

-- SEÇÃO X --

DESCRIÇÃO DE USO: LINGÜÍSTICA DE CORPUS

SIMPÓSIO*

PC AND SEXISM RELATED TERMS IN COBUILD'S BANK OF ENGLISH: A CORPUS STUDY OF MEANING

Aleksandra PIASECKA-TILL (UFSC; bolsista CAPES)

RESUMO: O presente trabalho relata os resultados de um estudo lexical efetuado no Bank of English do COBUILD (um corpus de 323 milhões de palavras) na Universidade de Birmingham. Usando o programa de colocação de Clear (1993) para analisar a frequência, colocação e concordância de dois grupos de palavras/frases -- termos associados ao conceito de politicamente correto e de sexismo-- comprovei a validade das seguintes hipóteses: 1. o uso dos termos do primeiro grupo (PC) muda diacrônicamente; 2. o campo semântico desse grupo incorpora conotações negativas; 3. existe uma interseção entre os campos semânticos dos dois grupos. Foi também detectado um deslocamento na intensidade do uso dos termos do primeiro grupo.

1. Introduction

In Brazil, the use of *politicamente correto/-a* and *o politicamente correto* (PC hereafter) -- the Portuguese equivalents of politically correct and political correctness respectively-- was brought to our attention by the media, appearing suddenly and intensifying in the early nineties. Simultaneously, the recent literature on PC in the Anglo-Saxon world has pointed out a media campaign against it. Neilson (1995: 60), for instance, writes about Newsweek announcing the arrival of the thought police, or calling PC a totalitarian philosophy in a clear association with fascism. In Time, he finds a new intolerance (id.: 61), while in the New York Times a McCarthyism of the left (id.: 62), which makes him draw the conclusion that the mass media depict PC as the leftist tyranny. Ohmann's (1995: 11) remarks on these same publications --with the inclusion of the New York Magazine-- are similar in tone. Also Cameron (1995) affirms that there appears to be a series of concentrated efforts in the media to discredit PC, which led to a state she describes as follows: "PC now has such negative connotations for so many people that the mere invocation of the phrase can move those so labelled to elaborate disclaimers, or reduce them to silence". (id.: 123)

Formulations like *campaign* or *concentrated efforts* made me think of negative comments on PC persistently repeated over a substantial length of time by the media. I also believed that explicit statements, similar to those quoted above (e.g., *thought police*), might have been replaced by a subtler lexical signalling.

To verify whether these expectations can be confirmed by the actual use of language, COBUILD's Bank of English, the largest corpus¹ of this language in the world, seemed to be the right instrument. By examining frequency, collocation and concordance, I tested three hypotheses about the PC related terms: the one of diachronic alteration in their occurrence; the other stating that their semantic fields embody negative assumptions; and the last one which predicts some overlapping between their semantic fields and those of the sexism related terms (since sexism is one of the forms of discrimination against which PC was created).

* Simpósio "Descrição de Língua: Lingüística de *Corpus*".

¹ A collection of texts, of the written or spoken word, which is stored and processed on computer for the purpose of linguistic research (Renouf, 1987/96: 1).

In this article, I report the findings of the study above, a small contribution to the discussion of the issue of Political Correctness in language use. I have looked at this issue as a critical discourse analyst, concerned with women, their position in society and gender relations, who focuses her investigations on the interdependence between language use and social practice. Since "Critical Discourse Analysis is... not a homogeneous method, nor a school or a paradigm but at most a shared perspective on doing linguistic, semiotic or discourse analysis" (Van Dijk, 1993: 131), I decided to carry out the present research on lexis by exploring possibilities offered in Corpus Linguistics to guarantee sufficiently dispassionate analyses, impossible to be refuted even by the opponents of CDA.

2. Some background notions

The word sexism, with its derivatives (e.g. sexist language), and all terms related to PC are recent, thus controversial. Sketchy definitions of these concepts I present below will help the reader to perceive why I stress interrelations between them.

2.1 Sexism and Sexist Language

On June 24th 1947 a strange object appeared in the sky over a little town in the USA, intriguing not only its inhabitants but also the American air force. Thanks to this military interest, we can precisely define when the term flying saucer entered the English language: on that day precisely it was used by one of the pilots to describe a UFO over Rosewell, and was recorded on tape.

In the case of sexism, the circumstances were different: although the social practice this word refers to had existed for centuries, there was no urge to name it in patriarchal societies in which only men were deemed significant. We cannot be exact in dating its first appearance and a statement that this word has only recently entered the English lexis must satisfy us. One in a series of neologisms coined by feminists in their attempt at describing-or rather renaming-the world according to women's views and necessities, sexism is defined as "particular manifestations of the bias in favour of males" (Spender, 1987: 15). These manifestations can be found in all areas of human activity, including language. Sexist language therefore is the one which does not refer clearly and fairly to both sexes (Martyna, 1983: 25), using lexical and grammatical means to reflect (but also actively construct) reality from the viewpoint of the male.

According to Firth, "language is a way of behaving and making others behave" (in Coulthard 1985: 1). When looked at from the perspective that links linguistics with social theory, this sentence's content is twofold. To interact with others and refer to the world we need to learn a classificatory code -- language. Although many concepts are acquired non-verbally (Carroll, 1964), when we finally learn the language, it begins somehow to limit us: we are able to perceive and describe the universe just according to arbitrarily named objects, facts, events or feelings. Hence, our social behaviour depends on the language we use, while the use of language results from our social practice. Consequently, the social practice of sexism must be seen as a cause of sexist language use and vice versa.

2.2 The Concept of PC

According to Cameron (1995: 127), for most people not engaged in radical political movements, PC is a neologism whose meaning has to be inferred from context. Molyneux (1993: 43) suggests that it is a cultural phenomenon that encompasses a series of attitudes, remnants of the movements for female, black and gay liberation. Extended to other areas of discrimination, those attitudes, being gradually transformed into social practice, attempt to change society by changing the way people think and talk. Such an understanding of PC can easily be associated with the Firthian idea above. Being an advocate of PC, I see it as a means of defending both women and minorities against the prejudice reflected in and expressed through the use of language (*inter alia*).

Outside the United States (where emphasis has been placed basically on multiculturalism and, in consequence, on university curricula considering cultural/literary heritage (c.f. Howe 1993, Ravitch 1993, Short 1993)), the promotion of language reform has been the most noticeable aspect of the 'PC phenomenon,' as it is usually labelled. However, the notoriety given to this prescriptive aspect, especially in the media (the core of my professional interest), makes us pay little attention to the complexity of social and cultural reform whose proposal is contained in the concept of PC.

Moreover, because of its link with the Left, PC has a negative status in popular understanding. Our intuition tells us that "PC is a dirty word in nineties Britain" (Dunant, 1994: vii) and people would prefer not to be described as 'politically correct.' Cameron (1995: 124) states that "the people who are characterized as adherents of the 'political correctness movement' strenuously deny that any such movement exists." They prefer to define themselves as supporters of affirmative action, non-discriminatory language and multicultural curricula, without explicit reference to PC. Since, as it has been claimed, the media have contributed to such perceptions of PC, I decided to verify how they did it.

3. Procedure

3.1 Data

COBUILD's Bank of English at the University of Birmingham to which I applied the collocata programme (Clear 1993) constituted my data.

As I learnt from Krishnamurthy (personal communication), COBUILD started to collect its data of written and spoken modern English in 1980. At the end of 1996 the Bank of English stood at 323 million words, organized into 16 sub-corpora, among which 3 exclusively spoken. For my research, it was important that the language of the corpus in daily use at COBUILD is basically a sample of journalistic data: its major part has been collected from British, American and Australian newspapers and magazines, as well as from transcribed international and national radio broadcasts. The remaining part consists of books, informal conversations, meetings, lectures, local radio phone-in programmes etc Such an arrangement of

COBUILD's data offered me a possibility of approaching the general corpus and the one consisting exclusively of the newspaper texts.

3.2 Methodology

CDA is based on a conception of interdependence between linguistic and social phenomena: they cannot be separated, since the way we think --and therefore speak/write and listen/read-- is socially determined. Thus, language, text and context are closely involved in the process of representation of human experience. Linguistic choices, among them the lexical ones, made by text producers carry specific social meanings.

In Fairclough's (1992) three-dimensional framework for CDA, text, discourse practice and social practice are approached in three analytical steps: description, explanation and interpretation. The present study of lexis, contained in the dimension of text, has basically a descriptive character.

As for the computer-assisted study of language, as Sinclair (1991: 4) explains, "the ability to examine large text corpora in a systematic manner allows access to a quality of evidence that has not been available before. The regularities of pattern are sometimes spectacular and, to balance the variation seems endless. The raw frequency of differing language events has a powerful influence on evaluation. [...] Indeed, the contrast exposed between the impressions of language detail noted by people, and the evidence compiled objectively from texts is huge and systematic".

According to him, human intuition about language use is so specific that it cannot be used as a reliable means to predict what happens in fact when the same people actually use the language (id.). Stubbs (1996) points out that intuition is often used to discredit results produced by descriptive methods for text analysis. Among people predisposed skeptically towards certain research methods, "when quantitative analysis supports a familiar belief, the comment is often that everyone knows that. When it provides an unexpected finding, the comment is often that you can prove anything with statistics" (Burrows, 1992, p. 183 in Stubbs).

Such criticisms are also extended towards certain research goals.

By using corpora of naturally occurring texts as a starting point of research, we can perform explicit comparative descriptions revealing clearer relations between lexico-grammar and text. This can trigger elaborating metamehtods, a way of "separating the correct and useful from the banal" (id.: 153), striving for objectivity in research, whose lack is a central argument against discourse analysis, especially CDA (e.g. Frawley 1987 in Stubbs, Widdowson 1995).

3.2.1 The COBUILD collocate programme

The computational tools I have used belong to the lexicographers' sphere. The programme enables us, basically, to identify new stereotyped word combinations from a large corpus, which constitutes one of the most important aspects of modern lexicography. The more specialized aspects of this programme concern "the retrieval of words within a range [...], the frequency of co-occurrence,

the treatment of inflected forms of a single lemma and the calculation of the significance value" (Clear 1993: 276).

In the case of the phrases under investigation, the identification of them as stereotyped word combinations has been established socially and is commonplace, therefore I used the collocate programme to:

- assess the overall frequency of occurrence of the combination of words politically +correct, political + correctness, and their distribution among the sub-corpora;
- assess the annual occurrence of the two phrases in the period of 1990-1996;
- sort the collocate lists of politically correct in descending order of significance; and
- assess the overall frequency of the word sexism.

Moreover, in the corpus compiled from the newspaper sub-corpora, the collocate programme helped me to

- assess the frequency of sexism, sexist and nonsexist; and
- elaborate collocate lists of these words.

The sequence of the steps of my research will be presented in the following section.

4. Analysis and discussion

First, the assessment of general and annual frequencies of politically correct and political correctness allowed me to evaluate diachronic modifications in the occurrence of these phrases. Next, by analysing the collocates of politically correct, I defined this phrase's semantic field and verified if it embodies negative assumptions.

Finally, the concordance of the words sexism, sexist and non-sexist was analysed. I used the general corpus to establish the collocates of sexism, and the corpus of newspaper texts to assess the collocates of all these three words. For the reasons already exposed, I aimed to find some overlapping between their semantic fields and that of politically correct.

I designed the research in the way described above to specify the lexical clues for the future selection of the complete articles to be included in my thesis corpus. The relevance of texts, whether selected manually (i.e., by eye rather than hand!) or electronically, needs to be identifiable through some objective, not arbitrarily chosen by me, criteria. I expected the lexical sets forming the semantic fields of the phrases and words under analysis to provide such tools.

4.1 PC

4.1.1 Frequency and distribution

The assessment of the frequency of occurrence of the phrases under investigation in the general corpus, produced the following tables:

POLITICALLY+CORRECT: 1292 lines

Corpus	Total Number of Occurrences	Average Number per Million Words
oznews	339	10.2
brmags	202	6.7
times	138	6.6
econ	74	6.1
guard	138	5.7
today	139	5.2
indy	66	3.4
npr	70	3.1
newsci	17	2.8
usbooks	60	1.8
brephem	5	1.1
usephem	1	0.8
brbooks	30	0.7
usnews	6	0.7
brspok	7	0.3
bbc	0	0.0

POLITICAL+CORRECTNESS: 965 lines

Corpus	Total Number of Occurrences	Average Number per Million Words
guard	183	7.5
times	141	6.7
oznews	207	6.2
econ	56	4.6
brmags	132	4.4
indy	63	3.2
today	73	2.7
npr	43	1.9
newsci	11	1.8
usnews	7	0.8
usbooks	23	0.7
brbooks	17	0.4
brspok	8	0.4
beephem	1	0.2
usephem	0	0.0
bbc	0	0.0

The lowest ranks of the tables are, basically, occupied by the spoken sub-corpora, which are British. This fact shows that the PC related terms are much less frequently used in general conversation by lay people, or by specialists during lectures or meetings, when compared to the use of 'journalese.' The top positions in the tables reserved for the newspaper sub-corpora (OZNEWS, TIMES, GUARD) and British magazines confirm this interpretation. Such a distribution can indicate an association of PC with the aspect of formality/informality in language use, thus an opposition between written and spoken respectively.

The same conclusion however, when applied to the BBC sub-corpus representing spoken journalese, becomes contradictory. As I pointed out elsewhere (Piasecka-Till 1994), most of the so-called 'spoken' discourse from radio and television is pre-scripted, thus written. There is no reason for the BBC World Service from which the data were transcribed to break

this rule. So, the surprisingly low position in the tables (no use of the PC-phrases) of this sample of definitely formal language must be attributed to another feature of the BBC: the World Service is renowned for its 'objective' fact-reporting rather than 'subjective' political comments. An additional factor contributing to the BBC's low position is the period from which the BBC data were collected: it is a finite corpus of 1990 - 91, when the Total Number of Occurrences of the phrases under investigation was low or equal to zero (see 4.1.2).

In terms of national distribution, among the newspaper sub-corpora, the Australian comes in top position, while the American appears at the bottom. When analysing this fact from a diachronic perspective, we could conclude that the interest in the PC phenomenon has been moving eastward: from the country where it acquired its label, the USA (texts started to enter the corpus in 1989), through the UK (first texts from 1990) to Australia (texts collected from 1994-1995). This conclusion demands further investigation, though.

The last comment refers to the unexpectedly high position in the rank of the TODAY sub-corpus which contains data from a tabloid newspaper. My reaction proves validity of Sinclair's stance on intuition: I approached the data with certain preconceptions based on formality vs. informality in language, and with some bias against the topics explored in tabloid newspapers. I was obliged to revise them when the evidence from the corpus laid my misconceptions bare. Hypothetically, had 1996 data been included into TODAY finite corpus of 1991-1995, it could have occupied even higher position, surprising us even more.

4.1.2 Annual occurrence

The analysis of the annual frequencies was based on the following figures, complementing the results above:

POLITICALLY+CORRECT: 1292 lines

Year	Corpus size (in Million)	Total Number of Occurrences	Average Number per Million Words
1990	24	0	0
1991	18	51	2,8
1992	23	64	2,8
1993	6	31	5,1
1994	14	80	5,7
1995	90	683	7,5
1996	14	76	5,4

POLITICAL+CORRECTNESS 965 lines

Year	Corpus size (in Million)	Total Number of Occurrences	Average Number per Million Words
1990	24	0	0
1991	18	11	0,6
1992	23	31	1,3
1993	6	33	5,5
1994	14	51	3,6
1995	90	550	6,1
1996	14	108	7,7

With the exception of the 1996 data, the absolute number of occurrences and the average rates in both the tables are progressive. Given the interruption of data collection at COBUILD in July 1996, we can confirm the initial assumption of the increase in the use of PC terms. It is especially transparent in the case of political correctness: when comparing 1994 with 1996, we notice that with corpora of the same size, 14 million, the average number per million words more than doubled in 1996.

By contrast, in the case of politically correct the same comparison will show us that both, the total number of occurrences and the average number representing 1996 are slightly smaller. This apparent reduction does not seem to be statistically relevant, when treated boldly. It is necessary to remember nevertheless, that we are comparing the whole year of 1994 with only 7 months of 1996: in fact, we face an implied increase.

The significantly bigger average rate per million words of the occurrence of political correctness when compared to politically correct could indicate a certain switch in the interest towards a concept. This observation refers specially to the British press since all 1996 COBUILD data come from the Times and Sunday Times. These occurrences analysed in larger contexts may substantiate this hypothesis.

4.1.3 Collocation

To assess collocation of political correctness, I selected MI-score from a group of the three most commonly used statistical methods available at COBUILD. All of them (frequency, T-score and MI-score), separately or in a comparable set, are relevant and revealing. An MI-score list of collocates, however, is more transparent and illustrates clearly the possibilities offered by Corpus Linguistics to other branches of linguistic studies.

In the theory of probability, MI (which stands for 'Mutual Information') is a measure of the strength of association between two events - in our case, words. When we use MI to order collocates, those having a very strong association with a node term appear at the top of the list, while, automatically, the most frequent collocates of any word in English (e.g., the, of or is) are discarded. Thus, the produced list enables us to focus on the most significant words.

Let us look at the first fifty (a standard COBUILD format) collocates of political correctness:

- | | |
|---------------------|-----------------|
| 1. discriminator | 26. advocates |
| 2. tokenism | 27. irrelevant |
| 3. multiculturalism | 28. bureaucracy |
| 4. priestess | 29. preference |
| 5. fanaticism | 30. mad |
| 6. political | 31. victim |
| 7. campuses | 32. climate |
| 8. epitome | 33. attitudes |
| 9. correctness | 34. era |
| 10. fad | 35. attacking |

- | | |
|-----------------|--------------------|
| 11. loony | 36. pc |
| 12. affirmative | 37. minded |
| 13. backlash | 38. notion |
| 14. feminism | 39. intellectual |
| 15. menacing | 40. gender |
| 16. rampant | 41. extreme |
| 17. inclusive | 42. discrimination |
| 18. feminist | 43. kingdom |
| 19. altar | 44. intervention |
| 20. trendy | 45. crazy |
| 21. accusations | 46. speech |
| 22. lobby | 47. fallen |
| 23. obsessed | 48. gone |
| 24. guilt | 49. attacks |
| 25. versus | 50. hate |

Even without entering a sophisticated semantic analysis, it can be easily noticed that the negatively-loaded words compose a quite long list. I organised it into two sub-groups which confirm my second expectation, that of a negative semantic field within which PC is embedded in the press. The first sub-group leaves no doubts and incorporates words such as discriminator, fanaticism, menacing, accusations, obsessed, guilt, victim, attacks or hate. Perhaps demanding a closer look at their immediate contexts, the words in the other sub-group have a negative connotation only: tokenism, fad, loony, backlash, trendy, lobby, bureaucracy or intellectual. By analogy to the latter, someone could insist on including in this group the words feminism and feminist. I prefer, however, to associate them with multiculturalism, campuses, climate or era, terms which define the place, moment and one of the core PC concepts.

In this way, we obtain slightly more than 10% (of the overall list of 50 items) of neutral terms, adding to the latter epitome, priestess, notion, kingdom and speech. On the other hand, we can also find two definitely positive words, affirmative and inclusive, which occupy relatively high positions among the collocates. Their presence within the strongest one third of collocates permits us to be more optimistic about the role of the press in society. Apparently, there is a sector in the press concerned with the one-sidedness of their coverage. Offering an alternative to such a view, they take a serious effort to achieve diversity and show that there are desirable aspects of PC. In this kind of lexical analysis it is impossible however, to determine who they are and why they considered the positive side of PC worth exploring.

4.2 Sexism, sexist and non-sexist

4.2.1 Frequency

Earlier in the paper I stated that it is impossible to define an exact moment when the term sexism and its derivatives were used for the first time in English. However, when we look at the bibliographical references of Vetterling-Braggin's (1981) anthology (essential for the discussion of sexist language for trying to provide an initial philosophical analysis of the claim that "ordinary language is

'sexist' and that use of such language ought to be altered or eliminated" (id.: Preface)), we find several publications from the early and mid-seventies. Academic publications reflect social interests and are produced usually after some consideration of the phenomena they describe. Hence, we can roughly estimate that the words of interest (mentioned in the title of this section) have been used for over three decades, especially when we remember their association with the feminist movement intensified in the sixties.

It is not much when evaluated from a historic perspective, but I expected that these words would be much more frequent than the terms linked to PC, a newcomer in the debate. The total number of instances of the word sexism in the whole corpus of Bank of English was 1399 though, so only 8.3% more than politically correct. Fairclough's (1994: 193) issue of overwording, a signal of "intense preoccupation pointing to peculiarities in the ideology", made me draw conclusions towards the opposite end of the continuum. Such a low number of sexism occurrences indicates that the problem of sexism does not occupy a prominent position among social problems reported or commented on in the nineties. On the other hand, it is necessary to emphasise the element of novelty, a characteristic of the media discourse (the Bank of English contains predominantly journalistic data) as a probable cause of comparably frequent use of politically correct.

4.2.2 Collocation

First, the strongest collocates of the word sexism in the general corpus were compared/ contrasted with those in the corpus of the newspapers. Here, I concentrate on the 'top ten':

General	Newspapers
1. classism	ageism (3.)
2. heterosexism	homophobia (4.)
3. ageism	racism (5.)
4. homophobia	misogyny (7.)
5. racism	hating
6. isms	stereotypes
7. mysogyny	blatant
8. racialism	rampant
9. patriarchy	spender
10. schoolroom	feminism

In the left-hand column we have seven examples of different forms of discrimination: homophobia and misogyny, and the five words ending in -ism. The plural form of this suffix, the neologism isms, ironically signals the focus of the latter group. According to the explanation given in Collins Cobuild English Guides 2: Word Formation (1994: 88), "-ism occurs in nouns which refer to particular beliefs, or to behaviour based on these beliefs. For example, 'feminism' is the belief that women should have the same rights, power, and opportunities as men".

The word patriarchy does not comply with the morphological form referred to above. But the system it defines is somehow based on a belief (that the fathers, i.e.

males, have more rights than the rest of society) which transforms this form of social organization into a kind of discriminating phenomenon: it caused unequal treatments between the genders.

Altogether, such the collection indicates that the texts in the general corpus concentrate on concepts or issues.

A similar group of phenomena is also present in the right-hand column, however, it is smaller: only five explicit forms of discrimination are there, among which four equal to the first column. Although these occupy different positions in the rank, they preserve the sequence: ageism precedes homophobia, and misogyny follows racism, which reflects the same degree of importance given to those collocates. It is interesting to see feminism entering the focus of attention, maybe substituting patriarchy, a word describing the system criticized and confronted by feminism.

While the left column shows a place, schoolroom (hopefully of discussions of the issues collected above and not practice of them), the newspaper column introduces words indicating evaluation: blatant, rampant and hating. The latter can be also seen as a process (as an expression of negative feelings I see it as a form of discrimination). The evaluative aspect of these words matches one of the theoretical presuppositions of CDA which states that media discourse not only reflects but also reinforces, creates and recreates social reality.

Now let us look at the strongest collocates of sexism, sexist and non-sexist in the newspaper corpus:

Sexism	Sexist	Non-sexist
1. ageism	misogynist	non
2. homophobia	ageist	correct
3. racism	homophobic	politically
4. misogyny	underbelly	racist
5. hating	racist	multicultural
6. stereotypes	disrespectful	language
7. blatant	dinosaur	provide
8. rampant	claptrap	but
9. spender	slob	be
10. feminism	condescending	its

The first feature in this group of words is the fact that the sexism collocates are predominantly nouns, while in the case of sexist the nouns are balanced with adjectives. This suggests that the terms from the first column appear in a string or pairs of words belonging to the same class, e.g., homophobia, racism and mysogyny; while those from the second would form classic noun phrases, e.g., mysogynist (and/or homophobic) claptrap. Such usage can contribute to a stereotypical perception of behaviour or a phenomenon referred to, especially when we notice that both these columns bring negative meanings to our attention, with the second of them exclusively negative (when we interpret the meaning of dinosaur metaphorically).

Neither of this is true when we look at the third group of collocates: there are different classes of words and the basically negative connotations do not surround non-sexist. (The fact that its strongest collocate is a prefix non- must be disregarded as an indication of a negative connotation. Since the analysed word is a two-element compound, the collocate programme automatically chose one of them as the node of search, and it was -sexist.) The only negative term is racist but when we look at the sample of concordances of non-sexist, we will see that in the texts it appears as non racist -, in perfect symmetry with non-sexist. Again, similarly to the two other columns, the strings and pairs of words commented above can be predicted.

It is worthwhile pointing out that for the first time some verbs emerge. One of them is be that belongs to the set of the most frequent verbs in the English language. However, the grammatical form of it, a complement of modals or an imperative, can be significant, suggesting a prescriptive or hortatory usage. This favourable climate is reinforced by provide which brings an idea of fulfilling somebody's needs, alluding to the desirability of non-sexist objects or behaviour - for example, of language, another important collocate in this column.

The most exciting finding, however, is the presence of politically correct. Although separated and presented in an inverted order, an explicit link is finally made between the two phenomena I am interested in, political correctness and sexism. At least in this case, I have obtained confirmation of topical interdependence. Moreover, the lexical neighbourhood of these terms in the third column is neutral.

Confronting the findings above with the collocation list of politically correct however, did not produce any exact matches, when I took into account only the first ten collocates. To acknowledge certain overlapping, it was necessary to analyse the lists of all 50 items (see Appendix), and thus it was possible to gather several groups of what I call 'key-themes.' Here, I quote only the most numerous groups:

- a) discrimination, discriminator, oppression, prejudice and harassment;
- b) sexism, sexist, macho, machismo and chauvinism;
- c) language, speech, writing, and terminology;
- d) racism, racist, racial, and racialism;
- e) attitudes, connotations, and climate;
- f) feminism, feminist, and gender;
- g) attacks, attacking, accusations and critique.

The lexical items, components of these groups confirm this time our expectations: when we hear or read about PC, exactly they come to our mind, being a classical example of overwording. Moreover, the presence of group c) brings to our special attention the question of expression (or communicating), an additional stimulus for a further linguistic investigation.

5. Final remarks

Throughout this article I have shown that Sinclair's (1991: 7) definition of a corpus function, i.e., "to identify what is central and typical in the language," can be narrowed down to a specific goal of any researcher. In my case, a large corpus (numbers convince, undoubtedly) was analysed to obtain objective results. These results will complement further text analyses, because I see language as a system of potential meanings (Halliday 1978), which makes me explore quantitative evidence to sustain qualitative outcomes about discourse.

Similarly, the importance of collocation for any lexical study has been recognised not only by lexicographers. Text analysts, who base their investigations on a conception that collocations demonstrate the connotations and assumptions words have, frequently apply this tool. I would like to emphasise however, the necessity of looking at collocates in a context, since this is the real factor that defines meanings, as is stated in the CDA principles.

By the same token, the confirmation of topical interdependence between PC and sexism, and the identification of a series of key-themes among collocates answered another of my questions: corpus linguistics serves the purpose of establishing the criteria for texts selection. According to Sinclair (1991), careful planning and composition of a corpus are essential for reliability of our findings. Had the periods of data collection at COBUILD been more systematic, the findings would have been more conclusive. On the other hand, a decade in the history of a language is not much even in our times of intense information flow. Such a large corpus as the Bank of English seems to be really representative as a sample of language used by different individuals or institutions in different places but in practically coinciding periods.

My comment on the time span above does not invalidate the findings of change in the use of PC related terms in recent years. The historical factor in language change is a complex question and it will be investigated elsewhere. By analogy, the geographical dislocation of PC intensity is worthy of further studying.

Finally, it was important to corroborate the idea that the media, by the use of specific lexical choices, create the negative semantic fields to embed PC. Specific terms do not create phenomena they refer to, but "naming something can bring it to consciousness, give it a social identity, and facilitate its identification" (Stubbs, 1997: 370). It is necessary to investigate who the texts' producers are and what are their (real) goals. One of the main objectives of my research is to indicate that by becoming aware of hidden institutional agendas, we can become less submissive to manipulation. As Cameron says, the so called 'politically correct' language "is no more of a threat to freedom of expression than any other set of linguistic norms [...]. It threatens only our freedom to imagine that our linguistic choices are inconsequential, or to suppose that any one group of people, in the guise of defending 'the language', has an inalienable right to prescribe them". (id.: 165)

Thus, a more rational debate on the nature of social norms for language use and their consequences not only for women but also for society as a whole is asked for.

REFERENCES

- BECKWITH, Francis J. and Michael E. Bauman (eds.). 1993. *Are You Politically Correct? Debating America's Cultural Standards*. Buffalo, New York: Prometheus Books.
- CAMERON, Deborah. 1995. *Verbal Hygiene*. London and New York: Routledge;
- CARROLL, John B. 1964. *Language and Thought*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- COULMAS, Florian (ed.). 1997. *The Handbook of Sociolinguistics*. Oxford: Blackwell.
- COULTHARD, Malcolm. 1985. *An Introduction to Discourse Analysis*. London: Longman.
- CLEAR, Jeremy. 1993. *From Firth Principles: Computational Tools for the Study of Collocation*. An offprint from Mona Baker, Gill Francis and Elena Tognini-Bonelli. 1993. *Text and Technology: In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins Publishing Co., pp. 271-292.
- DUNANT, Sarah (ed.). 1994. *The War of the Words: The Political Correctness Debate*. London: Virago.
- FAIRCLOUGH, Norman. 1994 (1989). *Language and Power*. London and New York: Longman.
- _____. 1992. *Discourse and Social Change*. Cambridge: Polity Press.
- HALLIDAY, Michael Alexander Kirkwood. 1978. *Language as Social Semiotic: The Social Interpretation of language and Meaning*. London: Edward Arnold.
- HOWE, Irving. 1993. *The Value of the Canon*. In Beckwith, F. J. and M. E. Bauman (eds.), 1993, pp. 133-46.
- MARTYNA, Wendy. 1983. *Beyond the He/Man Approach: The Case for Nonsexist Language*. In Thorne, B. et al. 1983: 25-37.
- MOLYNEUX, John. 1993. *The 'Politically Correct' Controversy*. In *International Socialism*, Winter 1993: 61, pp.: 43-74.
- NEILSON, Jim. 1995. *The Great PC Scare: Tyrannies of the Left, Rhetoric of the Right*. In Williams, J. 1995, pp.: 60-89.
- OHMANN, Richard. 1995. *On PC and Related Matters*. In Williams, J. 1995, pp.: 11-21.
- PIASECKA-TILL, Aleksandra. 1994. *Power and Gender Relations in Brazilian Television Interviews*. Unpublished MA dissertation. Florianopolis, Brazil: UFSC.
- RAVITCH, Diane. 1993. *Multiculturalism: E Pluribus Plures*. In Beckwith, F. J. and M. E. Bauman (eds.), 1993, pp.: 165-83.
- RENOUF, Antoinette. 1987/96. *Corpus Development*. In Sinclair, J. (ed.), 1987/96, pp.: 1-22.
- SINCLAIR, John. (ed.) 1994. *Collins Cobuild English Guides 2: Word Formation*. London: Harper Collins Publishers
- _____. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SHORT, Thomas. 1993. *"Diversity" and "Breaking the Disciplines": Two New Assaults on the Curriculum*. In Beckwith, F. J. and M. E. Bauman (eds.), 1993, pp.: 91-117.
- SPENDER, Dale. 1987. *Man Made Language*. London and New York: Routledge & Kegan Paul.
- STUBBS, Michael. 1996. *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell Publishers.
- _____. 1997. *Language and the Mediation of Experience: Linguistic Representation and Cognitive Orientation*. In Coulmas, F. (ed.), pp.: 358-73.
- THORNE, Barrie, KRAMARAE, Cheris and HENLEY, Nancy (eds.). 1983. *Language, Gender and Society*. Rowley, Mass.: Newbury House Publishers, Inc.
- VAN DIJK, T. (ed.). 1993. *Discourse & Society*. London: Sage, vol. 4 (2).
- VETTERLING-BRAGGIN, Mary (ed.). 1981. *Sexist Language: A Modern Philosophical Analysis*. Littlefield: Adams and Co.
- WIDDOWSON, H.G. 1995. *Discourse Analysis: A Critical View*. In *Language and Literature*, 4 (3): 157-72, Longman.
- WILLIAMS, Jeffrey (ed.). 1995. *PC Wars: Politics and the Theory in the Academy*. New York / London: Routledge.