). These strings occur 19.21 times per million words in the social science genres of the BNC.

by Word Sketch Difference is very similar for MAN and WOMAN, there are differences – recoverable by collocation analysis – in relation to the contexts in which these verbs are used. For example, the top two collocates for the search string MAN + GET are *drunk* and *car*; for WOMAN they are *men* and *help*.²³

Nevertheless, setting aside these caveats about the uneven distribution of collocational patterns and the focus on contrasts rather than correspondences, this paper has demonstrated the usefulness of Sketch Engine as a tool for the rapid extraction of social and cultural information from a corpus, particularly in its ability to sort collocates in ways which are intuitively appealing to the analyst. Earlier collocational studies exploring gender concentrated mainly on the adjectives modifying target items. Sketch Engine allows for a wider range of grammatical relations to be considered, including the potentially revealing roles of grammatical subject and object. Future research on gender representation in the BNC using Sketch Engine might involve extending the range of grammatical relations examined. Other possibilities include taking into account factors such as the gender or age of the speaker or writer, or the sex of the ‘target audience’ (indeed, BNC texts are tagged with this information). Further ways of building up a more nuanced picture of gender might also involve the examination of other gendered binaries (e.g., *boy/girl*, *lady/gentleman* and *male/female*).

References

- Ayres, M. and L. Murray. 2005. ‘Arrests for recorded crime (notifiable offences) and the operation of certain police powers under PACE England and Wales, 2004/05’, Home Office Statistical Bureau Issue 21/05. Accessed 13 June 2007, at: www.statewatch.org/news/2005/dec/uk-hosb-s-and-s.pdf
- Baker, P. 2005. *Public Discourses of Gay Men*. London: Routledge.
- Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, P. and T. McEnery. 2005. ‘A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts’, *Journal of Language and Politics* 4 (2), pp. 197–226.
- BNCweb (CQP-Edition) A Web-Based Interface to the British National Corpus (BNC). Available online at: <http://www.bncweb.info/>
- Bradley, H. 2007. *Gender*. Cambridge: Polity.

²³ Calculated with a Mutual Information statistic using BNCweb, based on words occurring in a +5 to –5 span of the node. MI scores are in the range 6.97–3.70. See: <http://www.bncweb.info/>

- British National Corpus (BNC). Available online at: <http://www.natcorp.ox.ac.uk/>
- Burnard, L. (ed.) 2000. Reference Guide for the British National Corpus (World Edition). Humanities Computing Unit at Oxford University Computing Services. Accessed 25 June 2007, at: <http://www.natcorp.ox.ac.uk/docs/userManual/>
- Caldas-Coulthard, C.R. and R. Moon. 1999. 'Curvy, Hunky, Kinky: Using Corpora as Tools in Critical Analysis', paper presented at the Critical Discourse Analysis Symposium, University of Birmingham, April 1999.
- Costa, P.T. Jr., A. Terracciano and R.R. McCrae. 2001. 'Gender differences in personality traits across cultures: robust and surprising findings', *Journal of Personality and Social Psychology* 81 (2), pp. 322–31.
- Department of Health Children and Mental Health Division and Home Office Violent Crime Unit. 2005. Developing Sexual Assault Referral Centres (SARCs) – National Service Guidelines. Accessed 11 July 2007, at: <http://www.crimereduction.gov.uk/sexual/sexual22.pdf>
- Diekman, A. and A.H. Eagly. 2000. 'Stereotypes as dynamic constructs: women and men of the past, present, and future', *Personality and Social Psychology* (26), pp. 1171–88.
- Goldberg, L.R. 1981. 'Language and individual differences: the search for universals in personality lexicons' in L. Wheeler (ed.) *Review of Personality and Social Psychology* Volume 2, pp. 141–65. Beverly Hills: Sage Publications.
- Golombok, S. 1994. *Gender Development*. Cambridge: Cambridge University Press.
- Hellinger, M. and H. Bußmann. 2001. 'Gender across languages' in M. Hellinger and H. Bußmann (eds) *Gender Across Languages: The Linguistic Representation of Women and Men*, Volume 1, pp. 1–25. Amsterdam: John Benjamins.
- Holmes, J. 1994. 'Inferring language change from computer corpora: some methodological problems', *ICAME Journal* (18), pp. 27–40.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- James, S. 2003. 'Feminisms' in T. Ball and R. Bellamy (eds) *The Cambridge History of Twentieth-Century Political Thought*, pp. 493–516. Cambridge: Cambridge University Press.
- Johnson, S. and A. Ensslin. 2006. 'Language in the news: investigating representations of "Englishness" using WordSmith Tools', *Corpora* 1 (2), pp. 153–85.
- Jutting, J.P., C. Morrisson, J. Dayton-Johnson and D. Dreschler. 2006. 'Measuring gender (in)equality: introducing the gender, institutions

- and development data base (GID). OECD Development Centre Working Paper No. 247. Available online at: <http://www.oecd.org/dataoecd/17/49/36228820.pdf>
- Kilgarriff, A. 2002. 'How to learn about fish', EL Gazette, August 2002. Accessed 12 September 2007, at: <http://www.macmillandictionary.com/articleskilgarriff.htm>
- Kilgarriff, A. and D. Tugwell. 2002. 'Sketching words' in M.-H. Corréard (ed.) *Lexicography and Natural Language Processing: A Festschrift in Honour of B.T.S. Atkins*, pp. 125–37. EURALEX.
- Kilgarriff, A., P. Rychly, P. Smrz and D. Tugwell. 2004. 'The Sketch Engine' in *Proceedings of EURALEX 2004*, Lorient, France. Accessed 30 August 2006, at: <http://www.sketchengine.co.uk/sketch-engine-elx04.pdf>
- Kjellmer, G. (1986) '“The lesser man”: observations on the role of women in modern English writings' in J. Aarts and W. Meijs (eds) *Corpus Linguistics II*, pp. 163–76. Amsterdam: Rodopi.
- McEnery, T., R. Xiao and Y. Tono. 2006. *Corpus-Based Language Studies*. London: Routledge.
- Partington, A. 2003. *The Linguistics of Political Argument*. London: Routledge.
- Piper, A. 2000. 'Some have credit cards and others have giro cheques: a corpus study of “individuals” and “people” as lifelong learners in late modernity', *Discourse and Society* 11 (4), pp. 515–42.
- Romaine, S. 2000. *Language in Society: An Introduction to Sociolinguistics* (second edition). Oxford: Oxford University Press.
- Schmitt, D.P. and D.M. Buss. 2000. 'Sexual dimensions of person description: beyond or subsumed by the Big Five?', *Journal of Research in Personality* (34), pp. 141–77.
- Sigley, R. and J. Holmes. 2002. 'Girl-watching in corpora of English', *Journal of English Linguistics* 30 (2), pp. 138–57.
- Sketch Engine. <http://www.sketchengine.co.uk/>
- Sketch Engine User Guide. Accessed 11 January 2007, at: <http://www.sketchengine.co.uk/Sketch-Engine-User-Guide.htm>
- Stubbs, M. 1996. *Text and Corpus Analysis*. London: Blackwell.
- Talbot, M. 2003. 'Gender stereotypes: reproduction and challenge' in J. Holmes and M. Meyerhoff (eds) *The Handbook of Language and Gender*, pp. 468–86. Oxford: Blackwell.
- Taylor, J.R. 2002. *Cognitive Grammar*. Oxford: Oxford University Press.

Appendix A

Shared verbs favouring MAN Verbs in boldface favour MAN between 14 and 20.9 saliency points ahead of WOMAN. Other verbs favour MAN between 7 and 13.9 ahead of WOMAN.	Shared verbs favouring WOMAN Verbs favouring WOMAN between 7 and 13.9 ahead of MAN.
abduct, abuse, admit, advance, appear, assault , attack , beat, believe, betray, bleed, brandish, break, burst, chase, chuckle, climb , desert, despise, dominate, dream, escape, fight, grab , grin, hide, invent, jump, kidnap, kill, kiss, lead, leap, march, murder, name, own, possess, ride, search, seize, shoot, shout, sleep, stagger, stand, stumble, surrender, swing, tread, try, wield, yell	attend, choose, cry, experience, faint, gossip, interview, participate, protest, scorn, scream, sob, smoke, study, undress, weep

Table 6: Subject verbs compared

Top 100 most frequently occurring verbs which collocate with MAN as subject but <i>not</i> with WOMAN (at 2+ frequency). Boldface, underlined words occur 20+ times in this relation. Words in boldface occur 10–19 times. All other words occur 2–9 times.	Verbs collocating with WOMAN as subject but <i>not</i> with MAN (at 2+ frequency). Boldface, underlined words occur 20+ times in this relation. Words in boldface occur 10–19 times. All other words occur 2–9 times.
abscond, amble, antagonise, await , beam, bludgeon, build , burgle, captain, check , cometh, con, conquer, contemplate, converse, court, cringe, crouch , crowd, curse, descend, dig , elbow, falter, father, feign, fiddle, frolic, gloat, groom growl, grumble, gun, hail, hammer, haul, heave, humiliate, hunt, inflict, infuriate, joke, leer, libel, lick, limp, lunge, mastermind, mistreat, moan, motion, muscle, mutilate, oppress, outrank, owe , parade, pen, perish, pinion, plough, pocket, poop, potter, pounce, putt, quail, race, raid, ransack, rape , reappear, rejoice, reprieve, saw, scowl, screw, scurry, seroconvert, sidle, sin, snarl, sneer, snore, snort, squint, stomp, straighten, strangle, strip, struggle, swear , sweat, thrive, toe, unload, urinate, walketh, waylay, writhe	account, acknowledge, affect, allow, annoy, anoint, apologise, arch, arrest, avoid, bath, benefit, berate, breastfeed, broaden, campaign, captivate, cease, centre, chain, chew, churn, cluck, consent , cuddle, damage, dedicate, <u>define</u> , delay, derive, dial, divide, file, flaunt, fold, fool, form , frighten, fuss, generate, grasp, gut, harvest, herd, hug, hum, illustrate, imitate, improve, increase, incur, indicate, infect, involve , knit, launch, mean, mention, migrate, mind, nag, narrow, note, ooze, patronize, place, preserve, presume, promote, rake, refer, review, service, shock, shoulder, stage, stake, stress, submit, sunbathe, survey, swamp, test, testify, tongue, underlie, urge, vary, wag, wail, wheel, wind

Table 7: Subject 'exclusive' patterns

<p>Shared verbs favouring MAN</p> <p>Boldface, underlined words favour MAN at 21+ saliency points ahead of WOMAN.</p> <p>Words in boldface favour MAN between 14 and 20.9 saliency points ahead of WOMAN.</p> <p>Other words favour MAN between 7 and 13.9 ahead of WOMAN.</p>	<p>Shared verbs favouring WOMAN</p> <p>Boldface, underlined words favour WOMAN at 21+ saliency points ahead of MAN. Words in boldface favour WOMAN between 14 and 20.9 saliency points ahead of MAN. Other words favour WOMAN between 7 and 13.9 ahead of MAN.</p>
<p>accommodate, accuse, accustom, allege, arm, arrest, bail, bind, blind, build, call, catch, challenge, charge, commemorate, convict, damn, drown, enlist, fine, forgive, gather, hang, hunt, inspire, jail, kick, kill, knock, lead, loathe, lose, lure, name, notice, order, pick, question, release, remand, resemble, reward, send, sentence, shelter, shoot, spot, station, summon, suspect, talk trace, wound</p>	<p>abduct, abuse, advise, affect, age, assault, book, cohabit, degrade, encourage, envy, exclude, help, labour, oblige, oppress, ordain, portray, procure, protect, rape, refer, register, relegate, scar, screen, segregate, subject, treat</p>

Table 8: Object verbs compared

<p>Verbs collocating with MAN as object but <i>not</i> with WOMAN (at 2+ frequency and 4+ saliency). Boldface, underlined words occur 20+ times in this relation. Words in boldface occur 10–19 times. All other words occur 2–9 times.</p>	<p>Verbs collocating with WOMAN as object but not with MAN (at 2+ frequency and all saliencies). Boldface, underlined words occur 20+ times in this relation. Words in boldface occur 10–19 times. All other words occur 2–9 times.</p>
<p>acquit, airlift, anoint, antagonize, apprehend, assemble, bait, baptise, beckon, befit, behead, bewitch, billet, bite, blindfold, boot, bowl, brief, burden, captivate, censure, charm, cheer, chuck, clear, cloak, clout, command, credit, crown, curse, dare, dazzle, deflate, demob, deploy, detest, devour, dine, disperse, displease, dispossess, dodge, drop, endow, engulf, enrol, enthrall, entice, entrust, esteem, exhaust, exile, fascinate, field, fight, fit, flatter, floor, frame, furnish, <u>glimpse</u>, handcuff, harbour, haunt, heed, humour, hurry, immerse, incarcerate, incite, intoxicate, knife, land, lecture, levy, line, march, martyr, milk, muster, nail, nominate, number, ogle, oust, outlive, overwhelm, pardon, part, perplex, persecute, pile, post, praise, predispose, press, profit, rack, raise, rally, recapture, refresh, rejoin, relieve, report, restrain, return, revere, ridicule, rouse, scald, scrutinise, shackle, slaughter, slay, smell, solicit, squander, stun, succeed, surpass, surprise, surrender, swallow, taunt, term, underestimate, unsettle, usher, victimize, vindicate, wake, witness</p>	<p>afford, assist, attend, bed, categorise, celebrate, coerce, compensate, conceptualise, construct, cushion, date, define, deliver, direct, discriminate, discuss, disempower, disguise, dislodge, downgrade, dump, empower, enjoy, equate, evolve, exhibit, fly, gag, groom, highlight, hoodwink, immunise, impregnate, integrate, interpret, interrogate, limit, marginalize, mistreat, monitor, nurse, objectify, omit, organize, penalise, perceive, prescribe, program, provide, ravish, recommend, regulate, restrict, saw, scorn, section, sexualize, shag, shame, sketch, stay, sterilise, suffocate, terrorise, trivialize, <u>use</u>, videotape, violate</p>

Table 9: Object ‘exclusive’ patterns

<p>Shared adjectives favouring MAN</p> <p>Boldface, underlined words favour MAN at 21+ saliency points ahead of WOMAN. Words in boldface favour MAN between 14 and 20.9 saliency points ahead of WOMAN. Other words favour MAN between 7 and 13.9 ahead of WOMAN.</p>	<p>Shared adjectives favouring WOMAN</p> <p>Boldface, underlined words favour WOMAN at 21+ saliency points ahead of MAN. Words in boldface favour WOMAN between 14 and 20.9 saliency points ahead of MAN. Other words favour WOMAN between 7 and 13.9 ahead of MAN.</p>
<p>24-year-old, 27-year-old, 37-year-old, 40-year-old, 43-year-old, 47-year-old, 55-year-old, 65-year-old, able-bodied, amiable, arrogant, ascetic, bald, best, best-dressed, better, big, bigger, <u>bird-like</u>, blind, brave, brilliant, broad-shouldered, broken, bulky, busy, charming, clever, clubbable, common, condemned, conscientious, contented, courteous, cruel, dangerous, dark, dead, distinguished, distinguished-looking, drowning, drunk, drunken, dying, earnest, embittered, eminent, evil, experienced, faceless, fair-haired, faithful, fastest, <u>fellow</u>, fit, fortunate, free, frightened, funny, garrulous, gay, generous, gentle, gifted, ginger, God-fearing, Godly, good, good-looking, grand, great, grey, guilty, handsome, happier, happiest, happy, hard, hard-faced, hated, hateful, hired, holy, homeless, homosexual, honest, honourable, humble, idle, influential injured, invisible, jolly, kind, kindest, kindly, landless, languid, leading, learned, lesser, likeable, listed, little, local, lonely, loyal, luckiest, lucky, macho, marked, masculine, merry, mighty, modest, mortal, mousy, nasty, neanderthal, new, nice, nicest, odd, old, ole, ordinary, outstanding, polite, powerful, practical, prudent, quiet, rational, reasonable, red-faced, religious, reserved, retired, rich, richest, right, right-hand, sandy-haired, sane, self-made, senior, sensitive, sexiest, shrewd, shy, sick, sinful, skilled, sleeping, solid, solitary, squat, stocky, stout, strongest, superstitious, tall, taller, tallest, thin, thoughtful, top, trapped, unarmed, unassuming, uniformed, upright, violent, wealthy, wee, well-built, wild, wise, wiser, wizened, worried, <u>wounded</u>, wrong, young, youngish</p>	<p>44-year-old, 59-year-old, 73-year-old, 74-year-old, 80-year-old, adult, African, African-American, Afro-Caribbean, American, Arab, Asian, attractive, Bangladeshi, battered, beautiful, Bengali, Biblical, blonde, British, butch, Catholic, celibate, childless, daft, desirable, disadvantaged, <u>distraught</u>, dumb, elderly, fallen, feminist, fertile, Filipino, French, gipsy, glamorous, grieving, heterosexual, ideal, immigrant, independent, Indian, infected, Irish, lone, mad, married, middle, middle-class, Muslim, neurotic, never-married, non-married, obese, older, Palestinian, part-time, pleasant-looking, plump, poorer, promiscuous, resourceful, respectable, round-faced, Salvadoran, Scottish, sensual, separated, Sikh, silly, single, slender, spirited, strong-minded, ugly, unmarried, upper-class, vulnerable, <u>widowed</u>, working-class</p>

Table 10: Attributive adjectives compared

<p>Attributive adjectives which collocate with MAN but <i>not</i> with WOMAN (10+ saliency, 2+ frequency).</p> <p>Words in boldface occur 10+ times in this relation. All other words occur 2–9 times.</p>	<p>Attributive adjectives which collocate with WOMAN but <i>not</i> with MAN (10+ saliency, 2+ frequency).</p> <p>Words in boldface occur 10+ times in this relation. All other words occur 2–9 times.</p>
<p>'orrible, 20-a-day, 45-year-old, 46-year-old, 48-year-old, 49-year-old, 50–cent 84-year-old, ablest, accused, advance, affable, amiable-looking, armed, armoured, arrested, astute, athletic, avuncular, back-row, bald-headed, balding, barrel-chested, beaky, bearded, beaten, beefy, beetle-like, bespectacled, best-looking, betting, black-bearded, black-skinned, blond, born-deaf, braver, bravest, broad-faced, brown-faced, bull-necked, burly, cappy, cautious, charismatic, choleric, civilised, civilized, clean-cut, considerate, convicted, cultivated, curious-looking, curly-haired, dapper, dark-faced, demented, devastating-looking, devout, dirty, disillusioned, dislikeable, dour, earphoned, evil-looking, ex-military, ex-navy, ex-red, ex-service, eyeless, fair-minded, falcon-headed, fantastic-looking, feckless, fittest, flush-faced, forgotten, front, funniest, gangly, good-natured, gorgeous-looking, green, green-coated, gregarious, grey-suited, grim-faced, gun-toting, hairy, half-starved, hanged, harlequin, harried, hawk-faced, head, heavy-set, homosexual/bisexual, horrid, humane, hunted, jailed, jovial, lanky, law-worthy, lecherous, likable, manly, marching, masked, masterless, medical, mild-mannered, military, monied, mounted, moustached, muscular, mustachioed, Neolithic, non-union, odd-job, olive-skinned, one-woman, palaeolithic, paralysed, patient, personable, picked, portly, prehistoric, primitive, proleptic, ram-headed, ratty-looking, retarded, righteous, rough-looking, rudest, running, scarab-headed, scholarly, sea-faring, seafaring, self-educated, self-important, silver-helmed, sincere, smartly-dressed, spare, starving, stone-age, stoutish, straight, strange-looking, strong-arm, strong-armed, suave, surly, swarthy, tallish, thick-set, thickset, thin-faced, thirsty, tic-tac, tolerant, trusted, trustworthy, truthful, tubby, two-coat, uncircumcised, uncouth, underground, unluckiest, untrained, unworldly, upstanding, urbane, vengeful, virile, vitruvian, wanted, weary-looking, weasel-eyed, wickedest, wild-looking, worried-looking, young-looking</p>	<p>10-stone, 81-year-old, 83-year-old, 85-year-old, 87-year-old, 89-year-old, 92-year-old, abused, arthritic, awfy, bare-breasted, barren, big-bosomed, black-shawled, blonde-haired, blowsy, bossy, burdened, buxom, chaste, chattering, childbearing, cleaning, Coptic, dumpy, efficient-looking, Euro-American, ex-care, ex-married, ferret-faced, first-century, frigid, gentle, giving-out-food, gossiping, grim-looking, headscarved, high-caste, hindu, hiv-positive, hysterical, incontinent, inter-senior, large-breasted, lesbian, liberated, lower-class, Iranian, memoried, menopausal, menstruating, motherly, motherly-looking, multiparous, near-naked, nineteen-year-old, non-asthmatic, non-pregnant, non-working, nulliparous, Pakistani, pear-shaped, pleasant-faced, plump-faced, plumpish, post-menopausal, postmenopausal, pregnant, pretty, primiparous, raped, remarried, Samaritan, Saudi, scarlet, searching, seductive, severe-looking, sexy, shallow-changing, sickly, subfertile, tired-looking, twenty-year-old, unescorted, veiled, vivacious, weeping, witch-like, working-age, worshipping</p>

Table 11: Attributive adjective 'exclusive' patterns

