

Received 21 April 2011 · Accepted 25 May 2011

"The golden rule of divine philosophy" exemplified in the Coruña Corpus of English Scientific Writing¹

Isabel Moskowich · Universidade da Coruña (Spain)

ABSTRACT

The present paper aims at presenting a new sub-corpus of the *Coruña Corpus of English Scientific Writing*. The corpus is a compilation of scientific texts published between 1700 and 1900 and has been compiled by strictly observing some principles that guarantee representativeness and balance. The *Coruña Corpus*, started in 2003, is organised in several sub-corpora which share mark-up language, structure and aim. Each of the sub-corpus covers one scientific discipline taking the UNESCO classification of the fields of Science and Technology as a starting point. The sub-corpus here presented, *Corpus of English Philosophical Texts (CEPhiT)*, is made up of 40 samples of texts on Philosophy and pertains to the field of Humanities so that it can be used for contrastive studies together with *CETA* (*Corpus of English Texts on Astronomy*).

Keywords: *Corpus linguistics, scientific register, Coruña Corpus, language of philosophy.*

¹ The research here reported on has been funded by the Ministerio de Ciencia e Innovación (grant number FFI2008-01649/FILO) and by the Universidade da Coruña. These grants are hereby gratefully acknowledged.

1. Introduction

The era of Modern Science, beginning sometime in the seventeenth century (Valle, 1999; Hoskin, 1999; Beal, 2004), entailed certain changes related to the way in which knowledge was transmitted. Along history knowledge of all sorts, either theoretical or practical, has been classified according to different taxonomies and has been accordingly named and renamed in different ways. The term Philosophy is defined in the *OED* as “advanced knowledge or learning, to which the study of the seven liberal arts was regarded as preliminary in medieval universities”. As a subject of study, philosophy was variously subdivided at different times. Many universities adopted a threefold division into natural, moral, and metaphysical philosophy. Depending on the institutions, *philosophy* could also include other elements or subjects that were necessary for the degree of M.A. During the eighteenth century this use of the term declines (*OED*) and Natural Philosophy was soon replaced by others such as Biology in the following century.

Philosophy is not a new field of science. On the contrary, as early as the first half of the fourteenth century the term was used to refer to the branch of knowledge that dealt with the principles of human behaviour; the study of morality and ethics as the example below from the *Ayenbite* perfectly shows:

Det is þe heȝeste wyt of man, wel to knawe his sseppere and him louie. Vor
wyþoute þise filosofie, alle oþre wyttes ys folye. (1340 *Ayenbite* (1866) 251)

From the understanding of the Universe, of everything surrounding human beings and human beings themselves, and having scientific knowledge as its subject of study, philosophy began to be conceived of as the rational thought that opposed any kind of knowledge revealed and subject to religious beliefs. However, it is in the eighteenth century that this last meaning was definitely adopted due to French influence². Only a part of this changing trend of mind can be seen in the samples compiled in the *Corpus of English Philosophy Texts* (*CEPhiT*), one of the subcorpora of the *Coruña Corpus of English Scientific Writing*³.

Philosophy, as any other discipline, has had its writing conventions. Such conventions may have not always been overtly expressed but scholars certainly know “how to write”, very often basing upon what others had done before. Changes in the way in which philosophical knowledge was transmitted certainly operated along history from the scholastic Middle Ages (when knowledge was a divine gift) to the very moment in which the rationalistic and empiricist movements advanced over Europe. During the Modern English period it is basically prescriptive tendencies we are going to find more or less overtly (Valle, 1999; Moessner, 2001) whereas nowadays the approaches adopted are more varied. We can still find some prescriptive viewpoints behind style sheets for prospective authors in scientific publications, but, at the same time, we can also find a more descriptive objective in the interest of corpus linguistics scholars. As a methodology, corpus linguistics offers an excellent opportunity to quantify findings and reach more reliable conclusions regarding the evolution of such conventions.

Like philosophy, every scientific field is likely to have its own traditions and restrictions in terms of writing. That is why the *Coruña Corpus of English Scientific Writing* (*CC*) is formed by a collection of several sub-corpora each of them containing samples of texts published between 1700 and 1900 and each

² One of the authors in *CEPhiT*, Burke (1770), uses the term precisely in this sense.

³ The *Coruña Corpus* (*CC*) is a long-term project that will be coming out little by little, its first part being the *Corpus of English Texts on Astronomy* (*CETA*) and the second the *Corpus of English Philosophy Texts* (*CEPhiT*).

corresponding to a different scientific discipline. After Astronomy, Philosophy is the second discipline selected for the compilation of scientific texts and, consequently, the second sub-corpus in the *Coruña Corpus*.

As already mentioned (Moskowich, 2011) the *CC* is intended to complement other corpora which share with it their diachronic nature and their specificity. Similar computerised corpora include *ARCHER*, *The Lampeter Corpus of Early Modern English Tracts*, *Middle English Medical Texts (MEMT)* and *The Helsinki Corpus of English Texts*. From a chronological perspective, all sub-corpora included in the *Coruña Corpus* cover a gap of 160 years after the scope of the *Lampeter Corpus*, (1640-1740). As for domain, *CEPhiT* (and all the parts of *CC*) is more specific than the *Lampeter Corpus*, which represents Science in general, and also more specific than the *Helsinki Corpus of English Texts*, which was not conceived of as a ‘specific’ corpus. From the point of view of the content, although both *CC* and *ARCHER* contain samples of scientific writing, they do not collide either since the latter has material extracted from the *Philosophical Transactions* whereas the former offers a representation of longer formats and different genres. Our aim in building the sub-corpus described here is that it will allow scholars to explore the negotiation of knowledge between authors and audience as well as to study the changing conventions as presented in different linguistic strategies.

Hyland (1998, p. 18) claims that the linguistic practices for expounding scientific knowledge are historical artefacts dating from the 1600s. In fact, contemporary authors discussed the necessity of establishing new discursive rules as well as a new textual organisation. Boyle and other members of the Royal Society proposed separating the exposition of hypotheses and of proven facts (Allen, Qin and Lancaster, 1994; Gotti, 1996) thus giving place to new formats. Formats, genres, reflect new modes of knowing.

The texts compiled in the *CC* reflect not only a specific use of English but a particular way of doing science in the modern period. In the case of texts on philosophy, although more timidly than in other disciplines, we can observe the importance of observation of phenomena as well as the use of the deductive method replacing authoritative statements. However, we must not think of an abrupt, sudden breach with the scholastic tradition, but rather a gradual abandonment of medieval practices. In fact, there are texts on moral philosophy

in our corpus that are deeply indebted to the scholastic tradition. It is also true that, at the other end, we have compiled samples like the one by Mary Astell or Wollstonecraft, which show a radically different way of thinking and champion more radical ideas. No doubt the Reformation accelerated the movement away from Scholasticism and favoured the opening to new approaches. This is the case of some other authors in *CEPhiT* such as Greene (1727), who are influenced by the increasing importance of observation and experimentation to confirm facts. As was the case with other sciences, the relationship between philosophy and society is also manifested in several works included in *CEPhiT*. The work *Philosophical principles of natural religion: containing the elements of natural philosophy, and the proofs for natural religion, arising from them* by George Cheyne (1705) constitutes an example of this need to have evidence of things instead of the sole word of ancient wise men.

2. Compilation principles in *CEPhiT*

Although from our present-day perspective it may seem that boundaries between scientific disciplines are clear, the truth is that there are always some overlapping or fuzzy areas. This constitutes a basic difficulty in the selection of representative samples of scientific language, mainly when it is not present-day science we are dealing with. Hence, instead of designing our own taxonomy of disciplines when compiling the *CC*, we resorted to the one published by UNESCO in 1988 as a starting point.

At the moment of writing this paper, only the disciplines in Table 1 have been chosen for compilation allowing for a re-allocation of some of them since there is no exact correlation between the present-day conception of scientific fields and the one existing in the period under discussion here. Today's increased specialisation in science leads to a degree of branching that we do not intend to reflect in our corpus compilation. Table 1 illustrates the distribution of disciplines proposed for the *Coruña Corpus of English Scientific Writing* and the different corpora being compiled:

Field	UNESCO disciplines	Coruña Corpus Discipline	Sub-corpus
Natural Sciences	Astronomy	Astronomy	CETA
	Biology	Life Sciences	CELiST
	Botanics		
	Zoology		
	Physics	Physics	CETePH
Humanities	Biochemistry, Chemistry	Chemistry	CECheT
	Philosophy	Philosophy	CEPhiT
	History of science and technology		
	History	History	CHET
	Archaeology, Numismatics, Palaeography, Genealogy		
	Modern languages	Linguistics	CETeL

Table 1. Disciplines and subcorpora in the CC

Each sub-corpus within the *CC* is devoted to one of the disciplines shown in the table. As already mentioned, the second of these disciplines, philosophy, has been compiled under the name *Corpus of English Philosophy Texts* (*CEPhiT*). This includes samples of texts on modern philosophy together with one metadata file per sample. As was the case with the first corpus compiled, *CETA* (*Corpus of English Texts on Astronomy*). Each text file contains a sample of around 10,000 words of prose text from which all items non-analysable from a linguistic point of view have been excluded.

As regards metadata files, they contain information on the life and socio-historical context of the author, and on the characteristics of the text compiled. The interrelations with other texts in the *CC* are mentioned. Factors relating to extra-linguistic variables such as age, sex, place of education and genre/text-type of each of the compiled samples are also part of the information in the metadata files (Moskowich and Crespo, 2010).

Two of the basic ideas behind the whole project are the concepts of balance and representativeness and they have been also taken into account for the compilation of *CEPhiT*. Since all the information regarding the principles of compilation applied to the *Coruña Corpus* and some of its sub-corpora have been already dealt with elsewhere (Moskowich 2004, 2007, 2009, 2010, 2011; Crespo and Moskowich, 2008; Moskowich and Parapar 2008; Moskowich and Crespo 2010) they will not be discussed here. As is the general practice for the project we have tried to compile to 10,000 words text files per decade, so that each of the centuries represented in *CEPhiT* contains approximately 200,000 words. May it suffice to say that *CEPhiT* shares the structure and mark-up conventions used for the whole project which have proved to be extremely useful and valid for linguistic research since the sampling methods avoid idiosyncrasies and interference due to translation.

Not only writing conventions, but communication in general, has changed in the last couple of decades. The irruption of the Internet in the academic world has come to be another scientific revolution probably comparable to that taking place at the beginning of the Modern Era. Under the slogan of “Publish or perish” modern science is a written one so that any unpublished idea simply does not exist. But it is not only the medium used to transmit science that has changed (electronic journals replacing paper ones) but the way in which words, structures and conventions in general are used in each scientific discipline. Writing practices have been subject to similar changes which are not necessarily random since such practices are historical artefacts dating back from the 1600s (Hyland, 1998, p. 18) and, as such, they are subject to discursive rules to meet the moment’s requirements (as Boyle and his colleagues did when they proposed to separate the exposition of hypotheses and that of proven facts). Different types of readership appeared as a result of the different discursive patterns and the negotiation of knowledge that may be observed from the seventeenth century onwards. The *Coruña Corpus: A Collection of Samples for the Historical Study*

of English Scientific Writing includes text samples belonging to different domains in which language and discourse can be seen to be used on the part of scientists as a way of negotiating knowledge.

The principles governing *CEPhiT* are those governing the *Coruña Corpus of English Scientific Writing*. The main principles of representativeness and balance are behind the compilation of samples of Philosophy texts as they are behind the whole *Coruña Corpus* project (McEnery and Wilson, 1996; Biber et al., 1998, pp. 251-253). We have included prose texts only and all of them edited and printed. As with the other sub-corpora, we have used two samples every ten years and have resorted to first editions whenever possible. When this was not possible, and assuming that language change can be observed within 30-year periods (Kytö, Rudanko and Smittenberg's 2000, p. 92), we have chosen those that were published within less than thirty years from the date the work was first published.

Our previous experience with *CETA* and the different pilot studies we have published using it have demonstrated that 1,000-word samples are not really enough for the study of variation within the scientific register (Biber, 1993) mainly because many of the samples contained in our corpus are not technical or scientific in the same sense as those we can find in present day English and the scientific register was not as standardised as it is nowadays.

We have tried to collect extracts from different parts of the works sampled excluding prefaces or dedications which are not scientific in their content. Introductions, central chapters and conclusions are more or less equally represented. We have also tried to compile a similar number of words and samples for each century. Therefore, we have obtained a total of 200,022 words for the eighteenth century part and 201,107 for the nineteenth-century one. However, not all genres/text types or other variables such as sex or place of education of the author are equally represented. Table 2 below shows the overall distribution in terms of word counts:

Eighteenth century	200022
Nineteenth century	201107

Table 2. Words in CEPHiT

Selection has often been determined by the availability of texts although in the last few years more and more copyright free images of them can be found.

3. Time-span represented

The time-span covered by *the Coruña Corpus* in general and by *CEPbiT* in particular is based on extra-linguistic considerations. As compilers, we have used landmarks in scientific thought rather than landmarks in language change to establish the dates limiting our text selection. In turn, we must bear in mind that changes in scientific thought imply changes in the way in which knowledge is conveyed and, therefore, in linguistic discourse (Moskowich & Parapar, 2007). *CEPbiT* has been compiled by selecting samples of texts published between 1700 and 1900 (that is to say, the Modern English period⁴). The time-span chosen is directly related with the outburst of the scientific revolution, the foundation of the Royal Society of London and, of course, with the publication of the basic guidelines on how to present scientific works to the members of the Society with the ideas of clarity and simplicity behind it all.

CEPbiT earliest texts date back to 1700 (Mary Astell) and 1705 (George Cheyne), a moment at which the old epistemological patterns of Scholasticism are suffering a radical transformation (Taavitsainen and Pahta, 1997) and, therefore, a moment we considered ideal to start our compilation. This starting point in our time-span coincides also with the new inductive method that one of the authors we have included in *CEPbiT*, John Stuart Mill (1845), systematised. Empiricism also promoted the development of Science outside Universities for the first time. These social and epistemological changes brought about the search for a new language to transmit science (Swales, 1990), a representative sample of which we have tried to compile in *CEPbiT*.

⁴ Alternative dates such as 1660, 1725, 1776 or even 1800 (Görlach, 1994, p. 22) have been pointed as the borders between early and late Modern English. It is true that from the 18th century English scholars tend to use prescribed forms regardless of their dialectal origin. Regional and social dialects are considered inferior (Freeborn, 1992, p. 180). Besides, it is in the eighteenth century that we observe the outburst of all sorts of pamphlets, grammars and articles aiming at linguistic improvement.

Our upper limit in time is 1900 due to the several events occurred around the turn of the century and which have proved really important for the History of Science. Among them, the discovery of the electron by J.J. Thompson in 1896, the crisis of the grounds of mechanical physics announced by Mach, Kirchhoff or Boltzmann in this same year, Planck's announcement of quantum mechanics, or Einstein's publication (be it his idea or Mileva Maric's) of a paper proposing what is today called the Special Theory of Relativity in 1905 have been already mentioned elsewhere (Moskowich and Crespo, 2010; Moskowich, 2011). All these discoveries, as had happened in the seventeenth century, were accompanied by a need to change the discursive patterns of science announced by Thomas Huxley at the 1897 International Congress of Mathematics.

4. Authors represented in *CEPhiT*

Our selection of samples is governed, as already explained, by the principles of the *Coruña Corpus*. Our sampling method and availability render, therefore, different results for the different sub-corpora. Table 3 below lists all the authors contained in *CEPhiT* as well as the title of their work and the year in which it was published.

Year	Author	Title of work sampled
1700	Astell, Mary	Some reflections upon marriage. London: John Nutt.
1705	Cheyne, George	Philosophical principles of natural religion: containing the elements of natural philosophy, and the proofs for natural religion, arising from them. London: printed for George Strahan.
1710	Dunton, John	Athenianism: or, the new projects of Mr. John Dunton.
1717	Collins, Anthony	A Philosophical Inquiry Concerning Human Liberty.

Year	Author	Title of work sampled
1727	Greene, Robert	The principles of the philosophy of the expansive and contractive forces. Or an inquiry into the principles of the modern philosophy, that is, into the several chief rational sciences, which are extant. In seven books. By Robert Greene. Cambridge: printed at the University-Press, by Cornelius Crownfield, and are to be sold by him, E. Jefferys, and W. Thurlbourne booksellers in Cambridge, and by J. Knapton, R. Knaplock, W. and J. Innys, and B. Motte, London, 1727.
1730	Kirkpatrick, Robert	The golden rule of divine philosophy: with the discovery of many mistakes in the religions extant.
1733	Balguy, John	The law of truth: or, the obligations of reason essential to all religion. To which are prefixed, some remarks supplemental to a late tract; entitled, Divine rectitude.
1736	Butler, Joseph	The analogy of religion, natural and revealed, to the constitution and course of nature. To which are added two brief dissertations: I. Of personal identity. II. Of the nature of virtue. Dublin: Printed by J. Jones. For George Ewing, 1736.
1740	Turnbull, George	The principles of moral philosophy. An enquiry into the wise and good government of the moral world: in which the continuance of good administration, and of due care about virtue, for ever, is inferred from present order in all things, in that part...London. Printed for J. Noon.
1748	Hume, David	Philosophical essays concerning human understanding. By the author of the essays moral and political.
1754	Bolingbroke, Henry	The Philosophical Works of the late Right Honorable Henry St. John, Lord Viscount Bolingbroke. Published by David Mallet, Esq; Volume I. London: printed in the year, 1754.

Year	Author	Title of work sampled
1755	Hutcheson, Francis	A system of moral philosophy, in three books. Glasgow, printed and sold by R. and A. Foulis.
1764	Reid, Thomas	An inquiry into the human mind, on the principles of common sense. Edinburgh: printed for A. Millar, London, and A. Kincaid & J. Bell, Edinburgh.
1769	Ferguson, Adam	Institutes of moral philosophy. For the use of students in the college of Edinburgh. By Adam Ferguson, LL.D. Edinburgh: printed for A. Kincaid & J. Bell, 1769.
1770	Burke, Edmund	Thoughts on the cause of the present discontents. Dublin. [Dublin]: London: printed for J. Dodsley. Dublin: reprinted for G. Faulkner, J. Exshaw, H. Saunders, W. Sleater, D. Chamberlaine, [and 8 others in Dublin], 1770.
1776	Campbell, George	The philosophy of rhetoric. London: printed for W. Strahan; and T. Cadell; and W. Creech at Edinburgh, 1776.
1783	Macaulay, Catharine	Treatise of the immutability of moral truth. London: Printed by Hamilton, Jun.
1790	Smellie, William	The philosophy of natural history.
1792	Wollstonecraft, Mary	Vindication of the Rights of Woman.
1793	Crombie, Alexander	An essay on philosophical necessity. London: printed for J. Johnson, 1793.
1801	Belsham, Thomas	Elements of the philosophy of the mind and of moral philosophy: to which is prefixed a compendium of logic.
1810	Stewart, Dugald	Philosophical Essays.
1811	Kirwan, Richard	Metaphysical Essays; containing the principles and fundamental objects of that science.

Year	Author	Title of work sampled
1820	Brown, Thomas	Lectures on the philosophy of the human mind.
1824	Phillips, Sir Richard	Two dialogues between an Oxford tutor and a disciple of the common-sense philosophy: relative to the proximate causes of material phenomena.
1830	Mackintosh, Sir James	Dissertation on the progress of ethical philosophy, chiefly during the seventeenth and eighteenth centuries.
1835	Hampden, Renn Dickson	A course of lectures introductory to the study of moral philosophy: delivered in the University of Oxford, in Lent Term, 1835, London.
1838	Powell, Rev. Baden	The connexion of natural and divine truth: or, the study of the inductive philosophy, considered as subservient to theology. The Saturday Mazine.
1845	Mill, John Stuart	An examination of Sir William Hamilton's philosophy and of the principal philosophical questions discussed in his writings.
1846	Combe, George	Moral philosophy, or the duties of man considered in his individual, domestic and social capacities.
1855	Lyall, William	Intellect, the Emotions, and the Moral Nature.
1860	Slack, Henry James	The philosophy of progress of human affairs.
1862	Simon, T. Collyns	On the Nature and Elements of the External World: Or, Universal Immaterialism, Fully Explained and Newly Demonstrated.
1866	Mansel, Henry Longueville	The philosophy of the conditioned: comprising some remarks on Sir William Hamilton's philosophy, and on Mr. J.S. Mill's examination of that philosophy.

Year	Author	Title of work sampled
1874	Woodward, Thomas Best	A treatise on the nature of man, regarded as triune; with an outline of the philosophy of life. London: Hodder & Toughton.
1874	Balfour, Arthur James	A defence of philosophic doubt.
1885	Seth Pringle- Pattison, Andrew	Scottish philosophy: a comparison of the Scottish and German answers to Hume.
1890	Mackenzie, John Stuart	An Introduction to Social Philosophy. Glasgow: J. Maclehose & Sons.
1893	Bonar, James	Philosophy and political economy in some of their historical relations.
1898	Hodgson, Shadsworth Hollway	The metaphysic of experience.

Table 3. Authors in CEPbiT

In the following paragraphs, all the extra-linguistic variables surrounding these authors and information relevant to the study of their way of writing will be presented since they delimit the nature of our corpus.

5. Genres and text types

Although academic writing may, at first sight, seem rather homogeneous or monolithic, variation can be seen to operate within each subject, among other things, to text type (the internal characteristics of texts) and to genre (as a way of socialising and, therefore, with certain external functions) (García-Izquierdo & Montalt, 2002). Therefore, texts belonging to the same domain are not necessarily similar but they show differences depending on text type/genre as

many different studies have already demonstrated (Nwogu, 1990; Myers, 1990; Bhatia, 1993).

Our classification of samples for the corpus is not based on linguistic features exclusively but on epistemological features and social factors too. As compilers, we have tried to include extracts from different epistemological levels which can be roughly compared to the three we can find nowadays (Fortanet et al., 1998):

- a) Highest epistemological level typical of research articles and *abstracts*.
- b) High epistemological level (*abstracts* in *abstracting journals* and informative scientific articles).
- c) A medium epistemic level for specialised non-academic articles.

Since they are socially determined, it is certainly difficult to delimit or define genres. In the case of the *Coruña Corpus*, we are dealing with paragenres, that is to say, genres belonging to one professional community (Monzó, 2002, p. 141). The taxonomy applied in *CETA* (*Corpus of English Texts on Astronomy*, based on Görlach, 2004) is the one to be found here as well.

The classification of samples according to genres offered some difficulties already discussed in earlier works (Moskowich, 2011) and will not be dealt with here in any detail. It is worth mentioning, though, that we have identified in *CEPhiT* a lesser number of genres than in other disciplines such as astronomy or life sciences. Table 4 below represents the number of samples compiled belonging to each genre:

Genres in <i>CEPhiT</i>	Samples
Treatise	22
Essay	10
Textbook	1
Lecture	5
Dialogue	1
Article	1

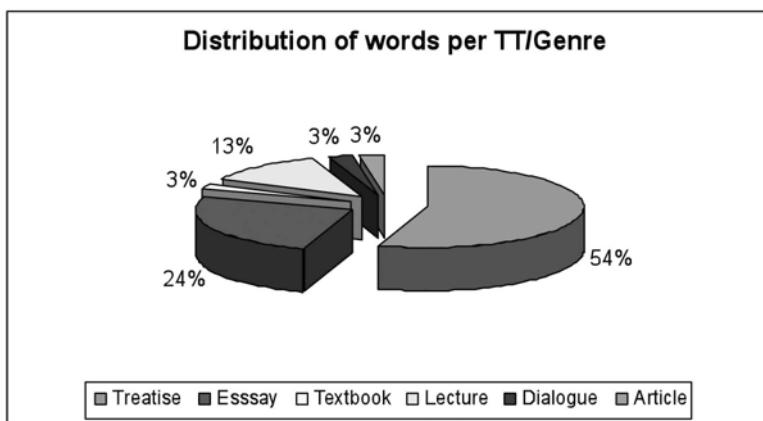
Table 4. Genres in Philosophy Texts

The ascription of a sample to one or another genre is arguable. As Fowler (1982, p. 41) puts it “genres may be considered as family members who are related in various ways without necessarily having any single feature in common by all”. The outline of the genres (understood as functional text categories) found in eighteenth- and nineteenth-century texts in *CC* does not coincide with the ones found in *CEPhiT*. Philosophy texts, as regards our samples, seem to be limited to six types whereas for other disciplines, such as Astronomy, we have found eight. This, once more, may be caused by the restrictions imposed by subject-matter: certain disciplines or domains seem to prefer just a few types of texts whereas others manifest themselves in a more varied way.

Contrary to what happens in *CETA*, textbook is a genre very scarcely represented with only 1 sample. Authors writing about Philosophy during the Modern period seem to prefer treatise by large as Table 4 above shows. Essays come next, which points to a real liking for more formal genres. We have included samples representing other categories: the informative function is the commonest, but the instructive, and even the entertaining functions are not uncommon either. Therefore, Lecture, Dialogue and Article can be found too. Görlach (2004, p. 88) has been taken into account for this more generic text-types taxonomy, since all these categories were already in use when our authors published their texts⁵.

A careful examination not only of the samples, but also of whole texts and their prefaces allowed us to conclude that *CEPhiT* contains samples of the six genres/text-types already mentioned above and whose proportions are shown in Graph 1:

⁵ Some of these formats were already used at the end of the fourteenth century. Such is the case of treatise, first recorded with its meaning of “a book or writing which treats of some particular subject” (*OED*). In modern times, however, the meaning includes also the idea of a book “containing a formal or methodological discussion or exposition of the principles of the subject”. Twenty-two of the forty samples contained in *CEPhiT* belong to treatises.

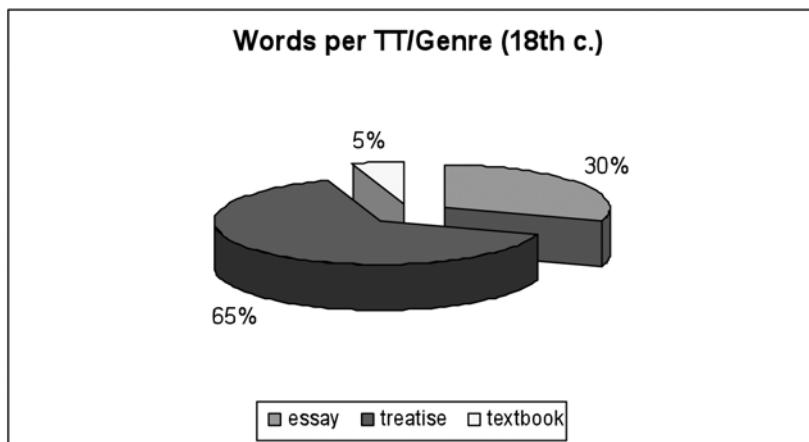


Graph 1. Proportion of words per genre

Graph 1 above illustrates the different genres gathered in all *CEPhiT* samples where 54% corresponds to treatise. However, on closer inspection, one can see that such distribution is not identical in the two centuries compiled. The tables and graphs below show the differences to be found that reflect the external reality affecting text production in the field.

Genre	Number of words
Essay	60213
Treatise	129745
Textbook	10064

Table 5. Words per genre in eighteenth century *CEPhiT*

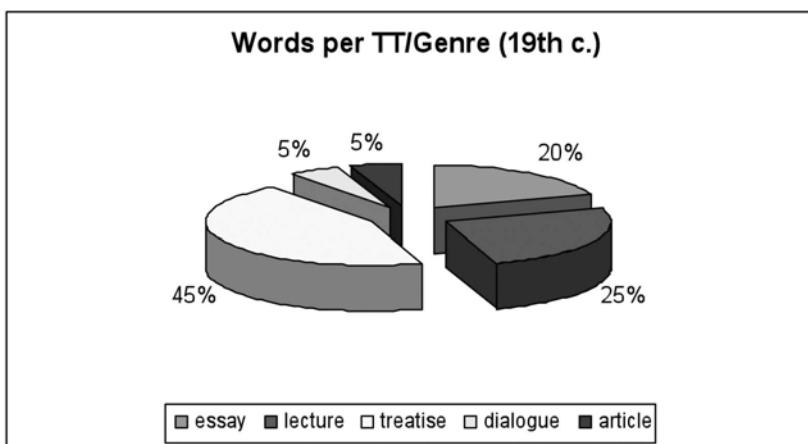


Graph 2. Words per genre in 18th c. Philosophy Texts

Both tables 5 and 6 and the corresponding graphs (2 and 3) show the existence of a wider variety of genres used in the nineteenth century as compared with those used by authors in the preceding century. This may be related to the fact that Philosophy, as a branch of scientific knowledge, was felt as something deserving dissemination at different social and cultural layers.

Genre	Number of words
Essay	40251
Lecture	50307
Treatise	90393
Dialogue	10084
Article	10072

Table 6. Words per genre in 19th c. CEPHiT



Graph 3. Proportion of words per genre in 19th c. philosophy texts

According to the data extracted from *CEPhiT* metadata files, the nineteenth century testifies to the opening of philosophy to a larger readership and does so by resorting to a wider range of genres.

6. Sex

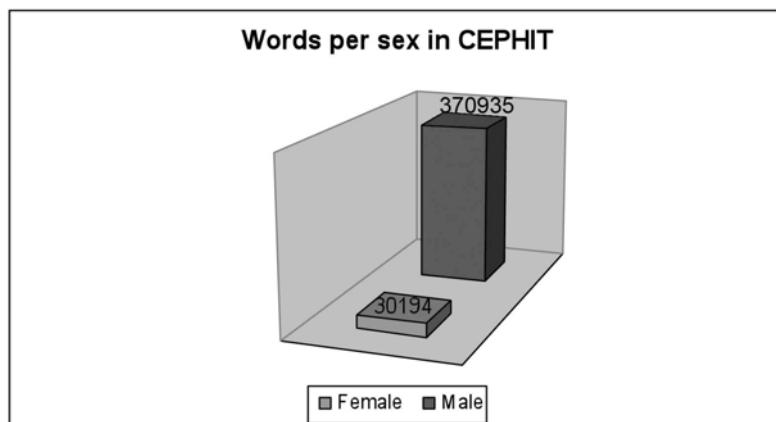
As could be expected, not many records written by women can be regarded as philosophical texts in the eighteenth and nineteenth centuries. The whole sub-corpus contains a total of three samples representing female writing, which represents only the 8% of all the words in it as Table 7 clearly illustrates. These women are Mary Astell (1700), Catharine Macaulay (1783) and Mary Wollstonecraft (1792).

Sex	Words in <i>CEPhiT</i>
Female	30194
Male	370935

Table 7. Number of words attending to sex

Women are seldom mentioned in books about the History of Science or in Biographical Dictionaries. Public female activity was not common in certain spheres of life and publishing in general, but publishing works on Philosophy in particular, was one of these uncommon activities. However, it must be admitted that other fields of science were regarded as even more masculine than philosophy. Women's work was often not taken seriously (Herrero, 2007, p. 75). Excluded from official science, the means women had to learn was by reading, by listening to other women, from mothers to daughters and, occasionally, by listening to men. Female authorship is difficult to establish. In certain fields of knowledge such as Astronomy, women did not sign their own works, as is the case of the Catalogue of Stars by German female astronomers in the seventeenth century. Although women participated intensively in science, their access to study and scientific work was limited to the role of mere assistants. Some scientific institutions, in fact, did not admit the first women until the second half of the twentieth century.

CEPhiT reflects this scarcity of overt female activity. No women writing philosophy in the nineteenth century have been included in *CEPhiT*; thus the 30194 words contained as female writing belong to the moments prior to the beginning of the suffragist movement. The information provided in Table 7 is more clearly shown in Graph 4 where the absence of women from the world of philosophical knowledge is made evident.



Graph 4. Words written by female and male authors

7. *CEPhiT* on the map

The *Corpus of English Philosophy Texts* serves as an instrument for the study of the evolution of English scientific writing in time as well as for that of variation depending on other different sociolinguistic variables, geographical origin being one of them. In order to provide data for this type of studies, we have resorted, when possible, to texts by authors whose linguistic habits could be traced⁶.

We have selected English-speaking authors writing in English, avoiding any translations even those made by the authors themselves. By geographical distribution of authors we refer not necessarily to the places where they were born but to those where they were educated, and where they acquired the linguistic habits to be found in their writings as sampled in *CEPhiT*.

Table 8 below shows the distribution of authors according to the geographical variable. As can be seen, no American authors have been included in this sub-corpus though they abound in other parts of *CC*. As a small-scale mirror of a reality, it was Europe that was producing most works on philosophy, whereas the North America had lived a convulsive eighteenth century and was, in the nineteenth, more worried about the practical application of scientific advances than about metaphysical ones. We have not been able to find information about the places where one of the authors (Kirkpatrick, 1730) was educated.

Year	Author	Place of Education
1700	Astell, Mary	England
1705	Cheyne, George	Scotland
1710	Dunton, John	England
1717	Collins, Anthony	England
1727	Greene, Robert	England

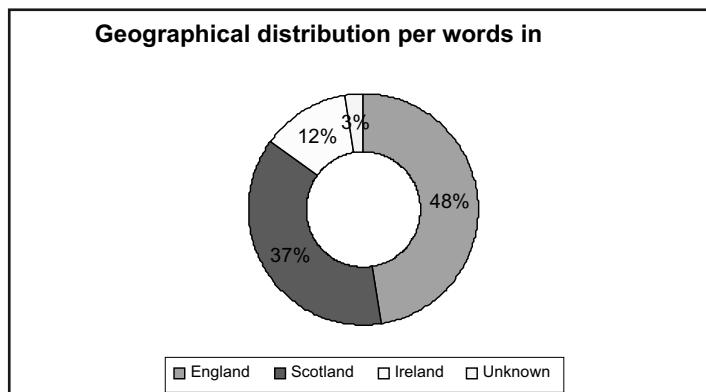
⁶ Biographical information has been compiled in a set of metadata files accompanying the samples themselves. The structure of such files is such that information inside them is also searchable by our search engine *Coruña Corpus Tool (CCT)*.

1730	Kirkpatrick, Robert	Unk
1733	Balguy, John	England
1736	Butler, Joseph	England
1740	Turnbull, George	Scotland
1748	Hume, David	Scotland
1754	Bolingbroke, Henry	England
1755	Hutcheson, Francis	Ireland/Scotland
1764	Reid, Thomas	Scotland
1769	Ferguson, Adam	Scotland
1770	Burke, Edmund	Ireland
1776	Campbell, George	Scotland
1783	Macaulay, Catharine	England
1790	Smellie, William	Scotland
1792	Wollstonecraft, Mary	England
1793	Crombie, Alexander	Scotland
1801	Belsham, Thomas	England
1810	Stewart, Dugald	Scotland
1811	Kirwan, Richard	Ireland
1820	Brown, Thomas	England/Scotland
1824	Phillips, Sir Richard	England
1830	Mackintosh, Sir James	Scotland
1835	Hampden, Renn Dickson	England
1838	Powell, Rev. Baden	England
1845	Mill, John Stuart	England/Scotland
1846	Combe, George	Scotland
1855	Lyall, William	England

1860	Slack, Henry James	England
1862	Simon, T. Collyns	Ireland/England
1866	Mansel, Henry Longueville	England
1874	Woodward, Thomas Best	Ireland
1874	Balfour, Arthur James	Scotland/England
1885	Seth Pringle-Pattison, Andrew	Scotland
1890	Mackenzie, John Stuart	Scotland
1893	Bonar, James	Scotland
1898	Hodgson, Shadworth Hollway	England

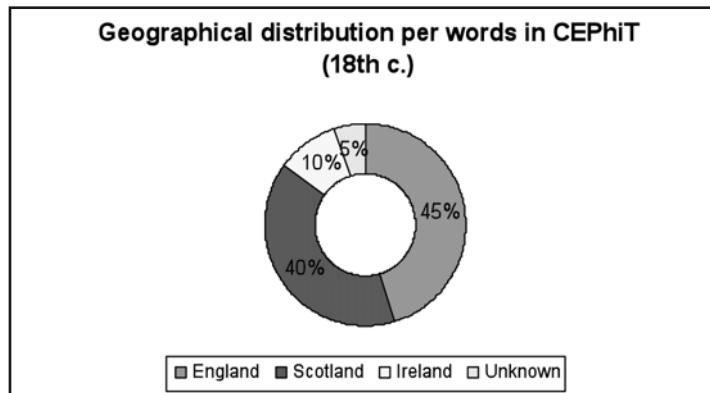
Table 8. Geographical origin of authors in *CEPhiT*

Graph 5 below illustrates more clearly the information contained in Table 8. It is evident that, once more, some external conditions have played an important role for this reality: philosophical movements have been more influential in and from Europe, American scientific writing outstanding in other fields. An overview of the different places where the authors contained in *CEPhiT* learned to write is the one offered in Graph 5.

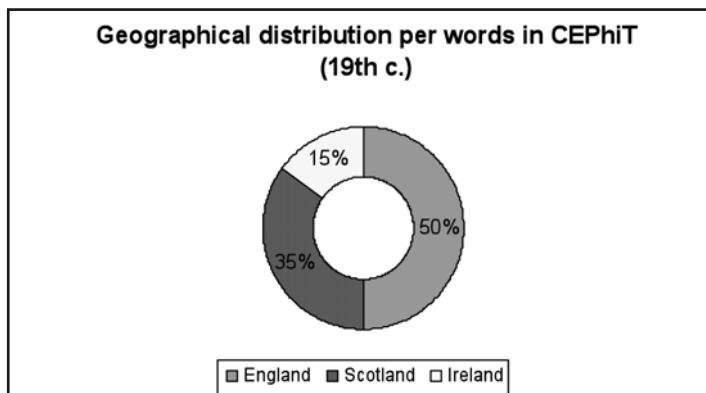


Graph 5. The provenance of authors in *CEPhiT*

Again, the distribution per centuries is slightly different to the overall one as reflected in graphs 6 and 7 below:



Graph 6. Geographical distribution in the eighteenth century



Graph 7. Geographical distribution in the 19th century

Social and political changes have a deep impact on the development of language. The way in which *CEPHiT* has been sampled represents social and political shifts. For instance, the fact that during the eighteenth century, Ireland

lived the Protestant Ascendancy implied that the native Irish population was excluded from power and public life (Claydon and McBride, 1999). Being England the coloniser, no wonder most scientific texts were produced there.

8. Editorial re-marks

Corpus compilation is an editorial task in itself. Many decisions had to be made before the selection of texts extracts and the application of the different representation conventions. Offering researchers the possibility of working with the information stored in the texts in a flexible and productive way implied taking some editorial decisions. The texts in the *Coruña Corpus of English Scientific Writing* of which *CEPbiT* is a part have been edited to represent even special graphemes in their XML format (visible in one of the windows of the *Coruña Corpus Tool* accompanying the corpus). As editors, we have avoided the representation of all those elements that did not constitute the language of the author. Therefore, quotations from other authors have been eliminated.

We have preserved old-fashioned characters such as <ſ> (long <s>), </ſ> (italicised long <s>) or the ligatured digraph <ct> in order to present a faithful representation of the evolution of spelling in the two centuries sampled. Since OCR under modern standards was completely unfeasible, manual typing was always needed at some stage.

For each sample we have included TEI-compliant headers with information about the file, full name of the research group behind this corpus, sponsors and director, name of this Philosophy sub-corpus (*CEPbiT*) and number of words in the file. The header box concludes with a reduced version of the full title of the text, pages selected the name of the author and the year of publication. We have kept the page numbers of the text, our only alteration being the centring of all page numbers on the screen in a bold font type between blank lines. In order to make the visual revision of texts more appealing we have used a bigger bold blue font on titles and chapters.

Editorial material –such as page headers, footers and margin notes– have been omitted since they do not represent the author's own language. A few spelling errors have been corrected because they are likely to have been made by the printer rather than by the author. We have considered the different spellings

across time and checked all the items in the *Oxford English Dictionary*. Those items impossible to identify or missing elements have been marked as [unclear].

Apart from the TEI tags we have included a set of editorial marks between square brackets. They contain information such as the location of quotations, figures, formulae, etc. in the original text. They are also used to disambiguate homographic forms that the *CCT* could consider a word⁷. Square brackets have been used for other strings of characters that could be ambiguous.

In *CEPbiT* files we have eliminated truncated words at the end of a line⁸. As for footnotes, their original form and location has not been respected not as the result of an editorial decision but due to TEI restrictions.

9. Some studies and contrasted validity of *CEPbiT*

Different pilot studies have shown that *CEPbiT*, as well as other brother corpora inside the *Coruña Corpus of English Scientific Writing*, is a reliable tool for the study of the evolution of scientific writing in the changing field of philosophy. The studies carried out can be grouped according to the various linguistic aspects they deal with. Since the corpus is still being finalised, not many of the research has been published but the following can be mentioned. The lexicon of science, the morphology of specialised terminology and other semantic implications have been explored from different perspectives in Camiña Riobóo (2010a and 2010b). Socio-linguistic variables included in *CEPbiT* have been also used for some studies such as Crespo García, (forthcoming), Moskowich (forthcoming) and Camiña Riobóo (2011). Other works such as Moskowich (2009), Monaco (2010), Lareo (forthcoming), Crespo (forthcoming) and Moskowich (forthcoming) revolve around different aspects of discourse.

⁷ For instance, the Roman number *I* has been enclosed in brackets to avoid the miscounting of the personal pronoun *I*. and the Roman number *I* will appear as [i].

⁸ Hyphens have been limited to compound words when they were hyphenated in the original. Therefore, when a hyphen has been used as a layout mark by the author or printer, an EM-dash has been placed instead.

Some aspects of the syntax of philosophy texts have been explored in Bello Viruega (2010).

Some MA and doctoral dissertations using *CEPhiT* have been or are also being written. All these works give *CEPhiT* an added value.

References

- Allen, B., Qin, J., & F. W. Lancaster.** (1994). Persuasive communities: A longitudinal analysis of references in the Philosophical Transactions of the Royal Society, 1665-1990. *Social Studies of Science* 24(2), pp. 279-310.
- Beal, J.** (2004). *English in modern times 1700-1945*. London: Arnold.
- Bello Viruega, I.** (2010). A diachronic study of nominalizations in Astronomy and Philosophy texts, *6th International contrastive linguistics conference*, Freie Universität Berlin, Berlin.
- Bhatia, V.** (1993). *Analysing genre: Language use in professional settings*. London: London.
- Biber, D.** (1993). Representativeness in corpus design. *Literary and Linguistic Computing* 8 (4), pp. 2243-257.
- Biber, D., Conrad, S. & Reppen, R.** (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Camiña Riobóo, G.** (2010a). New nouns for new ideas. In M.L. Gea-Valor, I.l. García, & M. J. Esteve (Eds.). *Linguistic and translation studies in scientific communications*. Vol. 86 (pp. 157-176). Bern/Berlin: Peter Lang.
- Camiña Riobóo, G.** (2010b). The language of women scientists in the 18th century: a morphological approach, *2nd International corpus linguistics conference (CILC10)*. University of A Coruña.
- Camiña Riobóo, G.** (2011). New nouns in the scientific register of late Modern English: A corpus-based approach. *3rd International corpus linguistics conference (CILC11)*. Universidad Politécnica de Valencia.
- Claydon, T. & McBride, I** (Eds.). (1999). *Protestantism and national identity: Britain and Ireland, c. 1650-c. 1850*. Cambridge: Cambridge University Press.
- Crespo García, B. & Moskowich-Spiegel Fandiño, I.** (2008). Advances in the Coruña Corpus. *29th ICAME Conference*. Ascona: University of Zurich.

- Crespo García, B.** (Forthcoming). Persuasion markers and ideology in eighteenth-century philosophy texts (*CEPhiT*). *Lenguas para Fines específicos* 17.
- Fortanet, I., Posteguillo, S., Palmer, J. C. & Coll, J.** (1998). Disciplinary variations in the writing of research articles in English. In Fortanet, I., Posteguillo, S., Palmer, J.C. & F. Coll, J. (Eds.), *Genre studies in English for Academic Purposes*, vol. 9 (pp. 59-78). Valencia: Universitat Jaume I.
- Fowler, A.** (1982). *Kinds of literature. An introduction to the theory of genres and modes*. Oxford: Clarendon Press.
- Freeborn, D.** (1992). *From Old English to Standard English*. London: Macmillan.
- Görlach, M.** (1994). *The linguistic history of English*. London: Macmillan.
- Görlach, M.** (2004). *Text types and the History of English*. Berlin/New York: Mouton de Gruyter.
- Gotti, M.** (1996). *Robert Boyle and the language of science*. Milan: Guerini Cientifica.
- Herrero López, C.** (2007). Las mujeres en la investigación científica. *Criterios*, 8, pp. 73-96
- Hoskin, M.** (Ed.). (1999). *The Cambridge concise history of astronomy*. Cambridge/New York: Cambridge University Press.
- Hyland, K.** (1998). *Hedging in scientific research articles*. John Benjamins.
- Kytö, M., Rudanko, J. & Smitterberg, E.** (2000). Building a bridge between the present and the past: A Corpus of 19-century English. *ICAME* 24, pp. 85-97.
- Lareo Martín, I.** (Forthcoming). Sociedad, educación y ciencia en los siglos 18 y 19: científicas británicas y americanas en el *Coruña Corpus*. In Crespo García et al (Eds.), *Mujer y ciencia: Historia de una desigualdad*. München: Lincom.
- McEnery, T. & Wilson, A.** (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Monaco, L. M.** (2010). Epistemic possibility and necessity in Modern scientific English: A first approach to modal auxiliaries in Philosophy texts. *II Congreso internacional de lingüística de corpus / 2nd International corpus linguistics conference*. Universidade da Coruña.
- Moessner, L.** (2001). Genre, text type, style, register: A terminological maze? *European Journal of English Studies*, 5(2), pp. 131-138.

- Monzó Nebot, E.** (2002). *La professió del traductor juridic i jurat: Descripció sociològica del professional i anàlisi discursiva del trasgenere*. Unpublished PhD Dissertation. Castelló: Universitat Jaume I.
- Moskowich-Spiegel Fandiño, I.** (2007). Exploiting the *Coruña Corpus. Encuentro de investigadores: Lingüística histórica inglesa (Genres, registers and text-types: A historical approach to English scientific writing.)* A Coruña: University of A Coruña.
- Moskowich-Spiegel Fandiño, I.** (2009a). Sobre la recopilación de corpus: la experiencia del Coruña Corpus. *I Congreso internacional de lingüística de corpus (CILC-09)*. University of Murcia.
- Moskowich-Spiegel Fandiño, Isabel.** (2009b). Científicos de la edad moderna en el *Coruña Corpus of Scientific Writing*: vida y obra, obra y vida. Lecture at the Instituto de Historia de la Medicina y de la Ciencia López Piñero (CSIC), Valencia.
- Moskowich-Spiegel Fandiño, Isabel.** (2010). “On the Coruña Corpus: Under construction.” *II Congreso internacional de lingüística de corpus (CILC-10)*. A Coruña: University of A Coruña.
- Moskowich-Spiegel Fandiño, I.** (2011). “A smooth homogeneous globe” in *CETA: Compiling late modern astronomy texts in English*. In Vázquez, N (Ed.), *Creation and use of historical English corpora in Spain*. Cambridge: Cambridge Scholar Publishing.
- Moskowich-Spiegel Fandiño, I.** (Forthcoming). Eighteenth-century female authors: Women and science in the Coruña Corpus of English Scientific Texts.
- Moskowich-Spiegel Fandiño, I. & B. Crespo García.** (2004). Presenting the Coruña Corpus: A collection of samples for the historical study of English scientific writing. *2nd Late Modern English conference (LMEC2)*. Vigo: Universidad de Vigo.
- Moskowich-Spiegel Fandiño, I. & Crespo García, B.** (2010). *CETA* in the context of the *Coruña Corpus. Literary and Linguistic Computing*. 25(2), pp. 153-164.
- Moskowich-Spiegel Fandiño, I. & Parapar López, J.** (2008). Writing science, compiling science. *The Coruña Corpus of English Scientific Writing*. In Lorenzo

- Modia, M. J. (Ed.), *Proceedings from the 31st AEDEAN conference* (pp. 531-544). A Coruña: Universidade da Coruña.
- Myers, G.** (1990). *Writing biology: Texts in the social construction of scientific knowledge*. Madison: University of Wisconsin Press.
- Nwogu, K. N.** (1990). *Discourse variation in medical texts: Scheme, theme and cohesion in professional and journalistic accounts*. Nottingham: University of Nottingham.
- OED Online.** (2009). Oxford: Oxford University Press. 28 April 2009. <<http://dictionary.oed.com>>.
- Swales, J.** (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Taavitsainen, I. & Pahta, P.** (1997). The Corpus of Early English Medical Writing. *ICAME Journal* 21, pp. 71-78.
- UNESCO.** (1988). *Proposed international standard nomenclature for fields of science and technology*. UNESCO/ROU257 rev. 1. Paris.
- Valle, E.** (1999). *A collective intelligence. The life sciences in the Royal Society as a scientific discourse community, 1665-1965*. University of Turku: Anglicana Turkuensia.

