

Student ID number	1304929
Module Number (1-6)	2
Title of Degree Programme:	MA Translation Studies
Title of Module:	CL and RM
Assessment Task No.	CL/13/01
First or second submission	First
Date Submitted	30 <sup>th</sup> May 2013
Name of tutor	Bogumila Pajak
Word Count	3,975

### **Assignment title**

Take a small number of words or phrases (between 2 and 5) and do a corpus study to show how they are used in similar or different ways. Choose words/phrases which are interesting in some way e.g. your students often confuse them; they cause problems for translators working with a specific language; you yourself have difficulty deciding when to use one or the other. Examples of words/phrases which have been studied in the past include: between and through; immense, enormous and massive; reason to and reason for; on the other hand and on the contrary. (You should not repeat these studies, which are mentioned as examples, but should choose different sets of words. You may choose words from a language other than English, if you have an available corpus, but make sure that a monolingual English speaker can understand your argument.)

1 Introduction.....	4
2 Corpus Linguistics .....	5
2.1 Word lists .....	5
2.2 Collocation .....	6
3 The Research Project .....	7
3.1 Approach .....	7
3.2 Research results.....	10
3.2.1 Analysis of <i>Por</i> .....	10
3.2.2 Analysis of <i>Para</i> .....	12
3.2.3 Similarities and differences .....	14
3.2.4 Detailed Case Study: <i>Por Siempre</i> and <i>Para Siempre</i> .....	15
4 Conclusion .....	16
References.....	17
Appendix A: Types of Corpora.....	18
Appendix B: Dictionary Meanings of <i>Por</i> and <i>Para</i> .....	20
Appendix C: List of Idioms using <i>Por</i> .....	22
Appendix D: Background Information and Research Findings.....	24
Appendix E: Tokens Viewed in Real Academia Española.....	27

# 1 Introduction

As a long-term student and part-time teacher of the Spanish language I have often found it difficult to explain adequately two near-synonymous prepositions, *por* and *para*. Each can be loosely translated as “for” but this is merely one of a myriad of uses for each word. The Collins English / Spanish Dictionary contains twenty-one entries for *por* and ten for *para* demonstrating the numerous ways these terms are used across the language. See Appendix B for a list of dictionary definitions. Despite appearing to share meaning at the simplest level, the two words are rarely synonymous. The following sentences show how each word could be correctly translated as *for*, demonstrating how easy it can be to confuse them.

Gracias **por** la información.

Thank you **for** the information.

Este regalo es **para** Pablo.

This gift is **for** Pablo.

In this assignment I shall explore a corpus, which is “a collection of naturally occurring language text, chosen to characterize a state or variety of language” (Sinclair, 1991:12) with the aim of understanding the similarities and differences that exist between the two terms. The principal source used in this study is the Real Academia Española (Spanish Royal Academy) Corpus consisting of 410 million recordings including both written and spoken information. The study of corpora can aid the translator or language learner enormously, allowing the researcher not only to single out specific words and analyse the syntactic frame they occur in, but also “to determine their contextual preferences, associated with other collocates” (Biber et al, 1998:24).

The following chapter introduces Corpus Linguistics as an area of study, outlining its importance in linguistic research and discussing corpus structure. Chapter three presents the research project, the approach and the findings from the study. The results section is broken down into two parts, looking at *por* and *para* individually. It includes a detailed look at how each of the terms is used resulting in a summary of use for each, and highlighting any parallels or discrepancies between the two words. The final chapter will conclude the study, describing the importance of corpora in a study of this nature.

## **2 Corpus Linguistics**

Corpus Linguistics is the study of language stored in a corpus, or electronic repository. Corpora may consist of quantities of written text or transcribed speech, which are “tagged” to facilitate targeted searching and sorting of data for linguistic study, and are often the foundation for the compilation of dictionaries. Data can be viewed as concordance lines which can be manipulated according to the user’s area of research.

Naturally corpora have not always been technical resources. Historically text was analysed manually, however, as Kennedy (1998:5) explains, despite the enormous contribution made by manual analysis, digital computers have revolutionised the ways in which scholars now work with large quantities of text.

There are many types of corpora which act as sources of evidence of natural language for linguists wishing to validate grammatical hypotheses, and in understanding what can and does occur in language. See Appendix A for more information on types of corpora.

Corpora are excellent tools for discovering patterns in language and learning about how a language works in reality, however, it is important to understand their limitations. A corpus is only as successful as the data it contains and cannot be viewed as the representation of an entire language. As Hunston (2002:23) explains, corpora allow the reader to view information as it is presented rather than as a statement of linguistic fact. In some cases, for example, oral transcripts, the results may contain grammatically incorrect information; however, this “bad” data still allows the researcher to perceive how language is used in reality. Therefore results based on corpus research are limited to what is contained therein, putting the onus on the researcher to interpret the data that is presented to them.

Let us now look at how corpora function. The next sub-sections will discuss the key concepts of statistical analysis, outlining the approaches when working with corpora.

### **2.1 Word lists**

As highlighted above, the majority of corpora are electronic, and this is the form used throughout this study. Corpus results are displayed in various ways, the simplest being frequency or word lists. Frequency is counted in “types” and “tokens”: the former describes the number of individual words, and the latter counts the total number of words regardless of how many times they are repeated. These are useful in the early stages of research to gauge the frequency of use of a particular word or expression within the corpus, but do not provide

any detail on context. Extremely common grammatical words such as *the* or *of* will yield a high frequency score. Clearly such searches must be refined in order to gain any clues as to how these words behave and to gauge the variety of meanings associated with them. Investigating how terms interact with lexical items around them can aid the researcher immensely; this is known as collocation.

## 2.2 Collocation

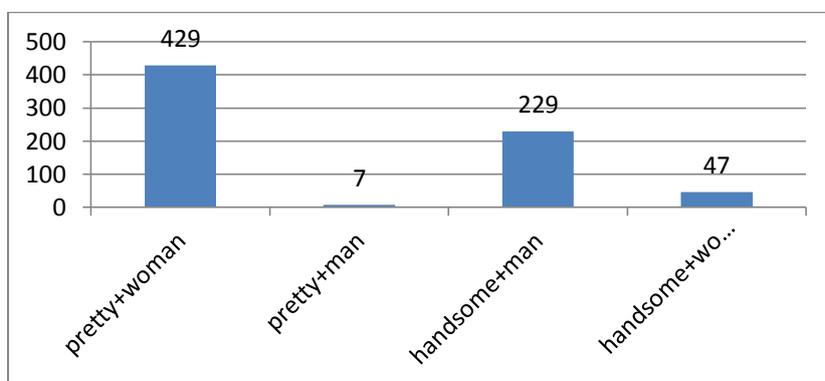
Collocation is a key concept within Corpus Linguistics, as the analysis of corpora enables linguistic patterns to be revealed in new ways.

“Collocation refers to the way that words are typically used together. In English, to borrow the well-known example from Leech (1981:17), *pretty woman* is a typical (or strong) collocation, and so is *handsome man*. This does not mean that *handsome woman* or *pretty man* is impossible, just that they are very unusual or marked.” Hatim and Munday (2004:249)

Concordance lines display the search word, or “node” in the centre, and its neighbouring words, or “collocates”, either side. These can then be sorted in a number of ways allowing collocation patterns may emerge.

To demonstrate the concept of collocation, the Bank of English (BoE) Corpus, jointly owned by HarperCollins Publishers and the University of Birmingham, was interrogated using Leech’s examples. Using the Key-Word-in-Context or KWIC view, the corpus yielded the following results, revealing the collocational strength between the four expressions.

Graph 1. Bank of English tokens demonstrating frequency of collocates



Clearly, *pretty woman* and *handsome man* are far more common than *pretty man* and *handsome woman*, although all are possible based on the information in this corpus.

There are three ways to examine collocation in the BofE: by frequency, t-score and mutual information (MI) score. The frequency count reveals words that regularly occur together but as many pairings will include common grammatical words it may be difficult to identify a true pattern. The MI score calculates the attraction between words, that is, words which commonly occur together. Pairings with the highest scores suggest that certain combinations are likely to arise. It is important to remember, however, that some couplings may be specific to a certain topic or corpus therefore will have a high MI rank but not be commonplace out of the context of the search. Finally, t-score measures the strength of the connection between the node and its collocate. This only works in one direction, unlike the MI score, therefore a combination of all three types of search may be useful to gather a balanced view of how particular lexical items function. The analysis in this study has been based predominantly on frequency lists, due to the large volumes of data analysed, and t-score, because the lexical items that follow these prepositions are usually the most significant.

Let us now look at the research into *por* and *para*. The following sections provide an overview of the uses of each word, the methodology used to research them and a full analysis of each preposition.

### **3 The Research Project**

The Collins English / Spanish Dictionary provides the first indication that the two prepositions have a number of meanings. Definitions for *por* include: *in order to, for, because of, from, by, in, on behalf of, however* and the mathematical use of *times*. It is also part of many common expressions. *Para* is defined as *for, towards, so, in order to, by* and *as*, but with little elaboration regarding usage. Sinclair (1991:7) clearly understood this predicament, highlighting the fact that despite the various meanings of words being presented in a dictionary, there is little guidance as to how one should use them.

Aiming for a thorough understanding of the two terms, the subsequent sections present the methodology deployed, followed by the results and analysis, concluding with an overview of the key similarities and differences between the two prepositions.

#### **3.1 Approach**

The Spanish Royal Academy corpus contains two corpora: CORDE (*Corpus Diacrónico del Español* or Diachronic Corpus of Spanish) containing entries from the beginnings of the

language until 1974 and CREA (*Corpus de Referencia del Español Actual* or Reference Corpus of Current Spanish) which houses more recent data from 1975 to 2004. The later corpus has been utilised as it provides a better reflection of contemporary language. In addition to the monolingual corpus two further corpora have been used in this study: the BofE and OPUS, the web based parallel corpus containing numerous sub-corpora in multiple languages.

The study began with a top level word count in CREA which consists of 90% written and 10% oral texts. The oral data originated as transcripts from television and radio broadcasts, and personal interviews. Understanding the frequency of use for each word is a valuable first step in gauging the importance of the terms, and data is then refined to focus on specific aspects of use. The CREA frequency list shows the following unsurprising results:

Table 1. CREA frequency list

Orden	Frec.absoluta	Frec.normalizada
1. de	9,999,518	65545.55
2. la	6,277,560	41148.59
3. que	4,681,839	30688.85
4. el	4,569,652	29953.48
5. en	4,234,281	27755.16
6. y	4,180,279	27401.19
7. a	3,260,939	21375.03
8. los	2,618,657	17164.95
9. se	2,022,514	13257.31
10. del	1,857,225	12173.87
11. las	1,686,741	11056.37
12. un	1,658,827	10879.95
13. por	1,561,904	10238.07
14. con	1,461,807	9711.74
15. no	1,465,503	9606.18
16. una	1,347,603	8833.36
17. su	1,288,687	8434.06
18. para	1,062,152	6962.26
19. es	1,023,803	6683.79
20. al	951,054	6234.03
21. lo	866,955	5682.77
22. como	773,465	5069.96

The first column lists the words in order of frequency, *de* (of) being the word that appears most frequently. The second shows the total number of tokens. To explain the third column in this image I defer to Biber et al:

“Normalization” is a way to adjust raw frequency counts from texts of different lengths so that they can be compared accurately. The total number of words in each text must be taken into consideration when norming frequency counts. Specifically, the raw frequency count should be divided by the number of words in the text, and then multiplied by whatever basis is chosen for norming.” (Biber, Conrad and Reppen: 1998:263)

The frequency with which both words appear is in line with the dictionary definitions, however both are clearly significant words, appearing in the top eighteen most frequently used words in Spanish. As Delbecque (1996:249-250) explains, grammatical morphemes such as these tend to offer less information as they cover such a wide variety of meanings, demanding more detailed analysis.

With over a million tokens available, the next step was to start to distil the results. CREA permits the user to filter on several levels including medium, region and theme. In order to procure a more refined group of concordance lines the results have been filtered as follows:

Table 2. Filters applied to searches

<b>Oral register</b>	<b>Written register</b>
Region: Spain	Region: Spain
Medium: Oral	Medium: Newspapers
Theme: Oral	Theme: Various Industries

The following table shows the number of tokens for each word in each register based on the filters above. It was necessary to refine the search criteria to the third level in the written register as the volume of results returned from the top two levels was still too large for analysis. The following two sub-sections provide a detailed analysis of each search word including the results gathered from CREA.

Table 3. Token counts for both words in both registers (CREA)

<b>Preposition</b>	<b>Register</b>	<b>Number of Tokens</b>
Por	Written	5,493
Por	Oral	27,839
Para	Written	4,589
Para	Oral	18,281

A sample of roughly 250 concordance lines was selected from these results for each of the four research areas, giving a total sample of over one thousand entries. As the lexical items that follow these prepositions are usually more significant than those preceding, it was decided to research collocations to one point to the right of the node. Each instance was categorised and logged in a spread sheet with the aim of building a picture of the key trends for each word. See Appendices D and E for a more detailed view of the line by line analysis.

### 3.2 Research results

The following two sections present the results of the analysis of the two prepositions. The third section aims to provide an overview of the similarities and differences between the terms and outline the linguistic rules associated with them.

#### 3.2.1 Analysis of *Por*

The results for *por* were similar across both registers with the majority of meanings appearing in both categories. In the oral corpus the idiomatic use of the word scored highest, perhaps due to the more fluid nature of the language used by the speakers. Phrases such as *por ejemplo* (for example) and *por lo mismo* (for the same reason) were scattered liberally through the transcripts of interviews and broadcasts. A list of common idioms using *por* can be found in Appendix B.

The table below shows the ways in which the preposition is used across both registers. The numbers represent actual entries in CREA and usage is ranked by the total number of tokens in descending order.

Table 4. Number of tokens for *por* by register (CREA)

<b>POR</b>			
Type	Oral	Written	Total
Idiom	92	28	120
By (agent)	23	55	78
Per	9	55	64
Because of	43	20	63
Through (place)	18	16	34
For (objective)	17	10	27

By (medium)	5	18	23
For (por+INF)	18	2	20
From	8	5	13
(With regard to) For	10	1	11
In exchange for	8	3	11
In, at (time)	4	6	10
To go after something	1	6	7
Approximately, around	0	7	7
In, by (mode)	4	0	4
(Substitution) For	2	1	3
(Representation) For, on behalf of	0	2	2
As, for	1	0	1
(Future) For	0	1	1

It is clear to see that the idiomatic usage is by far the most frequent in the oral register, with *by* (an agent) and *per* (usually as part of *per cent* or *per kilo*) holding equal top ranking in the written register. The frequent appearance of the “per” meaning may be due to the fact that the periodicals sampled were on the topic of “various industries” and there were many references to weights and volumes in the fishing and manufacturing industries. Nevertheless, this is an important meaning for the word so its presence is significant.

The use of *por* usually indicates cause or motive as demonstrated by the high scoring *because of* meaning. Based on this corpus research the most common uses of *por* could be summarised as follows.

- Set phrases: e.g. *Por lo general* (generally), *por supuesto* (of course)
- Indicating an action performed **by** someone, e.g. *Fue escrito por JK Rowling*. (It was written **by** JK Rowling)
- **Per**: e.g. *Gana 10 libras por hora* (He earns £10 **per** hour)
- Cause: e.g. *¿Por qué?* (Why?)

- **Through**, e.g. *Pasaremos por Madrid*. (We will pass **through** Madrid)
- **For**: an objective e.g. *Luchar por la patria* (To fight **for** one's country)
- Indicating a way of doing something, e.g. *Viajar por tren* (To travel **by** train)
- Preceding an infinitive to indicate a reason e.g. *Por no llegar tarde* (**So as not to** arrive late)
- In terms of evidence e.g. *Por lo que dicen* (**From** what people say)
- With regard to, e.g. *Por mí, que se va* (As far as I'm concerned, he can go)
- Acting on behalf of others e.g. *Hablo por todos* (I speak on behalf **of** everyone)
- In exchange for, e.g. *Compré el coche por 10.000 libras* (I bought the car **for** £10,000)

Less frequent uses include:

- Approximation, e.g. *Está por el Norte* (It's somewhere up North)
- Denoting the future, e.g. *Se quedarán por 15 días* (They will stay **for** a fortnight)
- To go after something or someone, e.g. *¡A por ellos!* (After them!)
- Representing substitution, e.g. *Hoy doy la clase por él* (Today I'm teaching his class **for** him)
- As or for, e.g. *Le dieron por muerto* (He was given up **for** dead)
- However, e.g. *Por difícil que sea* (However difficult it might be)
- Multiplication or measurement, e.g. *5 por 3, 15* (5 times 3 are 15)

“For” and “by” are the most frequent uses along with phrases denoting cause. However, these “simple” translations are not as straightforward as they seem as the subtleties of the situation influence the use and distinguish the meaning from *para* as we shall see below.

### 3.2.2 Analysis of *Para*

The table below shows the results for *para* in the oral and written registers. To the casual observer some definitions may look similar to those for *por* however the context differentiates them.

In addition to its prepositional use, the word *para* is also the third person singular conjugation of the verb *parar*, to stop. As this use has no relevance in this study these results have been disregarded.

Table 5. Number of tokens for *para* by register

<b>PARA</b>			
Type	Oral	Written	Total
In order to	96	138	234
(Intended, use) For	79	63	142
(Direction) Towards	24	14	38
So that	23	14	37
(Time) By	16	12	28
Why, for what purpose	8	6	14
(Interpersonal) To, towards	2	6	8
(Comparison) For	2	4	6
Only to do something	0	2	2
With enough, too much	0	0	0
Idiom	0	0	0

The trends are similar across both registers suggesting that the preposition is used consistently across both spoken and written language. In both the written and oral registers the use of *para* before an infinitive meaning *in order to* is the most common use. Naturally there are a number of ways of translating this structure as demonstrated below by the two concordance lines taken from OPUS. Despite being translated differently, the meaning conveying intent is clear in both examples.

Example 1.

p . En segundo lugar , hay que actuar internamente **para** garantizar la credibilidad . Sólo esa credibilida  
 Secondly , we must act within the Union in order to guarantee our credibility .

Example 2.

ción animal no pueda constituir una razón legítima **para** adoptar medidas comerciales restrictivas . Inclus  
 There is no reason why animal protection cannot be a legitimate reason for adopting trade restrictive measures .

The corpus research shows the following most common uses of this preposition:

- Before an infinitive, in order to, e.g. *Trabajo para comer* (I work in order to eat)
- To indicate purpose or need, e.g. *Ese regalo es para ti* (This present is **for** you)
- Movement towards, e.g. *Voy para Europa* (I'm heading towards Europe)

- So that, e.g. *Trabajo para que comas* (I work so that you eat). *Para que* always invokes the subjunctive as it usually indicates an element of doubt or that a subsequent action has not yet been completed.
- Expressing time, e.g. *Lo dejamos para mañana* (We left it until tomorrow)
- Why, for what purpose, e.g. *¿Para qué lo haces?* (Why are you doing that?)

Less frequent uses include:

- Towards: interpersonal sense e.g. *Es amable para todos* (He is friendly towards everyone)
- When making a comparison, e.g. *Para niña lo hace bien* (**For** a child she does it well)
- Training for a profession, e.g. *Estudia para dentista* (She is studying to become a dentist)
- Expressions, e.g. *Para siempre* (For ever), *Servir para nada* (To be good for nothing)

It is clear from these results that there are more differences than similarities between the two prepositions. With such a wealth of samples to study the results offered endless examples allowing a pattern to emerge, as explicated in the following sub-section.

### 3.2.3 Similarities and differences

The first noteworthy point is the degree of overlap that exists between the two prepositions. Both prepositions can precede an infinitive, they can sometimes mean “by” in a time bound situation, and each can signify “for” depending on the context. On the contrary, *por* is a key part of many idioms whereas *para* has a far weaker presence.

Although both appear before an infinitive the terms carry slightly different meanings. Where *por* is used in this way it can usually be translated into the English gerund or as “so as to”, for example, *por no incurrir en la censura* (by not engaging in censorship) whereas *para* followed by an infinitive typically means “in order to” with a sense of purpose or intent. Another frequent pattern for *por* involving the infinitive is when it precedes a verb and is followed by the same verb, for example *comer por comer* (eating for eating’s sake): this ordinarily denotes a lack of usefulness.

Each word can be translated as “by” however, the meaning alters in each case: *por* denotes an action by an agent or medium, whereas *para* is used more in a temporal sense. Although both

words are used to describe time the prepositions are not interchangeable as *para* has the sense of “for a certain time” whereas *por* denotes a sense of immediacy to a particular event, commonly signifying being at the point of doing something.

“For” is another common translation but again, the sense is different. *Por* suggests exchanging one thing for another and “with regard to”, whereas *para* yields more of a sense of movement, intended use and comparison.

Paradoxically, the similarities and differences, in many cases, are one and the same. Despite a number of shared translations, it is the context that defines which word must be used. This subtlety, in my opinion, is the main cause of confusion for many English learners. The case study below details a final example of how the two words may seem synonymous even in idiomatic use.

### **3.2.4 Detailed Case Study: *Por Siempre* and *Para Siempre***

*Por siempre* and *para siempre* both appear in the Collins English / Spanish Dictionary, with a single meaning: “for ever”. The headword definition for *siempre* is “always”. For me, *por siempre* is decidedly marked, as I would always favour the more usual *para siempre*.

However, with only a single line entry available I decided to run both searches through the corpus. *Para+siempre* yielded 5,530 tokens in 1,945 documents and *por+siempre* yielded 268 tokens in 192 documents. This validated the feeling that one was far more common than the other, confirming my belief but posed a new question: why are there two similar idioms apparently meaning the same thing?

By expanding the concordance line to “paragraph view” it is possible to see the node word in context. In this view, *por siempre* appeared in many theatrical, religious and romantic contexts, whereas *para siempre* was used across a broader range of themes. This additional information provided a fascinating insight into not only how the two expressions are used in current parlance, but also demonstrated how corpora can offer almost limitless opportunities for research.

To resolve the matter of when to use which expression, the dictionary facility within the Spanish Royal Academy Corpus provided the following definitions:

**para ~.**

1. **loc. adv.** Por todo tiempo o por tiempo indefinido. *Me voy para siempre*

**por ~.**

1. **loc. adv.** Perpetuamente o por tiempo sin fin. *Por siempre sea alabado y bendito*

Here *para siempre* is translated as “for all time or for an indefinite time” whilst the entry for *por siempre* is “perpetually or endlessly”, which ties in with the more expressive tones found in the corpus. For further reading on this topic, see Delbecque 1996: 297.

## 4 Conclusion

We have seen that *por* and *para* are united in some common definitions, albeit in different contexts. *Por* can be summarised as pertaining to a time, location, cause, reason, cost, exchange, an activity performed by someone and substitution, along with its dominant idiomatic senses. *Para* is used to signify purpose, movement towards something, effect, offering something, future dates, and professional training. Such subtle differentiations mean that even when they appear to have the same translation, in reality, they very rarely do. It is extremely uncommon for the two prepositions to be used interchangeably although it is easy to see how learners of the language could become confused.

It is clear that corpora are of significant value to the translator or language learner as research tools. The ability to study real language in context is an inestimable asset to linguists as such depth of detail is not available from traditional sources. The raw data in a corpus allows grammatical patterns to emerge, sometimes in unexpected ways. Always representing authentic language, corpus research may serve to validate a linguist’s instincts or perhaps challenge established beliefs. In the words of Hunston “It is no exaggeration to say that corpora, and the study of corpora, have revolutionised the study of language...over the last few decades”. (2002:1)

## References

- Biber, D., Conrad, S. and Reppen, R. (1998). **Corpus Linguistics: Investigating Language Structure and Use**. Cambridge: Cambridge University Press.
- Collins. (1998). **Dictionary Español>Inglés: English>Spanish**. 3rd edition. Glasgow and New York: Harper Collins.
- Delbecque, N. (1996). "Towards a cognitive account of the use of the prepositions *por* and *para* in Spanish". In Casad, E.H. (ed.) **Cognitive Linguistics in the Redwoods: The Expansion of a New Paradigm in Linguistics**. Berlin: Mouton de Gruyter. pp. 249-3180
- Hatim, B. and Munday, J. (2004). **Translation: An advanced resource book**. London and New York: Routledge.
- Hunston, S. (2002). **Corpora in Applied Linguistics**. Cambridge: Cambridge University Press.
- Kennedy, G. (1998). **An Introduction to Corpus Linguistics**. London: Longman.
- Leech, G. (1981). **Style in Fiction: A Linguistic Introduction to English Fictional Prose**. London: Longman.
- Sinclair, J. (1991). **Corpus Concordance Collocation**. Oxford: Oxford University Press.
- Sinclair, J. (2005). "Corpus and Text – Basic Principles" in Wynne, M. (ed). **Developing Linguistic Corpora: a Guide to Good Practice**. Oxford: Oxbow Books. pp 1-16. Available from Arts and Humanities Data Service (2004) [online]  
<http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm> [Accessed 20 May 2013].
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [online]. Corpus de referencia del español actual. <http://www.rae.es> [Accessed 20 May 2013]
- Parodi, G. (ed.) (2007). **Working with Spanish corpora**. London : Continuum, c2007.

## Appendix A: Types of Corpora

Type	Description	Purpose	Examples
General corpus	A wide variety of text types covering multiple subjects, registers and geographical areas.	A source of reference for the investigation of language as a whole.	Bank of English BNC Brown Corpus
Specialised corpus	A collection of texts of a particular type, often collated for a specific research project.	A representation of a distinct genre of text.	CANCODE MICASE
Comparable corpora	Texts in different languages or different versions of the same language e.g. British English and American English.	Used to compare different languages or varieties of languages in context. Provide information on similarities and difference in language.	ICE
Parallel corpora	Contain the same texts in multiple languages, either translated or simultaneously produced in more than one language.	A source of reference for translators to find possible equivalent expressions in previously translated works.	OPUS European Parliament Proceedings Parallel Corpus
Learner corpus	Texts produced by learners of a language for other students.	Used to compare the language used by students with native speakers and highlight	ICLE LOCNESS

		differences.	
Pedagogic corpus	Comprising all of the texts a learner has been exposed to during their studies.	Usually compared with a corpus of naturally-occurring English to ensure that the texts are representative of real language. Also a useful for source for revision.	Based on individual study
Historical or diachronic corpus	A collection of texts assembled over time to provide a view of language development.	Demonstrates the changes in language over time.	Helsinki Corpus
Monitor corpus	A regularly updated collection of texts used to capture current language trends.		Bank of English

Sources: (Hunston 2002:14-16 and Kennedy 1998:19-22)

**Key:**

BNC – British National Corpus

CANCODE – Cambridge and Nottingham Corpus of Discourse in English

ICE – International Corpus of English

ICLE – International Corpus of Learner English

LOCNESS – Louvian Corpus of Native English Essays

MICASE – Michigan Corpus of Academic Spoken English

OPUS – Open Parallel Corpus

## Appendix B: Dictionary Meanings of *Por* and *Para*

### *Por* PREP

- a) (+ *infin*: *para*) in order to
- b) (*objective*) for
- c) (*cause*) out of, because of
- d) (*evidence*) from
- e) (*regarding*) as far as
- f) (*medium*) by
- g) (*agent*) by
- h) (*mode*) in, by
- i) (*place*) by way of, through, via
- j) (*time*) in
- k) (*future*) for
- l) (*approximation*) around
- m) (*in exchange for*) for, in exchange for
- n) (*substitution*) for
- o) (*representation*) on behalf of, in place of
- p) (*distribution*) for, per
- q) (*maths*) times
- r) (*as*) for, as
- s) However
- t) To go for something
- u) ~ *qué* – why? ;¿~*qué*? why?

*Para* PREP

- a) (*destiny, purpose, use*) (intended) for
- b) ¿~ *qué?* why? for what purpose?
- c) ~*que* in order that, so that
- d) ~ *hacer algo* (to do something) (*purpose*) to do something, in order to do something
- e) ~ *hacer algo* (to do something) (*result*) only to do something
- f) (*with too much, enough, very*) to
- g) (*time*) by
- h) (*direction*) to, towards
- i) (*interpersonal*) to, towards
- j) (*comparison*) for

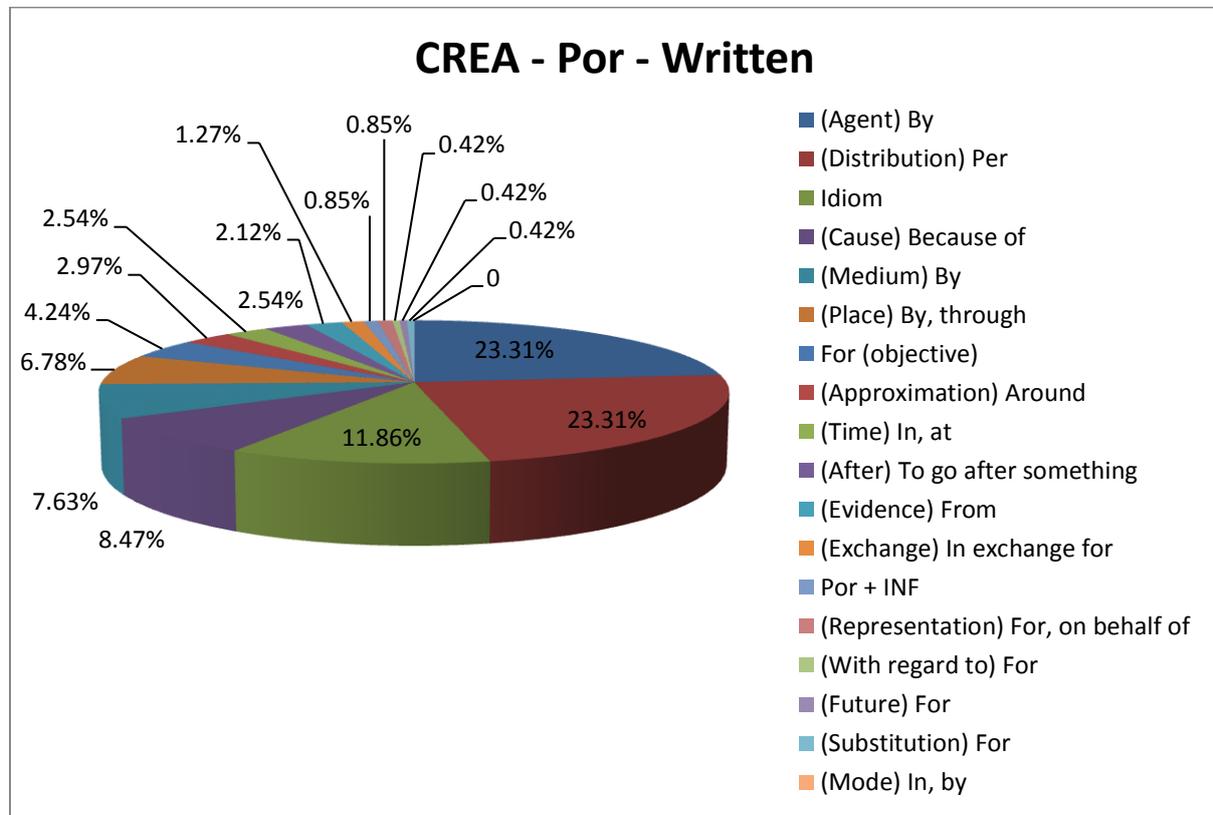
## Appendix C: List of Idioms using Por

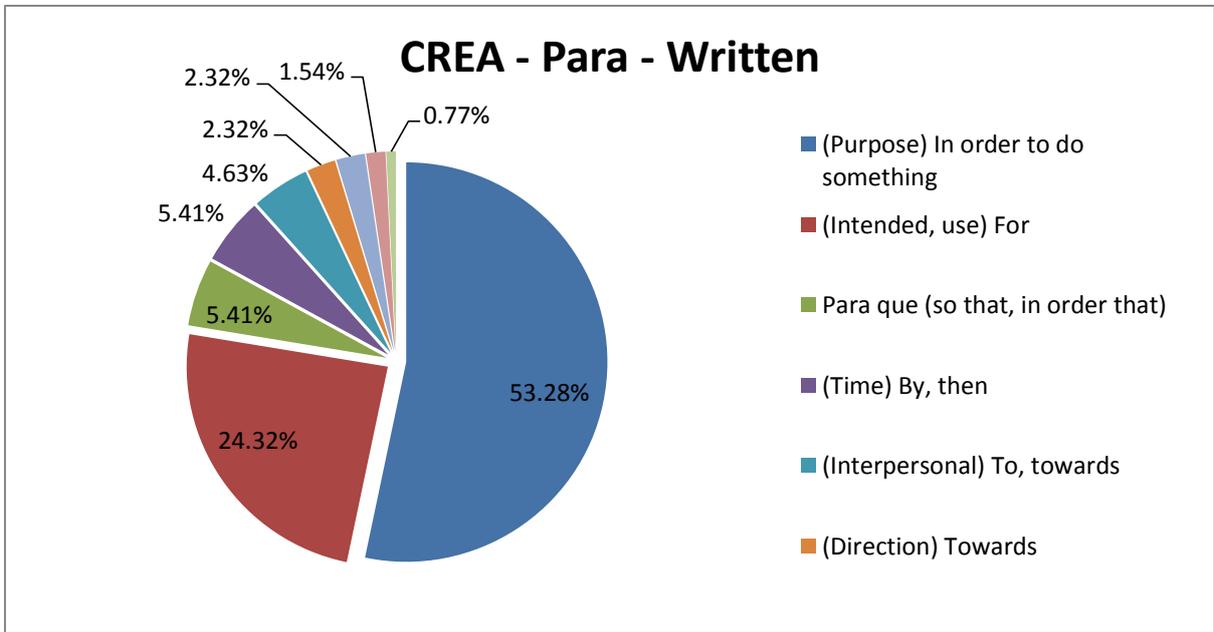
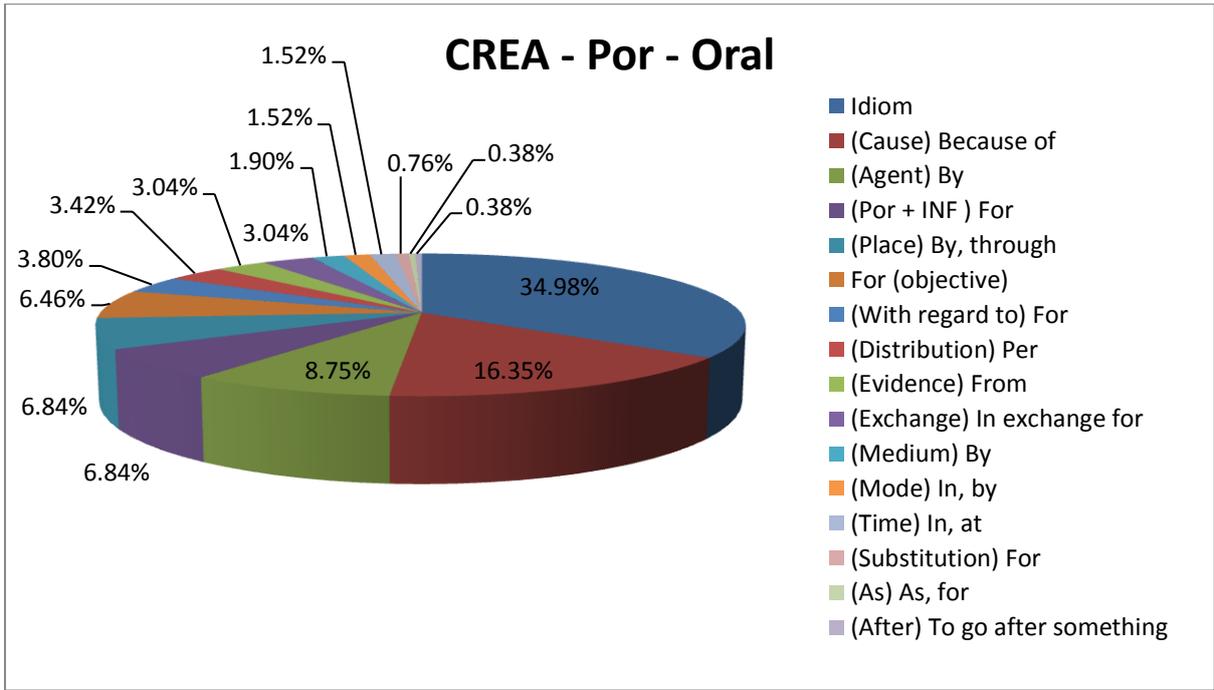
<i>Spanish</i>	<i>English</i>
Al por mayor	Wholesale
Hoy por hoy	These days, for the time being
Por ahí	Over there
Por ahora	For the time being
Por allá	Around there, that way
Por allí	Over there
Por aquel entonces	At that time
Por aquí	Around here, this way
Por avión	By plane
Por cerca	Hereabouts
Por ciento	Per cent
Por cierto	Certainly
Por completo	Completely
Por consecuencia	Accordingly
Por consiguiente	Consequently
Por cuestión de	As a matter of
Por culpa de	By fault of
Por delante	In front of
Por dentro y por fuera	Inside and outside
Por desgracia	Unfortunately
Por detrás	Behind
Por Dios	For Heaven's sake
Por ejemplo	For example
Por el estilo	Of that sort
Por encima de	On top of
Por esa época	At that time
Por escrito	In writing
Por eso	Hence
Por experiencia	From experience
Por falta de	Due to the lack of
Por favor	Please
Por fin	Finally
Por la mañana	In the morning
Por la noche	At night
Por la tarde	In the afternoon
Por lo contrario	On the contrary
Por lo cual	Whereby
Por lo demás	Furthermore
Por lo general	In general
Por lo menos	At least

Por lo mismo	For the same reason
Por lo pronto	For the time being
Por lo que a mí me toca	As far as I am concerned
Por lo tanto	Therefore
Por lo visto	Apparently
Por medio de	By means of
Por mi parte	As far as I am concerned
Por motivo de	On account of
Por otra parte	On the other hand, or again
Por otro (lado)	On the other (hand)
Por parte de	On behalf of
Por poco	Almost
Por regla general	As a rule
Por separado	Separately
Por supuesto	Of course
Por telefono	On the phone
Por todas partes	On all sides, all over, everywhere
Por todo	Throughout
Por todos lados	All over, everywhere
Por último	Ultimately
Por un lado	On one hand
Por vía	Via, by way of

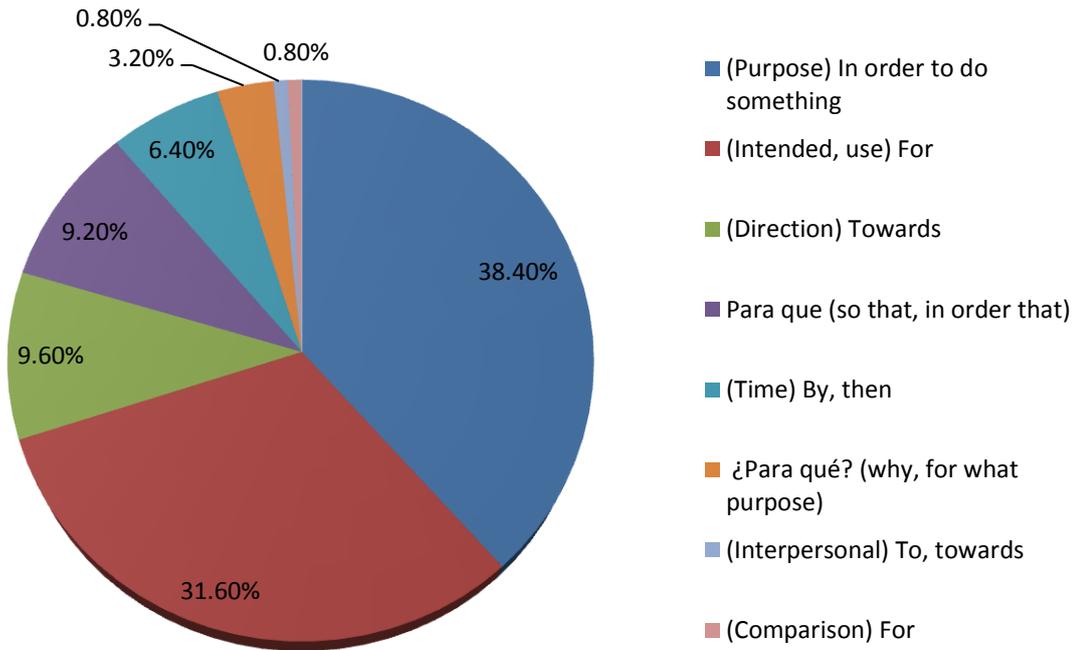
## Appendix D: Background Information and Research Findings

The following charts were created based on the data taken from CREA. Each concordance line was analysed and categorised using traditional dictionary definitions to build up a pattern showing how each word is used. The charts represent the use across each of the four categories which then allowed me to summarise my research findings.





## CREA - Para - Oral



## Appendix E: Tokens Viewed in Real Academia Española

*Por* – Written register – 5,493 tokens

REAL ACADEMIA ESPAÑOLA

Resultado de la consulta al banco de datos  
**Cómo citar el CORPUS**  
 Consulta: *por, en Periódicos, en Industrias diversas, en ESPAÑA*  
 Resultado: 5493 casos en 1148 documentos.  
[Ver estadística](#)

Filtros: Documentos  
 Ratio: 10  
 Mantener documentos (Solo para filtro sobre casos).  
[Filtrar](#)

OBTENCIÓN DE EJEMPLOS  
[Recuperar](#) Documentos Normal

Clasificación:    
 Agrupación:  Marcas:

*Por* – Oral register – 27,839 tokens

REAL ACADEMIA ESPAÑOLA

Resultado de la consulta al banco de datos  
**Cómo citar el CORPUS**  
 Consulta: *por, en Orales, en Oral, en ESPAÑA*  
 Resultado: 27839 casos en 795 documentos.  
[Ver estadística](#)

Filtros: Documentos  
 Ratio: 10  
 Mantener documentos (Solo para filtro sobre casos).  
[Filtrar](#)

OBTENCIÓN DE EJEMPLOS  
[Recuperar](#) Documentos Normal

Clasificación:    
 Agrupación:  Marcas:

*Para* – Written register – 4,589 tokens

REAL ACADEMIA ESPAÑOLA

Resultado de la consulta al banco de datos  
**Cómo citar el CORPUS**  
 Consulta: *para, en Periódicos, en Industrias diversas, en ESPAÑA*  
 Resultado: 4389 casos en 1092 documentos.  
[Ver estadística](#)

Filtros: Documentos  
 Ratio: 10  
 Mantener documentos (Solo para filtro sobre casos).  
[Filtrar](#)

OBTENCIÓN DE EJEMPLOS  
[Recuperar](#) Documentos Normal

Clasificación:    
 Agrupación:  Marcas:

*Para* – Oral register – 18,281 tokens

REAL ACADEMIA ESPAÑOLA

Resultado de la consulta al banco de datos  
**Cómo citar el CORPUS**  
 Consulta: *para, en Orales, en Oral, en ESPAÑA*  
 Resultado: 18281 casos en 778 documentos.  
[Ver estadística](#)

Filtros: Documentos  
 Ratio: 10  
 Mantener documentos (Solo para filtro sobre casos).  
[Filtrar](#)

OBTENCIÓN DE EJEMPLOS  
[Recuperar](#) Documentos Normal

Clasificación:    
 Agrupación:  Marcas: