

# The company *women* and *men* keep: what collocations can reveal about culture

Sara Gesuato

Dept. of Anglo-Germanic and Slavic languages and literatures,  
University of Padua, Italy

## 1. Introduction

Habitual ways of talking about given phenomena show people's attitudes to and conceptualizations of them: the meanings associated with certain events or situations determine the cultural significance of given topics within a group of speakers (cf Fairclough 1992, Hodge and Kress 1993). In particular, a language's keywords give insight into the culture of its speakers (e.g. Krishnamurty 1996, Partington 1998, Wierzbicka 1999); but actually all the words repeatedly circulating in a community form constellations of repeated meanings conveyed in typical patterns that produce conventional expressions and opinions. Therefore, any word recurrently used in interaction may carry social implications, even if these may go unnoticed among the interactants (see Tognini-Bonelli 2000: 125). To understand the social salience and semantic nature of the terms exchanged in verbal interactions, it is useful to look at their recurrent patterns of usage, namely their occurrences in co-text (see Stubbs 2001).

In this paper I examine collocations of WOMAN, MAN, GIRL, and BOY (from the *Usbooks*, *Ukbooks*, *Time* and *Today* components of the Cobuild on-line corpus) to determine if their denotational symmetry (identification of adult and non-adult human beings of the female and male sexes) is paralleled by similar patterns of usage (association with the same semantic fields and embodiment of the same semantic roles in given syntactic environments), i.e. if they belong to similar discourse domains.

## 2. Most significant collocates

The lexemes under examination do not share all of their most representative collocates (i.e. words they co-occur with in a statistically significant way), which reveals that they are part of different domains of discourse.

The top 50 significant collocates of WOMAN include terms that fall within the discourse domains of physical attractiveness (*beautiful, pretty, attractive*), age (*young, old, older, elderly, aged*), physical appearance (*body, dressed, looking, white, black*), family and personal relations (*married, pregnant, singles*), women's liberation (*rights, movement*), religion (*pontiff, ordination*), (groups of) people (*woman, women, girls, man, men*). Other collocates comprise *said, drowned, listening, named, division, percent, working, golf, title, doubles, lives, and won*.

The wordforms *man* and *men* co-occur with words exemplifying some of the semantic fields identified for WOMAN: age (*old, older, aged, young*), physical appearance (*white, black, tall, big, fat*), family and personal relationships (*married, singles*), and (groups of) people (*woman, women, men*). Other collocates are relevant to distinct domains of discourse: negative states (*blind, dead, poor*), the military (*officers, enlisted, squad*), sex (*gay, sexual*), negatively connoted physical force or action (*violence, arrested*), non-physical attractiveness (*rich, kind, nice, good, wise*). Among the remaining collocates, some are common to WOMAN as well (*division, percent, working; labeling: title, named*), while others are peculiar only to MAN (*burned, made, isle, saw, matc*).

The data show that both WOMAN and MAN are described in terms of such notions as age, physical appearance, family and personal relationships. On the other hand, WOMAN and MAN partly occupy different semantic space. First of all, MAN is relevant to a wider range of semantic fields, which points to a more versatile use of its wordforms. In addition, even when both lexemes are relevant to the same domains, these may be highlighted in different ways. For example, in the domain of physical appearance the collocates of MAN stress the notion of 'size' while those of WOMAN that of 'overall qualitative effect', as if to suggest that a person's visual impact on others is more relevant to women than men. Moreover, the euphemistic term referring to old age (*elderly*) is reserved to WOMAN, which signals a mild taboo in the co-occurrence of the notion of 'femaleness' and 'age'. Furthermore, WOMAN co-occurs with a greater variety of terms denoting people, indicating that relationships or interaction are more often discussed with reference to women. Finally, WOMAN and MAN occur in partially complementary distribution; the former lexeme is associated with the domains of physical attractiveness, religion, civil rights, and involuntary actions (*drowned, won*), the latter with those of non-physical attractiveness, the military, violence, and voluntary actions (*burned, made*), which shows that distinct areas of activity appear to be salient to the two gender groups.

The terms occurring with GIRL and BOY reproduce patterns very much like those outlined above. The collocates of GIRL are representative of the semantic fields of age (*little, year, young, teenage, baby, aged, old, adolescent*), personal relationships (*boy, girl, father, married, lover, boys, women, girls, men*), physical attractiveness (*lovely, pretty, beautiful, beauty*), negative state (*poor,*

*raped, died, dead, killed, murdered, screaming*), positive state (*nice, golden, jolly, pregnant*), labeling (*named, called*), and physical appearance (*black, white, haired*). The collocates for BOY refer to the domains of age (*little, old, year, young, baby, aged, teenage, older*), size (*small, big*), relationships (*girl, girls, boy, boys, men, brigade, club, posse, scouts, scout, mother, father*), negative state (*died, poor, dead, bad*), positive state (*golden, good*), activities (*school, game, play, jobs*), labeling (*named, called*), and outdoor events or fun activities (*beach, pet, shop, do*). The collocates reveal that GIRL and BOY show up in partially different discourse domains. The former lexeme is relevant to the topics of physical appearance and attractiveness, which suggests a young woman's outward characteristics are paid attention to. GIRL is also often associated with participles which denote dangerous or hopeless conditions and convey ideas of helplessness and victimization. Finally, GIRL may occur in contexts of talk focusing on *adult* relationships (see *married, lover, pregnant*), which indicates that such notions are considered appropriate even to *young* females and/or that GIRL is also used in partial contradiction with its denotational meaning, that is with reference to not very young women. The lexeme BOY, on the other hand, is more relevant to the notions of behavior (*good, bad*), bonding (*brigade, club, posse*), active involvement (*game, play, jobs*) and associated with outdoor contexts (*scout, beach*). This lexeme, therefore, occurs in contexts focusing on the idea of direct participation in concrete activities.

In general, the collocates of WOMAN, GIRL, MAN, and BOY show that these lexemes share certain contexts of use, but also that there are discourse domains pertaining only to WOMAN and GIRL, and others only to MAN and BOY. The *distinct* sets of collocates suggest that WOMAN and GIRL are more frequently associated with notions of passivity and physicality, and MAN and BOY with those of activity and general behavior.

### 3. Descriptive labels

The descriptive labels attached to a noun specify the characteristics of its referent and actually narrow it down by identifying a subset of it. A language's typical descriptive labels are attributive adjectives, but a similar categorizing function is also played by post-modifiers like relative clauses and *of*-headed prepositional phrases.

#### 3.1. Attributes

Adjective+noun combinations represent a type of colligation or association between grammatical categories. When examined in high numbers, they can be semantically revealing because they show what properties (as encoded in the pre-nominal adjective or other modifier) are frequently or by default ascribed to what entities (encoded as nouns). The following findings are based on a sample of the relevant Cobuild data (100 concordances per each wordform *woman, man, girl* and *boy*), systematically selected through the software running the corpus (one every a given number of occurrences determined by the program).

A comparison of representative samples of attribute+BOY vs. +GIRL combinations reveals that most of the attributes preceding both lexemes are of three types: they mention physical characteristics (*little, grown, blind*), refer to behavioral traits (*good, brave*) or identify the group membership of the relevant referent (*Jewish, factory*). But despite this basic similarity, the collocates of GIRL and BOY also reveal differences in the usage of those terms: GIRL is associated with the notion of beauty, encoded through various terms (*beautiful, lovely, pretty*), four times as frequently as BOY, which co-occurs with *pretty* only once; in addition, only GIRL is occasionally associated with attributes that signal its referent is thought of as an adult engaged or involvable in heterosexual relationships (*call-girl, flirtatious girl, and had a steady girl, is going to find a nice girl and marry her*). On the other hand, more varied terms are available for a description of the behavior and personality of the referents of BOY (*bully, charming, different, dull, fine, funny, nasty, problem, suitable, wonder*) than those of GIRL (*dangerous, hard, material, naughty, sympathetic*). Furthermore, if the attributes referring to physical attractiveness are excluded, it turns out that those referring to boys are also more evenly distributed between the categories of positive and negative connotations. Finally, BOY is accompanied by attributes referring to intellectual abilities (*clever, dull, sharp*) three times as frequently as GIRL (*smart*).

The attributes preceding WOMAN and MAN are more varied than those of GIRL and BOY, and identify, e.g., states or conditions experienced by their referents (*pregnant, dead, forgotten*), their area of origin (*Palestinian, Yale*), permanent properties (*Catholic, religious*), traits of behavior (*liberated, gambling*), physical and emotional characteristics (*tall, big, broken, loud, naked*), and group or category membership (*neighbor woman, company man*). However, although the general patterns of usage are similar, the specifics of these descriptive labels point out subtle differences in the ways people talk about women and men.

Attributes of WOMAN often signal negative or demeaning characteristics of the lexeme's referents — as revealed in other contexts as well (see sections 4.1. and 5.2.) — namely, women's past or

present role as victims (*liberated, helpless*), mental or emotional instability (*deranged, vengeful*), lack of control on their behavior (*loud, broken*), limited intellectual ability (*stupid*), socially marginalized love life (*divorced, spurned*) or low-qualified professional position (*cleaning woman*); the salience of their physical attributes (*beautiful, fair, fine, full*) and the non-prototypicality of their playing certain social roles (*woman boss, neighbor woman*) are also highlighted in the concordances. Neutral or positive attributes are infrequent (*black, tall, rich, real, sensible*).

Attributes of MAN signal neutral, positive, and only occasionally negative characteristics of its referents, which indicates that people's discourse practices provide a more balanced picture of men's personalities and behavior. Pre-modifiers of MAN emphasize men's importance and power (*big, powerful, wealthy*), appropriate conduct (*cautious, devout, helpful, proper, religious, spiritual, suitable*), general likeability (*great, fantastic, decent, lovely, handsome, dishy*), education (*cultured*), cleverness (*witty*), independence (*bold, free* [cf *liberated* of WOMAN]), modernity and authenticity (*modern, plain man of the people*), satisfaction (*chuffed*), and emotional success (*ladies man*). Negative attributes may refer to conditions the referent of MAN is not responsible for (*blighted, harassed, dying*), denote reprimandable behavior (*marked, violent, strange, gambling*) or point out lack of socially desirable qualities like out-goingness (*timid, shy*). Neutral attributes are those identifying types or groups of men performing specific functions in given contexts (*ball man*, in the context of badminton; *best man*, in the context of wedding, *radio man*, in the context of communication) or temporary or permanent physical characteristics (*naked, hooded, white*).

In general, while a variety of attributes are available for describing both females and males, those associated with the former are more frequently negatively connoted, restricted to fewer domains of discourse or focused on the decorative function of women. On the other hand, the attributive collocates of MAN are typically more numerous and varied, equally reveal positive and negative qualities of their referents, and highlight either socially important roles played by men or their intellectual values.

### **3.2. *Of-headed prepositional phrases***

More *of*-headed prepositional phrases modify BOY (100) and MAN (425) than GIRL (82) and WOMAN (218). Most can be grouped under the same headings, i.e. they classify their referents in similar ways. Those attached to *boy* and *girl* mostly refer to age (*When I was a boy of 13; A girl of four saved her family yesterday*); only 3 of the concordances of *girl* and 1 of those of *boy* refer to other characteristics of their referents (*Alex is a girl of very high morals; A Girl of the Streets; he was a boy of mercurial nature*). The phrases describing *boys* and *girls* are more varied and some are shared: they include reference to age (*Dr. Stuart Horner said girls of 12 could be donors; at Middlesex University found that boys of 13 or 14 who played computer*), place of origin (*Fox suggests that the girls of Jerusalem speak the first part; compared to the big boys of Hollywood*), and group membership (*and said the girls of the horde learned their walk; but even more the boys of his own class*). In addition, the concordances of *girls* exemplify only one other kind of descriptive phrase, which indicates what girls may possess (*a description of bombed houses in The Girls of Slender Means*). The other concordances of *boys*, on the other hand, specify a few more types of characteristics of their referents: character (*The right type of boys of a type - to appreciate her*), job (*the smart boys of the press*), epoch (*all the other little boys of his generation*), condition of being somebody's "property" (*They have four boys of their own*), identification with reference to a topic (*the seven Hard Flint boys of the Navajo myth; one of the boys of Old Woman Running*).

The concordances of both *woman* and *man* reveal the same types of classifications: both sets describe their referents in terms of their age (*anyone who thinks that a woman of 40 plus is undesirable; Just as a man of 70 probably can't lift as much*), good qualities and bad traits (*she was a woman of considerable literary ability; a woman of dubious loyalty; he was not a man of achievement, but a man of promise; a man of disgusting morals*), physical features (*You're a woman of medium height; if it is possible for a man of his size to be peripheral*), group membership (*a woman of the horde; a man of the tribe called the Westfolk*), possession (*the title of her book A Woman of Substance; would be a difficult matter to find a man of any property in the country*), character (*a woman of her word; a man of constant surprises*), origin (*a woman of England called Agnes; a man of the north*), importance (*a woman of rank; might well have become a man of destiny*); only *woman*, however, has concordances describing its referent as a man's "property" (*He's the sort that needs a woman of his own*), and only *man* has concordances mentioning its referent's work (*curator of the museum had to be a man of the pen and of the book; We might have had Msgr Montini a man of letters*) and epoch (*Only a man of Colonel Sartoris' generation*). In general, phrases that describe non-physical characteristics make up about 60% of the concordances of *man* and 42% of those of *woman*. Furthermore, of the concordances mentioning positive qualities, only 5 (i.e. about 8%) of those relevant to *woman* include intensifiers (*a woman of singular attainments; A woman of strong convictions!*), while 72 (i.e. about 45%) of those relevant to

men are pre-modified in this way (*a man of charismatic personality; a man of considerable intelligence; A man of evident refinement; A man of formidable intellectual gifts; A man of great courage and determination; a man of immense patience; a man of international calibre*).

The concordances of *women* and *men* reveal the same patterns as those identified above for *woman* and *man*: their embedded prepositional phrases describe the referents of those wordforms along very similar lines, but to different degrees. Age is mentioned 26 times (i.e. about 17%) with reference to women and 11 times (i.e. about 4%) with reference to men (*can be used on women of all ages; four units a day by men of all ages*); physical characteristics show up 4 times with reference to women (i.e. about 2.5%) and 3 times (i.e. about 1.1%) in the case of men (*This chapter is meant for women of every size; place the onus on men of Cleese proportions*); the time or generation a person belongs to is indicated more often in the vicinity of *women* (13 times, i.e. about 8%) than *men* (8 times, i.e. about 3%; *the women of that 1960s generation; men of an old generation*); the geographical, national and/or family origin of women is referred to more frequently than that of men (58 times, i.e. about 37% vs 57 times, i.e. about 21%, respectively; *All the women of the valleys had found some pride; labourers assumed that men of Anglo-American origins*); group membership, on the other hand, appears to be more salient to male than female referents, the frequency of occurrence being 31 times (i.e. 20%) for the former and 83 times (i.e. about 31%) for the latter (*that was unheard of among women of the aristocratic classes; to the men of the B company it felt as if*); the specification of people's non-physical qualities is seemingly equally relevant to women and men, since it is to be found in 19 concordances of *women* (i.e. about 12%) and in 44 concordances of *men* (i.e. about 16.5%; *it stresses the importance to women of good qualifications; was led by men of enduring stature*); however, there are also prepositional phrases accompanying *men* that mention other non-physical characteristics of their referents, like profession (13 times, i.e. about 5%; *we men of business*), type, personality and/or behavior (27 times, i.e. about 10%; *For men of the stamp of Scheer*), level of importance (9 times, i.e. about 35%; *You find it a lot with men of power*); overall, therefore, men are described in more varied detail than women through *of*-headed prepositional phrases. Finally, the positive characteristics of women are enhanced by intensifying adjectives slightly less frequently than those of men, namely 10 times (i.e. about 6%) vs 15 times (i.e. about 5%), respectively (*women of strong will; as free men of definite and sincere worth*).

In conclusion, the *of*-headed prepositional phrases post-modifying GIRL, BOY, MAN, and WOMAN present comparable, but not identical, kinds of specifications; for one thing, some of these do not occur with the same frequency across the concordances, that is physical attributes appear to be more salient to female than male referents, while non-physical characteristics are more tightly associated with male than female referents; in addition, certain types of prepositional phrases are not always available to describe females, for instance those mentioning jobs; finally, the positivity of certain attributes tends to be more strongly emphasized (through adjectives) when the words encoding them collocate with terms marked for maleness. In brief, GIRL, BOY, WOMAN, and MAN are not equally salient to the same types of concepts used to describe their referents.

#### 4. Roles and situations

A clause is the verbal representation of an event, that is the linguistic rendering of a phenomenon in the world. It pivots around a predicate, which represents what is going on, and may include nominals, phrases, and adverbials that refer to entities and circumstances that contribute to making up that event or situation. It is often possible to verbally represent the "same" event in various ways; the specific lexico-syntactic wording chosen will highlight certain aspects of it and obscure others (see Halliday 1994: ch. 5). Examining clauses in which WOMAN, MAN, GIRL, and BOY occur will help identify what events and situations they are most typically associated with and what roles their referents are represented as embodying.

##### 4.1. Relative clauses

A relative clause represents an event or situation in which the referent of the relative pronoun is a main participant. In the corpus, I looked for examples of *who*-headed relative clauses, in which, that is, the referent of the pronoun would most often be the subject or direct object of the clause, and thus likely to play the semantic role of agent or patient, respectively. From the output of each of my queries, I had the system select a random representative sample of 100 concordances.

The result of the *girl+who* search presents three interesting findings: (1) 34 of the concordances describe negative situations that the referent of the relative pronoun is affected by or, occasionally, responsible for (*strongly enough to want to murder the girl who'd informed them; the five-year-old girl who escaped the massacre; a troubled girl who felt pressures at home; I have a neurotic girl who is panic-stricken; a severely disabled girl who is unable to speak; I had a girl who was a Down's Syndrome's baby; the subject of the aggressive little girl who had had a 'perfect birth'; a young*

*Monk shelters an Albanian girl who is being hunted for the murder).* (2) Moreover, 7 concordances show that girls may be talked about with reference to their bodies, especially their physical attractiveness, occasionally seen as a source of problems (*period: puberty. What happens to a girl who begins to menstruate at 10; She was really pretty like that. A girl who feels pretty has a tendency; a girl with breasts and a girl who's skinny; in the case of a teenage girl who is ashamed of too little breast*), and 9 more associate girls with the notions of romantic and/or sexual involvement, that is in (a not always happy) relation to some man (*We found a Romanian girl who is married to a local man; for a little girl who loved him; a girl who fails to marry her suitor; a Mexican girl who eventually jilted him*). (3) Finally, 79 concordances reveal that the subject of the relative clause plays the role of an experiencer of situations, events or emotions, or the patient of others' actions, which may even be unpleasant (*a Colorado girl who becomes an opera singer; A girl who lived with her parents; She was a vivacious, outgoing girl who was always enthusiastic; a girl who will never forget a friend; a girl who shares his interest in sport; A teenage girl who collapsed; The girl who died was her friend; the handcuffed girl who was being pushed; A girl who was raped and strangled*); on the other hand, when playing the role of an agent, the situation the referent of *girl* is in charge of is often negative, vilifying or restricted to the field of love-relationships (*A heart-swap girl who battled against all the odds; did not bode well for the girl who had invited Kurt to lunch; Except that the girl who told the tale had been expelled; the black girl who'd killed herself; bordello in Lima, Noriega asked a girl who'd just finished making love; Paula shot to fame as the girl who ditched all her lover's gifts*).

The 100 concordances of *girls+who* reveal very similar patterns: 32 depict negative events or situations (*it may be that girls who are weak academically; his daughter was attacked by girls who are fellow pupils; to throw stones at the Chinese girls who broke world records; girls who develop eating disorders fail; This was very hard on the girls who had paid for their machines; the poor and unfortunate girls who walk the streets; child labor and raw immigrant girls who would work for next to nothing; nervous girls who never laugh*); 17 concordances mention girls in relation to sex, the emotions, or careers in the show business (*while girls who are sexually active are loose; more than half of the girls who become pregnant before age 18; and happily-ever-after endings, girls who buy into this script; suspicious of the girls who want commitment; one of the other girls who crossed his bachelor path; scores of local girls who ended up marrying; girls who are cheerleaders; where modeling is at now, for girls who could once only dream of owing; Hollywood's pick of the girls who look set to become tomorrow's*); 80 lines of concordance portray girls as experiencers or patients, sometimes in negative contexts, and not always in control of themselves (*girls who are average in development; girls who have good communication about; girls who had sat in that chair; three girls who arrived rather apprehensively; girls who looked pale and unhealthy; delicate girls who enjoy shopping; those girls who most fully accept; memories of the girls who'd drowned; girls who were adopted in the 1960s; girls who were sacked; quite young girls who must have been the victims; fair young girls who trembled; the three girls who were panicking*); when presented as agents, girls do not appear to always cut a good show (*girls who drop out most often do so; He caught two other girls who were cheating; they were middle-class girls who, to revolt against; there are far more girls who say they would rather be boys; ran a network of call girls who serviced French high society*).

Of the 200 collocations of the combinations *woman+who* and *women+who*, 49 present women as facing or, occasionally, causing difficult circumstances (*frustrating it must be for women who couldn't do what they wanted; I know there are many women who endure similar treatment; women who have abnormal Pap smears; only against illegal practitioners, women who obtained abortions; talking to the other women who shared some of their concerns; author of Women Who Love Too Much; mortally wounded a young woman who was watching the crowds; an unmarried, unemployed 24-year-old woman who is struggling to feed her four; warned against women who are always fawning; in fact, it was the woman who instigated the affair; a woman who is forever scaring; the woman who made his life a misery*); 157 concordances depict women as being passively involved, as experiencers or patients, in situations determined by external circumstances or run by more active participants; the verbs most commonly accompanying the wordforms *woman* and *women* are indeed *be* and *have* (*women who are infected with HIV; women who have had their ovaries removed; battered women who explode into loss of control; the number of women who are encouraged to do so; a young woman who disappears after meeting him; woman who had been infected*); however, when presented as agents, only occasionally are women presented as involved in negative situations (*the woman who'd rented Samelu his brickworks; the woman who had administered the test; the woman who tried to kidnap April; for two young women who have booked to go abroad; the older women who supported Mother Marie; Only one in ten women who runs a business works fewer*); 51 occurrences show women in association with the domains of emotions, sex, and physical appearance (*a young Roman woman who begins a passionate affair; was one of the women who adored their husbands; a woman who, along with sexual aversion; a fair number*

*of women who have never had an orgasm; lovely succulent woman who should be eaten as whole; a raven-haired, proud-breasted young woman who has risen from a humble; the woman who brings back the elegance; particularly for women who may be watching their weight).* Also, the notions of negativity and passivity often co-occur and the semantic fields that are specially “reserved” to women are negatively connoted (*women who are incapacitated; unmarried women who aren't prepared to occupy; women who have been widowed; women who have bulimia; a woman who discovers a breast lump; single woman who was attacked; woman who is afraid of flying*).

The 200 concordances of the wordforms *boy* and *boys* also include reference to negative contexts, but less frequently than those of *girl* and *girls*, namely on 30 occasions overall (*angry, needy and frightened boy who looks to women to mother him; Gregory was a loner, an Afrikaans farm boy who had been neglected; eight-year-old retarded quadriplegic boy who was murdered; motivate pupils, especially boys, who fail at school; prison sentences imposed on the two boys who killed James Bulger; neighbours also joined boys, who had set fire to curtains*); interestingly, these negative contexts occasionally involve females more directly than males (*it is girls rather than boys who are at greater risk; a boy who saw his grandmother killed*). In addition, only 10 are the concordances which somehow include reference to sex, physical appearance or feelings (*many teenage boys who feel homoerotic urges; for high school boys who are too skinny to play; A 16-year-old boy who's never had a girlfriend*). Finally, more numerous in these sets of data, namely 99, are the concordances that include representations of events in which boys are actively participating (*incredibly wiry boy, who could climb the highest pines; from a boy who first lifted a tennis racket; said the boy who had brought the coffee; I had trained a boy who had won a scholarship; punched and kicked by a boy who had knocked on her door; crowds are made up of boys who play sports; were moistened by two small boys who pressed sodden cloths against; and the boys who take the field will be very*); 15 of these portray boys as involved in or causing problems (*an eleven-year-old who has stolen some dried figs; a wild boy who poached rabbits; the boy who took the most drugs; care for boys who had committed offences; pension book stolen by boys who tricked their way into his East*).

The wordforms *man* and *men* display collocational patterns similar to those of *boy* and *boys*: 45 of the 200 concordances considered refer to negative situations (*A man who appeared to be drunk; What rung for the man who has been unemployed; Here we are in a room with a man who's been charged with murder; about the same frequency in men who are insulin-dependent; one of the families of the SAS men who died on the mission; in black balaclavas attacked two Asian men who were lying on the floor*). Reference to the domains of love or sex life and beauty are limited to 17 concordances (*he's the man who sleeps his way to the top; I am the type of man who can give to more than one woman; Clearly, a man who feels sexually attracted to; can't compete for women with men with flourishing forelocks; a troupe called The Hollywood Men, who dance in skimpy briefs; twice as likely to get divorced as the men who married later*). Furthermore, 97 are the concordances showing men taking an active role in the events being represented, and only a dozen of these refer to negative circumstances (*the man who brought you The Sound of Music; why I'm the man who helped fill the missing link; lost their lives trying to save a man who jumped in the water to rescue; a chapel man who later spoiled himself by drink; men who ally themselves with England; were written by men who had fought in it; half of the men who immigrated to Australia; prosecute the men who sexually harass me at work*); only 4 are the instances of passive clauses (*the man who had been certified dead; two men who had been killed in the quarry*). Finally, quite frequent, i.e. 40, are the concordances that show men in a positive light, whether they are presented as determining or simply experiencing the positive circumstances relevant to their lives (*This is a young man who believes in the future; the son of a man who had risen from beginnings; paid their last respects to the man who made Jaguar a world leader; and so is the man who now seeks supreme power; this is the man who won over the public; He was a loving man, a man who would do anything for anybody; the universal regard for men who had risked their lives to defend; Toward the men who held national power; eloquent men who stand above the crowd; two men who turned their dream into reality; the only two men who could win the title*).

The above data suggest that while WOMAN, GIRL, MAN, and BOY share similar contexts of usage, they do not occur in those contexts with the same frequency: it is more often the referents of WOMAN and GIRL than those of MAN and BOY that appear to be involved in difficult situations, affected by external circumstances or other agents' interventions, and associated with discourse domains that draw attention to a female's physical attractiveness and her role in relation to a male.

## 5. Semantic preferences

The notions people exchange in interaction when talking about females and males can be revealed in part by examining their collocates in syntactic environments likely to introduce topics for discussion. Two such constructions are the genitive and *for*-headed prepositional phrases.

### 5.1. Genitive

The semantic relation holding between ideas expressed by nouns can be of various types. When these nouns are grammatically linked by the genitive, they can express such notions as possession (e.g. *my neighbor's house*), personal relation (e.g. *Frank's son-in-law*), classification (e.g. *a children's story*), origin (e.g. *John's telegram*), authorship (e.g. *Wilbur Smith's latest bestseller*), agency (or subject-verb relation, e.g. *Mary's consent*), passive experience of an event (e.g. *the town's destruction*), predication of a state relative to an entity (or subject-complement relation, e.g. *Paul's happiness*), and more.

Of all the types of phrases examined in this study, the genitive is the one that reveals the least striking *categorial* differences between the collocates of the pairs of gender-marked lexemes. The concordances obtained for wordforms *girls* (92) and *boys* (86), and representative samples of 120 concordances for *girl*, *boy*, *woman*, *women*, *man*, and *men* show that the words following the lexemes in question (almost exclusively nouns or noun phrases) convey information about the group membership, physical characteristics, thoughts or attitudes, feelings, states, qualities of the possessee, and also about the objects possessed or actions performed by them, the events, experiences or abstract notions relevant to them or finally their relationships with others.

Neither the concordances of *boy* nor those of *girl* exemplify the notion of group membership. All the other concepts are represented, although to varying degrees within each set. For *girl* and *boy*, respectively, the frequencies of occurrence of the various notions, given in percentage, are as follows: event (about 3% vs 6%; *Girl's 24-hour van rape ordeal; Boy's 130mph joyride*), relationships (about 13% vs 7%; *Last night the girl's grandmother said; sent the boy's aunt Marlene in her place*), physical characteristics (about 42% vs 26%; *the desert area where the little girl's body had been found; He crossed to Mack and touched the boy's arm*), actions (about 14% vs 10%; *to detect a euphemism for the girl's 'welcoming' her lover into her; for his appointed visit. The little boy's coming back*), feelings (about 3% vs 0.8%; *the girl's love is said to be better; the latter start fighting for the boy's affections*), thoughts (about 4% vs 6%; *because of a girl's mistrust of one's courtiers; one boy's desire to become an actor*), states (about 2.5% vs 4%; *to ensure the girl's safety; she said the boy's presence made teaching the class*), qualities (about 0.8% vs 4%; *a fraction of a 13-year-old girl's energy and concern; If you can tell a boy's character from the company he*), objects (about 9% vs 13%; *Leaphorn checked the girl's duffel bag; The boy's bicycle was found in a pigsty*), and abstract notions (about 3% vs 9%; *only partially to do with a girl's psychological make-up; it became the boy's duty*). It appears from the above data that certain types of notions cluster more frequently around *girl* (actions, feelings, relationships, and especially physical characteristics), while others more typically accompany *boy* (events, conditions, thoughts, qualities and abstract notions, and especially possession of objects). In addition, more frequently associated with *girl* than with *boy* (i.e. about 26% vs 3%) is the reference to all of the following domains taken together: love (*may be rivals for the girl's affection; men's and boy's [sic] dominance in sexual relationships*), sexual attributes (*spending the night between the girl's breasts; keep secret a boy's description of his genitals*), clothing (*a girl's dress is displayed like an exhibit*). Finally, differently from other sets of concordances, reference to negative situations is only slightly higher in the case of *girl* than *boy* (i.e. about 26% vs 22%; *Moors murder girl's grave wrecked by 'sick' gang; Four charged after ecstasy boy's death*).

The frequency of occurrence of the semantic categories identified above is markedly different among the collocates of both *girls* and *boys*. The genitives of *girls* do not include reference to the notions of qualities or thoughts, and mostly concentrate on those of events (about 16%; *choir concert, night out, reunion, race, party*) and "possession" of objects (about 36%; *bedroom walls, cabin, dressing room, home, independent school, tent, toy*). On the other hand, the concordances of *boys* do not represent the category of state, while they focus on those of group membership (about 25%; *Brigade, Choir, club, network, soccer team*) and possession of objects (about 37%; *boarding school, camp, dormitory, house, locker room, old baseball caps, test room*). Also, the concordances of *girls* and *boys* represent negative situations much less frequently than those of *girl* and *boy* (both about 8%; *I saw the girls' tent had collapsed; too familiar with the terrors of the boys' locker room*). Finally, the association of the referents of the above wordforms with the notions of beauty, love, and sex is reduced to about 10% in the case of *girls* (*he knows that the girls' beauty will burn out a lot quicker; all Catholic girls dream of falling in love with; Girls' interest in sexuality*) and not attested at all in relation to *boys*.

The data about WOMAN and MAN reveal that the wordforms of these lexemes share only some semantic preferences. The concordances of both *woman* and *women* do not refer to the notions of feelings or states, but while the former focus on actions (about 12.5%; *actions, annual consumption, answer, becoming pregnant, behaviour, knitting, laugh, leading a caravan*) and especially physical appearance (about 56%; *acne, age, appearance, body, dark hand, face, lap, mouth*), the latter concentrate on the discourse domains of abstract concepts (about 8%; *health, issues, politics, rights*,

*roles), group membership (25%; City Club Members, Health Book Collective, Health Education Network, liberation movement, political organization, Political Caucus, teams), and especially sports events (about 37%; championship, lives, national indoors singles, polo tournament, World Cup downhill races).* On a parallel with the concordances of GIRL, those of WOMAN too show that representation of negative contexts and reference to physical appearance and sexuality is more frequent in association with the wordform *woman* (about 34% and 23%, respectively) than with the plural *women* (about 10% and 4%, respectively). Examples include: *the concentrated toxins in a woman's blood; women's abuse, the great waste of women's potential; skin care products improve a woman's appearance; sexual parts of a woman's body parts, possibly "vulva"; women's lingerie.*

The data about MAN show that its singular wordform disregards the notion of group membership, while it prefers those of action (about 9%; *abuse, activity, advances, bluff, historic quest*), abstract concepts (10%; *destiny, dignity, ego*), possession (about 19%; *books, country, house, land, name-tag, poster*), and especially physical attributes (about 28%; *abdomen, breath, cool eyes, face, hair, legs, shirt, shoulder, tonsured head*), only a few of which, however, suggest attractiveness. On the other hand, the concordances of the plural *men* do not refer to qualities or relationships, but draw attention to abstract concepts (about 8%; *title, greater access, sexual problems, studies*), groups (about 16%; *amateur squad, teams, Bible Study group, division, group*), actions (about 27%; *abuse, bidding, further domination, practice, projection and externalization, violence*), events, especially related to sports (about 27%; *400m hurdles, doubles, game, lives, singles*). These data thus reveal that both *man* and *men* show a preference for the categories of actions and abstract concepts, although only for the latter to a comparable degree. Another similarity has to do with the frequency of occurrence of negatively marked contexts, which is about 16% for *man* (*there was the problem of the man's face, distinctive even in death*) and about 17% for *men* (*within the complex relationship of men's sexual abuse and men's violence*), and which in most cases refer to problems caused, rather than experienced by, men. On the other hand, reference to men's attractiveness or sexuality is less frequent in the case of *man* (about 7%; *blocking the man's ability to get an erection; a means of bridling man's unbridled lust*) than in that of *men* (about 15%; *In the ballet world men's bodies are stunning, sinewy; an intro for some men's mag sex in the kitchen*).

In conclusion, in these sets of data it is still the referents of WOMAN and GIRL who are more frequently talked about in terms of sexuality, physical attractiveness, and romantic involvement; however, interesting correlations can also be found between their singular and plural wordforms and those of MAN and BOY. For instance, in the singular both GIRL and BOY are more often associated with the notions of physicality, sexuality, attractiveness, negativity, possession, and action. Similarly, both WOMAN and MAN show a preference for the notions of physicality, action, and negativity in the singular, and for group membership and events (mostly sports events) in the plural. Some conceptual categories turn out to be favored across wordforms; for example, that of possession is a “favorite” with *girl, girls, boy, boys, women, and man*; that of action with *woman, man, men, girl, and boy*; physicality with *girl, woman, man, and boy*; that of abstractness with *man, men, and boy*; and finally, that of group membership with *women, men, and boys*. Therefore it appears that in the concordances relative to the genitive both gender and number are relevant grammatical categories for the identification of similarities and differences in the patterns of usage of WOMAN, MAN, GIRL, and BOY.

## 5.2. *For-headed prepositional phrases*

The preposition *for* potentially encodes a number of semantic relationships between two concepts; for instance, it may indicate that something is intended to be given a specific person (e.g. *Let's save some cake for Phil*), that someone is in favor of a person (e.g. *Who are you going to vote for?*), that a characteristic is surprising when considering what a given person is like (e.g. *That was fast for you*), that something is an intended action (e.g. *The plan is for us to leave early*). *For*-headed prepositional phrases, therefore, may provide a convenient window to the range of semantic preferences linkable to the referent of the object of that preposition.

The Cobuild corpus reveals that *for*-headed prepositional phrases embedding the lexemes GIRL or WOMAN tend to be accompanied by words that can be grouped under a few broad notions: appropriateness of conduct, adequacy of conditions, problematicity of circumstances, unusuality of achievement, (cause of) disruption of relationships, and feelings. Such phrases are often preceded by adjectives or expressions conveying judgments about what a woman could, couldn't, should or shouldn't do (*advisable, indispensable, fitting role, acceptable norm, offence, unbecoming, unthinkable, risky*), expressing opinions about what is acceptable or condemnable in her behavior (*absurd, believable goal, lunatic*) or indicating restrictions on her actions (*unusual job, no place, only sphere, wasn't a job, what else is there, ain't the right kind of life, not the kind of garment for a girl to wear in a lonely search, too much*). Alternatively, such phrases are preceded by evaluative descriptions of the contexts in

which women act and of what they represent to them; often, these are negatively connoted (*torture* [reference is to age], *obstacle*, *unusual*, *urgency*, *important*, *nostalgic event*). Several of the circumstances mentioned in relation to such prepositional phrases are explicitly negatively marked (*cause of misery* [reference is to age], *to have her face smacked*, *complicated training*, *lonely life*, *ultimate time possible*, *problem*, *too small*, *dissatisfaction*, *concern*, *chances ... are remote*, *fears*), which reveals that problematic situations are frequently mentioned in association with women, and more frequently than neutral or positive ones (*conditions of life*, *pleasurable activity*, *splendid effect*). Moreover, when positive qualities are mentioned, these are presented as *exceptionally* related to women (*she could run well*, *professional*, *not bad*, *good question*, *convincingly*). A woman is also not infrequently mentioned as the “indirect” cause of the disruption of a relationship, thus being cast in the ambiguous role of a new beneficiary of the loving care that someone else is being deprived of (*left me*, *left his wife*, *leaves his male lover*, *left home*). Finally, when playing the role of a beneficiary, the referent of GIRL and WOMAN appears to be the object of affection rather than other kinds of attention (*affection*, *ideal mate*, *desire*, *love*, *hopeless love*).

Comparable prepositional phrases for MAN and BOY are more numerous and less negatively marked. For one thing, more frequent are the phrases in which the referents of those lexemes play the role of beneficiaries, depicted as profiting from situations in which they are not specifically or necessarily the object of someone’s affection, but of various forms of attention (*I would do this for a man*; *they had voted for a man who had promised much*; *very little a woman won’t do for a man who makes her look half her age*; *asking for immunity for a man who slaughtered*; *public affirmations of support for a man who was directly responsible*). Secondly, more numerous are the phrases that introduce or are part of mini-descriptions of sets of conditions in which the referents of MAN or BOY may find themselves acting; these are not meant to express value judgments, but rather to simply report on the state of their circumstances or to mention their salient characteristics: most of them are neutral or downright positively connoted (*for a man of Haig’s background*; *for a man like you*; *For a man running a campaign*; *For a man as careful as De Gaul*; *for a man in his position*; *For a man who has had little or no trouble*; *For a man who potentially had another life*; *For a man who could have taught Japanese*; *For a man with such power at his fingertips*; *for a man on his stag night*; *forty was a terrific age for a man*). Indications of the appropriate conduct for a man are scanty (*What greater shame was there for a man than to walk the streets cringing*; *You consider it right for a man of my years*) as is the reference to negative experiences or problematic circumstances (*The problems will be worse for a man if the disfigurement is on the left*; *an impossible position for a man of deeply skeptical inclinations*; *was the inevitable result for a boy whose disaffection from his family*; *scandal represented tragedy for a man*). More frequent are instead representations of positive situations applying to men: advantageous circumstances, personal achievements, solutions to possible problems (*wonderfully releasing experience for a man who, as a child, felt dominated*; *it will make it easier for a man to find and keep a partner*; *ease and comfort*; *it was a dream job, no mean achievement for a man of 50*; *Not bad for a man with an average voice, didn’t do badly at all - for a man hospitalized*). More generally, such *for-headed* prepositional phrases are set in text segments presenting positive or neutral descriptions or evaluations of the circumstances of events or situations regarding men (*which was normal for a man of his age*; *more creditable for a man to capture an animal alive*; *It is usual, I imagine, for a man to look for perfection*; *may be particularly intense for a man*; *a move to ... is a gamble for a man who has spent his whole career*; *jail would be another kick for a man who is down*); sometimes they mention untypical situations (*this is strange for a man to be turned into a woman*; *it’s somehow abnormal for a man with a partner*; *It seemed a queer thing for a man to do*; *Most unusual for a man to offer another man a seat*; *how out of the common it is, for a man to say something like that*); finally, although concordances occasionally re-present some of the semantic contexts already identified for WOMAN and GIRL (cause of disruption of relationship: *dumps her for a man*; object of affection: *would fall for a man serving a life sentence*; *yearning of a man*; problematic conditions: *it’s dangerous for a man knowing and feeling*; *not an easy assignment for a man*; appropriateness of conduct: *it is high treason for a man to violate the wife of the Sovereign*; *it’s unnatural for a man to show such inertia*), these do not constitute the focus of the various patterns of discourse characterizing men and boys.

## 6. Conclusion

An examination of the most significant collocates of WOMAN, MAN, GIRL, and BOY in four components of the Cobuild on-line corpus has revealed that these lexemes both share some contexts of use (e.g. age, physical appearance, personal relationships, positive and negative states), but not always to the same degree, and also that they are relevant to complementary discourse domains (e.g. religion, physical attractiveness, danger, and women’s liberation for females, and the military, non-physical attractiveness, activities, and physical force for males). A consideration of the concordances of the

above lexemes in selected syntactic environments has also shown that females and males may be associated with the same *types* of topics and notions, but also that the frequency of their co-occurrence is often dissimilar. Thus, for example, attributes describing females are more often negative or focused on the notion of beauty, while those modifying males tend to be more evenly distributed between positive and negative qualities or to stress men's social roles and intellectual qualities. Similarly, *of*-headed prepositional phrases preferably describe physical attributes of females but non-physical characteristics of males. In relative clauses, it is more often females rather than males who are represented as having to face difficult situations or in relation to their feelings, while men are portrayed in their professional roles. *For*-headed prepositional phrases typically refer to appropriateness of conduct with regard to females, and to advantageous circumstances or personal achievements in the case of males. On the other hand, concordances of genitive constructions have shown that comparable wordforms of different lexemes may occasionally share similar patterns of co-occurrence with certain semantic fields; thus, the singular of GIRL and BOY are both frequently associated with the notions of physical attractiveness, physicality, and negativity, while both WOMAN and MAN in their plural forms are similarly associated with the notions of group membership and sports events.

The findings indicate that in the discourse practices of English speakers the world of females is frequently associated with notions of passivity, negativity, and physicality, while that of males with notions of activity, positivity, and cognitivity. This suggests that socio-culturally salient concepts encoded in potentially symmetric terms may develop different ideological significance when they recurrently keep company with different sets of words. On the other hand, more data need to be collected to confirm or disprove these conclusions especially considering that different wordforms of the same lexeme may not occur in the same contexts of use (see Stubbs 2001: 27-28), and that collocations differ considerably in different text types and thus in different corpora components (see Nakamura and Sinclair 1995: 108-109). To this end, further pairs of gender-marked terms need to be examined (e.g. *sister, brother, wife, husband, female, male*) in more varied syntactic contexts (e.g. *in*-headed prepositional phrases, complements, passive constructions) and in additional corpora components (e.g. relevant to the spoken medium and to specific genres).

## References

- Fairclough N 1992 *Discourse and social change*. Cambridge, UK, Polity.
- Halliday M.A.K. 1994 *An Introduction to Functional Grammar*. London, New York, Sydney, Auckland, Arnold.
- Hodge R, Kress G 1993 *Language as ideology*. New York, Routledge.
- Krishnamurty R 1996 Ethnic, racial and tribal: the language of racism? In Caldas-Coulthard R, Coulthard M (eds) *Texts and practices: readings in critical discourse analysis*. London, Routledge, pp. 129-149.
- Nakamura J, Sinclair J 1995 The World of *Woman* in the Bank of English: Internal Criteria for the Classification of Corpora. *Literary and Linguistic Computing* 10(2): 99-110.
- Partington A 1998 *Patterns and meanings*. Amsterdam, Benjamins.
- Stubbs M 2001 *Words and phrases. Corpus studies of lexical semantics*. Oxford, UK, Blackwell.
- Tognini-Bonelli E 2000 *Corpus Linguistics at Work*. Birmingham, UK, TWC.
- Wierzbicka A 1999 *Emotions across languages and cultures: diversity and universals*. Cambridge, CUP.