



**HAL**  
open science

# Man/woman versus hombre/mujer: a contrastive analysis of compound nouns, collocations and collocational frameworks

Ramon Marti Solano

## ► To cite this version:

Ramon Marti Solano. Man/woman versus hombre/mujer: a contrastive analysis of compound nouns, collocations and collocational frameworks . Il, Elle: Entre Je(u), 2015. hal-01645072

**HAL Id: hal-01645072**

**<https://hal.science/hal-01645072v1>**

Submitted on 23 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## ***Man/woman versus hombre/mujer: a contrastive analysis of compound nouns, collocations and collocational frameworks***

**Ramón MARTÍ SOLANO**  
*Université de Limoges*

### **Résumé**

La présente étude prétend dévoiler un large éventail de combinaisons et structures lexicales associées aux bases *man* et *woman* en anglais et *hombre* et *mujer* en espagnol dans une perspective à la fois synchronique et diachronique. Nous avons repéré des stéréotypes liés au sexe aussi bien que de nouvelles collocations moins prédictibles dans les textes à l'aide de grands corpus généraux. Les résultats obtenus dans les corpus synchroniques anglais correspondent à des combinaisons telles que *family man*, *gay men* et *career woman*, qui sont les plus fréquentes et qui contrastent fortement avec d'autres combinaisons plus traditionnelles ou stéréotypées dans le corpus diachronique. Quant à l'espagnol, la base *mujer* s'associe principalement à l'apparence physique, l'âge et la personnalité tandis que la base *hombre* s'associe aux affaires, à la confiance et à l'action, ce qui confirme l'immobilisme des constructions mentales stéréotypées liées au sexe dans cette langue romane.

Mots-clés : *man/woman*, *hombre/mujer*, collocations, combinaisons lexicales stéréotypées, lexicologie contrastive, linguistique du corpus

### **Abstract**

This study aims to uncover the wide range of word patterns and combinations associated with the nodes *man* and *woman* in English and *hombre* and *mujer* in Spanish both from a synchronic and a diachronic point of view. The detection of gender stereotyping, as well as the emergence of new and less predictable collocations in texts, is achieved by the use of large reference corpora. Results in the English synchronic corpora include combinations such as *family man*, *gay men* and *career woman* as the most frequent, which contrasts strikingly with more traditional or stereotypical combinations in the diachronic corpus. As regards Spanish, the node *mujer* is mainly associated with physical appearance, age and character whereas the

Ramón Martí Solano

node *man* is associated with business, confidence and action, which corroborates the immovability of gender stereotypical constructs in this Romance language.

Key words: *man/woman, hombre/mujer*, collocations, stereotypical lexical combinations, contrastive lexicology, corpus linguistics

## I. Introduction

Gender stereotypes can be observed not only in the lexicon and the phraseology of a language but also in its morphology and pragmatics. Romance languages, for instance, use different endings to distinguish the feminine from the masculine gender and still retain the marked/unmarked grammatical notion which makes the masculine plural form usual and acceptable for a mixed company. Other uses have evolved and have given rise to what could be described, from a traditional and normative point of view, as ungrammatical. This is the case of the anaphoric use of the third person plural pronoun in English to refer to singular gender-unmarked nouns or pronouns such as *person, student, everyone* or *somebody*. However, we believe that it is in the field of lexis, and particularly in lexical collocations, where the evolution or the immovability of a number of stereotypes can be better observed and analysed—certain combinations seem to decline whereas others are still widely used or have developed lately and have even become pervasive and widespread.

## II. Languages and gender stereotypes

Although languages in general show clear and distinct degrees of stereotyping—usually racial or sexual—, not all of them make use of the same grammatical, lexical or phraseological strategies to convey stereotypical constructs. However, most languages seem to share an amazing large number of sayings, proverbs, aphorisms and other sentence-like units that speak ill of women and flatter men (Schipper 2004). The situation is not exactly the same if we look at smaller combinations such as compounds, collocations and other word patterns. Examples such as the approving *mujer de bandera* ‘a striking woman or a hell of a woman’ or the disapproving *mujer de armas tomar* ‘a formidable woman’ are definitely culturally-bound and even though translation equivalents can always be found, these highly lexicalised combinations seem to reflect specific idiosyncratic features of the Spanish culture. In contrast to the previous examples, both English and Spanish make use of a large number of word

Ramón Martí Solano

combinations to designate prostitutes. Examples such as *mujer de la vida*, *mujer de vida alegre*, *mujer de mala vida*, *mujer de mal vivir*, *mujer de la calle*, *mujer pública*, *mujer mundana* and *mujer perdida* are transparent for speakers of other languages as the set of references induced by the adjectives and nouns associated to the node are shared more or less universally. However, some combinations remain culturally-specific and show different degrees of opacity, as with *mujer del partido*, *mujer de punto* or *mujer del arte*.

### III. Word combinations and patterns

In this study I will be dealing with three main types of word combinations and patterns both in English and in Spanish. Firstly, compound nouns of the subtypes Noun + Noun (*career woman*) and Adjective + Noun (*gay man*) and secondly, recurrent collocations of the subtype Adjective + Noun (*married woman* or *hombre honrado* 'honest man'). Thirdly, I will consider collocational patterns that include the node, the collocate, a preposition and/or a determiner. The word pattern schemas under scrutiny are the following:

NOUN + NODE

ADJECTIVE + NODE

NODE + ADJECTIVE

NODE + OF THE + NOUN

NODE + PREPOSITION + NOUN PHRASE

The above pattern schemas are used, with different degrees, by both languages. For very well-known reasons concerning the syntactic structures of Germanic and Romance languages, Spanish seems to favour the collocational patterns and English, the compound nouns and collocations.

#### A. Compound nouns

It is not my intention here to develop a theory about compounding in general, or compound nouns in particular, but to address the issue from a practical point of view applied to the topic in question. I will claim that compound nouns are lexicalised word combinations and this means, among other things, that they have their own entry in dictionaries. For instance, the compound noun *career woman* is included in the nomenclature of the *Cambridge Advanced Learner's Dictionary 2008 (CALD08)* whereas it is absent in *Webster's Third New International Dictionary 1986 (WTNID86)*. This lexicographic comparison clearly accounts for the evolution of the social and professional status of women in the Anglo-American society. However,

Ramón Martí Solano

one could argue that if this combination has been lexicalised as a compound noun it is simply because the meaning of the whole cannot be retrieved by adding or putting together the meaning of either constituent. A *career woman* is defined in the *CALDO8* as “a woman whose job is more important to her than having children”. The combination is semantically specialised and therefore has been given full headword status.

### **B. Collocations**

The prototypical examples of collocations in English tend to be combinations of an adjective and a noun such as *strong tea* or *high winds*. Contrary to compounds, collocations are not lexicalised and as a result do not have their own entry in dictionaries. Nevertheless, lexicographers tend to include these recurrent word associations in at least one of the examples provided in the headword article.

Collocation is the co-occurrence of two items in a text within a specified environment. “Significant” collocation is regular collocation between items, such that they co-occur more often than their respective frequencies and the length of text in which they appear would predict. (Jones et Sinclair 1974: 19)

Collocations, as defined above, are extremely significant word combinations in a given language not only because they are recurrent but also because they carry in them a particular way of understanding the world around us and, especially, human stereotypes and relations. The recurrence of the collocation *hombre honrado* in the Spanish corpora is the reflection of the importance that the concept of *honra* ‘honour’ has always had in the Spanish culture. Recurrence is then a key criterion for regarding a word combination as a collocation, thus the importance of carrying out frequency studies by means of large reference corpora.

### **C. Collocational frameworks**

The term *collocational framework* was introduced in 1991 by Renouf and Sinclair in their article entitled *Collocational Frameworks in English*.

Our ‘frameworks’ consist of a discontinuous sequence of two words, positioned at one word remove from each other; they are therefore not grammatically self-standing; their well-formedness is dependent on what intervenes. (Renouf & Sinclair 1991: 128)

Collocational frameworks can be significant either grammatically or lexically. For this study I have looked at all the collocational frameworks that have the target words *man* or *woman* as nodes irrespective of the size of the framework. Table 1 shows the different types of collocational frameworks in English including *man* and *woman* whereas Table 2 deals with collocational frameworks with *hombre* and *mujer* in Spanish.

**Table 1: Types of collocational frameworks in English including the target words *man/woman***

schema	examples
man/woman + of + N	man of letters
man/woman + of + the + N	man of the people, woman of the streets
man/woman + of + NP	woman of childbearing age
man/woman + about + N	man about town
man/woman + about + the + N	man about the house

**Table 2: Types of collocational frameworks in Spanish including the target words *hombre/mujer***

schema	examples
<i>hombre/mujer + de + N</i>	<i>hombre de provecho, mujer de mundo</i>
<i>hombre/mujer + de + det. + N</i>	<i>mujer de la vida</i>
<i>hombre/mujer + de + su + N</i>	<i>mujer de su casa</i>
<i>hombre/mujer + de + NP</i>	<i>mujer de armas tomar</i>

The schemas in Tables 1 and 2 are quite limited as I am only interested in equivalent frameworks in the two languages. The combinations illustrated in Spanish have also been described as *composición sintagmática* ‘syntagmatic compounding’ (Val Álvaro 1999: 4825) although the examples provided by Val Álvaro usually include idiomatic, non-compositional and opaque or semi-opaque combinations such as *diente de leche* ‘milk tooth’, *bautismo de fuego* ‘baptism of fire’, *telón de acero* ‘iron curtain’ or *ojo de buey* ‘porthole’. Only by revealing and putting to the test the whole range of combinational structures can one have a comprehensive view of the multi-layered and intricate network of lexical and cultural associations of both terms in each language.

#### IV. The corpora and the method of analysis

Three general corpora have been used in order to detect and analyse the main word combinations and patterns in English, namely the *British National Corpus (BNC)* and the *Corpus of Contemporary American English (COCA)* for a synchronic examination and the *Corpus of Historic American English*

(COHA) for a diachronic perspective of American English since the early 19<sup>th</sup> century to date.

As concerns Spanish, I have used the *Corpus de Referencia del Español Actual (CREA)* for synchronicity, the *Corpus Diacrónico del Español (CORDE)* for diachronicity and the *Corpus del Español (CdE)* for both a synchronic and diachronic approach. The latter is a 100-million-word collection of texts from the 13<sup>th</sup> to the 20<sup>th</sup> century, which comes in useful when the analysis of the development of a given combination or pattern is needed.

**A. Corpus results for English combinations and patterns with *man* and *woman***

The BNC is a corpus of contemporary British English from the second half of the 20<sup>th</sup> century including texts up to the year 1993.

As a general rule, I have decided to use the lexemes *man* and *woman* as the nodes, taking into account for frequency counts the plural and the genitive forms. As regards collocates, I have intentionally disregarded the adjectives *young* and *old* and the determiner *other* as they do not combine with the nodes to create significant collocations.

For all the corpus searches I have systematically extracted the four most frequent Noun + Noun combinations. As for compounds, the results are presented in Table 3.

**Table 3: Most frequent compounds with *man* and *woman* in BNC, COCA and COHA**

compounds	BNC	COCA	COHA
<i>woman</i>	<i>career woman</i> <i>peasant woman</i> <i>mystery woman</i> <i>cleaning woman</i>	<i>cleaning woman</i> <i>career woman</i> <i>medicine woman</i> <i>spider woman</i>	<i>Negro woman</i> <i>peasant woman</i> <i>cleaning woman</i> <i>business woman</i>
<i>man</i>	<i>family man</i> <i>business man</i> <i>field man</i> <i>mystery man</i>	<i>family man</i> <i>con man</i> <i>point man</i> <i>hit man</i>	<i>business man</i> <i>newspaper man</i> <i>medicine man</i> <i>grocery man</i>

I have highlighted the first two compounds in the COCA column for *woman* as they have nearly the same amount of hits, 161 for *cleaning woman* and 160 for *career woman*. On the other hand, it should be noted that the COHA has not yielded either *career woman* or *family man* as frequent compounds, which corroborates the evolution of these modern constructs in

contemporary society when compared to the four most frequent in the diachronic corpus.

It should also be noted that the only compound with *woman* found in the three corpora is *cleaning woman*.

As far as collocation is concerned, the most frequent Adjective + Noun combinations in the *BNC* are shown in Table 4. The most frequent significant collocation in the *BNC* is *married woman*. This is without the shadow of a doubt a clear sign of the association that still exists between being a woman and marriage in most cultures including the British society as a whole. Interestingly enough, the most frequent adjective that collocates with *man* is *gay*. The collocation tends to occur more recurrently in the plural (190 tokens for *gay men* versus 16 for *gay man*).

**Table 4: Most frequent Adj. + N collocations with *man* and *woman* in *BNC***

<b>[j*] man</b>	<b>[j*] men</b>	<b>[j*] woman</b>	<b>[j*] women</b>
<i>young man</i>	<i>young men</i>	<i>young woman</i>	<i>young women</i>
<i>old man</i>	<i>other men</i>	<i>old woman</i>	<i>other women</i>
<i>other man</i>	<i>old men</i>	<i>other woman</i>	<b><i>married women</i></b>
<i>dead man</i>	<b><i>gay men</i></b>	<b><i>married woman</i></b>	<i>older women</i>

#### **B. Corpus results for Spanish combinations and patterns with *hombre* and *mujer***

I have also used three different corpora for searches in Spanish. The same methodological approach as the one adopted above has been used for the combinations in Spanish.

As regards collocations, I have searched the combination *hombre* + Adj. and *mujer* + Adj. in the *CdE*. The results speak for themselves and are set out in Table 5.

**Table 5: Most frequent collocations with *hombre* and *mujer*, in the singular and in the plural, in *CdE***

<b><i>hombre</i> + Adj.</b>	<b><i>hombres</i> + Adj.</b>	<b><i>mujer</i> + Adj.</b>	<b><i>mujeres</i> + Adj.</b>
<b><i>hombre</i></b>	<b><i>hombres</i></b>	<i>mujer joven</i>	<i>mujeres jóvenes</i>
<b><i>honrado</i></b>	<b><i>honrados</i></b>	<b><i>mujer hermosa</i></b>	<b><i>mujeres</i></b>
<i>hombre joven</i>	<i>hombres</i>	<i>mujer amada</i>	<b><i>honradas</i></b>
<i>hombre</i>	<i>armados</i>	<i>mujer honrada</i>	<i>mujeres</i>
<i>moderno</i>	<i>hombres</i>		<i>hermosas</i>
<i>hombre solo</i>	<i>políticos</i>		<i>mujeres casadas</i>
	<i>hombres libres</i>		



Ramón Martí Solano

It is interesting to note that when the research is carried out in the plural, results may change considerably. This is something that should be done systematically in order to capture the whole range of possibilities and corroborate the initial hypotheses.

Women are most often described in Spanish texts as young, beautiful, honest and married whereas men are depicted as politicians or as fighting men and are described as honest and free. The association with marriage is only to be found with women in stark contrast with the association that exists between freedom and men.

Bigrams are much less frequent in Spanish than in English as Romance languages tend to favour the syntagmatic Noun + Noun combinations by linking them with a preposition: *un hombre de letras* is 'a man of letters' whereas 'a business man' is *un hombre de negocios*. In order to account for the presence, frequency and extent of the compound nouns of the type N + de + N in contemporary Spanish I have carried out a search in the 20<sup>th</sup> century sub-corpus of the CdL. The nouns associated with the node *mujer* are, in order of frequency, *aspecto* 'appearance', *edad* 'age' and *carácter* 'character'. As for the node *man*, the associated nouns are *negocios* 'business', *confianza* 'trust' and *acción* 'action'. The conclusions that can be drawn could not be more straightforward and, at the same time, less encouraging for women and above all for the way women are regarded or considered as well as for the role they play in today's society. If language is, among other things, a reflection of the society that uses it—and I strongly believe in this—then women are not held in high regard as age, appearance and character seem to be the most important aspects, at least quantitatively, with which women are associated in the Spanish-speaking world. There is a striking coincidence between these empirical lexical results and the popular belief, which demonstrates that, unfortunately in some cases, stereotypes are alive and well.

## V. The question of substitutability

Gender substitutability stands here for the possibility of replacing *man* with *woman* in those common phrases in which *man* has been traditionally the only option. The CALD08 registers *the man/woman in the street*, *a man/woman of the people*, *a man/woman of the world*, to mention just a few. However, some of them resist substitutability as can be seen with *a man of action* or *be man enough to do something*. Spanish seems to be more conservative in this respect and even though some combinations show an equivalent number of occurrences in corpora—*hombre de principios* 'a man of principle' has 25 occurrences in the CREA and *mujer de principios*,

23—the general combinatorial tendency is clearly biased towards the masculine noun as shown in Table 6.

**Table 6: Very low or inexistent degree of substitutability *hombre/mujer***

	number of tokens in CREA		number of tokens in CREA
<i>hombre de palabra</i>	37	<i>mujer de palabra</i>	4
<i>hombre de confianza</i>	299	<i>mujer de confianza</i>	1
<i>hombre de provecho</i>	24	<i>mujer de provecho</i>	0

On the contrary, the pair *hombre/mujer de mundo* ‘a man/woman of the world’ seems to show a medium degree of substitutability if one looks at the results for both variants in the CREA—62 for *hombre* and 22 for *mujer*. This difference in the degree of substitutability can be explained if we examine the diachronic corpora. *Hombre de mundo* is not a very old phrase as most tokens in the *CORDE* are from the 19<sup>th</sup> century—the earliest example I have found is from 1813.

*Hombre de palabra* ‘a man of his word’ and *hombre de confianza* ‘a right-hand man’ are much older phrases. Both the *CORDE* and the *CdE* give *Guzmán de Alfarache* by Mateo Alemán as the earliest written record of the former phrase:

Y si lo habéis de dar y con ellos no andáis tan relojeros, que un solo momento faltáis a lo puesto, si no les pagáis al justo lo prometido, si se lo dilatáis un hora, ni sois **hombre de palabra** ni de buen trato. (*CORDE* 1604)

The earliest example of *hombre de confianza* in the *CORDE* is from 1545 and is to be found in a letter from the Spanish conquistador Pedro de Valdivia to the emperor Charles V.

As a possible explanation we could hypothesise that for the older combinations the use with *woman* is extremely low or non-existent. However, with the more recent combinations the use with *woman* seems to be more widespread. It could be inferred from the above that the longer a combination with *man* is lexicalised and therefore used the smaller the degree of substitutability is and vice versa.

In its turn, the Spanish *hombre de provecho* does not lend itself to a straightforward translation into English. The combination is frequent and well-known among Spanish speakers and although it can be actually applied both to male and to female referents, no examples with *mujer* have been found in the *CREA*. *Hombre de provecho* encapsulates the stereotype of the honest, hard-working and successful man that all parents would like their sons to become. A search in the *CREA* has yielded a total of 24 tokens of this

Ramón Martí Solano

phrase whereas the *CORDE* has got 63, which shows that the combination is a traditional one and that is little by little becoming less pervasive and widespread.

## VI. Conclusions

Results in the English synchronic corpora yield combinations such as *family man*, *gay men* and *career woman* as the most frequent, which contrasts strikingly with more traditional or stereotypical combinations in the diachronic corpus. The role of the modern man as happily in charge of home and children in the Anglo-Saxon societies is well established at least as far as word combinations is concerned. As regards Spanish, the node *mujer* is mainly associated with physical appearance, age and character whereas the node *man* is associated with business, confidence and action, which corroborates the immovability of gender stereotypical constructs in this Romance language.

## References

- British National Corpus*, <<http://www.corpus.byu.edu/bnc/>>.  
*Cambridge Advanced Learner's Dictionary*, Cambridge, Cambridge University Press, 2008.  
*Corpus de Referencia del Español Actual*, <<http://corpus.rae.es/creanet.html>>.  
*Corpus del Español*, <<http://www.corpusdelespanol.org/>>.  
*Corpus Diacrónico del Español*, <<http://corpus.rae.es/cordenet.html>>.  
*Corpus of Contemporary American English*, <<http://corpus.byu.edu/coca/>>.  
*Corpus of Historical American English*, <<http://corpus.byu.edu/coha/>>.  
*Diccionario de americanismos*, Lima, Asociación de academias de la lengua española, 2010.  
Susan Jones et John Sinclair, « English lexical collocations », *Cahiers de Lexicologie*, n° 24, 1974, p. 15.  
*Oxford Collocations Dictionary for Students of English*, Oxford, Oxford University Press, 2003.  
*Oxford English Dictionary*, Oxford, Clarendon Press, 1989.  
Antoinette Renouf et John Sinclair, « Collocational Frameworks in English », *English Corpus Linguistics*, études réunies par Karin Aijmer et Bengt Altenberg, New York, Longman, 1991, p. 128-143.  
Mineke Schipper, *Never Marry a Woman with Big Feet. Women in Proverbs from around the World*, Amsterdam, Amsterdam University Press, 2004.

Ramón Martí Solano

John Sinclair, *Corpus Concordance Collocation*, Oxford, Oxford University Press, 1995.

José Francisco Val Álvaro, « La composición », *Gramática descriptiva de la lengua española*, Ignacio Bosque et Violeta Delmonte, Madrid, Espasa, 1999, p. 4757-4841.

*Webster's Third New International Dictionary of the English Language Unabridged*, Chicago, Merriam-Webster, 1986.