NOUN FORMATION IN THE SCIENTIFIC REGISTER OF LATE MODERN ENGLISH

A CORPUS-BASED APPROACH

Gonzalo Camiña-Rioboo

Department of English Philology

Faculty of Philology

University of A Coruña

2013

Supervised by

Prof. Isabel Moskowich-Spiegel Fandiño

Doctorado Internacional, año 2013.

Programa: Novas aproximación aos estudos ingleses

Directora: Isabel Moskowich-Spiegel Fandiño

Departamento de Filoloxía Inglesa

RESUMEN

NOUN FORMATION IN THE SCIENTIFIC REGISTER OF LATE MODERN ENGLISH: A CORPUS-BASED APPROACH // LA FORMACIÓN DE NUEVOS NOMBRES EN EL REGISTRO CIENTÍFICO DEL INGLÉS MODERNO TARDÍO: UNA APROXIMACIÓN BASADA EN CORPUS.

Gonzalo Camiña Rioboo

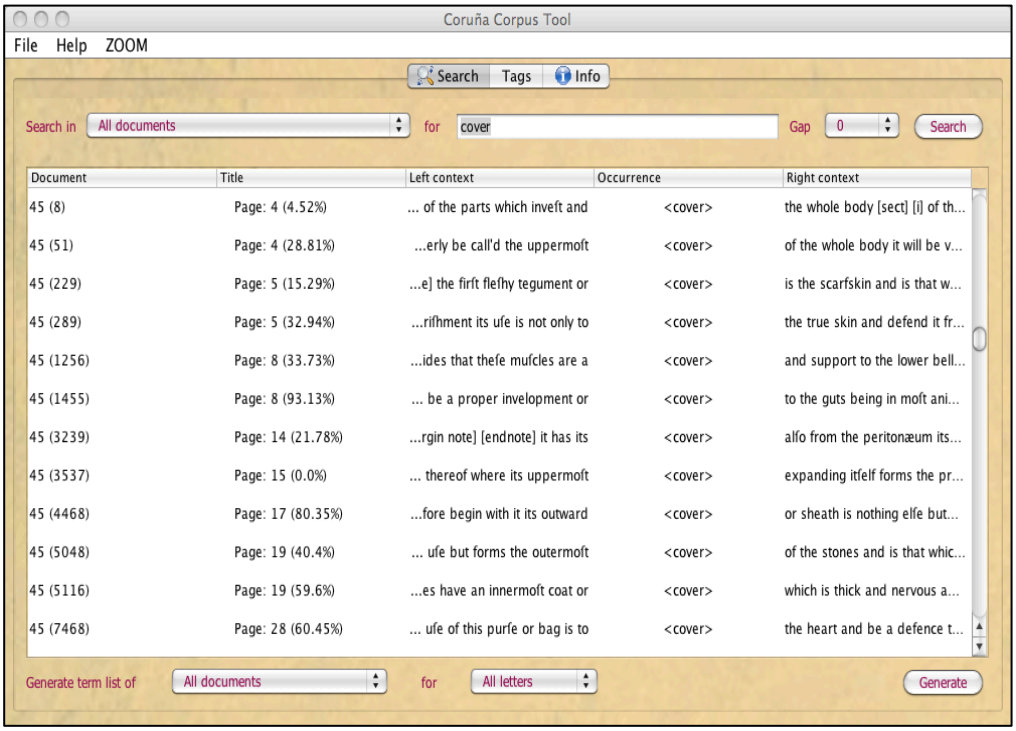
Esta tesis doctoral presenta un análisis empírico y cuantitativo de los procesos de formación de nombres que se observan en el registro científico del inglés del siglo XVIII, después de la llamada 'Revolución Científica' en Gran Bretaña, y establece los patrones lingüísticos que siguen los nombres acuñados en el ámbito de la ciencia por los científicos de la época para denominar los

nuevos descubrimientos, herramientas, avances tecnológicos, teorías e ideas predominantes en la época. Además, la investigación aquí presentada revisa los períodos de la historia de la lengua inglesa en los que los diferentes procesos morfológicos y unidades involucradas en esos procesos adquieren más o menos relevancia, y analiza la veracidad de las diferentes teorías lingüísticas actuales, las cuales presentan datos contradictorios: unas sostienen que en el siglo en cuestión se produjo una deceleración intencionada en la producción de palabras para conferir al inglés un status semejante al del latín y griego clásicos, lenguas que gozaban de gran prestigio en Europa y que no se veían afectadas por el cambio lingüístico por el hecho elemental de no contar con hablantes nativos vivos. Eran además las linguas francas de la ciencia y utilizadas en publicaciones y reuniones científicas hasta ese momento. Otras teorías aducen que no hay un número suficiente de estudios que se centren en el inglés moderno tardío como para determinar cuantitativamente la deceleración citada con anterioridad. Sea cual fuere el motivo, el interés de este período radica fundamentalmente en el deseo y voluntariedad por parte de los investigadores de la época, así como de instituciones científicas tales como la Royal Society of London, de abandonar el uso de las lenguas anteriores para diseminar el conocimiento científico en inglés, ya que esta era la única lengua que conocía la mayoría de la población y, por tanto, el mejor vehículo para la transmisión del saber. Este se considera el momento crucial en el que la ciencia sale de las universidades y se hace universal.

En este punto se enmarca esta tesis, y para demostrar todo lo citado en el párrafo anterior he utilizado la metodología de la lingüística de corpus, es decir, la combinación de un corpus lingüístico anotado conforme a los estándares 'eXtended Mark-up Language' y 'Text Encoding Initiative', y diversas herramientas informáticas para extraer y cuantificar los datos morfológicos relevantes al estudio, con el fin de analizarlos y encontrar patrones lingüísticos que expliquen amplios rangos de ocurrencias que corroboren, contradigan o complementen las teorías lingüísticas existentes. Como principal fuente de consulta del origen de los nuevos nombres, los elementos que los componen y de las fechas de primer uso tanto de esos elementos como de las palabras resultantes, he utilizado el prestigioso *Oxford English Dictionary* (*OED*) en su versión online, en sintonía con otros trabajos significativos, como el publicado por Culpeper y Clapham (1996) sobre el sobre el préstamo lingüístico, y Camiña Rioboo (2010, 2012) sobre la formación de palabras, o el de Caso (1980) sobre el cambio semántico, aunque este último utilizó el diccionario *Chambers* de ciencia, más conocido como *Collocott*.

Con el objetivo de ofrecer una variedad de resultados que abarque disciplinas científicas en claro contraste, he estudiado textos que provienen de dos ciencias diferentes, una experimental basada en la observación y las matemáticas, y la otra basada en el razonamiento abstracto. Para este estudio

he utilizado dos subcorpora de astronomía y filosofía de muy reciente creación pertenecientes al *Coruña Corpus of English Scientific Writing* (de aquí en adelante, *Coruña Corpus*) desarrollado en la Universidade da Coruña por el Research Group for Multimedia Studies in English, concretamente el *Corpus of English Texts on Astronomy* (*CETA*, publicado en 2012 por John Benjamins Publishing Company), y el *Corpus of English Philosophy Texts* (*CEPhiT*, pendiente de publicación por John Benjamins), respectivamente. Ambos han sido analizados desde un punto de vista morfológico con la herramienta desarrollada por IRLAB, *Coruña Corpus Tool*, cuya apariencia se muestra en el gráfico siguiente:

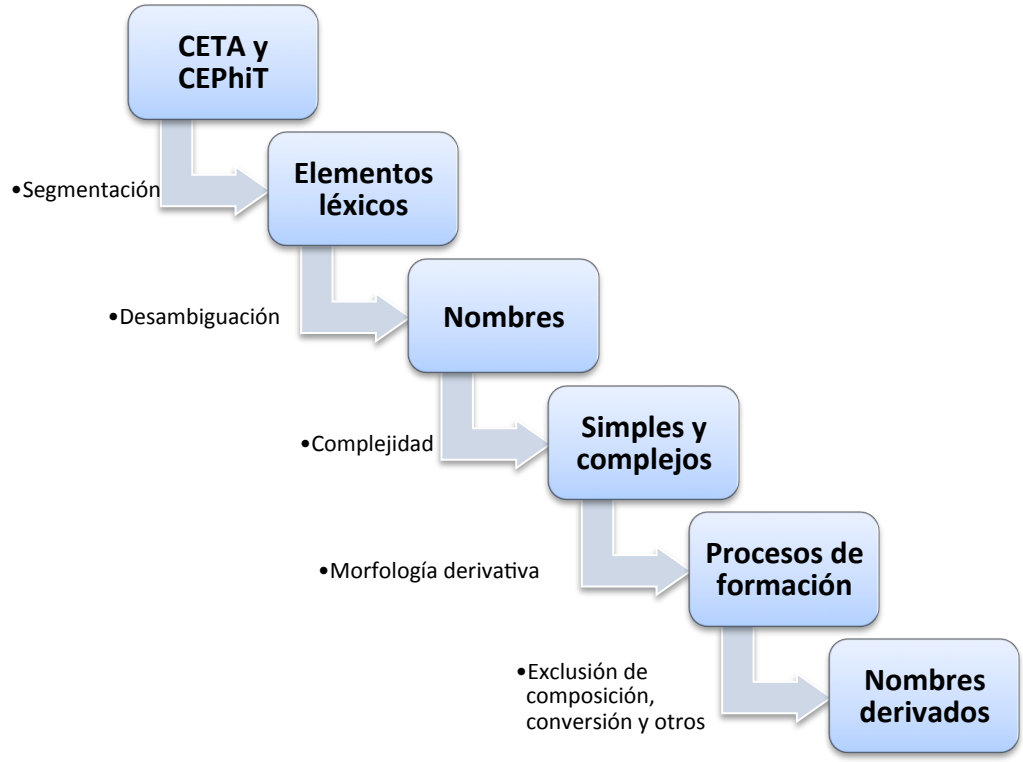


La tesis está dividida en cuatro capítulos principales, y cuenta también con una introducción, un capítulo de conclusiones y un apéndice. Los cuatro capítulos tratan en profundidad los aspectos intralingüísticos y extralingüísticos más relevantes en la historia de la lengua, detallando aquellos que teóricamente intervienen en el cambio lingüístico, y que dan forma al inglés que se escribía en los textos científicos entre 1700 y 1799, mi objeto de estudio.

Así pues, el capítulo 1 describe la historia externa de la lengua. En él se destacan la importancia de las sucesivas invasiones de Gran Bretaña, con la consiguiente influencia de las lenguas invasoras en los angloparlantes, así como los cambios y revoluciones culturales y sociales que tuvieron impacto en el conocimiento en general, y en la lengua en particular. Igualmente, se valida el estudio de eventos históricos y científicos del pasado desde la perspectiva del siglo XXI, aunque se advierte de los peligros que esto conlleva, haciendo hincapié en que algunos de los términos empleados, como el de 'ciencia' o 'científico' todavía no existían en el momento de la historia en la que se basa este estudio. Además, se recalca que la división entre astronomía y filosofía no era como la entendemos hoy día. Para obtener más información sobre la composición del corpus habrá que referirse a los múltiples estudios publicados por Isabel Moskowich, Begoña Crespo, e Inés Lareo, entre otros (ver bibliografía al final de esta tesis).

El capítulo 2 explica teorías lingüísticas actuales sobre morfología y formación de palabras, intenta definir los límites entre la morfología flexiva y la derivativa, define los procesos principales de esta última, al igual que los elementos que intervienen en la acuñación de nuevos nombres, y compara las teorías lingüísticas de hoy con las teorías gramaticales imperantes en el siglo XVIII, cuando el término 'morfología' ni siquiera se aplicaba en el ámbito lingüístico, sino que estaba restringido a las ciencias naturales. Es de destacar que los ejemplos citados en este capítulo para explicar estas teorías han sido extraídos del corpus prácticamente en su totalidad.

El capítulo 3 se centra en la metodología empleada, la lingüística de corpus, y el conjunto de muestras analizado. Se describen las condiciones establecidas por MuStE que deben cumplir las muestras de texto para que puedan ser incluidas en un corpus lingüístico, en este caso el *Coruña Corpus*, condiciones que han sido ahora también adoptadas por importantes lingüistas de corpus a nivel internacional, y se presentan los contenidos de los dos subcorpora analizados. En este capítulo se muestra el proceso en cascada descendente del análisis realizado en el capítulo 4, con el que se espera restringir y focalizar el estudio con cada paso. El gráfico siguiente muestra el proceso:



En primer lugar se han compilado los subcorpora de astronomía y filosofía utilizando cuarenta y una muestras, todas ellas de diez mil palabras, excepto en dos casos que corresponden a artículos científicos más breves. Todas las muestras han sido escritas por autores diferentes con un uso de la lengua variado, y se han utilizado prácticamente en todos los casos primeras ediciones. Las muestras son trabajos originales en inglés, es decir, no proceden de traducciones de otras lenguas, y ni siquiera de adaptaciones al inglés escritas por los mismos autores que también publicaban trabajos en latín o griego. Todos los autores son hablantes nativos que han sido educados en las islas británicas, tanto del sexo masculino como femenino. El corpus presenta una

variedad de géneros literarios/tipos de texto que permite contrastar usos lingüísticos diferentes en ellos.

En segundo lugar he segmentado el corpus, que contiene cuatrocientos ocho mil novecientos ochenta y un elementos léxicos, y he cuantificado todos los nombres, y agrupado todas las ocurrencias en tipos, obviamente teniendo en cuenta tanto las diferentes grafías como los casos de singular y plural. El gran número de nombres obtenido, cuatro mil quinientos cincuenta y ocho, me ha obligado a crear una base de datos informatizada que pueda manejar la cantidad ingente de información a manipular y el uso de fórmulas matemáticas complejas que comparan y contrastan los datos desde el punto de vista de una multitud de variables, tales como el número de ocurrencias de cada nombre en los textos, el proceso lingüístico asociado a su creación, en caso de ser afijación también he analizado el subproceso, las bases y afijos que intervienen, la fecha de acuñación de estos y su origen, y la fecha de primer uso de la palabra resultante de los procesos anteriores. Esto puede observarse en el gráfico que se muestra más abajo.

En esta enorme base de datos de más de trescientas veinticinco mil celdillas que contienen información he cuantificado aquellos nombres que pertenecen a *CETA* y a *CEPhiT* por separado, a fin de comparar ambas disciplinas. El siguiente paso ha sido separar los nombres simples de los que han sufrido algún proceso lingüístico y se han analizado los últimos. He

cuantificado los casos de derivación, composición y conversión, además de otros procesos menores explicados con profundidad en los capítulos 2 y 4 de la tesis, y he hecho un estudio exhaustivo de los nombres derivados y de las unidades que los componen.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	TYPES	TOKENS	TokAstr	TokPhil	a02a	a02b	p00	p05	a15	a19	p10	p17	a26a	a26b
2682	nail	1	0	1										
2683	name	249	176	73	1	10	6	1		5	6		2	4
2684	narration	15	0	15										
2685	narrative	1	0	1										
2686	nation	60	12	48			1			1				1
2687	native	2	1	1	1									
2688	nativity	1	1	0	1									
2689	naturalist	3	1	2										
2690	nature	692	82	610	7	3	10	16	3	8	10	11	1	
2691	navigation	8	6	2					1					1
2692	navigator	10	9	1	1	1				2				
2693	navy	1	1	0										
2694	nearness	7	7	0										
2695	nebula	1	1	0									1	
2696	nebulosa	4	4	0		4								
2697	necessarian	3	0	3										
2698	necessity	72	4	68			2	8	2		4	22		1
2699	neck	7	5	2	2	2					1			
2700	need	10	3	7	1		2		2		1	1		
2701	needle	9	8	1	1	1				2			1	
2702	negation	1	0	1										
2703	negative	1	0	1										
2704	neglect	9	3	6			1			2		1		1
2705	negligence	1	0	1										
2706	negro	1	0	1										
2707	neighbour	7	2	5			3							

El capítulo 4, el más voluminoso y central de la investigación, expone el análisis de los datos según los procesos lingüísticos que intervienen en la creación de nombres nuevos, con especial énfasis en la derivación. Así pues, en este capítulo se mostrará todo lo referente a nombres simples, derivados, compuestos, aquellos obtenidos sin la adición de material lingüístico (‘zero-derivation’), y también los obtenidos por medio de la sustracción de materiales

(‘clipping’). Se mostrarán numerosos gráficos y tablas de unidades y elementos tales como bases y afijos, se valorará su origen, cuándo se utilizaron por primera vez en la lengua inglesa y con qué frecuencia intervienen en la acuñación de nuevos nombres. Las variables intralingüísticas del estudio intentarán aclarar los siguientes puntos: 1) qué procesos son los más utilizados en la creación de palabras nuevas; 2) qué materiales lingüísticos, bases y afijos, son los más productivos; 3) cuáles son las lenguas donantes, de entre el latín, griego, anglofrancés, francés o inglés antiguo, que más han contribuido a la conformación del inglés moderno tardío; 4) en qué período de la historia de la lengua se han acuñado más nombres y cuáles son las razones para ello; 5) qué impacto tiene la revolución científica en la evolución de la lengua inglesa a partir del siglo XVII; 6) cómo afecta al cambio lingüístico la supuesta intención de “congelar” la producción de palabras en el siglo XVIII. Además, se considerarán cinco variables extralingüísticas: a) una contrastiva entre la astronomía y la filosofía para decidir qué disciplina científica presenta una mayor variedad en el uso de nombres y, como consecuencia, una influencia en el cambio lingüístico; b) una comparativa del género/tipo de texto científico para determinar si los resultados son diferentes para cada uno de ellos; c) el sexo del autor será tenido en consideración para determinar si existe una diferencia notable en el uso de la lengua entre hombres y mujeres en un momento en que estas tenían vetado el acceso a las universidades; d) también se estudiará el uso individual de la lengua por parte de los autores según su

lugar de educación: Inglaterra, Escocia, Irlanda o Irlanda/Escocia; e) finalmente se ofrecerán datos sobre el impacto en la versatilidad y complejidad del uso de la lengua de cada autor según su edad.

En la sección de conclusiones se expresan detalladamente los resultados obtenidos en el análisis cuantitativo y se explican con profundidad las razones por las cuales se han producido estos hechos lingüísticos en determinados períodos de la historia de la lengua y en qué medida. Finalmente, es de esperar que cada nueva tesis doctoral contribuya a la ampliación del conocimiento y a la optimización de las herramientas disponibles para el análisis y estudio de la lengua. Es por ello que en el apéndice se facilitan un buen número de nombres para los que el *OED* ofrece una fecha posterior al texto analizado en el corpus, cuyo desfase llega a ser en algunos casos de más de cien años. Sin duda, estos hallazgos ayudarán a mejorar la fuente de información para futuros trabajos similares a este.

BIBLIOGRAFÍA MÁS RELEVANTE DE CADA SECCIÓN

Historia de la ciencia

Birch, Thomas (1756). *The History of the Royal Society*. London: A. Millar.

Boyle, Robert (1661). A Proëmial Essay, wherein, with some considerations touching experimental essays in general, is interwoven such an introduction to all those written by the author, as is necessary to be perus'd for the better understanding of them. In *Certain Physiological*

Essays Written at Distant Times, and on Several Occasions. London: Henry Herringman.

Butterfield, Herbert (1965). *The Origins of Modern Science*. New York: The Free Press.

Crombie, Alistair C. (1969). *Augustine to Galileo: The History of Science A.D. 400-1650.* London: Penguin.

Gotti, Maurizio (2005). *Investigating Specialized Discourse.* Bern/Berlin: Peter Lang.

Gotti, Maurizio (2001). The Experimental Essay in Early Modern English. *European Journal of English Studies*, *5*(2), 221-239.

Grant, Edward (2007). *A History of Natural Philosophy: From the Ancient World to the Nineteenth Century.* Cambridge: Cambridge University Press.

Greenwood, James (1722). *An Essay towards a Practical English Grammar, Describing the Genius and Nature of the English Tongue.* (2nd ed., with additions). London: printed for John Clark.

Halliday, Michael A. K., & Webster, J. J. (2004). *The Language of Science.* London/New York: Continuum.

Hård, Mikael & Jamison, Andrew (2005). *Hubris and Hybrids: A Cultural History of Technology and Science.* London: Routledge.

Henry, John (2002). *The Scientific Revolution and the Origins of Modern Science.* New York: Palgrave.

Jacob, Margaret C. (1988). *The Cultural Meaning of the Scientific Revolution.* Philadelphia: Temple University Press.

Kuhn, Thomas S. (1996). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Moskowich, Isabel & Crespo, Begoña (Eds.) (2012). *Astronomy 'Playne and Simple': The Writing of Science Between 1700 and 1900.* Amsterdam: John Benjamins.

Priestley, Joseph (1768). *The Rudiments of English Grammar, Adapted to the Use of Schools; With Notes and Observations, For the Use of those who Have Made some Proficiency in the Language.* London: printed for T. Becket and P.A. De Hondt, and J. Johnson.

Snyder, Laura J. (2009). Hypotheses in 19th Century British Philosophy of Science: Herschel, Whewell, Mill. In Michael Heidelberger & Gregor Schiemann (Eds.), *The Significance of the Hypothetical in the Natural Sciences* (pp. 59-76). Berlin: Mouton De Gruyter.

Subbiondo, Joseph L. (2007). Competing models for a 17th century universal language: A study of the dispute between George Dalgarno and John Wilkins. In Douglas E. Kibbee (Ed.), *History of linguistics 2005* (pp. 112-119). Amsterdam: John Benjamins.

Taavitsainen, Irma (2000). Scientific language and spelling standardisation. In Laura Wright (Ed.), *The Development of Standard English, 1300-1800. Theories, Descriptions, Conflicts.* (pp. 131-154). Cambridge: Cambridge University Press.

Valle, Ellen (2006). Reporting the doings of the curious: authors and editors in the *Philosophical* Transactions of the Royal Society of London. *News discourse in early modern Britain: selected papers of CHINED 2004,* 71.

Teorías gramaticales del siglo XVIII

Coote, Charles (1788). *Elements of the Grammar of the English Tongue.* London: printed for the author.

Bevis, John (1765). *A Pocket Dictionary; Or, Complete English Expositor.* London: J. Newbery.

Bicknell, Alexander (1790). *The Grammatical Wreath; Or a Complete System of English Grammar: Being a Selection of the Most Instructive Rules from All the Principal English Grammars, Etc.* London: printed for the author.

Elphinston, James (1765). *The Principles of the English Language Digested; Or, English Grammar Reduced to Analogy.* London: James Bettenham.

Maittaire, Michael (1712). *The English Grammar: Or, an Essay on the Art of Grammar, Applied to and Exemplified in the English Tongue.* London: printed by W. B. for H. Clements.

Tuite, Thomas (1726). *The Oxford Spelling-Book; Being a Complete Introduction to English Orthography. In a Method much more Clear and Intelligible than any Book of this Nature, hitherto Extant.* London: printed for J. Hazard.

Ward, William (1967). *An Essay on Grammar, 1765.* Scolar Press: Menston, England.

Wilkins, John (1668). *An Essay towards a Real Character, and a Philosophical Language.* London: Samuel Gellibrand and John Martin.

Lingüística general y morfología:

Adams, Valerie (1973). *An Introduction to Modern English Word Formation (English Language Series).* London: Longman.

Adams, Valerie (2001). *Complex Words in English, English Language Series.* Harlow: Pearson.

Bailey, Richard W. (2010). Variation and change in eighteenth-century England. In Raymond Hickey (Ed.), *Eighteenth-Century England: Ideology and Change* (pp. 182-199). Cambridge: Cambridge University Press.

Bauer, Laurie (1983). *English Word Formation.* Cambridge: Cambridge University Press.

Bauer, Laurie (1988). *Introducing English Morphology*. Edinburgh: Edinburgh University Press.

Bauer, Laurie (2001). *Morphological Productivity.* Cambridge: Cambridge University Press.

Bauer, Laurie (2004). *A Glossary of Morphology.* Edinburgh: Edinburgh University Press.

Beal, Joan C. (2010). Prescriptivism and the suppression of variation. In Raymond Hickey (Ed.), *Eighteenth-Century English: Ideology and Change* (pp. 21-37). Cambridge: Cambridge University Press.

Carstairs-McCarthy, Andrew (2002). *An Introduction to English Morphology: Words and their Structure.* Edinburgh: Edinburgh University Press.

Ciszek, E. (2008). *Word Derivation in Early Middle English.* Frankfurt am Main: Peter Lang.

Durkin, Philip (2008). Latin loanwords of the early modern period: How often did French act as an intermediary? In Richard Dury, Maurizio Gotti, & Marina Dossena (Eds.), *English Historical Linguistics 2006* (pp. 185-202). Amsterdam/New York: John Benjamins.

Görlach, Manfred (2004). *Text Types and the History of English.* Berlin: Mouton de Gruyter.

Kastovsky, Dieter. (2006). Vocabulary. In Richard M. Hogg & David Denison (Eds.), *A History of the English Language* (pp. 199-270). Cambridge: Cambridge University Press.

Labov, William (2001). Principles of Language Change: Social Factors. Chichester: Wiley.

Lieber, Rochelle (2010). *What is morphology?* Cambridge University Press: Cambridge.

Marchand, Hans (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach.* Munich: C. H. Beck.

Millward, Celia M. (1996). *A Biography of the English Language: Principles and Applications* (2nd ed.). Boston: Thomson Learning.

Raad, B. L. (1989). Modern trends in scientific terminology: morphology and metaphor. *American Speech*, *64*(2), 128-136.

Stein, Gabrielle (2007). *A Dictionary of English Affixes.* Munich: Lincom.

Tieken-Boon van Ostade, Ingrid (2009). *An Introduction to Late Modern English.* Edinburgh: Edinburgh University Press.

Lingüística de corpus

Adolphs, Svenja (2006). *Introducing Electronic Text Analysis.* London: Routledge.

Adolphs, Svenja (2008). *Corpus and context: investigating pragmatic functions in spoken discourse.* Amsterdam: John Benjamins.

Banks, David (2005). Emerging scientific discourse in the late seventeenth century: A comparison of Newton's *Opticks* and Huygen's *Traité de la lumière*. *Functions of language*, *12*(1), 65-86.

Banks, David (2012). Thematic structure in eighteenth century astronomical texts: a study of a small sample of articles from the Corpus of English Texts on Astronomy. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The Writing of Science between 1700 and 1900* (pp. 221-238). Amsterdam: John Benjamins.

Biber, Douglas, Conrad, Susan & Reppen, Randi (1998). *Corpus Linguistics.* Cambridge: Cambridge University Press.

Camiña, Gonzalo (2010). New nouns for new ideas. In M. Lluisa Gea-Valor, Isabel García, & M. José Esteve (Eds.), *Linguistic and Translation Studies in Scientific Communications* (pp. 156-176). Bern/Berlin: Peter Lang.

Camiña-Rioboo, Gonzalo (2012). Accounting for observations of the heavens in the 18th century: New nouns to explain old phenomena. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The Writing of Science between 1700 and 1900* (pp. 93-122). Amsterdam: John Benjamins.

Crespo, Begoña & Moskowich, Isabel (2009). The limits of my language are the limits of my world: the scientific lexicon from 1350 to 1640. *Skase Journal of Theoretical Linguistics*, *6*(1), 45-48.

Crespo, Begoña & Moskowich, Isabel (2010). *CETA* in the context of the *Coruña Corpus*. *Literary and Linguistics Computing*, *25*(2), 153-164.

Harrington, Jonathan (2010). *Phonetic Analysis of Speech Corpora.* Oxford: Wiley-Blackwell.

Information Retrieval Lab (2012). *The Coruña Corpus Tool* [computer software]. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The Writing of Science between 1700 and 1900*. Amsterdam: John Benjamins.

Lareo, Inés (2010). New trends exploring the language of science: The *Corpus of English Texts on Astronomy* (*CETA*) and its tool (*CCT*) in the context of the *Coruña Corpus.* In M. Lluisa Gea-Valor, Isabel García-Izquierdo, & M. José Esteve (Eds.), *Linguistic and Translation Studies in Scientific Communication* (pp. 131-156). Bern: Peter Lang.

Meyer, Charles F. (2002). *English Corpus Linguistics: An Introduction.* Cambridge: Cambridge University Press.

McEnery, Tony & Wilson, Andrew (2001). *Corpus Linguistics* (2nd ed.). Edinburgh: Edinburgh University Press.

Moskowich, Isabel (2012). *CETA* as a tool for the study of modern astronomy in English. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The Writing of Science between 1700 and 1900* (pp. 35-56). Amsterdam: John Benjamins.

Popescu, Ioan-Iovitz *et al*. (2009). *Word frequency studies*. Berlin: Walter de Gruyter.

Taavitsainen, Irma & Pahta, Päivi (Eds.). (2004). *Medical and scientific writing in late medieval English.* Cambridge: Cambridge University Press.

Acknowledgments

The writing of this dissertation has taken nearly ten years of my life. Over this time I owe several people for their support. I would like to thank my dear friend and supervisor Prof. Isabel Moskowich, without whose great ideas, unconditioned support and infinite patience when writing tended to slow down due to the natural course of life, I would have never carried out this kind of research. I would also like to thank Dr. Inés Lareo for her help during the long and tedious compilation and organisation of the corpus, and data collection. Her constant encouragement, long conversations and shared confidences guided me through this process. I also thank Dr. Begoña Crespo for her contribution to the metadata of the Coruña Corpus, which gave shape to the core of my study, and those who have given me the jobs that kept me alive while I was writing: Prof. David Mackenzie and Prof. Nuala Finnegan at University College Cork, Prof. Stanley Black and Mr. Alvaro Jaspe at University of Ulster and Prof. María Jesús Lorenzo and Prof. Isabel Moskowich at University of A Coruña. I am also very grateful to my partner Silvia for all her love and support, especially during these hard last months.

To three incredible women: my mother, grandmother and grandaunt.

Resumen

Esta tesis doctoral analiza procesos morfológicos de formación de nombres en el registro científico del inglés moderno tardío usando metodología de lingüística de corpus. Mediante el análisis de cuarenta y una muestras escritas en el siglo dieciocho, después de la llamada 'revolución científica', extraídas de dos subcórpora del *Coruña Corpus of English Scientific Writing* —el *Corpus of English Texts on Astronomy* y el *Corpus of English Philosophy Texts*—, se revisarán cerca de medio millón de palabras. Mi objetivo es determinar cuáles son los procesos más productivos no sólo desde un punto de vista sincrónico en dicho siglo, sino también desde un punto de vista diacrónico a través de la historia de la lengua. Además, los nombres complejos se dividirán en sus constituyentes básicos, que serán a continuación analizados para establecer cuantitativamente patrones de las bases y afijos más productivos, en qué fecha se utilizaron por primera vez en la lengua, y las fuentes etimológicas más abundantes del inglés dependiendo de los períodos. El estudio se complementa con variables extralingüísticas que comparan los recursos empleados según la disciplina científica, el género del texto, el sexo de los autores, el país donde se formaron, y su edad cuando publicaron sus trabajos.

Resumo

Esta tese de doutoramento analiza procesos morfolóxicos de formación de nomes no rexistro científico do inglés moderno tardío usando a metodoloxía da lingüística de corpus. Mediante a análise de corenta e unha mostras escritas no século dezaoito, despois da chamada 'revolución científica', extraídas de dous subcórpora do *Coruña Corpus of English Scientific Writing* —o *Corpus of English Texts on Astronomy* e o *Corpus of English Philosophy Texts*—, revisaranse perto de medio millón de palabras. O meu obxectivo é determinar cales son os procesos máis productivos non só desde un punto de vista sincrónico no devandito século, senón tamén desde un punto de vista diacrónico a través da historia da lingua. Ademáis, os nomes complexos dividiranse nos seus constituintes básicos, que serán analizados a continuación para establecer cuantitativamente padróns das bases e afixos máis productivos, en qué data se producíu o seu primeiro uso na lingua, e as fontes etimolóxicas máis abundantes do inglés dependendo dos períodos. O estudio ven complementado con variables extralingüísticas que comparan os recursos empregados según a disciplina científica, o xénero do texto, o sexo dos autores, o país onde se formaron, e a sua idade cando publicaron os seus traballos.

Abstract

This PhD. dissertation analises noun forming processes in the scientific register of late Modern English using corpus linguistics methodology. By revising forty-one samples written in the eighteenth century after the so-called 'scientific revolution', extracted from two subcorpora of the *Coruña Corpus of English Scientific Writing —Corpus of English Texts on Astronomy* and *Corpus of English Philosophy Texts*—, I will analyse nearly half a million words. My aim is to determine which are the most productive processes, not only from a synchronic point of view in the above-mentioned century, but also from a diachronic standpoint throughout the history of the language. Furthermore, complex nouns will be decomposed in their basic constituents, which will be subsequently analysed to establish quantitatively patterns of the most productive bases and affixes, the dates when they were coined and used for the first time in the language, and their source languages across different periods. This research work is complemented by several extralinguistic variables that help comparing the linguistic resources depending on the scientific discipline studied, genre/text-type, sex of the author, place of education and age when the work in question was published.

Table of contents

List of tables

List of figures

Abbreviations

AE	American English
Arab.	Arabic
BE	American English
Brit.	Britain
CCT	*Coruña Corpus Tool*
CETA	*Corpus of English Texts on Astronomy*
CEPhiT	*Corpus of English Philosophy Texts*
eModE	esrly Modern English
FCF	Final Combining Form
Fr.	French
Ger.	German
ICF	Initial Combining Form
Lat.	Latin
lModE	late Modern English
ME	Middle English
ODNB	*Oxford Dictionary of National Biography*
OE	Old English
OED	*Oxford English Dictionary*
U.S.	United States

Introduction

> The current explosion of scientific knowledge is having considerable linguistic effect as the need to communicate new concepts forces the devising of new terms. The question of what methods are being employed to meet this need is one of importance to anyone concerned with lexicology.
>
> Caso, The production of new scientific terms (1980)

Men of science of all times have been overwhelmed by a seemingly limitless number of new facts discovered and theories necessary for the explanation of these new discoveries. In each such case a new word must be found to express the new idea, to refer to a newly invented mechanical instrument, or to identify a recently discovered celestial body. These words in English have been traditionally derived or imported largely from Greek and Latin (Jespersen 1933), languages that conveyed a high prestige in academic environments, were taught at schools and universities, and ultimately remained the main vehicle of communication and the primary source of names for new concepts and phenomena in science for centuries.

A first corollary can be drawn from above: we must presume the existence of a 'type' of English different from the vernacular, which is used mainly by learned men. Not only are they masters in the traditional and modern sciences but also they show a good deal of concern for the shaping of the language in order to suit their specific and, at first sight, restricted needs. This technical —scientific— register of the language would deviate from the vernacular at all levels and, as regards word formation, its differences would be founded essentially on the sources employed to produce the new elements coined to enrich the English lexicon at a highly specialised level.

This doctoral dissertation[1] analyses the scientific register of English from a morphological standpoint. In order to explain how the names given to scientific discoveries were born —coined, rather— I will review over four hundred thousand words contained in a scientific corpus. My first aim is to establish patterns of noun formation across the history of the language, so first of all I will extract all nouns contained in the corpus. Later I will assess their complexity, then I will determine how complex nouns were formed, that is to

[1] The research here reported on has been funded by Consellería de Educación e Coordenación Universitaria, (2007/000145-0), Xunta de Galicia (pgidit07pxib104160pr) through its Dirección Xeral de Investigación e Desenvolvemento and Departamento de Recursos Humanos and by Ministerio de Ciencia e Innovación (ffi2008-01649/filo). These grants are hereby gratefully acknowledged.

say, what linguistic processes are involved in their coinage. Finally, I will decompose them into their basic constituent elements, which I will later analyse individually. By paying attention to their origin, date of first use and productivity throughout the history of the language, among others, I intend to map the most creative periods and determine the linguistic devices and materials used to enlarge the English vocabulary.

It is a well-known fact that scientists had a relatively significant knowledge of the classical languages. At least this had been the situation until the eighteenth century. For this reason new technical terms usually had followed a word-formation pattern based on classical grounds, and they had made the effort to express as accurately and consistently as possible the form and treatment of the new facts by virtue of Greek or Latin root and affix combinations, or even by importing foreign terms especially for the occasion (Nybakken 1959).

Three hundred years later, however, the tendency used to coin new words based on classical languages' word-formation patterns seems to have lost vigour, to such an extent that some authors affirm that in the twentieth century new trends have emerged, by means of which new scientific terms are primarily generated by the compounding of elements with purely English

origins (Raad 1989) or by the semantic change of existing English items[2] (Caso 1980). This may be due to two factors at least, the first of which would involve the long decline in teaching Greek and Latin in secondary education and universities, especially in the second half of the twentieth century. Consequently, scientists have simply become unacquainted with their use, and they simply lack the knowledge necessary to maintain this type of device productive. The second and probably foremost factor can be that English itself has become the international language of science in the twentieth century, overpowering other widely spoken languages, such as Spanish, French or German, among others. This status of privilege provides present-day writers in English with a degree of confidence never achieved before as regards the production of new technical terms that involve English bases only.

A second and more complex corollary can be drawn from above: the boundaries between vernacular and technical English seem to have narrowed at a slow but steady rate during the past three centuries. It can be argued, however, that the gap between them was never as wide as it has been judged in

[2] Caso (1980) studied 4300 scientific terms found in the Chambers *Dictionary of Science and Technology* (Collocott 1971) from physics and earth sciences, on the grounds that these two branches of science have developed enormously in the twentieth century. He states that "semantic change was by far the most productive of the categories, and the most common type of semantic change was the widening of the meanings of established words" (103).

the past. Or rather, their limits were not as clear as we believed them to be. As Van Dyke (1992: 383) puts it,

> Aside from its magnitude, […] the contribution of science and technology to the English vocabulary has not been established. In particular, we do not know whether or how modern scientific terms differ from other words etymologically or semiotically.

To make matters worse, then, something as basic and apparently simple as defining a word as 'scientific' is not an easy task, since a word by itself does not carry any intrinsic meaning. However, such a distinction is not necessary at this point since all the samples analysed have been extracted from scientific texts, and for the sake of simplicity all the terms and word-formation processes included in them have been regarded as scientific. A good number of them can nonetheless be also found in non-scientific texts, especially in later periods. Furthermore there has been a kind of abstruse symbiosis between vernacular and technical English, along with social, historical, economical and political reasons, that may have contributed to the present predominant status of the English language in scientific contexts. On the one hand, technical English has benefited from vernacular English and its unrestricted freedom to make use of native and foreign elements in order to enlarge the lexicon, and the latter has also provided the former with an expedient path towards the acceptance of Greek- and Latin-based terms, given that around seventy-five per cent of the

vocabulary of English had already originated on classical grounds. Throughout history a good share of this existing vocabulary had come into English directly from Greek or Latin, or via French. Also, a present-day high-status vernacular English even retains enough self-assurance to supply science with new technical words resulting from word-formation processes involving English-only elements, such as *smog*, a blend of *smoke* + *fog*, or the compounds *red giant* and *blackbody* (Raad 1989), used in astronomical discourse. On the other hand, even though "the contribution of science and technology to the English vocabulary has not been established" (Van Dyke 1992), there can be no doubt that vernacular English has benefited greatly from technical English, and terms fundamentally scientific, such as *aspirin*, *gene*, *neurosis*, *penicillin*, or *vitamin* (Raad 1989) have become part of the daily language, and have consequently lost the tag of 'technical' that they used to hold in the past.

There must be a logical explanation for the quantum leap of English, from being regarded as a second-class language only three centuries ago, to the universal recognition that it has acquired in the present. Also, another explanation is needed for the progressive narrowing of the line between the scientific register and the vernacular. And we need to look for these explanations in the period following the so-called Scientific Revolution, a key moment in the history of Britain in which several dissimilar elements converged to transform the approach to science and language beyond recognition: the scientific method, the English language, and a growing interest

in the dissemination of science. Scientific knowledge had been until then based on induction, and it was substituted by experiment; English replaced Latin and Greek as the vehicle to convey knowledge and, consequently, scientific works were available to a wider audience that could not read Latin or Greek. The members of the Royal Society of London for the Improvement of Natural Knowledge, the first scientific society in England, can be set as examples of this innovative interest in the promotion of science, the use of English in scientific works, and also the advocates of a reform in high education that might lead to an improved quality in teaching and learning science.

For this reason I have chosen to analyse the writings of scientists after the scientific revolution, more specifically texts written in the eighteenth century (from 1700 to 1799). I intend to review the status of word formation in the late Modern English period, which will provide a wide perspective of the whole situation of morphology in the Middle and Modern English periods. Besides, it may be interesting to analyse the vocabulary choices of different authors to compare and contrast their style. Additionally, it is my purpose to prove the validity of linguistic theories of the times, which suggested a 'freezing' of the language to confer it with a status similar to that of Latin and Greek. This would explain, for example, the difference in vocabulary growth between the eighteenth century and the previous or subsequent centuries, apparently lower in the former. However, some modern linguists (Görlach 2001: 45) also attach this difference to the apparent lower number of studies

focused on the eighteenth century. In any case, my analysis will offer a full review of noun coinages in the eighteenth century, along with word formation processes that will, I hope, clarify the subject further with reliable data.

If we are to agree that the Scientific Revolution took place around the mid-seventeenth century we may be wondering why I have not started my study fifty years earlier. There are a number of reasons, methodological and practical, for this intentional delay. As will be explained in the following chapter, changes in language do not take place immediately after changes in history and/or society, so I expect that allowing a fifty-year gap between the revolution and the first texts studied will be a reasonable decision in order to observe significant changes in texts. Besides, compiling a corpus of texts from the seventeenth century that could meet the strict rules established by the *Coruña Corpus: A Collection of Samples of English Scientific Writing*[3] —henceforth the *Coruña Corpus* made their inclusion merely impractical. These rules, which will be explained in chapter 3, were devised by Moskowich & Crespo (2007).

[3] The Research Group for Multidimensional Corpus-Based Studies in English (MuStE), to which I am honoured to belong since its foundation in 2003, is currently developing the *Coruña Corpus* at University of A Coruña. For more information, see http://www.udc.es/grupos/muste.

Since the outbreak of the scientific reformation in the previous century English had already consolidated as a developed language[4], suitable to express scientific knowledge, but as Gotti (2005) observes, scientific production written in English or the translation of classical works will not appear until the end of the century. For that reason, the eighteenth century may represent a suitable starting point for this study. Moreover, a hundred years encompass a period long enough to achieve not only synchronic results, when comparing the data from one discipline with that from the other, but also diachronic, when contrasting the data in 30-year-long time stretches (Kytö, Rudanko & Smittenberg, 2000). Modern linguists such as Nybakken claim that word formation processes based on classical languages dominated earlier periods. They also state that English has become self-sufficient in the twentieth century, regarding the formation of new technical terms, and has relinquished classical grounds altogether. In view that the time span of this corpus comprises a full century, and before its extensive analysis can refute or confirm any established word-formation theories, it seems fairly interesting to observe the rate at which the transformation from the first state to the second took place, if any such transformation existed at all in the scientific register of English.

[4] Halliday defines a developed language as "one that is used in all the functions that language serves in the society in question" (1978: 194).

Thus, the time span selected follows unambiguous motives. In general, the late Modern English period is characterised by antagonistic tendencies regarding language. A movement that purported the customary use of Latin in scientific texts coexisted with another that suggested that disseminating knowledge in the vernacular would undoubtedly reach a wider audience. Also, to make matters worse a third movement intended to create a universal scientific language from scratch. Though the date for the vernacularisation of science has been set as early as 1375 (Taavitsainen & Pahta 2004), it is nonetheless certain that we cannot talk about an outburst of texts written in English until the turn of the seventeenth century, with a consolidation in the eighteenth, since even the most important promoters of the scientific revolution published their works in Latin in the first place (Bacon's *Novum Organum*, Newton's *Principia*). In sum, the eighteenth century seems, therefore, a fairly reasonable moment in the history of the English language to begin an approach to the morphological devices employed in the period to coin new linguistic elements. We must bear in mind, however, that changes happening in the eighteenth century are not likely to be largely observed until the following century, especially those closer to the turn of the century. Future studies using the nineteenth century section of the *Coruña Corpus* will definitely address this issue.

Considering the magnitude of my study I have resorted to corpus linguistics methodology, and my approach will be from the standpoint of

computational linguistics —or rather, computational morphology (Trost 2003)— to examine two of the key disciplines in the development of knowledge in general and science in particular. In the following pages I will use computerised corpus analysis, to verify established theories that claim affixation in late Modern English to be essentially the combination of classical bases and affixes. Indeed, the use of computer hardware and software will allow me to manage the huge amounts of data required for my analysis. Crystal (2003: 112), among others, considers that this methodology is valid "as a starting-point of linguistic description or as a means of verifying hypotheses about a language", and suitable to provide valid statistical patterns of affixation for the period examined.

I intend to examine the technical register of English, therefore I have used a 'specialised corpus' (McEnery *et al.* 2006), and I have chosen two subcorpora from the *Coruña Corpus* that I have recently helped compiling and tagging, namely, the *Corpus of English Texts on Astronomy* (Moskowich *et al.* 2012) and the *Corpus of English Philosophy Texts* (Moskowich *et al.*, forthcoming) —henceforth *CETA* and *CEPhiT, respectively*. The period covered by these two subsets of the *Coruña Corpus* spans from 1700 to 1900, but as explained above, I have only selected the subsets of samples belonging to the eighteenth century for this study. There are a good number of reasons for my choice, and they will be explained fully in the following chapter. But mainly I have chosen these two disciplines because they represent two different

approaches to science, a more mathematical one in astronomy and a more social one in philosophy. Other authors have followed similar paths in their studies. In the example reviewed above, for example, Caso (1980) focused on physics and earth sciences in his study of scientific terms for the sake of contrast.

Lastly, nouns have been chosen as the objective of our study because they make up the word category that increased the most in late Modern English, either by borrowing (Nevalainen 2006a: 53-6), or affixing. In the current research my aim is to analyse the word-formation processes used to enlarge the noun inventory in scientific writing. Word-formation processes will be identified and explained, and statistics based on experimental data will reproduce patterns of productivity. I will pay special consideration to affixation as a means to expand the coinage of nouns in the eighteenth century.

Chapter 1 revises a series of interdependent and intercalated linguistic and historical factors taking place along the history of English, which had an effect on the evolution of the language until the eighteenth century. It analyses the impact of invasions that derived in language contact, the cultural and social exchanges and migrations that led to the rise of the standard, and the view that linguists of the times had on language in general and English in particular. The influence of Latin, Old Norse and French on English are ascribed to the first group, the Renaissance and the Scientific Revolution embody the second group. I will pay special attention to the importance that the Royal Society of

London had on science and language after the mid-1700s. I will comment on the threefold transformation that its members attempted to carry out, i.e. a new research method based on experiment, a new scientific essay, and a new language, while keeping in mind their ultimate goal: a deep renovation in the education system in Britain. Besides, since astronomy and philosophy are the core disciplines in my research, in this chapter I will also deal with their status and their degree of interdependence in the period revised, while I intend to assess their roles as contributors to science and language. The last section of the chapter reviews contradictory theories about the language that should be used to write science, be it Latin, English or a new language devised by scientists. And if we agree that it must be English, are there any devices that must be used to create new words that may bear the label 'scientific', which will make them different from 'common' words in the vernacular?

Chapter 2 looks over the axioms and boundaries of morphology —with special emphasis on derivational morphology— and gives an account of the most productive processes to coin new nouns in English, bearing in mind that the main focus of this study will be on affixation. Accordingly, I will attempt at yet another definition of 'word' and explain the linguistic units taking part in the coinage of new nouns. Other main word-formation processes such as compounding and conversion will be outlined; however, I will concentrate chiefly on affixation in order to explain not only valid theories in the twenty-first century, but also those put forward by grammarians in the period under

revision in this study. In the final section of the chapter I will retake the controversial issue of 'productivity'. In this respect I intend to draw a line among disputing theories so that I can use a rather uncompromising version of the concept that may suit my data without getting into unnecessary trouble.

Chapter 3 contains all the relevant information on the corpus and the two disciplines analysed. Besides, it attempts at validating the use of the *Oxford English Dictionary* —hereafter *OED*— as a reliable resource for the analysis of word formation processes, as the valuable information that it holds about bases, affixes, and their origins makes it suitable for linguistic research. Although, the evidence of first usage provided in the samples contained in the 'date chart' section must be handled with caution, because its compilers have not had access to all existing texts, and then it will be used as mere reference. Undoubtedly, the existence of an increasing number of computerised corpora like the one introduced here will contribute to updating the *OED* in terms of vocabulary enlargement and correct dating of entries. And in fairness, I hope that my research will provide supplementary information on coinage dates, so that some of my eighteenth-century samples may contain nouns that, according to the *OED*, were coined at a later date.

Likewise, this chapter defines the term 'corpus' and explains why I have chosen corpus linguistics as the methodological approach to my analysis, and why I have selected the standpoint of computational morphology. Also, it reviews the forty-one samples and authors, twenty-one from *CETA* and twenty

from *CEPhiT*, which make up the corpus, and introduces the computer tools employed to extract and classify the data contained in it. Similarly, the chapter describes the criteria based on quantitative analyses, in line with other research works carried out in recent times, and the variables that I have evaluated so that I can establish reliable patterns of word formation and compare disciplines and authors from different angles. Finally, it explains the reasons for the selection of nouns as the core of this study.

Chapter 4 presents the heart of my analysis. It starts by describing the virtues and limitations of the *Coruña Corpus* at its early stages, especially before the *Coruña Corpus Tool* (*CCT*), the piece of software bundled with it, was fully operative. This is when I carried out the heaviest load of manual and computer tasks. Here I will attempt to filter the material stored in my wide-ranging database, so that it will offer information on a good number of variables, both intra- and extralinguistic. Examples of the first kind will be the rate of noun types and tokens in the corpus, the rate of simple and complex nouns, the processes used to coin them, their elements and their origin, coinage dates, among others. Extralinguistic variables include the comparison and contrast of the two disciplines, the genres/text-types, and the authors from the standpoints of their places of education, sex, and their ages when writing the texts. Considering the amount of numerical data presented in this chapter, it will be laid out not only in writing, but also represented visually in the shape of tables and figures, for ease of reference.

Finally, I will present the conclusions to the objectives that I have described here. I hope that the data contained in the analysis explained in chapter 4, by meeting the requirements laid out in the previous chapters of this dissertation, will allow me to obtain significant results for most of the variables introduced above. Besides, I hope that these results may shed some light on the status of English morphology in the eighteenth century since they are based on reliable quantitative data.

1. A revolution in science and language

1. Introduction

To understand the characteristics of eighteenth-century scientific writing, it is necessary to pay careful attention to earlier stages of the language. Therefore this chapter, apart from defining the terminology employed, also describes several historical and sociolinguistic facts that substantially affected language use in Britain up to the eighteenth century. It tackles a diachronic approach to the development of the scientific register of English in late modern times, and seeks to outline the language-external and language-internal factors —succinctly the former and more comprehensively the latter— that made an impact on the vernacular at various stages of its history. Due to the approach to English morphology used in this work, section 2 revises historical events such as the Scandinavian and Norman invasions that led to language contact, alongside the social evolutions and advances in knowledge that crystallised in the successive transformations undergone by science, namely, the Renaissance and the Scientific Revolution. Bearing in mind the impact of interrelated social and scientific facts on language use in Britain, I will discuss the rise of a standardised variety of English as a consequence of increasing migrations from

all corners of the isle, especially from the Midlands (Keene 2000), and the prescriptivism that originated in the second half of the eighteenth century (Beal 2010).

As stated in the introduction, I conceive of the elements intervening in the period as interdependent and intercalated. The relation among such elements can be better represented graphically in Figure 1.1 below. As we can see, the structure of this chapter mirrors a tree representing broadly the sections included in it. Its main branches are further subdivided into narrower sub-branches that illustrate the topics under discussion. At the top I have placed linguistic change as the underlying feature throughout the chapter. And then I have followed a descending line of thought to understand each of the issues that, in my opinion, contributed to make of English what it was at the beginning of the late modern period. Thus, historical, social and linguistic events that bring about changes in language are reviewed in the lower branches, and at the bottom I have placed the ultimate goal of my research, that is to say, the morphological analysis of the scientific register of English in the eighteenth century[5].

[5] See section 3.

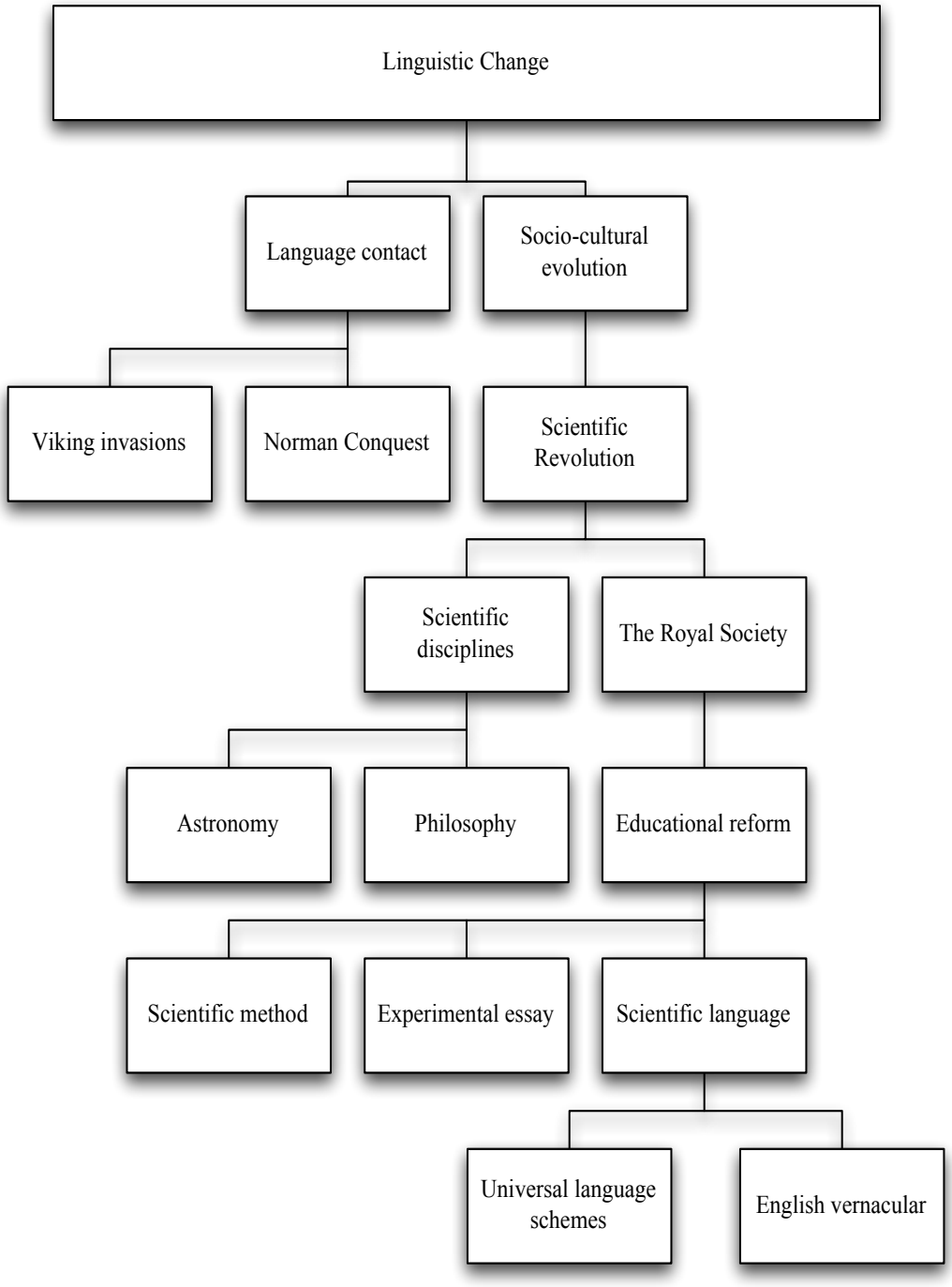


Figure 1.1: Structure of chapter 1.

Therefore, linguistic change will be explored from two major viewpoints, i.e. language contact and socio-cultural developments. In order to achieve this goal we are compelled, occasionally, to look back in the remoteness of time and space to find explanations for the roots of changes taking place so many years, sometimes centuries later, that they might seem totally disconnected from their origins at first sight. Some of these explanations are related to crucial historical events derived from migrations, warfare and military conquest, whereas others are the product of new discoveries, inventions, cultural movements and social evolution. Generally, their effects on language are felt almost immediately when it comes to the first kind, and more gradually when it comes to the second, especially in societies with a high rate of illiteracy, as is the case of the population of Britain in the periods revised here. Scientific English falls into this second category.

Aside from minor exchanges derived from commercial trade, and the undying influence of Latin in religious and scientific environments since the roman invasion, there have been two significant moments of language contact under very different conditions in the history of English. The Scandinavian colonisations in the eighth and ninth centuries brought in Old Norse on a parallel level with English, whilst the Norman Conquest in the eleventh century introduced French on a dominant position with respect to the conquered vernacular. Assessing the impact of contacts happening a millennium ago poses great complexity, due to an obvious total lack of oral

evidence and the fragmentation of written texts. Thus changes happening in this period cannot be generally observed until centuries later. For example, Moskowich (1995, 2002) maintains that the linguistic effects of Old Norse were not appreciated until the eleventh century. In the case of French, Dalton-Puffer (1996: 12) presents relative frequencies that prove its low impact on the English lexicon until almost two centuries after the Norman Conquest. With respect to my core topic, I will try to provide evidence of a similar phenomenon concerning the morphological changes undergone by the English language in the eighteenth century, as seen in astronomy and philosophy texts. I will examine whether these texts reveal a higher occurrence of lexical items affixed in previous periods or, on the contrary, they show more innovative derivations happening in their own time.

Section 3 explains the second approach, which focuses on socio-cultural development in general and, considering that I am going to analyse the scientific register of English, I will deal with the scientific revolution in particular. I will analyse the long-term consequences of the revolution that transformed the way in which natural knowledge was acquired, presented and disseminated. Besides, Section 4 will comment on the importance of the Royal Society of London in seventeenth-century England, its role in the gathering of scientific information on a first stage after its constitution, and the subsequent diffusion of that information (Hall 1975: 173). Some prominent members of the Society conceived a grander design to radically reform the existing

scholastic-based education at universities. Therefore they devised new methodologies for the acquisition of knowledge based on experimentation —inspired by Bacon's scientific method—, new formal structures for the presentation of the results achieved —represented by Boyle's experimental essay—, and a new linguistic vehicle to convey the previous two premises to an ever-increasing audience —including both Wilkins's unsuccessful analytical language and Newton's use of the vernacular. This intellectual restlessness was complemented by a social struggle led by astronomers and mathematicians to demand social recognition in scientific and academic hierarchies, reserved only to thinkers so far. In a way, empirical philosophy was confronting natural philosophy.

Notwithstanding the impact that some medical texts had in the vernacularisation of science (Taavitsainen & Pahta 2004: 11), I have considered that this discipline has already been dealt with in depth. Besides, as will be explained in the chapter devoted to my corpus, medical texts are not included in the *Coruña Corpus* because a good selection of samples from this discipline can be found already in the corpus compiled by Taavitsainen *et al*. (2005) *Middle English Medical Texts*. Therefore I have resorted to materials extracted from other two significant disciplines to build the subject matter of our study, i.e. astronomy and philosophy. Given that my intention is to analyse morphological patterns and measure the degree of relevance that these two disciplines had in the development of science written in English, I feel

compelled to summarise their status in the early modern period, and their contribution to the development of the scientific revolution itself. In short, on the one hand astronomy implements the latest advancements in mechanics and allows the most outstanding astronomers of the times to provide a comprehensive picture of the skies. The starting point had been the inheritance received from the bold asseverations issued by Copernicus, and the culmination of the period would be represented by Newton, artificer of the mathematisation of astronomical observation. On the other hand, philosophy attempts to synthesise the new changing reality of nature and reinvents itself completely, diverting from the scholastic approach based on Aristotelian metaphysics and embracing a modern empirical conceptualisation of the phenomena defining the system of the world, a system that should in turn leave enough room to allocate the existence of God.

The prevailing attitudes towards vernaculars across the early and late Modern English periods will be explained in Section 5. These will consist of disregard until the mid-seventeenth century, tolerance in the second half of that century and the first half of the eighteenth, and prescriptivism from 1750 onwards. Attempts to elaborate a pan-cultural *philosophical language*[6]

[6] This is the label that Wilkins and his colleagues applied to the analytical language they intended to elaborate, and it made perfect sense in their times. Since *science* was still

(Subbiondo 2001, 2007) will be counterbalanced by scholars that recommend the use of English, first in literature and then in scientific environments (Nevalainen 2006a: 39). This discussion will provide the language-internal content for the transition between this and the next chapter, which is devoted exclusively to linguistic issues.

2. Language contact

It is widely acknowledged that languages naturally change over time (Millward 1996; Labov 2001), and English is no exception to this rule. Structural changes may affect all levels of a language, namely, phonology, morphology, semantics, syntax, and pragmatics, but this research work will focus exclusively on the changes that involve the coining of new words from a morphological point of view. Alongside the increase in the lexicon that any language undergoes across its history, counteracted by the loss of old vocabulary, there have been outbursts of new words flooding into English at various stages, coined to denominate new inventions, actions, feelings and realities. It may sound self-evident to point out here that the external history of a language runs parallel to the internal, and that the former can generally explain many of the phenomena affecting the latter. Since language cannot be

considered *natural philosophy*, the new language dealing with natural philosophy should therefore be termed *philosophical language*.

physically detached from human speakers —so far and until science may prove my statement wrong—, any major changes affecting the welfare of human beings at a specific moment in time will most likely have an impact on their language as well, be it by means of an undesired face-to-face contact with the enemy's tongue, or by bloodless new discoveries and technical developments. And, as Beal (2004: 13) emphasises, "the influence of 'external' historical events on the 'internal' structure of a language is nowhere more transparent than in the lexicon". Bearing the above self-evidence in mind, two types of external processes, dramatic enough to alter the regular use of the language, will be briefly presented below, and their consequences analysed. Firstly, warfare and military conquest that mark the end of old eras and the beginning of new ones, and bring together the languages of the conquerors and the conquered, which interact and affect each other; and secondly, non-violent social developments, evolutions or revolutions in science and thought that may usually lead to modifying attitudes and perspectives towards life and reality in general, and language in particular.

Early examples of the first type were the Scandinavian colonisations in the ninth and tenth centuries and the Norman Conquest of Britain in the eleventh, which brought about a flood of lexical items that permeated the core of Old English (OE) in the first case, and of Middle English (ME) in the second about a millennium ago. Quantifying Scandinavian influence turns out to be rather complicated given the fact that Old Norse and Old English had a

very similar structure. Consequently, detecting whether words were native or borrowed becomes a somewhat complex task on numerous occasions. Besides, the low literacy rate of the society of the period extends the time lag between the first use of a term in oral speech and its first occurrence in writing. As a consequence, Scandinavian influence can be better observed in a later period in Middle English, especially between 1150 and 1250 (Millward 1996: 196).

The defeat of the Anglo-Saxon king Harold Godwinson against the Normans at the Battle of Hastings in 1066 resulted in the introduction of French as the official language in Britain for almost three centuries, but as it happens in the case of Old Norse, it is complex to attest the actual use of French words by English speakers, especially because the imposed language was restricted to Court —and the law courts, where it remained for centuries, since English peasants were obviously not willing to learn the language of the conquerors. Proving its use in writing is especially difficult before the year 1200 given the scarcity of texts written in English, but it becomes no easier after that year due to the time lag between the acquisition and actual use of foreign terms in either speech or writing. It might seem plain that the conquering French-speaking minority was keen on imposing its language over the defeated English-speaking majority. Besides, it is evident that numerous foreign terms intermingled with the vernacular and were phonologically adapted by the natives for their own use in their own tongue, until "the entire nature of the English lexicon had been transformed by the flood of loanwords

from French" (Millward 1996: 198), but it is also true that this transformation was very slow and asymmetric. In this respect, Dalton-Puffer (1996: 36) makes it clear that the presence of Romance words "is much larger in some styles and text-types than in others and greatest in texts that combine the features formal, written, scientific, or technical".

The Scandinavian colonisations and the Norman Conquest may appear too far back in time, and perhaps only marginally connected with the period studied here, but as will be shown in Chapter 4, which deals with the analysis of data, the influences that Old Norse and French had on the English morphology and lexicon have lasted centuries, and make them compulsory references here and now. French, moreover, had a dual influence on English. Not only did it supply new lexical items and morphemes to further develop the language, but also served as a bridge to the acquisition of yet more Greek and Latin learned terms.

Another military conquest, a distant one on this occasion, resulted in the fall of Constantinople to the Turks in 1453. Again, though this may seem too remote an event to mention here in connection with the evolution of the English language, the truth is that the ulterior flight of Byzantine scientists to Italy marked the beginning of a process that radically transformed all areas of knowledge, and ultimately reformulated the concept and perspective of science in the western world (Taton 1964). This significant historical feat provides one link between history and culture. This process embodies the second type of

external factors having an impact on the language, a revolution in knowledge that swept away the old foundations of scholastic education, first in Italy, and then spread to the rest of Western Europe, including Britain. This trend has continued to develop exponentially on the time-line until our days, with its peaks and valleys occurring within narrower and narrower intervals of time. Some of these peaks have acquired enough significance to be awarded brands by historians (Cronin 1992; Patrick 2007) as the Renaissance, which spans roughly across the fifteenth and sixteenth centuries, and the Scientific Revolution, traditionally set in the middle of the seventeenth century. It is worth noting that the emergence of technical inventions such as the printing press or the telescope contributed to the highest degree to emphasise the notoriety of these periods when compared with the Middle Ages. While the telescope provided a new wide-open window into the cosmos, the press synthesised the new discoveries in book form by means of a mass-production process, fast and reliable, that accelerated the dissemination of knowledge to an extent unforeseeable before it had been invented. But first of all the terms used to explain three hundred years old history and historical linguistics from a twenty-first century perspective need to be clarified in order not to fall into anachronistic explanations.

3. Socio-cultural developments: the Scientific Revolution

Henry (2002: 2) defines the term 'Scientific Revolution' as a brand made up by historians of science to refer to a period in which "the conceptual, methodological and institutional foundations of modern science were first established." The dates provided to delimit the period may vary from historian to historian, and some of them trace back its first stage to the fourteenth century, when the Oxford Calculators put forward their theory of impetus, which challenges Aristotle's explanations of motion (Butterfield 1965: 20). There is, nevertheless, a generally agreement on the mid-seventeenth century as the main focus of the revolution, with antecedents in the sixteenth and consolidation in the eighteenth. But before proceeding expeditiously to give a detailed account of the causes and consequences of the Scientific Revolution, it would be sensible to point out the difficulties that embracing such an enterprise may have to face with. Dealing with previous periods of history always entails a few hidden obstacles not observable at first sight, and we may need to resort to the advice of expert historians to overcome them. The first obstacle is wisely uncovered and identified by Henry (2002: 2) as whiggism, or "to judge the past in terms of the present". If due attention is paid to Henry's words and the paragraphs written above are re-read, we must acknowledge that there are at least three whiggish discrepancies in this study that need to be addressed at this point, before repetition may result in misleading arguments or incoherent

explanations. The reason for these discrepancies is rather obvious: using twenty-first century terminology to designate eighteenth-century facts might not be the most sensible choice. The word 'science' provides the first example. Though attested in written records in the fifteenth century (c1430, *OED*), its modern use was coined in the nineteenth, so it becomes an unfortunate inaccuracy to call 'science' the disciplines brought together under the canopy of natural philosophy at the time of the scientific revolution. And this, to some extent, accounts for the heterogeneity of the topics covered in *CEPhiT* (*vid*. chapter 3). The same explanation can be applied to the adjective *scientific* (coined in 1589, *OED*), which, rather strikingly for our present idea of science, excluded mechanical achievements from its ambit, and was only used in connection with the liberal arts. The second mismatch corresponds to the use of the brand name 'scientist' to refer to the learned men of science that made the revolution possible. Unlike the previous case, this word simply did not exist until the philosopher William Whewell coined it in 1833, in response to a demand made by the poet Coleridge, who requested Whewell to find a new name to substitute the previous *natural philosopher*[7]. Linking scientific and

[7] William Whewell and his work *The Plurality of Worlds* (1858) are included in *CETA* as well but, since his work was published in the nineteenth century, they fall outside the scope of this study. Besides coining the word *scientist* "to describe a cultivator of science in general" (*OED*), Whewell also invented for Faraday the nouns *anode* and *cathode* for the electrodes in

linguistic fields presents the third inconsistency as regards terminology. The term *morphology* was coined in the 1830s for its use in the field of biology. The *OED* defines it as "that branch of biology which is concerned with the form of animals and plants, and of the structures, homologies, and metamorphoses which govern or influence that form". The evolutionary model that Darwin proposed in his *On the Origin of Species*, published in 1859, provided the connexion between biology and the science of language precisely when the philologists of the times were willing to solve the riddle of the origin of language by studying the evolution of words from Indo-European down to its daughter languages. As a consequence, the term 'morphology' started being employed in the 1870s also in linguistic contexts with the meaning of "the branch of grammar which is concerned with the form of words (including word-formation and inflexion)" (*OED*). Therefore, referring to eighteenth-century word-formation processes under the label of morphology might appear somewhat anachronistic. This does not imply, however, that attention was not paid to the devices employed to enlarge the vocabulary. Most authors in the eighteenth century dealt with *derivation* or *etymology, simples* or *primitives* (Mackintosh 1797: 36-8)*, derivatives* and *terminations*, to refer to word-

electrolysis, and *ion* as a general term of *anion* and *cation* (ODNB). In a letter written to Faraday in 1838, Whewhell tells him "I have considered the two terms you want to substitute for *eisode* and *exode*, and recommend instead of them *anode* and *cathode*. [...] for the two [*anion* and *cation*] together you may use the term 'ion'." (*OED*)

formation processes and elements. Jones (1701: 125) uses the term *derivative*, explained as "Words, that come from other Words, that they agree with (more or less) in Sound and Signification, having generally (tho' not always) more Letters or Syllables than the Primitives". Watts (1753: 663) defines the same term as "one word and a syllable coming after it, which also is called a termination". Gough (1754: 67) focuses on root origin and states that "Etymology is that part of grammar which treats of the derivation of words". The famous grammarian Elphinston (1765: 218) also deals with "etymology, or the formation of words", and defines derived words as *formatives* (1765: 276). Towards the end of the century, Fogg (1796: 77) blends compounding and derivation in one, by stating that "in compound words the order is frequently reversed, and sometimes one or more particles to be supplied. Sometimes the sense is totally changed as in *understand*; and sometimes partially as in *remove*."

The rather complex situation we are left at after explaining the inconsistencies caused by the use of contemporary terminology requires some compromising decisions. As already stated above, it is not the aim of this work to provide a theoretical approach to these controversies, which would result in a laborious effort of considerable extent and, no doubt, falls more into the field of historiography than that of linguistics. Consequently the terms 'natural philosophy' and 'science' will be employed indistinctly, considering not only that the meaning of the former in early modern times is the most similar to that

of the latter after the nineteenth century, but also given the fact that the process of transformation endured by natural philosophy in the seventeenth and eighteenth centuries gave birth to the concept of science that we currently use in modern times. It may be objectionable from a strictly historical point of view, but it is not likely to jeopardise the consistency of the argumentation. Henceforth, and in consonance with this decision, the older terms 'men of science', 'natural philosopher', and the modern 'scientist' will also be treated as synonyms and used quite interchangeably in this work.

As regards the label 'morphology', we should bear in mind that biology is not part of this analysis[8], so the use of the modern term will not convey any ambiguous connotations either, and will therefore be used more or less freely across this dissertation. Further explanation on the disambiguating of old and new uses of terminology will be offered in chapter 2, when eighteenth-century grammatical theories regarding word-formation are compared with those of modern linguistics in general and morphology in particular.

The second obstacle that we may have to take on when exploring the past is that of applying labels to the temporal divisions of general history and history of the English language, as well as defining their temporary boundaries, and it usually leads to bitter disputations among historians and linguists.

[8] Biology is another discipline present in the *Coruña Corpus*, grouped together with zoology and botanics under the label 'Life Sciences'.

Whereas most have agreed on the seventeenth century to put a date to the revolutionary movement, it has likewise been argued by some continuist historians such as Crombie (1953, 1969) or Duhem (1991) that the so-called Scientific Revolution was not a revolution in the literal sense of the word simply because it did not entail a break with the past. They rather suggest that it was the result of an uninterrupted trend that had already started in the Middle Ages and continued evolving during the Renaissance. Therefore the term *revolution* is misconceived here. Taton (1964: xix) goes still further in the preface to his general history of the sciences, and extends this historical continuum until our days, by affirming that "contemporary science, which is universal in the sense that its findings are accepted and taught throughout the world, is the direct outcome of a trend that the Renaissance inaugurated". Nevertheless, those defending a non-continuist view of the history of sciences agree that the amount of changes, their magnitude, limited time span, and influence on all aspects of knowledge contain all the features necessary to consider this Revolution a tangible one. For example, Henry (2002: 2) affirms that the Scientific Revolution refers to "a very real process of fundamental change", and Kuhn (1996: 92) justifies the use of the label by pointing out some characteristics inherent to revolutions, like the observance of "non-cumulative developmental episodes" and the replacement of older paradigms "in whole or in part by an incompatible new one". All in all, if we measure some of the key features of the period, such as writing in the vernacular instead

of Latin, the use of the experimental method instead of scholastic precepts, the application of mathematical principles to all branches of knowledge, and the separation of science from university environments, and even from ecclesiastical influence, then it may be concluded that those features are indeed revolutionary in essence. In contrast, other authors such as Hård and Jamison (2005) and Patrick (2007) prefer the brand 'scientific reformation', presumably because one of the main goals and consequences of this process was the reform of the educational system in universities. In my work both terms 'revolution' and 'reformation' will be used synonymously, because both can be regarded equally legitimate, perhaps complementary, to define the spirit of the movement.

As far as the history of the English language is concerned, there is a general agreement on the labels used to term historical periods, but the dates demarcating limits between one another pose another source of controversy. I am particularly interested in the boundary between early Modern English (eModE) and late Modern English (lModE), because my materials were written in the eighteenth century, so it would be interesting to identify to which period they belong to before developing my research any further. Excepting Lass (1999), who sets the establishment of the printing press in 1476 as the start of the eModE period, most authors agree on the year 1500 as a broad start for the period (Barber 1976, Görlach 1991). Regarding the year or years that mark the boundary between eModE and lModE the agreement is not so general.

Moskowich (2001) gives several alternative dates coinciding with key events in history, such as the end of the civil war in 1660, the transition from the Stuarts to the House of Hannover around 1700 (William III died in 1702), or the United States declaration of independence in 1776, among others. She also mentions other reputed linguists that stretch the eModE period until the end of the eighteenth century, such as Adams (1973) and Görlach (1994). Since I intend to establish this border exclusively on linguistic grounds, I will consider that the number and relevance of changes happening in the language across the sixteenth and the first half of the seventeenth centuries consolidated in the second part of the seventeenth century, and these alone may be used to separate both periods. Among these changes we can include: (i) the most significant growth in the vocabulary in the history of English, "especially in the domain of learned and technical vocabulary" (Kastovsky 2006: 256), which contrasts with a steep descent in the eighteenth century, if we take into consideration recent studies on the subject (Camiña-Rioboo 2010, 2012); (ii) The Great Vowel Shift, which affected the pronunciation of the language, especially in the south of England (Jespersen 1956; Labov 1972); (iii) the establishment of a dual level in word formation, one following Germanic models and the other based on Latin and French patterns; (iv) the rise of the standard in the seventeenth century, which gave way to the prescriptivism in the eighteenth century, another feature that differentiates both periods in the history of the language; and (v) the vernacularisation of knowledge, which fuelled and was fuelled by

the increase in the literate population. Therefore, if we consider all these elements intrinsic to the eModE period, we will classify our samples under the lModE label, which will enable us to compare the latter period to the former, from which it inherits all of the changes that took place in language during two hundred years, from 1500 to 1700.

4. Science and scientific disciplines: astronomy and philosophy

Before I continue my explanation for the choice of the two disciplines that build up the nucleus of my research, as will be explained in Chapter 3, devoted to the corpus material, I may need to provide an explicit justification for the use of the scientific register of English in the corpus, and also for the selection of the eighteenth century as the time-span studied in this morphological review. I am aware of the fact that numerous studies on present-day morphology are based on recordings of live spoken conversations and/or a multitude of written texts belonging to any genre and text type in particular that any linguist may think fit. Unfortunately, it is patently obvious that I cannot resort to the words uttered by living speakers when dealing with past periods of the language. Likewise it is also plain that the amount of texts available for study, let alone genres and text types, decreases considerably the further we delve into previous centuries, mainly because the degree of literacy and the

production of texts was significantly lower than in our days, especially prior to the invention of the printing press.

Conversely, the register of scientific writing has provided materials since Old English times (Voigts 1979), and displays "almost unbroken continuity from the earliest periods to the present" (Taavitsainen & Pahta 2004: 1). Furthermore, scientific texts have incorporated the majority of the upcoming linguistic elements and fashions across many centuries, and the register has been shaped by integrating some sort of patterns that can be used as a structured framework for further investigation. It seems reasonable to believe, therefore, that the scientific register provides a solid ground on which my linguistic study can be built. The justification for a study on the morphological and lexical levels of the English language can be found in Beal's words (2004: 13):

> Interaction between speakers of English and those of other languages, whether through trade, colonization or warfare, brings words from those languages into English, whilst innovations in science and technology create the need for new words, sometimes introducing new morphological patterns into the language.

Nowadays distinguishing astronomy from philosophy may seem a relatively straightforward practice, because the topics dealt with by the two disciplines can be considered utterly different in our time. But attempting this division in an eighteenth-century environment remains a rather complex task, if

not impossible, since astronomy and philosophy did not exist as isolated disciplines in early modern times. They were regarded simply as two more elements in the multidisciplinary matrix of Natural Philosophy. This philosophy, considered by some as "the great mother of the sciences" (Grant 1996: 192, Bacon *et al.* 2000: lxxix), grouped all disciplines and attempted to give explanations about every kind of physical phenomena. An example of this unifying 'superscience' is represented by the mathematical principles published by Newton in 1687, illustratively entitled *Philosophiae Naturalis Principia Mathematica.*

Astronomy was then an indivisible part of the whole, considered one of the 'middle' or 'exact sciences', together with optics and mechanics (Grant 2007: 321), that is to say, those sciences that applied mathematical principles to nature, which were subsequently reinterpreted from a metaphysical perspective. In fact, since Ptolemy's times the mathematical analysis of astronomy had coexisted with its physical —philosophical— counterpart, which remained anchored to Aristotelian thought until Copernicus's age. To complicate things more, further investigation carried out on the astronomy and philosophy texts written in the early modern years that make up my corpus shows that the boundaries between them are sometimes more methodological and theoretical than factual. Then again, I may have made a whiggish decision as regards discipline choice for my corpus, but on this occasion it is a deliberate decision that intends to take advantage of both ancient and modern

divisions of science at the same time to fulfil my purpose of uniformity in the analysis. The modern distinction will thereby present a clear-cut division between two different disciplines, validating them as suitable subjects for our contrastive analysis. Additionally the old distribution of natural philosophy will confer the body of the investigation with the unity required to obtain more regular patterns of linguistic processes, rather than offering a disparity of elements of disciplines totally disconnected from one another, given the fact that my corpus has been shaped by incorporating mathematical and philosophical branches of knowledge on equal terms, so I hope that this may prove my choice fairly reasonable.

The eighteenth century beheld a relentless evolution of Astronomy and philosophy. Astronomy's scientific methodology began to develop as a consequence of the observational data obtained by Tycho Brahe, the discoveries of notorious astronomers such as Kepler and Galileo in the sixteenth and seventeenth centuries, and ultimately by Newton, who introduced the mathematical treatment of the phenomena studied in this field in the second half of 1700s. Butterfield (1965: 72) emphasises the importance of astronomy by stating that "the scientific revolution is most significant, and its achievements are the most remarkable, in the fields of astronomy and mechanics". We could add, however, that the development of astronomy as an independent science in itself could not be possibly understood without a parallel progress in optics. Thanks to the latter, astronomy has continued to

evolve rapidly over the past four hundred years, during which we have witnessed an unremitting modernisation of indispensable instruments, i.e. more powerful telescopes that pierce deeper in the skies and still more precise measuring devices. These, alongside the invention of new ones, enhanced astronomers' capabilities to discover more phenomena that require yet further explanation, fortunately less and less conditioned by conservative respect for pre-established ancient precepts or by the interference of biased religious beliefs. In sum, astronomy symbolised the archetype of exact science[9], originated on the grounds of observation, but methodically transforming this observational approach into accurate computable data by the application of mathematical formulae.

When it comes to explain the type of philosophy that I intend to analyse, I need to specify that I am understandably not referring to the 'superscience' mentioned above, since we have agreed that it was a huge container that encompassed all the other disciplines, including astronomy. Whitehead (1933: 73) explains that its origin is "analogous to that of science, and is contemporaneous". If we analyse his words carefully we may infer that the modern discipline that we call 'philosophy' was disconnected from the

[9] Henry (2002: 9) calls it 'mixed' science because of its dual components, mathematical and physical.

main trunk of knowledge when ‘natural philosophy’ became ‘science’. Agassi (1973: 613) considers that the segregation took place at two different stages:

> Whereas the seventeenth century is characterized by the spread of science and radicalism in the natural sciences, the eighteenth century is characterized by the spread of the same into the fields of the social sciences, "moral science," or "moral philosophy."

And I consider that his denomination for the resulting philosophy is very appropriate for my work, since it matches the topics covered in the samples: liberty, understanding, truth, women’s rights, and so on.

Until the seventeenth century philosophy had traditionally played a major role in the process of validating scientific discoveries in general, either accepting or denying them by the exercise of metaphysics, logic and common sense. Its evolution in the seventeenth century is certainly not derived from the unconditional acceptance and interpretations of new scientific findings by the rigid hierarchy of natural philosophers, most of them members of the clergy, who made use of lecture halls and pulpits to filter new discoveries and theories through religious prejudices and Aristotelian precepts. They retained the role of bearers of truth and knowledge, and were considered the “purveyors of the written and spoken word” (Jacob 1988: 11). The evolution of philosophy is derived, consequently, from other internal and external factors. On the one hand, some radical changes took place in the core of natural philosophy, and

these altered the sheer nature of the science. While Bacon set out to revise the epistemology of natural philosophy in his early works, *The Advancement of Learning* (1605) and *Novum Organum Scientiarium* (1620), Newton culminated it with his *Philosophiae Naturalis Principia Mathematica* (1684). On the other hand, social struggles within the scientific community shook and ultimately transformed the established organisation of natural philosophers. Thus, advancements in mechanics and groundbreaking discoveries achieved by scientists without formal education granted them a status comparable with that of natural philosophers. Boyle (1661: 193) shares his views on the conflict between traditional philosophers and modern experimentalists as follows:

> Whilst this vanity of thinking men obliged to write either systems or nothing is in request, many excellent notions or experiments are, by sober and modest men, suppressed, because such persons being forbidden by their judgment and integrity to teach more than they understand, or assert more than they can prove, are likewise forbidden by custom to publish their thoughts and observations, unless they were numerous enough to swell into a system.

Furthermore, private individuals began to sponsor the study of the heavens, and institutions alien to university environments, such as the Royal Society of London, were founded. Two main consequences can be drawn from the facts mentioned above: scientific knowledge is not exclusive of universities any longer, and we are observing a slow but steady process of laicisation and

universalisation of science (Camiña-Rioboo 2012). In essence the Scientific Revolution is not only a revolution based on inventions and discoveries, but also, perhaps primarily, a transformation of thought. This transformation is deeply rooted on scientific —experimental— grounds, rather than on philosophy or religion, and its ultimate origination may be ascribed to the new study of motion, both in the sky and on earth (Butterfield 1965: 129).

5. The Royal Society of London

Europe witnessed an explosive development of science in the seventeenth century, and scientific associations, such as the Academie des Sciences in France or the Accademie dei Lincei in Italy were founded all over the continent. But probably none of them managed to achieve the degree of significance that the Royal Society of London for Improving Natural Knowledge accomplished in Britain. According to Hall (1975: 173) one of the reasons for its success is that the association "encouraged and honoured the best scientific brains in a seminal period in science, when an extraordinary number of these brains were English". Besides, the renovation of scientific epistemology has been ascribed to the Society in the figure of its spiritual guide Francis Bacon, as well as the fatherhood to new methodologies in the communication of scientific facts, such as the experimental essay and the scientific article and journal, represented respectively by Boyle's prolific

works (Gotti 2001) and the *Philosophical Transactions* (Valle 1999, 2006). The Society had also a substantial impact on the linguistic level and, in spite of the fact that the *Transactions* still contained some material published in Latin, some linguists (Halliday & Webster 2004: 145) consider Newton's *Opticks* (1704) the true origin of the English scientific language. Finally, it is particularly interesting to include here this brief account of the significance of the Royal Society, given that its fellows wrote a good number of the texts that build up my corpus.

Its origins have been comprehensively explained in two extensive works written by two historian fellows to the Society, Thomas Sprat (1667), commissioned in the early years to publicise its methods and achievements, and Thomas Birch (1756), who was also secretary to the Society in the second half of the eighteenth century. According to these accounts we learn that it originated in a group of scientists who first met in Oxford and London around 1645, and consolidated in 1660. The new association was heavily impregnated with Baconian philosophy and, at the time of its foundation, the first and foremost concern was the collection of data (Hunter 2007: 1), which later evolved into a willingness to disseminate the scientific achievements derived from the observation and analysis of the data collected for experiments. These two practices, however, were only partial stages of a more ambitious plan devised originally by Bacon, who was later joined by Webster and Wilkins, among others, to reform science and education in Britain. Bacon called it the

'Great Instauration' or 'Great Renewal' of learning (Bacon *et al.* 2000: viii) and it has been defined and explained by Subbiondo (2001: 274-5) in the following terms:

> A new paradigm of education emerged out of the seventeenth-century empiricism. This paradigm was as an alternative to the traditional method of teaching: the scholastic disputation. […] Wilkins and his colleagues challenged not only course content, but also the prevailing styles of teaching and learning. Most importantly, this new paradigm integrated science, educational reform, and philosophical language.

The ambitious enterprise to reform science and education purported by the members of the Royal Society was founded on three pillars: a) the methodology employed to deal with scientific facts, b) the vehicle to disseminate the results of the experiments performed and the knowledge acquired, and c) the language employed to communicate those experiments and knowledge. The scientific method, the experimental essay and the philosophical (scientific) language represented those pillars, respectively. Bacon provided a full explanation concerning the first one in his *Novum Organum*, which established the epistemological foundations of the new reformed science. These involved the abandonment of philosophical introspection and the embracement of experimental induction, in order to provide theories, principles and data based on evidence, from which correct

generalisations could be derived (Gotti 1996: 172). Accordingly, this change of attitude towards the perception of science brought about a need for renovating the way it should be transmitted, i.e. scientific texts. Bacon had already tried to institute the aphorism as the new expository form to fulfil these needs, but he was only partially successful. The new writing device was useful enough for short comments based on personal observations, but too short to describe experiments appropriately (Gotti 2001: 222).

Conscious of the limitations of the aphorism, Boyle undertook the task of renovating one of the longer traditional genres. He decided on the essay, because treatises and dialogues[10] were either too long or too rhetorical for his purposes. But his attempt may have rather meant the birth of a new one, the experimental essay (Moessner 2006: 60), engendered in his *Proëmial Essay...with Some Considerations Touching Experimental Essays in General* (1661). He gave directions on structure and linguistic style and form, and anticipated some of the linguistic features present in Newton's *Opticks* (1704), which has been in turn considered by some authors as the birth of scientific English (Halliday & Webster 2004). The following sections will expand the

[10] In spite of the fact that essays, treatises and dialogues contribute fifty per cent of my material, other scientific genres/text-types have been included in the corpus, i.e. textbooks (11), lectures (1), dictionaries (1), articles (2), letters (1) and tracts (1). These genres/text-types will be dealt with in Chapter 3.

three main elements of the scientific reformation, relating them to the main figures that contributed to their development.

5.1. The scientific method

The question of a general scientific method had already drawn the attention of many conscientious men of science in the sixteenth century, but in the seventeenth it became one of the key issues. Francis Bacon remains for many, including his contemporaries and present-day writers, the pioneer of the scientific method. When he set out to present a starting point for scientific investigation in his *Novum Organum*, not only did he reformulate the concept of science, which had until then progressed along the path of scholasticism, but also contributed to establish the methodological infrastructure for the new science. Earlier scientists had made use of observation and experiment as a means to anticipating conclusions based on induction, sometimes merely to complement existing theories. Inductive reasoning anticipated probable conclusions starting from true premises, but relied on preconceived assumptions instead of experimentation to achieve those conclusions, which made it unsuitable for science (Gower 2002: 14). Bacon —and Newton after him— attempted to eradicate induction from science, because he considered that its method of hypotheses was founded upon probability rather than truth. Contrary to the customary trend, he viewed observation and experiment as

indispensable steps for deductive reasoning, "pre-requisite for the construction of scientific theory itself" (Bacon *et al.* 2000: xv). In other words, generalisations could be drawn out of experiments. Therefore he proposed an alternative based on deduction, his laureate *scientific method*, defined in the *OED* as a "method of procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses". In essence the scientific method is based on four solid foundations, i.e. the systematic observation of natural phenomena, the elaboration of a hypothesis derived from observation, the tests or experiments with data to validate or invalidate the hypotheses proposed, and the communication of the conclusive remarks, whether the experiment is successful or not, modifying the hypotheses formulated if this was necessary. The communication of results must include the methodology and data employed to carry out the experiment, so that other members of the scientific community may perform a test on the same data by means of the same methodology in order to obtain the same results. Otherwise the experiment will not meet the requirements set by the scientific method.

Many historians and disciples have praised Bacon's role as renovator of science. Among the first, Sprat ([1667] 1772: 35) acknowledges that there have been only a few philosophers that dared to pursue the right path of experimenting, diverting from the repetitive path of ancient philosophers,

> And of these, I shall only mention one great man, who had the true imagination of the whole extent of this enterprise, […] and that is the Lord Bacon; in whose books there are every where scattered the best arguments, that can be produc'd for the defence of experimental philosophy, and the best directions, that are needful to promote it.

Among his most reputed disciples we can mention Whewell, Herschel or Mill (Snyder 2009: 59), all of them included in the Coruña Corpus, who also acclaim the relevance of their instructor. For example, Herschel (1830: 105) somewhat extravagantly links the birth of natural philosophy with the publication of Bacon's *Novum Organum*, which will be looked upon in future times as the seed of the reform in philosophy. Whewell (1840: 392) concluded that "if we must select some one philosopher as the hero of the revolution in scientific method, beyond all doubt Francis Bacon must occupy the place of honor". Finally, Mill (1843: 378) awards Bacon the title of "founder of inductive philosophy".

Beyond the praises issued by colleagues and followers, a few considerations must be observed to ascribe so much significance to Bacon in the reformation of science. Firstly, by questioning the divorce between observation and explanation, he proposed a consistent formulation for the conception of the experimental method, which tightened the bond between experiments and theory. Secondly, he introduced the idea of collaborative work

to develop science, which he put forward in *The New Atlantis* (1626), and which materialised in the meetings of the Royal Society to present new experiments that would be debated by the attendants. Thirdly, he proposed the use of the vernacular as the language of science to reach a wider audience. At the turn of the seventeenth century, Newton, another prominent fellow of the Royal Society, put forward a new approach to science, which replaced the Baconian research agenda (Valle 1999: 97). The method devised by Bacon still regarded mathematics as a 'servant' to natural philosophy (Gower 2002: 44), whereas Newton's system of the world was explained according to his mathematical principles, as seen in his *Principia*.

5.2. The experimental essay

As stated above, Bacon contributed greatly to the establishment of a new methodology to carry out scientific experiments based primarily on the observation and systematic verification of hypotheses (Lareo & Montoya 2007: 121). His ideological contribution set the grounds for a new scientific epistemology that involved a shift from metaphysical rhetoric, based on introspection, or, as some authors put it, author-centred (Atkinson 1999: xxvi), to a more modern object-centred discourse, relying solely on empiricism, which has remained customary till our days. As a consequence of the radical change in the perspective and focus of scientific facts, a parallel evolution was

needed to adapt the method to communicate adequately the results of the experiments carried out in the light of the experimental method. At this point Boyle provides the link between ideas and the physical world by materialising Bacon's ideals on paper.

The fellows of the Royal society were conscious of the fact that "essays written up to then mainly followed principles and employed techniques of a prevalently literary type" (Gotti 1996: 55), so they agreed on the imperative necessity of renovating also the way in which contemporary scientific knowledge should be conveyed to the world, as is made explicit in the *Charters and Statutes* (1728: 48) elaborated at its foundation. Chapter V deals with the nature of scientific communication, by stating that "in all Reports of Experiments, to be brought into the Society, the Matter of Fact shall be barely stated, without any Prefaces, Apologies, or Rhetorical Flourishes". Robert Boyle is perhaps the most influential member in the founding group, and he lay down the grounds for the experimental essay. For this reason he is considered the father of the new genre (Gotti 2001, 2003, 2005). His *Proemial Essay* follows Bacon's guidelines on objectivity and empiricism and concisely defines the proper approach to natural phenomena by scientists, "who having sufficiently conversed with books, are now desirous to begin to converse with things themselves" (1661: 192). Boyle (1661: 193) criticises the lack of objectivity and all-comprising nature of the essays written so far by stating that,

> When men, by having diligently studied either chymistry, anatomy, botanicks, or some other particular part of physiology, or perhaps by having only read authors on these subjects, have thought themselves thereby qualified to publish compleat systems of natural philosophy, they have found themselves, by the nature of their undertaking, and the laws of method, engaged to write of several other things than those, wherein they have made themselves proficient.

Many times, he continues, readers have been mislead by the generality of the title and grandeur of the undertaking, offering comprehensive methods of natural philosophy, but in essence the topics covered are usually far from what has been promised. He also censures the lack of originality found in traditional essays by stating that authors had so far "been reduced, either idly to repeat what has been already, though perhaps but impertinently enough, written by others on the same subjects" (1661: 193).

Given the linguistic nature of this research work, I am primarily interested in the linguistic characteristics of the experimental essay. They will be dealt with generally now, touching only stylistic features, and more in depth later in the section referring to the status of language in the eighteenth century. Boyle himself provided a set of recommendations in his *Proemial Essay... with Some Considerations Touching Experimental Essays in General* (1661: 195) on how experiments should be presented in writing, (1) avoidance of repetition as "the reader needs not be clogged with tedious repetitions of what others

have said already", because authors tended to compose long treatises in which they revised the theories of previous writers, sometimes on topics in which they did not particularly excel, advancing little of the own material, which needed to be contrasted with approved versions already published. (2) Solid foundations based on true experiments. Scientists must concentrate on the results obtained from their experiments, "for let his opinions be never so false, his experiments being true". He also advances some directions on style, which Moessner (2006a: 61) considers "scattered and imprecise", but relevant enough to summarise them as follows:

> Rhetorical embellishment should be avoided, but some adornment helps to keep the reader's attention, expressions should be short, but sufficiently explicit that the intended meaning is unambiguously conveyed; loanwords should be avoided, but those with a wider currency are available.

5.3. Attitudes towards Language in modern times

The communication of scientific achievements is considered to be reliant on three main elements, namely, the nature and method of the experiment, the means to convey the results achieved, and the language employed. Once depicted the empirical method to conduct experiments on nature, and the characteristics that essays must meet to describe those experiments and explain their results, it seems interesting to give an account of the status of language in

the period studied here. Dalgarno *et al*. (2001) stress the highly critical attitude that natural philosophers showed towards language in general. Philosophical and religious factors alike bestowed their influence on these theories, by which nature was a consequence of God's design and, as such, it impersonated perfection, whereas languages were the outcome of men's doings, therefore defective. For that reason linguistic activity was considered inferior to the knowledge of nature and, as a result, words were also inferior to the things they intended to denote. Until the seventeenth century most works on linguistics had been visibly biased by religious influences. Language change, a key element to modern linguistic studies, was viewed as a mere corruption and, consequently, the changes undergone by vernaculars over time diverted them relentlessly from the original 'common language' of the Scriptures to which all of them could be traced back, that is, Hebrew and its prestigious interpreters, Greek and Latin (Hickey 2010: 3). It must be noted that derision is not devoted exclusively to English; in the preface to his grammar Elphinston (1765: viii) describes Italian and Spanish as degenerate daughters of Latin. Besides, there was a general belief that learned languages were more suitable to express knowledge, because they "have now ceased to be vulgar, and remain only in books, by which the purity of them is regulated, [and] may, whilst those books are extant and studied, continue without change" (Wilkins 1668: 6). As a consequence of the previous, Latin patterns were imported into their texts

written in English, and this can be seen in the astronomers' and philosophers' writings in our corpus, as will be shown in Chapter 3.

But this veneration for dead languages was not unanimous. It coexisted with a long-lasting debate known as the *inkhorn controversy*, which spanned from the mid-sixteenth to the mid-seventeenth centuries. Scholars adhering to the controversy demanded the use of plain English and the elimination of Greek and Latin terms from the language and the adoption of Anglo-Saxon roots for new coined words (Vos 1979; Beal 2004: 19; Nevalainen 2006a: 39). In view of the widespread disregard for language, there can be little doubt that the main reasons for the development of the scientific register of English are purely pragmatic, as it was regarded the best means of expression to reach a wider audience. Again, the Royal Society pioneered its use in their *Philosophical Transactions*, the monthly journal born in 1665 that included summaries of their latest meetings and letters sent by prominent scientists. The language employed for publication was chiefly English, though texts in Latin sent by foreign contributors were equally accepted for publication, and a great deal of translation work was done to and from the vernacular in order to favour the diffusion the information to an ever-increasing foreign readership (Hall 1975: 185-6).

5.3.1. The 'philosophical' language

Immersed in this chaotic but mostly derogatory context regarding vernaculars, seventeenth-century scientists engaged in enthusiastic debates on the necessity to create a universal language that could go back in time to restore the biblical connection between words and things, a language in which "the expression of our conceptions by marks […] should signifie things, not words" (Wilkins 1668: 21), or rather, in which concepts should express the semantic characteristics of nature. They strongly believed that learning was hindered by two main factors in connection with the linguistic domain. On the one hand, a multiplicity of languages that prevented understanding across countries, consequence of Babelian times. And on the other hand, an excess of ambiguity and redundancy in existing languages that disqualified them as the right means to describe nature truthfully. Therefore they needed to seek a more perfect language in order to achieve a likewise more perfect natural philosophy (Stillman 1995). Descartes, Mersenne and Comenius were famous advocates of the universal language scheme on the continent, whereas George Dalgarno and John Wilkins emerged as the most fervent campaigners of the movement in England. Dalgarno and Wilkins collaborated in their early years, but later followed a divergent line and eventually ended up criticising bitterly each other's different approaches. Dalgarno's main objective was simplicity. He aimed at designing a language that could be easily learnt, as can be deduced from his *Art of Signs* (1661). In contrast, Wilkins intended to create a language

suitable for science, which could in turn become one of the angular stones of his greater plan to reform high education in Britain (Subbiondo 2007). This plan found support in the bosom of the Royal Society, of which he was a member. Boyle and other notorious members of the Society shared with him the idea of the necessity to cleanse the human fallacies and vagueness imprinted on language across time by continual use (Botvina 2005).

It took a long time to discard these theories. Almost a hundred years after the publication of Wilkins's *Essay* John Bevis wrote in the introduction to his *Pocket Dictionary: or English Expositor* (1765: 7) that the history of the country and the diverse origins of the peoples inhabiting it across centuries had furnished English with tools necessary to stand out as the most significant language in Europe, "adapted to all subjects, and expressive of every sentiment with elegance and propriety." But it seems remarkable that still one century later we can find instances of criticism towards the vernacular in the writings of some members of the Society. Bevis (1765: 4) contradicts himself when he writes "our mother tongue has within half a century been much refined and changed, whether by discharging antiquated words, coining new ones, or adopting them from abroad". It may be supposed that the refinement referred to by the author consists of some sort of purification of the language, discarding Germanic terms ('antiquated words') to import Latin, Greek and French counterparts, which would suggest a closer bond with the dead languages praised by Wilkins. But the following lines are even most striking, and take us

back to a debate the seemed departed years ago, when the author shows (6) some sort of melancholic determinism for the lack of a universal language,

> The idea of an universal language, cou'd such a one possibly obtain, should seem to imply something highly beneficial to the human race: But eternal unerring wisdom, either for our advantage or punishment, has determin'd against it, and appointed to every nation a particular tongue, and to each district a peculiar dialect.

We must still keep in mind that one of the main objectives of the Royal Society was the dissemination of science. But prejudiced as they were by the prevailing theories that regarded vernaculars virtually as second-class languages, they considered that English simply lacked the quality required to communicate scientific facts. It is no wonder, therefore, that they embraced cheerfully the task proposed by Wilkins, since the prospect of developing a language that might contribute to their ultimate goal seemed very appealing. Viscount Brouncker, second president of the Society, commissioned Wilkins in 1662 to create a *philosophical* —scientific— language. And he accepted the task gladly, because it fulfilled his ultimate aspirations to contribute to educational reform. He undertook the commission with the assistance of his mentor Seth Ward, and also John Ray, Francis Willoughby and several other members of the Society, and in 1668 he published his *Essay towards a Real Character, and a Philosophical Language*. Stillman (1995) claims that Wilkins's project

represents "the Restoration's most massive effort to create a new language and a new logic on the promise of securing universal benefits in nature and culture" (185).

If we analyse the *Essay* from a present-day post-Saussurean perspective, when the arbitrariness of the linguistic sign with respect to its referent has become widely accepted, the naivety of the attempt seems almost embarrassing. But some linguists review the *Essay* in a more positive light, declaring that Wilkins actually managed to create "a comprehensive philosophical language based on an extensive classification of the knowledge of his day, including the scientific learning that emerged from groundbreaking research in astronomy, biology, and zoology" (Subbiondo 2001: 273). And even his modern detractors acknowledge that "Las palabras del idioma analítico de John Wilkins no son torpes símbolos arbitrarios; cada una de las letras que las integran es significativa, como lo fueron las de la Sagrada Escritura para los cabalistas." (Borges 1993: 104). Despite the failure of his analytical language as a means to convey knowledge across cultures, the *Essay* accomplished at least two significant feats. Firstly, by undertaking the project together with other colleagues, Wilkins established an innovative model of collaborative investigation inspired by Bacon's *New Atlantis*, in contrast with the customary individuality that had prevailed in scientific research until then. Secondly, by building this new language on scientific tables, he integrated in one single volume, and in English, most of the existing scientific knowledge

and achievements in various key scientific disciplines in the seventeenth century, especially in astronomy and what we have labelled 'life sciences' in our corpus.

5.3.2. The vernacular as the language of science

The transition from Latin to English as the language of science was neither smooth nor undisputed. But perhaps a clear-cut line must be drawn to differentiate the situation of the vernacular in the seventeenth century from that of the eighteenth century. Once again we may need to resort to Bacon, who had already proposed the use of English in the process of learning in his *Novum Organum*. His proposal may seem somewhat contradictory as he did it in Latin, but it is not less true that contradiction is a habitual characteristic of transitional periods. Already in the seventeenth century we can perceive a divergent process occurring in the linguistic domain, in part because of the disappointment derived from the futile efforts to create a universal language, in part as the outcome of the growing recognition that writings in Latin would not reach the vast majority of the population. As a consequence men of science started to look back to English as a suitable means to convey knowledge. The vernacular was still imperfect in their view, but it could be furnished with the terms that were lacking to implement their experimental theories. To fill this gap numerous publications on language tried to give solutions to some of the issues that natural philosophers needed to address when attempting to express

scientific knowledge in English. Hickey (2010: 3) summarises the most important of them as "(i) the difficulties in English vocabulary caused by the many creations and borrowings from classical languages, (ii) the pronunciation and orthography of English and (iii) the nature of English grammar compared to classical languages, above all Latin."

The so-called vernacularisation of science, however, poses another source of disagreement among modern linguists, who attach a disparity of dates to what they consider the most precise date of birth of the scientific register in English. Halliday (2004: 145), for example, affirms that "for registering the birth of scientific English we shall take Newton's *Opticks* (published 1704; written 1675-87). Newton creates a discourse of experimentation". Conversely, Taavitsainen & Pahta (2004) acknowledge that English had already been used to express scientific knowledge of some sort since OE times, and they regard the year 1375 as the starting date for the emergence of the scientific register. They observe a change in the pattern of scientific —medical— texts from 1550 onwards and stress the significance of the *Philosophical Transactions* in the instauration of a new tradition in scientific writing (Taavitsainen & Pahta 1997: 73-4). They also contemplate the influence that Latin, in its role as lingua franca of science in Europe until the eighteenth century, had on scientific English at the lexical and discursive levels. Indeed, the first grammars written in English were elaborated to assist students in learning Latin, but by the eighteenth century there was a general

recognition of English not as a corruption of Latin, but as a language with its own linguistic idiosyncrasy (Bailey 2010: 191).

While controversial and, more often than not, disparate tendencies had dominated the linguistic activity in the seventeenth-century, some authors (Tieken-Boon van Ostade 2009) have defined the eighteenth century as the "age of prescriptivism". As we have mentioned above, men of science replaced Latin with English mainly for pragmatic purposes, but they were still anchored in the old-fashioned linguistic model of perfection. This meant that the vernacular needed to accomplish the revered 'purity' of Latin to be considered suitable for scientific communication, and this could only be achieved by retaining some of the characteristics that had made Latin successful for so many centuries. Bailey (2007: 31) observes a peculiar connection between role models in society and perfection in language built upon linguistic immobility, by stating that "exemplary users of the language need to be identified, celebrated, and imitated. Through their example, English may be prevented from changing and raised to an ideal of purity." Intellectuals of the times feared language change above all evils, as they considered that change and corruption where closely linked, so they were convinced that it was a must to prevent further changes in the language, and that it could be achieved (Bailey 2010: 182). Therefore, measures were adopted to put forward the best variety of the language. These involved a standardised use of a language in which, auspiciously for them, the major changes had already taken place (Hickey

2010: 2), the last of which, the Great Vowel Shift, had been completed by 1750. Besides, Taavitsainen (2000) explains the temporal coincidence in the processes of standardisation of the English language in general and of spelling in the Central Midlands, as seen in some text types studied by her, so she suggests that the Central Midland Standard played a major role in the standardisation of the national language. It seems interesting to note that, for some authors (Hickey 2004: 352) standardisation —or supraregionalisation, as he calls it— is a process of language change "despite the fact that scholars that he consciousness for it has not been very high".

In an attempt to eradicate 'imperfections' from the language, prescriptive grammars and dictionaries were regularly published throughout the eighteenth century, among the most influential of which were Samuel Johnson's *Dictionary of the English Language* (1755), Robert Lowth's *Short Introduction to English Grammar* (1762), and Thomas Sheridan's *A Rhetorical Grammar of the English Language* (1781). At this time there was a great deal of geographical mobility, and floods of immigrants from all over the country poured into London in search of employment or education, and visitors attended social events, did business or simply went shopping in the big stores. The capital became a melting pot from a social, cultural and linguistic point of view, and the speech of its inhabitants came to be looked up to. This does not mean that it was a London-born standard, since the majority of immigration came from the Midlands, it was indeed this variety that took over the rest. It is

not very clear why dialects from the Midlands managed to permeate the speech of Londoners, but some scholars (Pahta 2001: 208) emphasise the nationalist nature of the Midlands Standard, which can certainly be related to the beginnings of the standardisation based on the nationalist politics of the Lancastrian kings in the fifteenth century. Other commentators on language (Samuels 1963) claim nevertheless that the rise of the English Standard is based on the writings of the clerks at the Chancery. Meanwhile Wright (1994, 1996) establishes a link between the development of the standard and the *macaronic* business writing used by London merchants, caused by the change of "trade patterns, especially that of the marked increase in the commercial influence of London on the country as a whole" (Wright 2001: 189). Nowadays, however, there is a common agreement among researchers, who reject previous theories relying on single origins or unitary processes to explain the birth of the standard (Hope 2000). Also, Moskowich & Montoya (2003: 14) outline a multifactor approach to the question of the standard. Following Haugen (1996: 76) they apply four criteria in their study that "any variety must fulfil in order to say it is fixed, or in other words, that it is a standard". These are: (i) the selection of a particular variety over the rest; (ii) the codification and (iii) the elaboration of that variety to fulfil all linguistic purposes, and (iv) its wide acceptance as the only variety for any language use.

Leaving aside how it came to happen, the fact is that the language spoken and the pronunciation used by the established —also known as

'polite'— classes in English society in the metropolis became the paradigm of correctness. Pronouncing dictionaries were published, adding more prescriptive rules to grammars and lexicographic dictionaries, and there was a general concern among scholars for spelling reform to harmonise sets of pronunciations from a phonetic point of view. Citizens intending to climb the social ladder would be willing to imitate the language usage of their betters, and other regional varieties were regarded as inadequate already in the first half of the century, and increasingly disdained in the second half, when prescriptivism grew exponentially. To make this point clear, Hickey (2010: 13) indicates traces of "deliberate neglect of regional features" and "severe condemnation of all traits of language which do not correspond to 'standard' usage". Moreover, Beal (2010: 30) quotes Sheridan's *Lectures on Elocution* (1762: 30) when he refers to the "difficulties of those who wish to cure themselves of a provincial or vicious pronunciation". Dialectal regionalism or any other variety that deviates from the standard is considered as "vulgar", one of the keywords in late-eighteenth-century prescriptive times.

English Standard and scientific English have undergone similar processes and have contributed to each other's welfare, the former providing a prescriptive background for the secure establishment of the latter, and the latter contributing a learned register to enlarge the prestige of the former. In the seventeenth century the members of the Royal Society made a conscious effort to create a learned register of English, which needed to be supplied with

vocabulary to rival more prestigious scientific languages in use. Kearney (1964: 151) considers that,

> Newtonian science, by which is meant both the work of Newton himself and of those who successfully applied similar methods in the eighteenth and early nineteenth centuries, explained the natural universe by means of a limited vocabulary of terms. As observation intensified and the field of exploration widened, it gradually became clear that this vocabulary was insufficiently flexible to deal with the great stream of new phenomena in which scientist were interested.

According to Taavitsainen (2000: 132), this kind of efforts derive from nationalistic policies, so we may infer that by creating a scientific variety of English, they were also helping to fix and strengthen the standard. Also, the relationship of the non-scientific and the scientific register has been abstrusely symbiotic throughout history. On the one hand technical English has benefited from non-technical English and its unrestricted freedom to make use of native and foreign elements in order to enlarge the lexicon, and the latter has also provided the former with an expedient path towards the acceptance of Greek- and Latin-based terms. On the other hand, even though some authors (Van Dyke 1992: 383) claim that "the contribution of science and technology to the English vocabulary has not been established" there can be little doubt that non-technical English has benefited greatly from technical English, and terms

fundamentally scientific, such as *aspirin*, *gene*, *neurosis*, *penicillin*, or *vitamin* (Raad, 1989: 131) have become part of the daily language, and have consequently lost the tag of 'technical' that they used to hold in the past. If we pay attention to the opening sentence by Keene (2000: 93) in his chapter on migration and social values, when he claims that "ideas of language entwine with those which shape the identity of peoples", we must also agree to this reasoning by Taavitsainen (2000: 132) in the same volume that says: "creating a learned register involves a conscious effort of a nationalistic language policy"

5.3.3. Scientific vs. non-scientific language

As I advanced above, the object of my research is to analyse word-formation processes in the language present in scientific texts. We ought then to assume that there is a set of characteristics that differentiate the scientific variety of English from Standard English at the morphology level. It must be so then, that when scientists achieve new discoveries, invent new mechanical instruments, and identify recently discovered celestial bodies or put forward new theories, they can resort to devices exclusive to science to coin new words. We must believe in the existence of a scientific English that differs from the vernacular, and it is scientists that have the tools to create new words in order to suit their specialised needs. If such a variety exists, which are those features that make it specifically scientific? Taavitsainen & Pahta (1997: 71) state that "modern scientific writing can be characterized by its lexicogrammatical features".

Basically, technical language deviates from the vernacular insomuch as the sources employed to produce new elements on a highly specialised level are different from the non-scientific registers. Already in his *Proëmial Essay* Boyle (1661: 196) hinted at the implicit validity of introducing technical terms from learned languages —he refers to them as *exotick words*— "whose energy cannot be well expressed in our language, at least without a tedious circumlocution". These learned terms, however, must only be used in English scientific writing "when custom has once made them familiar and esteemed" (Boyle 1661: 196). It is generally agreed (Jespersen 1933) that the major sources of new words in eighteenth-century English were Greek and Latin, "languages that had been plundered by intellectuals since the Renaissance" (Bailey 2010: 186). We may doubtlessly add French to the equation, and with a double relevance as donor of lexical forms and affixes, and as interpreter or vehicle of the other two, especially Latin. The reason for the use of these classical languages is very simple, as most scholars had knowledge of either of the two, or both. Consequently, not only did they import loanwords from them, but also many new terms coined followed word-formation patterns based on classical grounds. Nybakken (1959: 12-23) explains somewhat vaguely the characteristics that make vernacular terms unsuitable for scientific production. When compared to their technical counterparts, he says, common words are less descriptive, less precise and stable, emotionally loaded, and generally polysemic. Sometimes a whole set of popular names can be used for the same

entity, which renders the meaning inconsistent. He emphasises that technical and non-technical terms should not be employed to prevent unpleasant mixtures, and dismisses the use of vernacular terms in order to convey meaning to a foreign readership. Conversely, he continues, technical terms must meet strict properties such as descriptiveness, specificity, linguistic correctness, economy, purity, euphony, economy, adaptability and stability. Unfortunately, this explanation does not allow us to understand the connotative differences between scientific and vernacular terms. Raad (1989: 129) specifies that,

> In assessing the degree to which a term is considered scientific or technical rather than vernacular, one cannot search for any intrinsic quality in the term itself, for the signification of words does not lie in any objective trait […] but rather in convention and accumulation of sense over time. The intention of the scientist has been denotative —that is, to produce a one-to-one correspondence between the term invented and the object or fact to be described with a minimum of interference or ambiguity.

Even though this research deals with scientific terminology, I consider that a distinction is not necessary at this point. Since all the samples analysed have been extracted from texts regarded as scientific beforehand, all the terms and word-formation processes included in them can be regarded as scientific as well, although many of them may be also found in non-scientific texts, especially in later periods.

2. Units and processes

> Any discussion of word-formation makes two assumptions: that there are such things as words, and that at least some of them are formed.
>
> Laurie Bauer, *English Word-formation* (1983)

1. Introduction

In the previous chapter I observed how innovations in science in early modern times brought about a necessity for new terms to designate new discoveries, and to explain the theories derived from them put forward by natural philosophers. I analysed some language-external factors that triggered linguistic change, paying special attention to the scientific revolution, and commented on the emergence of the register of scientific writing, even though authors do not generally agree on starting dates. This chapter deals with the fundamentals of derivational morphology, and revises the word-formation devices employed to coin new terms for those discoveries and theories. I will attempt to define the linguistic units taking part in word-formation processes in general, and more specifically the devices employed to coin new nouns in the

scientific register of English in the eighteenth century, as observed in astronomy and philosophy texts of the period. I will also try to delimit the boundaries between the inflectional and derivational branches of morphology, and those between morphology and other levels of analysis of the language, i.e. phonology, syntax and semantics. These boundaries may seem well defined at first sight, but deeper study shows some overlapping areas between them. Other elements relevant to word formation will also be explained, such as productivity and blocking.

My focus will be on affixation, as it has always been the most productive word-formation process to enlarge the lexicon, but I will also deal with other processes not involving exclusively affixes, or no affixes at all, such as compounding, conversion and various types of reanalysis. Affixation will be studied from the perspective of current linguistic theories on language change and variation, and the terminology dealing with units and processes has been extracted from the latest theories on morphology. However, it is my intention to respect the nature of the body of the texts under study. Therefore I will also take into consideration the theories endorsed by seventeenth- and eighteenth-century grammarians. These were aimed at keeping the status of the language unchanged in order to preserve its purity, whereas change implied 'vulgarisation' or degenerating the language. For their discussion I will often refer to the study on linguistic terminology between 1600 and 1800 by Sundby (1995). A mass terminology is explained in this chapter, and it may appear

somewhat confusing, therefore a diagram summarising some of the main information is provided at the end of the chapter.

2. Units

2.1. Word

Throughout this research work I will be frequently referring to words, so it seems necessary to explain what is assumed by my use of this label. Defining the term *word* from the standpoint of literacy may be considered a fairly simple task nowadays. Sapir (1921: 34) justifies this ability by means of some sort of "psychological validity" intrinsic to speakers, which allows even those without reading or writing skills to recognise words as such in an intuitive way. Most speakers of a language are aware of their continual use and generally take them for granted, as they know many words —45,000 to 60,000 on average (Plag 2003: 4)— that enrich their linguistic competence, and they can learn new ones on a daily basis, which are automatically added to their mental lexicon and used in their discourse almost immediately after they learn the objects or ideas that those words denote (Lieber 2010: 5). This indulgent awareness may spring from the fact that literate speakers tend to imagine a word as a visual unit in print, rather than as a sequence of sounds (Ong 1982: 121). Thus, words are delimited by means of blank spaces before and after them, and their degree of isolation has been amplified across history chiefly because of two inventions,

i.e. writing and the printing press. Both innovations, but especially the latter, have contributed greatly to attaching meaning to sight rather than to sound by providing regular spaces between words. We might add that this awareness among speakers has been enhanced relatively recently by yet another invention, computers, which have conferred spaces with an additional significance, to such an extent that they have become meaningful themselves[11]. As a consequence computer software can exploit spaces as clear-cut boundaries to extract words from texts and generate wordlists for further processing, as has been the case in this study (see chapters 3 and 4). We can nonetheless find instances in which speakers may not have the conviction of whether they are uttering one or two words: *alright/all right, everyday/every day, alot/a lot*, and so forth. However, these inconsistencies are more related to their level of linguistic competence than to the nature of the words themselves. It is likely that words must belong to our lexical background, so that we can identify them individually. In addition, if we bear in mind the manner in which children learn a language, it seems evident that their progress consists of the ability to combine an increasing number of words in a correct way rather than breaking down sentences (Carstairs-McCarthy 2002: 4), and they "readily learn to break utterances up into words when learning to write" (Aronoff & Fudeman

[11] The space has its own value as a unique character in informatics (ASCII code 20), and as such it is used in processing information further.

2005: 37). But the above-mentioned reasons only account for what some linguists denominate *orthographic word* (Plag 2003; Katamba 2005; Booij 2007).

Explaining the meaning of *word* from an exclusively linguistic perspective poses nevertheless a rather complex challenge to scholars. In the seventeenth century, nouns, or names of things rather, stood at the centre of linguistic study, and were treated almost as synonyms of words themselves, as can be seen in the results of the analytical language elaborated by Wilkins, already described in the previous chapter. An interesting psycholinguistic definition of the term can be found in a sermon preached on May 1686 by Robert South (1720: 282) that reads "as conceptions are the images or resemblances of things to the mind within it self; so are words or names the marks, tokens or resemblances to the minds of them whom we converse with". In the eighteenth century, commentators of language did not look too much into defining something that seemed so obvious either. Rather laconically, Johnson (1755) defines *word* in his celebrated dictionary as "a single part of speech", and Bevis (1765) introduces a phonological perspective, by explaining it as "an intelligible sound expressed, in order to declare the sentiments of the mind". Fogg (1796: 14) adds a morphosemantic alternative approach by affirming that "words, in their first origin, must have some foundation in [the signification of] things; though in many instances no doubt a slight one: in their secondary application they must be derived from other

words at first thus founded". At the beginning of the twentieth century Saussure (1989: 147) put forward his theory on the arbitrariness of the linguistic sign and established the dichotomy between signifier and signified by claiming that "une image acoustique est associée à un concept" (1989: 148) and "le signifiant est auditif; le signifié, conceptuel" (1989: 150). Onomatopoeias and phonaesthemes (Firth 1930: 184; Adams 2001: 121) seem to overrule Saussurean precepts, since their sounds intend to reflect their meaning more or less faithfully, as is the case of *rattle* (Wollstonecraft 1792: 66), but as Carstairs-MacCarthy (2002: 6) puts it, even onomatopoeic words are subject to a great degree of convention. He gives examples of animal cries, but the sound of every animal differs across languages. Subsequent schools after Saussure diverted their attention to the degree of independence of words to explain them. On the grounds of structuralism Bloomfield (1933: 178) describes *word* as "a free form which does not consist entirely of (two or more) lesser free forms; in brief, a word is a minimum free form". This description might serve us to explain derived words such as *absurdity* (Collins, *CEPhiT* 1717: 57), which consists of only one free form, the adjective *absurd*, and a bound form, the affix *-ity*. Although it will not encompass compound words such as *bowstring* (Cheyne, *CEPhiT* 1705: 27), formed by two free forms, *bow* and *string*, and this would furthermore inhibit additional deriving as in *houswifery* (Harris, *CETA* 1719: 24), since this noun is not the outcome of *house*+**wifery*, but otherwise the consequence of adding the affix *-ry* to the

compound noun *hous(e)wife*, which in this case in particular does need to be considered as a single word. In line with Bloomfield, Hockett (1958: 166) affirms that words may be determined by the features of "pause and isolability" and defines word as "any segment of a sentence bounded by successive points at which pausing is possible". However, he concedes that asking speakers to pause and isolate words in an utterance has its risks, because "the pausing habits of a literate speaker of English are doubtless conditioned by his literacy" (Hockett 1958: 166). The renovator of morphology Hans Marchand (1969: 1) commits to define *word* from a morphological point of view, as "the smallest, independent, indivisible, and meaningful unit of speech". But this again would neither account for PDE cases such as *pre- and post-war period*, in which a prefix stands independently from the base to which it attaches, nor to phrasal and prepositional verbs like *build up*, *take over* and so forth, which are composed by more than one item, "though they often form a semantic unit, and may be equivalent to a single-item verb" (Adams 1973: 9).

Lately morphologists have opted for acknowledging the existence of words as units, and the difficulty in trying to provide a definition of their psychological nature, which I will venture to explain broadly as the basic unit of expression that has instinctive recognition by native speakers. Therefore they have focused on words from two different approaches. On the one hand, some attempt to determine what a word is or is not by performing empirical tests. On the other hand, some others give an explanation based on the

morphological units that make up words, i.e. morphemes —this approach will be fully developed in section 2.2. Aronoff & Fudeman (2005: 36-8), for example, claim that words must fulfil three conditions at least to be considered as such and differentiated from bigger units such as phrases and sentences:

1) Fixed order of elements: a complex noun like *admeasurement* (Hill, *CETA* 1754: 17; Nicholson, *CETA* 1782: 119) cannot be uttered as **measurementad* or **admentmeasure*. This property cannot be applied to sentences.

2) Non-separability —also termed uninterruptibility or cohesiveness of words (Crystal 2003: 501)— and integrity: in theory, words are non-separable because other linguistic material cannot break them up to be inserted in them, but as we will see in section 3.2.2, this condition is at times circumscribed by the use of infixes like in *absobloominglutely*, *kangabloodyroo*, and *funfuckingtastic* (Camiña-Rioboo 2005). Besides, the historical evolution of some words can affect their internal structure. Such is the case of *bucketful*, a compound word coined in Old English in the way of *mouthful* or *handful* from a noun (*bucket*) and an adjective (*full*). It achieved its quality as a full noun in Middle English, and may show either the plural form *bucketfuls* —which would probably be the most correct one— or, after a subsequent analysis of its forming constituents as Noun + Adjective, the form *bucketsful* can be obtained. Therefore, the word becomes interrupted by the plural *-s* (*OED*). Integrity means that "syntactic processes cannot apply to pieces of words [...] Adjectives

and adverbs, for example, modify words, not morphemes. Words and phrases are often displaced to the beginning of a sentence or questioned, but not morphemes" (Aronoff & Fudeman 2005: 37).

Finally, 3) stress may used to determine whether some words are compounds or phrases. This can be seen in the following examples extracted from my corpus[12]:

(1) a. gréyhound (Hill, *CETA* 1754:10) > A variety of dog used in the chase (*OED*)

grey hóund > A grey dog

b. géntleman (Curson, *CETA* 1702: 337) > A man of gentle birth (*OED*)

gentle mán > A man of gentle character

c. hót-house (Smellie, *CEPhiT* 1790: 11) > A greenhouse kept artificially heated (*OED*)

hot hóuse > A house with high temperature

[12] More information about the differentiation between compounded words and phrases will be provided in section 3.3. Likewise, the arbitrariness in the use of hyphens in compound words, as well as my own marginal approach to compounding in the corpus will be dealt with in section 3.3.

d. níghtshade (Smellie, *CEPhiT* 1790: 10) > A plant with narcotic properties (*OED*)

night sháde > The darkness of night

Carstairs-McCarthy (2002: 59) explains that "there is a difference in sound corresponding to the difference in meaning". In the first example of (1) the stress is on the first element *grey*, while in the second example, it is on *hound*. The same pattern can be extrapolated to examples (b), (c) and (d). English compounds are generally stressed on the first element, so we may assume that the first cases from every example are compounds, whereas the second are phrases. It is to be observed, though, that this condition of stress used to differentiate compound words from phrases is not applicable to written texts, as is the case of our corpus.

2.2. Word-form, lexeme

My corpus-based study will emphasise the importance of the number of words classifiable under different parameters in order to prove linguistic theories and principles. It is necessary, therefore, to clarify which elements will be computed and how they will be treated for further processing. If we consider the following sentence:

(2) My friend and I walk to class together, because our classes are in the same building and we dislike walking alone. (Lieber 2010: 4)

And we need to know how many words there are in it, we may count them under two different methods: firstly, counting every item in the sentence, so the final count will be twenty-one; secondly, we might reconsider counting the conjunction *and* twice, because it is repeated; besides, *class* and *classes* may be regarded as two different realisations, singular and plural, of the noun *class*; and finally, we may also want to include *walk* and *walking* as two different tenses in the paradigm of the same verb, *walk*, in which we also expect the form *walk* and *walked*. Both methods are acceptable, and both provide valuable, but different, information about the linguistic processes under examination. The first count of the total number of words, repeated or not, singular or plural, will supply the word-forms in the text; by means of the second we will obtain the number of lexemes. While the latter will allow us to analyse the degree of productivity or creativity of some processes, the former will outline the number of occurrences of those processes, allowing us to attest patterns of vocabulary choice regarding those types, if we are to agree to the fact that "lexicalisation is chiefly a matter of frequency" (Plag 2003: 103). The word-formation processes revised in chapter 4 can be paralleled with those on lexeme formation dealt with by the authors mentioned above. Additionally, the

terms *word* and *lexeme* will be used quite interchangeably in the following sections.

As stated above, if we are to accept the fact that, in general, all speakers are conscious of the existence of an entity such as word, we must ask ourselves: does the same happen with word structure? Is word structure something intuitive in some way so that a speaker can decompose words and isolate every constituent? Bloomfield (1933: 208) considers that "since the speaker cannot isolate bound forms by speaking them alone, he is usually unable to describe the structure of words". The examples provided by Bloomfield in order to clarify this statement are evident. Therefore, *goose*, *gosling* and *gooseberry* seem to be closely related as far as their meaning is concerned. A speaker may certainly recognise a morphological relation between *goose* and *gosling*, and he may even identify *gosling* as a diminutive of *goose*, is spite of phonological change from [gu:z] to [goz], due to the application of a somewhat productive pattern (vid. section on productivity below) to make diminutives, either without a phonological change, as in *duckling*, *fopling* or *yearling*, or with a change, as in *darling* (< *dear·ling*) or *heed·ling* (< *head·ling* 'darling'). Nevertheless, the element *goose-* in *gooseberry* does not share a meaning with the previous examples. A deeper knowledge in the history of the English language shows that the word *gooseberry* could be the result from the Dutch word *kruisbezie*, from German *krausbeere* or from the native forms **gorseberry* or, most probably,

**groseberry*, the origin of which could reside on the term *gozell*, ultimately derived from French *groseille* in the Early Middle English period. Besides, it seems very difficult that a speaker can establish a semantic relation between the feminine *goose* and its masculine *gander*, based only on the coincidence of their initial sound [g] or between *duck* and its masculine *drake* referring to the initial and final sounds [d], [k]. We can conclude, then, that an analysis of the structure of words requires a deeper study, because it cannot be observed without a basic knowledge of morphology and its units. In the following sections the elements that constitute words will be described and their internal relations explained.

2.3. Formatives, morphemes, morphs, allomorphs

Grammarians in the eighteenth century generally used the term 'formative' in a rather indiscriminate way. It would refer practically to all the elements present in words and even to processes. Sundby (1995: 54-5) quotes several authors to illustrate the multiplicity of theories across the century. For example, Greenwood (1711: 186) states that "[our Ancestors] not only cut off the formative Terminations, but even the Heads or Beginnings of Words". Browne (1700: 99-100), Tuite (1726: 76) and Dilworth (1751: 90) relate formatives to "endings". Elphinston (1765) uses the term to refer to derivatives in general, and also with the modern sense of 'formative element'. In present-day

morphology Trask (1993: 175) defines morpheme as "the smallest unit which plays any part in morphology and which cannot be further decomposed except in phonological or semantic terms". In spite of some hazy boundaries between morphology and other parts of speech like phonology and syntax, which will be explained later, Trask's definition indicates the relationship of independence that morphology keeps with respect to phonology, bearing in mind that its units —sounds, syllables and rhythmic units— lack meaning individually. The number of morphemes that makes up a word subdivides them into simple or monomorphemic, if they contain only one, like in *azimuth* (Curson, *CETA* 1702: 343), *crab* (Morden, *CETA* 1702: 11) or *enthusiasm* (Balguy, *CEPhiT* 1733: 12), and complex or polymorphemic if they contain more tan one, like in *cob·web* (Balguy, *CEPhiT* 1733: 26), *abstract·ed·ness* (Hume, *CEPhiT* 1748: 18) or *dis·agree·able·ness* (Collins, *CEPhiT* 1717: 73). On the other hand, morphemes do not have a determined length: there are long and simple words: *caterpillar* (Smellie, *CEPhiT* 1790: 6)*, crepusculum* (Fuller, *CETA* 1732: 20)*, phenomenon* (Morden, *CETA* 1702: 2); and short and complex words: *four·th·s* (Morden, *CETA* 1702: 14).

The classification of morphemes can be studied from several points of view: (a) if we contemplate their level of independence within the word, they can be divided into free, if they can be found isolated in a phrase; and bound, if they need to be accompanied by, at least, another morpheme. In the word *fixedness* (Crombie, *CEPhiT* 1793: 49), for example, *fix* can appear isolated as

an independent word-form, therefore it is a free morpheme, whereas *-ed* and *-ness* can only appear in conjunction with other morphemes, therefore they are bound. (b) If we consider the function that bound morphemes perform within the word, we can distinguish between inflective and derivative. In the verb form *viewed* (Wilson, *CETA* 1773: 15), the bound morpheme *-ed* offers a predictable variant included in the paradigm of the regular verb *view*, therefore, *-ed* is an inflective morpheme. In the case of the noun *mismanagement* (Burke, *CETA* 1770: 4), both *-ment* and *-mis* elaborate on the verb *manage* to coin two new lexical items: *-ment* transposes the verb *manage* into a noun, *management*, coined in 1598 (*OED*); and *-mis* is used to coin its antonym in 1668. As noted in the definitions of type and token, this distinction between inflective and derivative processes is essential for my analysis, since inflective morphemes contribute to elaborate word-forms or tokens, whereas derivative morphemes are used to create new lexemes or types, remaining these the key elements in word-formation studies.

Despite the arbitrariness of the linguistic sign, the meaning of complex words must be predictable, so that speakers can understand the resulting one. In order to achieve this, their constituent morphemes must fulfil two conditions (Carstairs-McCarthy 2002: 17): (i) they need to be identified in the words they form, and (ii) they must somehow contribute to the meaning of those words. The following examples illustrate what has just been expounded:

(3) *compact·ed·ness* (Cheyne, *CEPhiT* 1705: 13)

[free, bound, bound]

co·sine (Watts, *CETA* 1726: 23)

[bound, free]

change·able·ness (Watts, *CETA* 1726: 7)

(free, free?, bound)

Two bound morphemes can be united to coin a new word as well. In our corpus we have observed two different sources of materials to create new lexical items.

(i) *percuss·ion* (Greene, *CEPhiT* 1727: 9), *pestil·ence* (Burke, *CEPhiT* 1770: 3), and *marri·age* (Astell, *CEPhiT* 1700: 56) could be considered as borrowing; however, I have taken into consideration the fact that it is not clear whether the bound base was already free at the time of coining the derived word, therefore they have been included in the category of word-formation[13]. Ŝtekauer (2000: 366) explains this process from the perspective of morphological lexicalisation, and provides the contrast between the pair of examples *ed·ible* and *eat·able*, where only *eat-* remains to be productive.

(ii) Some authors (Adams 1973, Bauer 1983, Plag 2003), additionally, explain another process that involves the combination of bound morphemes

[13] In my corpus this has been considered as Derivation Class III (*vid*. chapters 3 & 4).

taken from Latin and Greek, or 'initial' and 'final combining forms' (ICFs and FCFs henceforth). They classify this process under the category of compounding, and label it 'neoclassical compounding'. In her latest major reformulation of word formation and complex words, Adams (2001: 118) renames these words as 'stem compounds', so she still discusses this process from the standpoint of compounding. Examples of this variant are *zoo·phyte* (Campbell, *CEPhiT* 1776: 3), *tele·scope* (Curson, *CETA* 1702: 348) and *micro·meter* (Hodgson, *CETA* 1749: 109). Notwithstanding the respectable reasons explained by the above-mentioned linguists, in my opinion these words can also be treated from the perspective of word-formation, because some of these morphemes had already evolved —or evolved later on— into free forms, such in the cases of *scope* and *meter*[14]. Regarding the first example, *zoophyte*, we must not mistake the contemporary noun *zoo* as an evolution of the Greek combining form *zoo-*, but rather as the shortening of the adjective *zoological*, coined in 1807 (vid. section 3.5.1). Meanwhile, the noun was coined later in 1847 (*OED*).

A morph can be defined broadly as a segment of a word form that stands for a particular morpheme. Katamba (2005: 32) gives a similar definition as "any physical form that represents a morpheme". In the examples (3) given above *co-*, *-ed*, *-able*, *-ness* are all morphs representing derivative

[14] This is my Derivation Class IV (vid. chapters 3 & 4).

bound morphemes. Morphs can also represent inflectional morphemes with grammatical function. For example, in the word *stars* (Whiston, *CETA* 1715: 5), the morph -*s* represents the morpheme {plural}. Also, morphs can perform the function of several morphemes at the same time. In the following sentence,

(4) But becauſe the Moon is nearer to us, ſhe appears as big as the Sun (Charlton, *CETA* 1735: 14).

As we can see in the case of -*s* in the verbal form *appears* (Charlton, *CETA* 1735: 14), it represents the morphemes {present}, {singular}, {third person} and {indicative}. In these situations this morph is called *portmanteau* morph.

We must not confuse morphemes with syllables, although sometimes they coincide in form, like in *childhood* (Butler, *CEPhiT* 1736: 23), where the two syllables match the morphemes *child* and -*hood*, other cases like *correctness* (Bryan, *CETA* 1797: 117) demonstrate that one morpheme may contain more than one syllable, *correct*. Additionally, morphemes may not contain any syllables at all, as can be illustrated by the noun *motions* (Lacy, *CETA* 1779: 4), in which the plural morpheme -*s* does not constitute a syllable itself. Indeed, the nature of morphemes is purely semantic and that of syllables purely phonological. Both can be decomposed phonologically into their minimal units —phonemes—, but the routes to achieve that decomposition

must be different. The following figure shows the phonological decomposition of a word by Harrington (2010: 99):

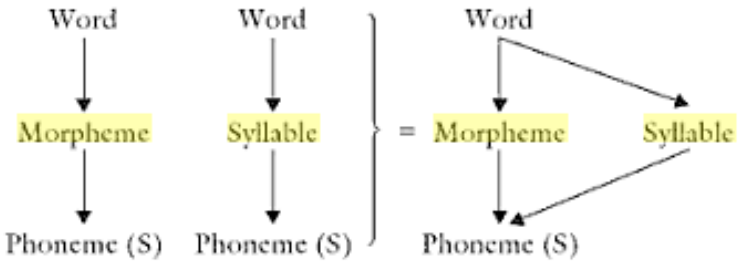


Figure 2.1: Phonological relationship between morphemes and syllables.

Although Harrington's approach (Figure 2.1) seems understandable and clear, morphemes are not made up of phonemes, because we can identify morphemes that are not pronounced in the same way in every context (Katamba 2005: 32). Some of them show different realisations, as it happens in the case of the morpheme {plural}, the pronunciation of which is conditioned by the base to which it is attached. Therefore, we can see that in *ships* (Whiston, *CETA* 1715: 4), the plural morpheme is [s]; in *stars* (Lacy, *CETA* 1779: 2), [z]; in *uses* (Bolingbroke, *CEPhiT* 1754: 41) [iz]; and in *sheep* (Collins, *CEPhiT* 1717: 54), [Ø]. Under these circumstances [s/z/iz/Ø] are considered to be allomorphs of the morpheme {plural}. For my own convenient use, which will contemplate phonological variants and etymological origins alike, my analysis of the corpus will treat different allomorphs separately, meaning that *im-*, *ir-*,

in- (allomorphs of the privative morpheme), and *-er, -or*, *-our* (representing the agentive morpheme) will be listed independently, all of which will be seen in chapter 4.

Bauer (1988: 40) describes the existence of a type of morph (which he calls unique morph) that is only present in fossilised expressions. Occasionally these morphs can be free morphemes, such as *kith*, used nowadays only as part of the expression *kith and kin*; however, they are generally bound morphemes. Besides, Spencer (1991: 40) detects a form-function problem in the use of the so-called cranberry morphemes, because they do not conform to the rule prescribed by Carstairs-McCarthy on the predictable nature of complex words. If the morpheme is the minimal unit with meaning, what is the meaning of *cran-* in the noun *cranberry*, or of *rasp-* in *raspberry*? Jackendoff (2010: 425) labels the compounds explained by Bauer as 'lexicalised compounds' and also subdivides cranberry morphemes in two groups. The first one includes those words like *cranberry* or *nightmare*, *iceberg* and *linchpin* in which the underlined morphemes have no meaning by themselves, and the second one, which he terms 'strawberry morphemes', include "real words within compounds that play no role in the compound's meaning" (2010: 425). Instances of the second group, he adds, would be *strawberry*, *horseradish*, *gangway* and *dogwood*.

2.4. Root, stem, base

Nowadays these terms are used in morphology to designate the remaining part of the word after different types of morphemes have been subtracted. I am unaware of any uses of the terms *stem* and *base* in eighteenth-century grammars, but *root* was used extensively with the meaning of 'primitive' or 'original' (Sundby 1995: 94). I have resorted to Elphinston (1765: I.276) to analyse his treatment of the term, which he defines as follows: "To the English scholar…every word is a root or primitive, which is not formed or derived from some other English word". For Mackintosh (1797: 36), *radical word* expresses the same idea that Elphinston applies to *root*. Bauer (1983: 20) defines a root as a word-form that does not allow a more exhaustive analysis, neither from inflective nor derivative morphology perspectives, because it refers to that part of the word that remains after all morphemes, both inflective and derivative, have been removed. It is, therefore, the basic part that is always present in a lexeme. If we consider his famous example of the complex word *untouchables*,

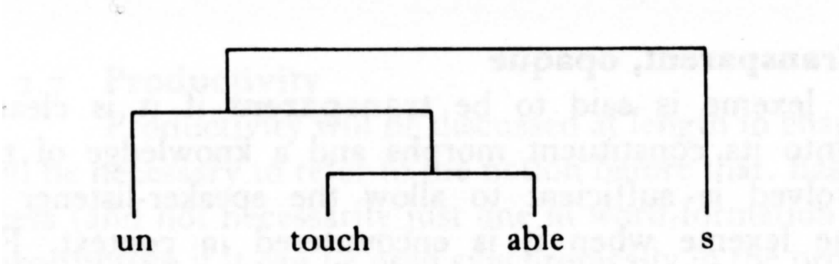


Figure 2.2: Decomposition of a complex word.

This complex word is formed by four morphemes, the root is the free morpheme *touch*, followed by the derivative suffix -*able*, preceded by the derivative prefix *un*-, and all of it followed by the inflective suffix -*s*. The previous figure shows the morphological relations among the different constituents of the word, and the analysis must follow this order for its decomposition, because we can find forms such as *touchable* or *untouchable*, but other forms such as **untouch* or **touchables* are not possible. A simple word is formed by one single root, whereas complex words such as *football* contain two: *foot* and *ball*. The term 'stem' has validity only in the field of inflective morphology, because it refers to that part of the word obtained after removing all inflective morphemes, but not the derivative. If we use the previous example, the stem would be *untouchable*. Also, stems can be (i) simple, as in the case of *stars* (Lacy, *CETA* 1779: 2), whose stem is *star*; (ii) complex, if they contain derivative morphemes, as in *inequalities* (Costard, *CETA* 1767: 277), where the stem is *equal*; and (iii) compound, because they may contain two roots, as in the case of *earthquakes* (Greene, *CEPhiT* 1727: 8), where the stem is *earthquake* and the roots are *earth* and *quake*. Finally, the term 'base' refers to any word form to which affixes of any kind can be added, both inflective and derivative. This means that any root or stem can be termed base as a generic tag. In the example provided by Bauer *touch, touchable* and *untouchable* can be considered as bases, because they serve as sources to make new words by adding more affixes (in spite of the fact that *touch* fulfils

likewise all the requisites to function both as root and stem). Finally, *touchable* and *untouchable* can be considered as bases and stems, but not roots.

3. Processes

As indicated in the introduction to this chapter, all processes of word-formation will be reviewed in this work. Some of them will be mentioned briefly and examples extracted from my corpus will be provided, where possible. Major word-formation processes such as affixation, compounding and conversion will be explained more comprehensively, since they cover most nouns in the corpus. Before I deal with them, however, it may be necessary to delimit derivative from inflective processes, because at times the boundaries between them are not as clear as it might seem at first sight. Also, early- and late-modern theories on morphology or etymology will be revised.

3.1. Morphology ('etymology'): inflection and derivation

Morphology is one of the major levels of linguistics descriptive theory that deals with the study of forms of words (Matthews 1974: 3). The term was coined in the nineteenth century, as already explained in chapter 1 (section 2), and before the 1860s linguists usually referred to etymology when dealing with word structure and word formation. As Sundby (1995: 49-55) illustrates diagrammatically, many grammarians in the 1700s considered the terms

etymology and *derivation* as synonyms (Collyer, 1735; Martin 1748; Gough 1754; Fogg 1792). The most detailed and comprehensive account of morphological processes in the eighteenth century is provided by Elphinston (1765: I. 218-392), as is schematised in figure 2.3 below. Here I have omitted the term *word-formation* intentionally because it cannot be found in his descriptions.

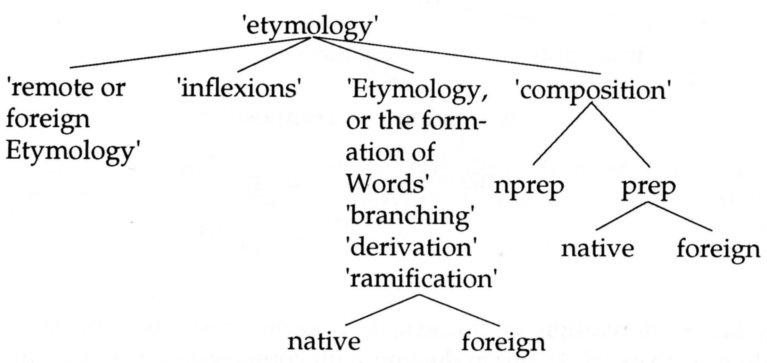


Figure 2.3: Elphinston's theory on etymology (Sundby 1995: 54)

Morphological processes such as loan ('remote or foreign Etymology'), inflection, derivation and compounding can be extracted from Sundby's diagram. Likewise, Elphinston distinguishes native and foreign base origin for derivation and compounding. Clear distinctions between inflection and derivation were not common and some treat inflectional endings as part of derivation, whereas others (Baskerville 1765) defend the universality of inflection versus derivation, something that Bauer (1983) explains as the

existence of gaps in the paradigm of derived words, contrasting with a higher degree of regularity in the inflective paradigms. In the nineteenth century studies on morphology acquired greater importance, as Aronoff (1983: 355) reminds us:

> Morphology was central to nineteenth century linguistics for two reasons, First, traditional grammar, out of which modern linguistics grew, had been morphologically based, as all of us know too well who have learned their Latin declensions and conjugations. Second, the comparative method of historical linguistics, which provided the most spectacular successes of nineteenth century linguistics, which indeed made linguistics into a respectable modern academic discipline, depends to a great extent on morphology.

He also stresses the difficulties in separating morphology from the other three systems —syntax, semantics and phonology— at the word level because, as Kastovsky (1977) and Dressler (1979) have emphasised, these three systems interact (Aronoff 1983: 358). As for the description of inflection and derivation from a present-day perspective, we may resort to minimalistic instances like the one provided by Hockett (1958: 209), who asserts that "inflection is that part of morphology which involves inflectional affixes. The remainder of morphology is derivation". But it seems quite incomplete, because it barely describes the first, and ignores the second almost completely. Lyons (1977: 521) gives an extended definition of inflection, and states that "inflection

produces from the stem (or stems) of a given lexeme all the word-forms of that lexeme which occur in syntactically determined environments". However, recent works published on morphology have introduced brevity and clarity when describing both branches of morphology. For example, Aronoff & Fudeman (2005: 45) affirm that,

> Inflection involves the formation of grammatical forms — past, present, future; singular, plural; masculine, feminine, neuter; and so on — of a single lexeme. The use of these grammatical forms is generally dictated by sentence structure. Thus *is, are,* and *being* are examples of inflected forms of the lexeme BE, which happens to be highly irregular not only in English, but in many other languages as well. [...] Derivation involves the creation of one lexeme from another, such as *selector* or *selection* from *select*.

Inflective morphology, then, studies the variation in the form of words to perform specific grammatical functions. The main inflective morphemes in the English language are: (i) the plural mark, expressed with *-s* in *star/stars* or *planet/planets, -es* in *bus/buses* or *church/churches*, *-en* in *ox/oxen*, sometimes it is expressed by means of a change in root vowels, as happens in *foot/feet* or *mouse/mice*, and it is some other times represented by Ø, as in *fish/fish* or *giraffe/giraffe*; (ii) genitive -*'s*, as in *Darwin's theory*; (iii) the verbal ending *-s* of the third person of the singular indicative, *-ing* for the present participle, *-ed* (in regular verbs such as *devise/devised* or *discover/discovered)* and the strong

forms (in irregular verbs such as *see/sawn* or *brought/brought)* for the past and past participle; and iv) the endings *-er* and *-est* of the comparative and superlative degree of adjectives, such as *taller/tallest* (or their corresponding *more/most* with two-syllable adjectives).

As a contraposition to inflectional morphology, which specifies the grammatical functions of the words in phrases without altering their meaning, derivative processes aim at coining new words with a new meaning as well. Marchand (1969: 2) defines word formation as "that part of the language which studies the patterns on which a language forms new lexical items". Kastovsky (1996: 95) in turn explains that "the basic function of word-formation is to provide new lexical items, when a speaker, and ultimately a speech community, feels the need to have a name for some segment of extralinguistic reality rather than a description". Plag (2003: 15) puts emphasis on the fact that inflectional morphology preserves the meaning and the category of the original word in the inflected form, whereas derivation generally changes the semantic characteristics of the lexemes obtained, and these may either alter the category of the word, as is the case of suffixes, or preserve it, as with prefixes. These two characteristics, named class-changing and class-maintaining were already observed by grammarians in the Late Modern English period, as pointed out by Sundby (1995: 67): "the grammarians recognize the class-preserving nature of inflection […] Conversely, the fact that derivation normally brings about a change of word-class is taken more or less for granted". Early grammarians

also detected an overlapping between inflection and derivation in the case of *-ed* participles, to which we can add *-ing* forms, because they may belong to both inflective and derivative paradigms. Adams (2001: 6) has defined this dual character as "troublesome". In my case, given that my focus in the corpus is on nouns, *-ed* cases did not pose any difficulty, because they may form regular past/participle tenses and also adjectives; however, *-ing* forms needed to be manually disambiguated, because *CETA* and *CEPhiT*, so far and until they are tagged further, do not allow for the categorising of progressive tenses, gerunds, adjectives or nouns. Adams (2001: 7) continues by saying that both "have some claim to be lexeme-forming suffixes in words with apparently nominal bases", and she provides examples like *alarmed* in *the building was alarmed*, derived from the noun *alarm*, as compared to *John was alarmed*, inflected from the verb *to alarm*. Other examples related to nominal bases including *-ing* are *scaffolding* and *shirting*.

Stump (1998: 14-8) proposes five empirical criteria for distinguishing inflection from derivation, which will be expounded below. Although his main line of argument has been respected, I have reorganised his criteria by grouping semantic features[15] and detaching grammatical categories as follows:

[15] When dealing with meaning, Stump (1998: 15-7) separates change and regularity, as commented in his criteria (i) and (v).

(i) Change in meaning: we do not observe difference in the lexical meaning of elements belonging to the same inflectional paradigm, whereas this characteristic is typical of derivational morphology. This explains cases such as *counterpart* > *counterparts* (Hume, *CEPhiT* 1748: 44). The inflective morpheme {plural} does not change the meaning of the singular word. There are instances, though, in which derived words do not show a clear difference in meaning as happens in the cases of pairs of affixes such as *-ic/-ical* in *geographic/geographical*, *mythic/mythical*, *classic/classical*.

(ii) Change in part of speech[16], meaning that derived elements may belong to a different word category, whereas inflected elements do not. Generally prefixation is class-maintaining, as in *declination* (n) > *codeclination* (n) (Watts, *CETA* 1726: 23), whereas suffixation is class-changing, as in *astonish* (v) > *astonishment* (n) (Bonnycastle, *CETA* 1786: 43). The use of the inflective morpheme *-s* that denotes third person singular indicative forms does not cause a change in the part of speech of *observe* (v) > *observes* (v).

(iii) Syntactic determination. A lexeme is not required to belong to a particular class of derivatives in syntactic contexts, whereas it does require a particular element in the paradigm of inflected words. If we take the verb *observe* mentioned above, when the action is present and performed by a third

[16] Some authors like Adams (2001: 15) speak of 'transposition' instead of class-changing processes. She includes within transposition "all cases in which the base of a complex word belongs to one word class and the derived word to another".

person it requires the form to be *observes* and no other. As regards the derivational paradigm, this is not so, as can be seen in the following example:

(5) The Earth is the place from whence we view the heavenly bodies (Long, *CETA* 1742: 61)

We could substitute the adjective *heavenly* with *spatial* and the syntactic structure of the sentence would not be affected.

(iv) Productivity[17]: inflection shows a more productive activity than derivation, in the sense that virtually every countable noun has a plural counterpart, for example. Also, regular verbs have past and past participle forms ending in *-ed*. On the contrary, not all adjectives have a related causative verbs, as is the case of *harden*, *deafen*, **colden*, **braven* (Stump 1998: 16). Besides, derivative affixes from different origins coexist in English, thus competing among them and making paradigms less stable, as is the case of suffixes to coin abstract nouns, such as *-acy*, *-age*, *-ance*, *-ence*, *-ism*, *-ity*, *-ment*, *-ness*, -(*t*)*ion*, *-ship*, among others. We shall see in chapter 4 that the choice of one affix over another when coining new nouns may be quite arbitrary and, in many cases subject to fashions, rather than relying exclusively on the origin of bases. For example, Adams (2001: 14) emphasises that the use

[17] This criterion is also explained by Katamba (1993: 79-81)

of the diminutive affix *-let* and the denominal adjective-forming *-ish* were "at their most productive in the nineteenth century". The phenomenon by which a derivational element inhibits the use of another one is termed *blocking*, explained by Aronoff & Fudeman (2005: 219) as "an economy principle that can be thought of informally as an injunction to avoid coining synonyms".

(v) Closure: inflectional morphemes close words to further derivation, whereas derivation, within certain limits, does not. This means that even if we can derive the adjective *wonderful* from the noun *wonder*, should we add an inflectional plural morpheme {-s} the to noun, *wonders*, that would prevent further derivation, which renders **wondersful* not possible. Regarding the limits for derivation, although words such as *anti·dis·establish·ment·arian·ism* (Spencer 1991: 3) are recorded in English by means of valid derivative patterns, they are usually exceptional. Hockett (1958: 179) claims that English words "rarely achieve great morphological complexity. *Ungentlemanliness* and *impressionistically* are about as far as English goes".

3.2. Affixation

Once the characteristics of inflectional and derivational morphology have been explained —word-formation or lexeme formation, depending on the authors—, and overlapping areas where both branches of morphology cannot be distinguished from one another very clearly, I will now proceed to describing

the units involved in the coining of new lexical items in English, though some references to other languages will be provided. I have already defined root, stem and base as building materials for new words; therefore I will deal with the elements that attach to these: affixes. According to the *OED,* before 1865 the term *affix* did not apply inclusively to all types of affixes. Indeed, Martin (1748), Priestley (1762) and Wood (1777) usually refer to affixes with the meaning of 'suffix' or 'termination' (Sundby 1995: 22). Plag (2003: 72) defines affix as "a bound morpheme that attaches to bases". This seems a sufficiently clear explanation. However, there are some instances in which an affix is not really bound. Sometimes a free morpheme may have become bound across the history of the language, as is the case of *-ful* —developed from the free form *full*— in words like *beautiful*, if we are to consider it an affix. Other examples provided by Plag (2003: 72) include *-free*, *-less*, *-like* and *-wise* in expressions such as *error-free text*, *carless* and *lawless*, *prison-like school gates* and *education-wise*. He resorts to semantic arguments to exclude *-free*, and *-like* from the list of affixes, given that the pairs *-free*/*free* and *-like*/*like* mean exactly the same. Therefore, for him these would be clear cases of compounding, not derivation. Conversely, the free morpheme *wise* means 'clever', whereas the bound morpheme *-free* that we find in complex words 'related to'. Also, the free morpheme *less* stands for the opposite of *more*, while the bound morpheme *-less* means 'without'. From the phonological point of view *less* and *-less* are pronounced differently, so that supports the theory

that regards them as different morphemes. Finally, from a syntactic point of view *wise* is an adjective, but -*wise* produces new adverbs, meanwhile *less* is and adverb and -*less* creates adjectives (Plag 2003: 73).

When phonology and morphology interact we may talk about neutral and non-neutral affixes (Aronoff 1983: 362), which complement and contrast with the traditional theories of primary and secondary affixes (Whitney 1889). Neutral affixes do not interact phonologically with the base to which they are attached, whereas non-neutral affixes cause changes to the base. Cases of neutral affixes are -*ness* and -*ment*, and non-neutral, -*ity* and -*ic*(*k*), for example, in pairs of words such as *awful*/*awfulness* (Balguy, *CEPhiT* 1733: 27), *incroach*/*incroachment* (Reid, *CEPhiT* 1764: 22), and *active/activity* (Turnbull, *CEPhiT* 1740: 14), *drama*/*dramatic* (Hume, *CEPhiT* 1748: 20). Katamba (1993: 91) refers to strong (#) and weak (+) stem boundaries to separate stems from neutral and non-neutral affixes. He also states that Germanic affixes are neutral, whereas Greek and Latinate, "having entered the language with loanwords from Greek, Latin and French" (1993: 92) are non-neutral. He stresses that "many affixes from foreign sources will only combine with bases borrowed from the same foreign language" (1993: 92) but as we will see in the complex words in our corpus Germanic bases can combine with Latinate affixes, as happens with *amazement* (Dunton, *CEPhiT* 1710: 17), *behaviour* (Balguy, *CEPhiT* 1733: 28), and *drudgery* (Dunton, *CEPhiT* 1710: 10), and Latinate bases can combine with Germanic affixes, as seen in

consciousness (Balguy, *CEPhiT* 1733: 28), *courtship* (Dunton, *CEPhiT* 1710: 1) and *intruder* (Astell, *CEPhiT* 1700: 22). More information on the interaction of phonology and morphology will be provided in section 3.2.2 dealing with infixes.

For my research on nouns, I will focus basically on prefixes and suffixes, as I have not come across any instances in which other affixes have been used to coin new nouns in any of the texts selected, but I will also explain other minor classes of affixes. Special attention will also be paid to their rise and period of validity as productive in the language, the kind of bases they attach to and their origin, in an attempt to measure the productivity of affixes and affixation patters and their periodisation. I intend to attest whether affixes may 'die' from lack of use, or excess, as was the case of OE *ge-* justified by Millward (1996: 123) as follows:

> The most common prefix in OE is *ge-*, so widely used and in so many different ways that it came to be virtually meaningless and lost from the language. It was a marker of the past participle of verbs, but it was also used throughout the entire conjugation of many verbs, usually to indicate perfective aspect (completion of an action). Sometimes it distinguished a special meaning of the verb. For example *gān* meant "to go", while *gegān* meant "to conquer". Often ge- was attached to a verb with no discernible change in meaning at all: both mænan and gemænan meant "to mean". And ge- was used with other parts of speech as well. Attached to nouns, it often signified association; for example

> brōðor meant "brother", while gebrōðor meant "a member of the community, a monk". But when attached to a noun or adjective, ge- often meant no more than that the word was derived from a verb […]

The life of an affix can also be subject to social developments, so movements like feminism can also affect the decay in use of particular affixes, as is the case in present-day times with the derivative suffix *-ess*[18].

3.2.1. Prefixation and suffixation

Only a few instances of the use of *prefix* as a noun can be found in early grammars. Martin (1748: 103-5) refers to "Saxon prefixes" when listing words such as *ashore*, *overflow* and *unable*. Priestley (1768: 141) makes reference to the "prefixes *re* and *pre*" and Coote (1788: 180-1) says that "the initial syllable of *imperfect*, *infertile* is a prefix'" (Sundby 1995: 88-9). Prefixes can be plainly defined as affixes that precede bases. As we have already explained, prefixes are normally class-maintaining and stress-neutral, so the word belongs to the same category and keeps the stress on the same syllable regardless of the presence or absence of prefixes (Katamba 2005: 58). However Lenski (2000: 4) mentions a few exceptions, such as:

[18] My own experience in Ireland as a Spanish teacher of university drama students has shown me that women actors do not consider themselves as (potential) *actresses*, but *actors*, exactly as their male counterparts.

> The deadjectival verbalizing prefix *en-*$_2$ (as in *enlarge*, *embody*, etc.), the denominal verbalizing prefix *de-* (as in *debark*, *debone*) and the denominal or deverbal prefix *a-* which forms predicative adjectives (as in *aglitter*, *ablace*), and the semantically unpredictable verbalizing prefix *be-*.

The following examples extracted from our corpus serve to illustrate Plag's (2003: 98-101) semantic classification of prefixes, considering that:

(i) They quantify the base that they precede:

uni- (='one') → *unicorn* (Hill, *CETA* 1754: 10)

bi- (='two') → *bissextile* (Fuller, *CETA* 1732: 17)

multi- (='many') → *multiform* (Reid, *CEPhiT* 1764: 8)

poly- (='many') → *polytheism* (Greene, *CEPhiT* 1727: 9)

semi- (='half') → *semicircle* (Fuller, *CETA* 1732: 1)

omni- (='all') → *omnipotent* (Cheyne, *CEPhiT* 1705: 41)

micro- (='small') → *microscope* (Cheyne, *CEPhiT* 1705: 21)

macro- ('large') → *macro-economics* (no instances found in the corpus)

hyper- ('excess') → *hyperbolical* (Curson, *CETA* 1702: 377)

over- ('excess') → *over-measure* (Dunton, *CEPhiT* 1710: 332)

under- ('not sufficiently') → *undertaking* (Dunton, *CEPhiT* 1710: 1)

(ii) They confer bases with a locative meaning:

circum- (='around') → *circumgyration* (Whiston, *CETA* 1715: 16)

counter- (='against') → *counter-motion* (Macaulay, *CEPhiT* 1783: 36)

endo- (='internal') → *endocentric* (no instances found in the corpus)

epi- (='on, over') → *epicycle* (Morden, *CETA* 1702: 2)

inter- (='between') → *intermarriage* (Hutcheson, *CEPhiT* 1755: 167)

intra- (='inside') → *intramuscular* (no instances found in the corpus)

para- (='along with') → *paraphrase* (Campbell, *CEPhiT* 1776: 49)

retro- (='backwards) → *retrogradation* (Curson, *CETA* 1702: 355)

trans- (='across) → *transposition* (Morden, *CETA* 1702: 33)

(iii) They attach temporal meaning to bases:

ante- (='before') → *antediluvian* (Bolingbroke, *CEPhiT* 1754: 33)

pre- (='before') → *predominance* (Crombie, *CEPhiT* 1792: 55)

fore- (='before') → *foreknowledge* (Kirkpatrick, *CEPhiT* 1730: 21)

post- (='after') → *posthumous* (Butler, *CEPhiT* 1736: 26)

neo- (='new') → *neoclassical* (no instances found in the corpus)

(iv) They add negative connotations:

a(*n*)- → *achromatic* (no instances found in the corpus)

de- → *degradation* (Crombie, *CEPhiT* 1793: 29)

dis- → disaffection (Balguy, *CEPhiT* 1733: 35)

in- → inconclusiveness (Greene, *CEPhiT* 1727: 1)

non- → non-entity (Bolingbroke, *CEPhiT* 1754: 9)

un- → unquietness (Dunton, *CEPhiT* 1710: 334)

(v) Other:

mal- (='wrong, evil') → *malfunction* (no instances found in the corpus)

mis- (='badly, wrongly') → *misapprehension* (Wilson, *CETA* 1773: 17)

pseudo- (='false, deceptive) → *pseudo-martyr* (*OED*, no instances found in the corpus)

co- (='together, jointly') → *colatitude* (Watts, *CETA* 1726: 23)

vice- (='in place of') → *vice-president* (Vince, *CETA* 1790: 6)

Marchand (1953: 246) defines a suffix as "a derivative final element which is or formally was productive in forming words. A [suffix] has semantic value, but it does not occur as an independent speech unit". Sundby (1995: 105) claims that the term does not appear in early grammarians' works. Most of them referred to 'termination', coming from *terminatio* in Latin grammars, and 'ending' for both inflection and derivation. Nevertheless, there is an instance of its early use in the *OED*, perhaps the first one ever. Lowth (1778: 243) makes use of the word in a note to his translation of *Isaiah*: "These being all the places, where this word occurs without a suffix" (*OED*). Again, I will resort to Plag's list of nominal suffixes (2003: 86-92):

-*age* → from Old French origin, then a living English suffix, it derives either collective nouns, such as *concubinage* (Hutcheson, *CEPhiT* 1755: 162) or nouns that represent the outcome of an action, as in *assemblage* (Campbell, *CEPhiT* 1776: 42)

-*al* → from Latin origin (Lat. -*alis*). It derives nouns that symbolise the result of an action, as in *proposal* (Dunton, *CEPhiT* 1710: 8).

-*ance* → from French origin. It derives action nouns such as *attendance* (Hutcheson, *CEPhiT* 1755: 150).

-*ant* → from French origin. It derives nouns referring to persons, as is the case of *disputant* (Greene, *CEPhiT* 1727: 11).

-*cy* → from Latin origin (Lat. -*cia*, -*tia*). It derives nouns related to state, property, quality or fact, such as *celibacy* (Greene, *CEPhiT* 1727: 7), *legacy* (Collins, *CEPhiT* 1717: 72), *accuracy* (Whiston, *CETA* 1715: 9) and *fallacy* (Reid, *CEPhiT* 1764: 28).

-*dom* → from OE origin, and semantically related to -*hood* and -*ship*. It derives abstract nouns "that can be paraphrased as 'state of being X'" (Plag 2003: 88), as in *whoredom* (Dunton, *CEPhiT* 1710: 333).

-*ee* → from Anglo-Norman origin, it is one of the few autostressed affixes in English, and was originally an adaptation of -*é*, present in certain Anglo-French past participles. Examples with -*ee* denote "sentient entities that

are involved in an event as non-volitional participants" (Plag 2003: 88), as is the case of *trustee* (Hutcheson, *CEPhiT* 1755: 165).

-*eer* → from French origin, representing the French suffix -*ier*. Nouns derived with this suffix denote persons concerned or dealing with something, as *in scrutineer* (Kirkpatrick, *CEPhiT* 1730: 21).

-*er* → from OE origin, it derives agentive nouns from action verbs. It can be considered the counterpart of -*ee* as regards actions. Whereas nouns ending in the latter indicate persons suffering the consequences of an action, nouns derived with the former indicate the person that performs the action. Examples abound: *traveller* (Fuller, *CETA* 1732: 12), *adulterer* (Dunton, *CEPhiT* 1710: 334), *supporter* (Wollstonecraft, *CEPhiT* 1792: 51), and so forth. Apart from coining deverbal nouns, -*er* can also derive nouns from nouns, as are the cases of *Laplander* (Hume, *CEPhiT* 1748: 25) and *biographer* (Hume, *CEPhiT* 1748: 36).

-(*e*)*ry* → from French origin, it provides a locational aspect to the derived word, such *as nursery* (Harris, *CETA* 1719: 24), or collectivity, as in *machinery* (Hutcheson, *CEPhiT* 1755: 153).

-*ess* → from French origin, it forms nouns referring exclusively to female persons or animals. It is not very productive, the *OED* listing twenty-three coinages only (Plag 2003: 90), and in the twenty-first century is also menaced by its sexist connotations (vid. footnote 7). Some examples

are *enchantress* (Campbell, *CEPhiT* 1776: 45), *adulteress* (Hutcheson, *CEPhiT* 1755: 176), *poetess* (Dunton, *CEPhiT* 1710: 6).

-*ful* → from OE origin, it can be found in other Teuton languages too. It is derived from the free form *full* as explained in section 3.2. It can be an adjectival suffix and when used as a nominal suffix it derives measure partitive nouns (Plag 2003: 90), as in *handful* (Bolingbroke, *CEPhiT* 1754: 33).

-*hood* → from OE origin, it derives abstract nouns referring to state, as in *childhood* (Butler, *CEPhiT* 1736: 23) and *manhood* (Crombie, *CEPhiT* 1793: 28), or collectivity, such as *neighbourhood* (Gordon, *CETA* 1726: 102).

-(*i*)*an* (and its variant -*ean*) → from Latin origin (Lat. -*ianus*) meaning 'of or belonging to' (*OED*). It is usually attached to proper names, indicating 'support of', as in *epicurean* (Collins, *CEPhiT* 1717: 58) and *Pythagorean* (Greene, *CEPhiT* 1727: 12). Other examples such as *mechanician* (Campbell, *CEPhiT* 1776: 3) *or civilian* (Ferguson, *CEPhiT* 1769: 82) are more related with the meanings of 'person having to do with something' and 'person from X origin', respectively (Plag 2003: 90).

-*ing* → from OE origin, this is a very productive suffix because it can attach virtually to any verb. It has posed a lot of extra work in our research,

given the multiple possibilities for *-ing* words to belong to different parts of speech, so manual disambiguation was an absolute must. We will pay special attention to the evolution of this suffix across the history of the English language due to its versatility. The original function of this affix, as the *OED* states, was to coin nouns of action, like *dancing* (Wollstonecraft, *CEPhiT* 1792: 43) but already in OE they coined nouns that described a completed action, like *breeding* (Reid, *CEPhiT* 1764: 8). They started off as abstract nouns but later developed into concrete nouns with the possibility of adding pluralised forms, as in *dwellings* (Harris, *CETA* 1719: 31). These uses developed greatly during the ME period, "and in the 14th c. the formation became established (*OED*).

-ion → from Latin origin, this suffix has several allomorphs depending on the base. Apart from cases such as *connect·ion* (Turnbull, *CEPhiT* 1740: 16), verbs ending in *-ify* require the allomorph *-ification*, as in *modif·ication* (Collins, *CEPhiT* 1717: 47). In cases such as the verb *dissolve*, it requires a phonological change and the allomorph *-tion* to become *dissolu·tion* (Butler, *CEPhiT* 1736: 14), meanwhile verbs like *compete* are derived into *compet·ition* (Balguy, *CEPhiT* 1733: 42). This suffix bears another significant phonological characteristic: all resulting words derived by *-ion* have their primary stress on the penult syllable, which means that the

suffix can cause a stress shift. In spite of the fact that most bases to which *-ion* attaches are verbs, Plag (2003: 91) emphasises that,

> There is also a comparatively large number of forms where *-ation* is directly attached to nouns without any intervening verb *in -ate*. These forms are found primarily in scientific discourse denoting chemical or other substances as bases (e.g. *epoxide-epoxidation*, *sediment-sedimentation*).

-ism → from French *-isme* (Lat. *-ismus*), it forms abstract nouns from either concrete nouns or adjectives to denote "state, condition, attitude, system of beliefs or theory" (Plag 2003: 90) as in *idiotism* (Hutcheson, *CEPhiT* 1755: 167), *favouritism* (Burke, *CEPhiT* 1770: 5), *scepticism* (Hume, *CEPhiT* 1748: 13), *paganism* (Greene, *CEPhiT* 1727: 9), and *Athenianism* (Dunton, *CEPhiT* 1710: 1), respectively.

-ist → from French origin, it derives agentive nouns from either nouns, as in *botanist* (Macaulay, *CEPhiT* 1783: 45) and *canonist* (Hutcheson, *CEPhiT* 1755: 173), or adjectives, as in *naturalist* (Campbell, *CEPhiT* 1776: 15) and *optimist* (Wollstonecraft, *CEPhiT* 1792: 79). From a semantic point of view, nouns derived in *-ist* are of the kind 'follower of', as in *dogmatist* (Ferguson, *CEPhiT* 1769: 54), 'devoted to', as in *anatomist*

(Hume, *CEPhiT* 1748: 8) and *novelist* (Wollstonecraft, *CEPhiT* 1792: 63).

-ity → from French origin (Fr. *-ité* < Lat. *-itas*), it derives abstract nouns from comparatives and adjectives in general, and resulting nouns usually denote qualities or states. Examples of qualities are *authenticity* (Reid, *CEPhiT* 1764: 5), *sincerity* (Dunton, *CEPhiT* 1710: 9), and of states, *security* (Hume, *CEPhiT* 1748: 7) and *solemnity* (Balguy, *CEPhiT* 1733: 27). From a phonological point of view, *-ity* shares with *-ion* the potentiality to alter the stress of the base, so bases preceding *-ity* are stress on the antepenult syllable. Considering the previous cases, *authéntic* > *authentícity*, *sincére* > *sincérity*, *sólemn* > *solémnity* and so on.

-ment → from Anglo-Norman/Old French origin, it derives abstract action nouns mainly from verbs and less commonly from adjectives. Indeed in our corpus we have found only one instance of *-ment* following an adjective: rudiment (Campbell, *CEPhiT* 1776: 11). All other examples are derived from verbs, as *astonishment* (Bonnycastle, *CETA* 1786: 43), *disengagement* (Hutcheson, *CEPhiT* 1755: 182) and *retrenchment* (Burke, *CEPhiT* 1770: 17).

-ness → from OE origin. Plag (2003: 92) defines it as "perhaps the most productive suffix of English". We shall see in the chapter devoted to our

analysis that this is not quite so in the scientific register of eighteenth-century English or, at least, in the restricted body of texts that makes up our corpus. *-ness* derives abstract nouns from almost any other word category, mainly from adjectives. The following examples illustrate the use of the suffix in our corpus with adjectives, as in *agreeableness* (Collins, *CEPhiT* 1717: 70) and *duskishness* (Charlton, *CETA* 1735: 20), and from participial adjectives as well, as are the cases of *abstractedness* (Hume, *CEPhiT* 1748: 18) and *interestedness* (Ferguson, *CEPhiT* 1769: 97).

-ship → from OE origin, it shares noun-forming characteristics with *-age*, *-hood* and *-dom*. It is usually added to nouns and adjectives to derive abstract nouns that denote state or condition, such as denominal *partnership* (Hutcheson, *CEPhiT* 1755: 150) and *sonship* (Kirkpatrick, *CEPhiT* 1730: 8), and deadjectival *hardship* (Balguy, *CEPhiT* 1733: 16).

In his classification Plag groups *-ance* together with its variants *-ence*/*-ancy*/*-ency*, *-ant* with its variant *-ent*, *-cy* with *-ce* and *-er* with *-or* but, as we will see in the analysis in chapter 4 I have decided to keep them separate, because my analysis focuses also on the etymological origin of affixes —this information has been extracted from the *OED* in the previous list—, and all of these come from different sources and enter the English language at different stages in its evolution. Besides, my analysis has included more suffixes found in the

corpus, which have not been listed by Plag. All of this will be explained in chapter 4.

3.2.2. Infixation

There is a special kind of affixes, if they can be considered as such, which contravenes the rule of uninterruptibility of the word. Also, if we are to accept Trask's (1993: 175) definition of morpheme as indecomposable as explained in section 2.3 above, we will see that cases such as *absobloominglutely*, *fanfuckingtastic* and *kangabloodyroo* breach both. The inclusion of infixation as a marginal word-formation process will be briefly revised from morphological, phonological, syntactic and semantic points of view. I will try to prove whether it is a single unified process and whether it should be studied within the framework of morphology. Camiña-Rioboo (2005: 77-8) proposes several reasons to validate or invalidate infixation from the morpho-phonological and morpho-syntactic standpoints. Firstly, if we are to follow my previous definition of affixes as bound morphemes in section 3.2, the examples mentioned above show that -*blooming*-, -*fucking*-, and -*bloody*- are free morphemes. Therefore, we could be dealing with something other than affixes and affixation, some sort of compounding perhaps. In cases such as *infuckingcredible* and *uneffingbelivable* we could consider them as prefixes that follow other prefixes, but they are never used alone before a base, so they are definitely not prefixes. Secondly, from the point of view of phonology, two

premises must be stated beforehand: (i) infixes must have more than one syllable, and (ii) there are no specific rules for infixes to break up words. In general, they are inserted before the stressed syllable, as is the case of *bloody*, which can only precede stressed syllables (Burridge 2004: 11), as in *imbloodypossible*, *guaranbloodytee*, but other epithets like *effing* do not, as is the case of *uneffingbelievable* mentioned above. And thirdly, they behave syntactically like adjective modifiers, whereas affixes do not bear any syntactic function. There are instances in which a only one word is used as an infix, as seen in the previous cases, whereas on some other occasions a phrase is used as an infix, as in *indegoddampendent*. Other phonological restrictions of infixation have to do with the length of the base, which must have at least three syllables or more. Plag (2003: 103) states that "infixation in English is determined by the metrical extension of the base", and as we mentioned above, by the metrical extension of the infix as well, so it can be "regarded as a case of prosodic morphology, i.e. a kind of morphology where prosodic units and prosodic restrictions are chiefly responsible for the shape of complex words". Other authors like Burridge (2004: 11) add one more semantic constraint to infixation. An infix like *bastard*, she claims, cannot be used in every situation, and speakers would refer to their television as their *telebastardvision*, because it is something that one might call one's television if it does not work properly, but they would never use *bastard* in the context of *bloody* or *fucking*, like **fanbastardtastic*.

Infixation is a relatively new phenomenon in the history of the English language, generally restricted to a more informal register of the language, and mainly oral. Therefore I have not come across any instance of infixed words in our study. However there is a process similar to infixation called *tmesis*, defined by the *OED* as the "separation of the elements of a compound word by an intervening word or words, and it is used mainly in informal speech". My corpus is formed by scientific texts, so I will look for instances of *tmesis* in a highly literary register of English, as can be seen in the following examples taken from the corpus:

(6) And *how* loath *soever* one is to own it (Astell, *CEPhiT* 1700: 60)

Now of *how* little force *soever* this objection may be in other respects (Astell, *CEPhiT* 1700: 61)

In *what* proportion *soever* it is given (Bolingbroke, *CEPhiT* 1754: 27)

Therefore let the Earth be in *what* point *soever* of her orbit (Fuller, *CETA* 1732: 2)

Most occurrences, as we can see, contain the compounds *whatsoever* and *howsoever*. They are sometimes interrupted a noun or a noun phrase. These examples show that two elements, an adjective and a noun, form the maximum complexity achieved by the noun phrase interrupting the compound.

3.2.3. Combining Forms and interfixation

Adams (1973: 128-33) and Bauer (1983: 213-6) give special importance to a specific group of affixes or pseudo-affixes that they label "combining forms". Their combination form what they call "neoclassical compounds", whereas other authors (Carstairs-McCarthy 2002: 66) denominate them periphrastically 'compounds containing bound combining forms'. They are explained in the *OED* in the following manner:

> In Latin and other languages, many words have a special combining form which appears only in compounds (or only in compounds and derivatives)... The foreign-learned part of the English vocabulary also shows a number of special combining forms; cf. *electro*, combining form of *electric*, in such compounds as *electromagnet*.

Examples of this type would be *astro-*, *hydro-*, *-crat*, *-naut*, *-phile*, *-phobe*, etcetera, which can be considered affixes, since they are occasionally added to lexemes as if they were proper affixes:

an·electric	*photo·electric*
music·al	*music·ology*

If we consider *an-* to be a prefix, we may be tempted accept that *photo-* is also a prefix. Likewise, if we consider *-al* to be a suffix, we can take *-ology* as a

suffix as well. However, if we follow this line of argument we can find instances —as I have already mentioned in section 2.1 above— in which a word is formed by a prefix and a suffix without a root of any kind, words such as *diacosm* (Greene, *CEPhiT* 1727: 13), *perimeter* (Whiston, *CETA* 1715: 5), *zoophyte* (Campbell, *CEPhiT* 1776: 3), *ontology* (Bolingbroke, *CEPhiT* 1754: 31) or *apology* (Astell, *CEPhiT* 1700: 74) would belong to this group. Nevertheless, I can notice some behavioural differences between those I consider to be "pure" suffixes and those that Bauer (1983: 214) denominates Final Combining Forms (FCFs). Firstly, only FCFs can be combined with Initial Combining Forms (ICFs), originating combinations such as *electroscope*, or *electrolyte*, whereas examples such as **electroization*, **electroness* or **electroesque* do not exist at all. At the same time, it seems that FCFs can only be combined with ICFs, among other reasons because ICFs usually end in a vowel, unless they are combined with a formative that also starts with a vowel (*chronology*).

Prefixes are different from ICFs in the sense that the former cannot combine with FCFs. If prefixes end in a consonant (*arch-*, *circum-*, *dis-*, *en-*, *ex-*, *in-*, *mis-*, *non-*, *sub-*, *un-*, *vice-*, etcetera) the impossibility lies on phonological grounds since, as I have explained above, FCFs require a first element ending in a vowel. In spite of this, a lexeme can become an ICF if it ends in a vowel, in which case it can be added directly to the FCF, as it happens in *negrophile*, or it will need the addition of *-o-* if it ends in a

consonant, as happens in *jazzophile*. Finally, prefixes ending in a vowel (*anti-*, *extra-*, *co-*, *infra-*, *intra-*, *meta-*, *mini-*, *pre-*, *retro-*, *semi-*, *supra-*, *ultra-*, etcetera), cannot apparently combine with FCFs: **prephile*, **co-ology*. The 'linking element', as Plag (2003: 157) labels it, has been defined by Bauer (2004: 57) as an interfix, or "an affix which occurs between two elements, linking them together". Bauer provides an example in German to illustrate de use of the affix: *Liebesbrief* (love-letter), in which the *-s-* is an interfix between *Liebe* and *Brief*. Examples of this kind, where the interfix may be originated on inflectional grounds, can also be found in English, as is the case of *kinsman* (Hutcheson, *CEPhiT* 1755: 174), in which the *-s-* stands for the genitive case, in line with the German examples given by Bauer above. Anyway, and in case that Bauer's arguments can be considered as valid today, there is no doubt that the fast evolution of the English language may make any reasoning, no matter how accurate it happens to be, lose validity if we pay attention to contemporary cases such as *lunarnaut*, *selfcide*, *supercrat*, *bio-energy*, *geoelectric* and so on.

3.2.4. Circumfixation

In some cases a prefix and a suffix are used together in order to enclose a base. When none of the two affixes is used in isolation but, on the contrary, they seem to constitute one single morpheme, they are usually called circumfixes. In German, for example, the past participle of weak verbs is formed by the

simultaneous addition of the prefix *ge-* and the suffix *-t*, as it was in Old English.

Ger. *film·en*	'to film'	*ge·film·t*	'filmed'	**ge·film, film·t*
Ger. *frag·en*	'to ask'	*ge·frag·t*	'asked'	**ge·frag, frag·t*
Ger. *lob·en*	'to praise'	*ge·lob·t*	'praised'	*ge·lob, lob·t
Ger. *zeig·en*	'to show'	*ge·zeig·t*	'shown'	*ge·zeig, zeig·t

This process cannot be observed in either Present-day English or in late Modern English, but it is relevant enough to other Germanic languages to devote some attention to it here.

3.2.5. Transfixation

There is a special class of infixes (Bauer 1988: 24-5) that involves discontinuous bases and affixes. The vocabulary of Semitic languages is built on a group of roots, mostly trilateral and consonantal, which can never occur in isolation. Along with them there is a group of affixes, generally vocalic —though they may also be formed by vowels and consonants—, which are combined with the roots by means of certain rules. For this reason they are called transfixes. Sometimes transfixes are also accompanied by prefixes and suffixes, which modify their meaning and serve as inflections, in the same way as other languages make use of suffixes for the same purposes. Each transfix occurs in a fixed position in respect to the root, but this position can vary from

one transfix to another. Besides, the incorporation of affixes may affect the structure of the root: it may result in, for instance, the reduplication of some of the consonants; some of them allow consonant grouping, whereas others will not. Some examples are:

Arab. k-t-b (= *inscribing, marking, writing*)

Arab. *katîb* *writer*

Arab. *kitaba* *the act of writing*

Arab. *kitâb* *book*

Arab. *kutubî* *bookdealer*

This process has never occurred in any period of the history of the English language. It has been given some attention here because it may be interesting to observe how different can be the affixation processes in different languages. The next sections will deal only with processes that can happen in English.

3.3. Compounding

The second word-formation process that I will comment on is compounding. Bauer (1983: 28) defines a compound as "a lexeme containing two or more potential stems". In later years he has somewhat expanded this definition in his *Glossary of Morphology* (2004: 32), in which he adds to the previous definition the condition that the compound "does not have any derivational affix which

applies to the combination of stems". We could add to these descriptions that since a stem must contain at least a root, it can be affirmed that a compound is formed by at least two roots. Early grammarians were conscious of the great capacity of the English language to coin new compounds on a regular basis, so they "make much of the richness and diversity of composition as a process of English word-formation" (Sundby 1995: 31). Some of them defend vehemently the use of compounding to counteract the invasion of Latin words entering the English language in Early Modern times, as I already mentioned in chapter 1 (vid. information on the *inkhorn controversy* in section 5.3). Among these early grammarians, Gill (1619: 109-10) justifies its use in the following manner:

> Because a large number of our words are monosyllables, and are freely adaptable to compounding, writers should devote themselves more eagerly (with greater advantage) to this matter, so that they may express their ideas by an apt compounding of our words, and may therefore enrich (almost to excess if they should wish it so greatly) a language already most copious in itself, rather than hide its native elegance under a foreign discourse.

The theory behind compounding seemed to be clear, as can be drawn from the definition given by Greenwood (1711: 195): "a Compounded Word is, when two or more Words goe [sic] to the making up of one. Words in English are compounded, either with a Preposition, or with some other Part of Speech".

The truth is, however, that not all of them considered compounding outside etymology (affixation), so sometimes the terms are slightly mixed up, as can be seen in some of the following examples quoted by Sundby (1995: 32-5). For instance, Jones (1701: 125-8) considers compounds from the phonological point of view, distinguishing derivatives from primitives by the addition of syllables, before or after it, or both. For Jones *cleare·th*, *clear·ly*, *safe·guard*, *ad·judge*, *judge·able* and *ad·judge·able* are all compounds. We may infer that compounding and both inflectional and derivational morphologies are present in Jones' conception of compounds. Other examples considered as compounds by Martin (1748: 26) and Murray (1795: 16), such as *non·sense*, *un·used*, *name·less*, *grace·ful*, *never·the·less* etc. suggest that confusion in terminology was customary in the seventeenth and eighteenth centuries.

Nowadays linguists debate on the possibility of studying compounding from the standpoint of syntax, rather than morphology. They agree on the fact that noun+noun compounding is a very productive device to coin new words in English, but propose to study this process as a nominal phrase rather than a morphological item. Carstairs-McCarthy (2002: 59) provides the following examples that serve to distinguish compounds from noun phrases from a phonological standpoint by means of stress:

(7)	*green hóuse*	*gréenhouse*
	‘a house that is green’	‘a glass structure for growing vegetables’
	black bóard	*blackboard*
	‘board that is black’	‘bóard to write on’
	silk worm	*sílkworm*
	‘worm made of silk’ (e.g. a toy)	‘ caterpillar that spins silk’
	hair nét	*háirnet*
	‘a net maid of hair’	‘a net for covering hair’

In English it is the first element of a compound that generally carries the stress, so compounds tend to be right-headed (Carstairs-McCarthy 2002: 61). Therefore the sequences on the left column are noun phrases, whereas the ones on the right column are compounds. Adams (1973: 65) states that compounds of the type V(-*ing*) + Subject(N) can be distinguished semantically from free phrases, because the present participle does not show the progressive aspect typical of free phrases. Therefore *working man* is not a free phrase because it means ‘a man who works’, and not ‘a man who is working’. Still from the semantic point of view it is agreed that the meaning of compounds is less predictable than that of noun phrases, or absolutely not predictable at all. If the head of the compound is the first element, then we have endocentric compounds and the meaning is related to one or two of the elements; if, on the

contrary, there is no explicit head in the compound we are dealing with exocentric compounds, and the meaning needs to be inferred. Plag (2003: 146) indicates that exocentric compounds are generally restricted to denoting "human beings or higher animals" but, as we see in the exocentric compounds found in our corpus, they may deal with many other fields too. The following examples illustrate cases of endocentric and exocentric compounds:

(8) Endocentric

burying-place (Kirkpatrick, *CEPhiT* 1730: 27): 'place to bury people'

cannon-ball (Bonnycastle, *CETA* 1786: 37): 'projectile for cannons'

fly-trap (Smellie, *CEPhiT* 1790: 8): 'artifact to catch flies'

love-visit (Dunton, *CEPhiT* 1710: 14): 'a visit to show love'

Exocentric

fox-glove (Smellie, *CEPhiT* 1790: 43): 'an ornamental flowering plant' (*OED*)

Milky-Way (Gordon, *CETA* 1726: 66): 'the external disk of our galaxy' (*OED*)

Nosegay (Smellie, *CEPhiT* 1790: 47): 'a small bouquet of flowers'

Whitlow (Smellie, *CEPhiT* 1790: 11): 'suppurative inflammatory sore' (*OED*)

Plag (2003: 161) mentions another characteristic typical of compounds that differentiates them from phrases. He summarises the theory on compounds put forward by Stockwell and Minkova (2001) who claim that "compounds are often lexicalised, a property not typical of syntactic phrases" (Plag 2003: 161), which would strengthen the potentiality of compounds to be considered morphologically valid. I have considered Marchand's (1969) lists on the different types of part-of-speech combinations for compounds, being noun + noun and adjective + noun the most productive patterns. Since I observe only nouns in my corpus, the combinations proposed by Marchand that do not result in nouns have been obviated and two more classes have been added (pronoun + noun and name + noun), probably as a consequence of nonce-formations. These are as follows:

(9) Noun + Noun

dog-star (Curson, *CETA* 1702: 352)

bell-polypus (Smellie, *CEPhiT* 1790: 47)

sap-vessel (Smellie, *CEPhiT* 1790: 32)

gall-insect (Smellie, *CEPhiT* 1790: 6)

goose-grass (Smellie, *CEPhiT* 1790: 37)

ground-work (Campbell, *CEPhiT* 1776: 1)

gunpowder (Crombie, *CEPhiT* 1793: 15)

load-stone (Greene, *CEPhiT* 1727: 7)

quagmire (Reid, *CEPhiT* 1764: 35)

quick-silver (Greene, *CEPhiT* 1727: 3)

rope-dancer (Collins, *CEPhiT* 1717: 66)

sea-plant (Smellie, *CEPhiT* 1790: 38)

time-piece (Bryan, *CETA* 1797: 118)

mob-government (Burke, *CEPhiT* 1770: 15)

slave-trade (Hutcheson, *CEPhiT* 1755: 161)

Adjective + Noun

all-suficiency (Kirkpatrick, *CEPhiT* 1730: 16)

long-suffering (Wollstonecraft, *CEPhiT* 1792: 64)

back-stairs (Burke, *CEPhiT* 1770: 9)

burying-place (Kirkpatrick, *CEPhiT* 1730: 27)

brute-animal (Cheyne, *CEPhiT* 1705: 28)

commonwealth (Burke, *CEPhiT* 1770: 6)

free-will (Cheyne, *CEPhiT* 1705: 7)

good-breeding (Astell, *CEPhiT* 1700: 64)

hot-house (Smellie, *CEPhiT* 1790: 11)

ill-nature (Astell, *CEPhiT* 1700: 59)

vain-glory (Ferguson, *CEPhiT* 1769: 75)

equinoctial-circle (Fuller, *CETA* 1732: 4)

cometic-moon (Costard, *CETA* 1767: 290)

Name + Noun (not in Marchand)

London-measure (Fuller, *CETA* 1732: 9)

Frome-spy (Dunton, *CEPhiT* 1710: 5)

Verb + Noun

pastime (Reid, *CEPhiT* 1764: 31)

helpmate[19] (Wollstonecraft, *CEPhiT* 1792: 39)

stumbling-block (Wollstonecraft, *CEPhiT* 1792: 66)

drawing-room (Harris, *CETA* 1719: 2)

Verbal phrase

well-being (Turnbull, *CEPhiT* 1740: 26)

well-doing (Balguy, *CEPhiT* 1733: 27)

welfare (Ferguson, *CEPhiT* 1769: 71)

Noun + Verb

sun-rising (Watts, *CETA* 1726: 42)

sun-setting (Watts, *CETA* 1726: 42)

Pronoun + Noun (not in Marchand):

she-wit (Dunton, *CEPhiT* 1710: 17)

Preposition + Noun

by-stander (Astell, *CEPhiT* 1700: 45)

[19] It could also be considered as a Noun + Noun compound.

Noun + Adjective

cousin-german (Hutcheson, *CEPhiT* 1755: 171)

sum-total (Astell, *CEPhiT* 1700: 56)

captain-general (Harris, *CETA* 1719: 52)

Preposition + Noun

behalf (Balguy, *CEPhiT* 1733: 27)

afternoon (Curson, *CETA* 1702: 339)

Adverb + Noun

inside (Reid, *CEPhiT* 1764: 39)

insight (Curson, *CETA* 1702: 364)

offspring (Cheyne, *CEPhiT* 1705: 2)

well-doing (Balguy, *CEPhiT* 1733: 27)

Adverb + Verb

inlet (Hume, *CEPhiT* 1748: 25)

Although some attention will be paid to compounding as a word-formation process in the next chapters, especially to analyse its rise or decay in productivity —or creativity, rather—, my view of compounding will be quite marginal. Since it is not the main focus of my study, I have paid attention only to those compounds that the *Coruña Corpus Tool* interprets as single items, that is to say, either compound words that resemble a word-form, or hyphenated compounds, respecting the criteria employed by eighteenth-century

scientists. Therefore, compounds for me will be not only words such as *brimstone* (Dunton, *CEPhiT* 1710: 332), *cobweb* (Balguy, *CEPhiT* 1733: 26) and *jawbone* (Curson, *CETA* 1702: 353), but also *map-maker* (Watts, *CETA* 1726: 21), *land-measurer* (Campbell, *CEPhiT* 1776: 1), *weather-cock* (Collins, *CEPhiT* 1717: 114), listed in the *OED* as established compounds (the last one without the hyphen), together with others not listed such as *ninepin-bowl* (Harris, *CETA* 1719: 7) and *joint-education* (Hutcheson, *CEPhiT* 1755: 159), which can be taken as nonce formations, and even 'exotic' cases which will never be lexicalised for obvious reasons, such as *sun's-beam* (Fuller, *CETA* 1732: 27), *woman's-man* (Harris, *CETA* 1719: 21) and *new-year's-day* (Fuller, *CETA* 1732: 17). These last ones confirm the lack of patterns to use hyphens for the compounding of words in the Modern English period, and *new-year's-day* represents, with its three elements, the longest pattern of compounding in my corpus, the rest being composed by only two. The fact is that the use of the hyphen in compound words has been a matter of great controversy across the history of the English language. Still in our days the situation has not been properly regulated. In this respect, Fowler & Crystal (2009: 243) emphasise the anarchic situation in present-day English as regards the use of hyphens by stating that,

> The chaos prevailing among writers or printers or both regarding the use of hyphens is discreditable to English education. Since it sufficiently proves by its

> existence that neither the importance of proper hyphening nor the way to set about it is commonly known. (Fowler & Crystal 2009: 243)

Back in the seventeenth and eighteenth centuries the situation was, if possible, still more complicated, given the fact that grammarians of the times did not agree on the definition of compounding, derivation, etymology and the rest of the concepts already explained above. As regards hyphens in compounding, some agree on their use in temporary compounds, called 'half-compounds', and no use when they have become familiar with use (lexicalised), and then the hyphen is not needed any longer. Lane (1700: 19) gives a prosodic reason for using or not using hyphens as follows:

> When we compound two or more Words, without putting them under one Accent, we only join them with a Hyphen or Mark of Union, and such may be called Half-Compounds, as a *Water-Spider*. But if the Custom of the Language has put them under one Accent, we must write them in one Word without a Hyphen, as a *Shoomaker*, not a *Shoó-Maker*, a *Highlander*, not a *H[i]gh-lander*.

Also Elphinston (1765) and Priestley (1768) agree on the use of the term half-compounds or 'occasional compounds' referring to those requiring hyphens, as a midpoint or middle state between 'absolute compounds', those in which the two elements are united, and absolute separation (Sundby 1995: 62). Fowler and Crystal (2009: 244) define hyphens as follows:

> A hyphen is a symbol conveying that two or more words are made into one; the union may be for the occasion only (as in most of the examples above), or permanent (as in *fire-irons, committee-man*); the commonest for of temporary union is that in which a phrase (say *Home Rule*) is to be used attributively, i.e. as an adjective to another noun; to this end it must be marked as one word by the hyphen (*the Home-Rule Bill*).

And among their rules for proper use they suggest not to hyphen words that do not expressly need it if they can work separately. They also state that lexicalisation will render the use of the hyphen unnecessary, and the word will become what ancient grammarians used to term 'absolute compounds', if no phonological, morphological or semantic constraints are against it.

3.4. Conversion or zero derivation

Together with affixation and compounding, conversion is nowadays one of the most productive processes to coin new words in English. Before the twentieth century the term *conversion* was not used to encompass this process. As Sundby (1995) meticulously explains, early grammars speak of 'transformation' (Bicknell 1790; Staniford 1800), 'transmutation' (Elphinston 1765; Ash 1775), 'enallage' (Ward 1765; Priestley 1768) etc., but the overall feeling of the resulting words becoming 'derived' from the original is present in most authors' writings, as can be seen in Fenning's words (1771: 13):

"Sometimes [verbs] are derived from [substantives] without any change at all: as *to sail*, *to salt*, *to taste*, from the substantives *a sail*, *salt*, *taste*". In later times, Bauer (2004: 36) describes conversion as "the presumed derivational process which takes place when a word which normally occurs in one word-class takes on the characteristics of a different word class without any change of form." It is noticeable that he uses the word 'presumed' to describe the process, which can be analysed from two perspectives. On the one hand, if we consider word-formation processes as processes in which linguistic material in the form of an affix is 'added' to a base to form a new lexical item, then Bauer's definition would probably exclude conversion from my study. On the other hand, if 'presumed' refers to 'derivational' only, then conversion could be treated from the standpoint of inflection (Myers 1984). Other linguists study conversion under the label 'derivation without affixation" (Plag 2003: 107). But I would like to deepen a little more in this matter, before discarding conversion altogether. In the following paragraphs I will try to justify the inclusion or exclusion of conversion as a derivative process by providing phonological, morphological (affixational) and semantic evidence.

From a phonological point of view, are there any visible differences between converted words and the originals they are derived from? Contrary to my opinion, some authors such as Aronoff & Fudeman (2005: 49) include the function 'shift-stress' and a quality inherent to conversion between nouns and verbs since OE times. Examples of such noun-verb pairs are the following:

(10) *ádvance*$_{\text{Noun}}$ – *advánce*$_{\text{Verb}}$ (Hill, *CETA* 1754: 2)

íncrease$_{\text{Noun}}$ – *incréase*$_{\text{Verb}}$ (Curson, *CETA* 1702: 356)

cóntest$_{\text{Noun}}$ – *contést*$_{\text{Verb}}$ (Reid, *CEPhiT* 1764: 22)

óverthrow$_{\text{Noun}}$ – *overthrów*$_{\text{Verb}}$ (Kirkpatrick, *CEPhiT* 1730: 8)

cónvert$_{\text{Noun}}$ – *convért*$_{\text{Verb}}$ (Collins, *CEPhiT* 1717: 62)

These pairs show changes in stress accompanied by other changes in the pronunciation of some of the vowels. But this characteristic cannot be generalised, since I have also found other instances in which changes cannot be observed, as in:

(11) *addréss*$_{\text{Noun}}$ – *addréss*$_{\text{Verb}}$[20] (Astell, *CEPhiT* 1700: 57)

mistáke$_{\text{Noun}}$ – *mistáke*$_{\text{Verb}}$ (Whiston, *CETA* 1715: 22)

concérn$_{\text{Noun}}$ – *concérn*$_{\text{Verb}}$ (Hill, *CETA* 1754: 14)

pólish$_{\text{Noun}}$ – *pólish*$_{\text{Verb}}$ (Ferguson, *CEPhiT* 1769: 52)

cóver$_{\text{Noun}}$ – *cóver*$_{\text{Verb}}$ (Dunton, *CEPhiT* 1710: 16)

Besides, monosyllables do not obviously show any change in their patterns, as neither do nouns converted from adjectives or adverbs. Therefore, and even though there are some explicit phonological features applicable to conversion,

[20] Both in BE and AE the pronunciation for the verb are the same, i.e. /əˈdrɛs/, however, it can differ in some realisations of the noun in AE: Brit. /əˈdrɛs/ , U.S. /əˈdrɛs/ , /ˈæˌdrɛs/ (*OED*).

too many exceptions seem to thwart any attempt to establish this as a rule from the phonological point of view.

If we continue this line of argumentation above and we are to accept Plag's definition of conversion as "the derivation of a new word without any overt marking" (2003: 107) we must acknowledge that, from the standpoint of derivational morphology, it is rather difficult to imagine the coining of new words without the use of an affix, or at least without the use of an 'overt' one. Indeed, one of the major controversial points when dealing with conversion has to do, for some authors, with accepting the existence of a so-called 'zero-morpheme'. As Balteiro (2007: 40) expounds, some other linguists plainly reject the existence of the zero-morpheme and conversion as a derivational process. She claims that Štekauer refers to conversion as "a unique, specific word-formative process, based upon principles different from those that characterize the concept of derivation" (1992: 43), and "those of zero-derivation (1992: 86). Conversely, Kastovsky (1989) explains that conversion belongs to word-formation founding his statement on two principles, one syntactic and one semantic, and he confers capital importance to the following: "(1.) The fact that the relation between the base and the derived form is pattern-forming and not isolated, and that (2.) a semantic relationship exists between the two words which may be interpreted as a derivational one" (Balteiro 2007: 40). The following example provided by Kastovsky illustrates this point and contributes to provide our semantic view on conversion.

(12) *sing*$_{\text{Verb}}$ – *singer*$_{\text{Noun}}$ 'someone who sings'

cheat$_{\text{Verb}}$ – *cheat*$_{\text{Noun}}$ 'someone who cheats'

The suffix *-er* attaches to the verb *sing* to derive the agentive noun, whereas in the second case, no suffix is needed to derive the performer of the action of cheating. Both cases are parallel, at least semantically speaking, but only the first one shows an overt change in the resulting word, which is labelled by some authors as the 'overt analogue criterion' (Sanders 1988, Plag 1999, 2003)

In English the most common types of conversion are verb to noun, noun to verb, adjective to noun and noun to adjective, "though other types can also be found, but seem to be more marginal (e.g. the use of prepositions as verbs, as in *to down the can*)" (Plag 2003: 108). Examples of conversion into nouns from verbs, adjectives and adverbs, which can be found in my corpus, are listed as follows:

(13) Adjective > Noun

alternative	(Ferguson, *CEPhiT* 1769: 148)
credential	(Hutcheson, *CEPhiT* 1755: 179)
exponent	(Crombie, *CEPhiT* 1793: 50)
fluid	(Costard, *CETA* 1767: 274)
secant	(Whiston, *CETA* 1715: 8)

Verb > Noun

affront	(Harris, *CETA* 1719: 52)
increase	(Curson, *CETA* 1702: 356)
mistake	(Whiston, *CETA* 1715: 22)
glitter	(Astell, *CEPhiT* 1700: 56)
result	(Ferguson, *CETA* 1756: 148)

Adverb > Noun

hereafter	(Butler, *CEPhiT* 1736: 31)
piecemeal	(Kirkpatrick, *CEPhiT* 1730: 35)
wherewithal	(Burke, *CEPhiT* 1770: 12)

Another difficulty that conversion poses for its analysis is trying to guess its random directionality, that is to say, to infer which is the original word and which is the derived one. In order to sort out this problem, we may need to look into the history of the language, but in the period under study here, written records are not always reliable. Lately numerous scholars have resorted to the invaluable help of the online version of the *OED*, and in general they take for granted the coinage dates provided by this dictionary. If we examine, for example, the noun *display* in the following sentence:

(14) And thus, by a very judicious *diſplay* of omnipotence, he confirms the belief of the attentive mind (Macaulay, *CEPhiT* 1783: 45)

And we need to decide on whether the noun is derived from the verb or vice versa, we would look it up in the *OED*, which states that the verb was coined first, around the year 1320, whereas the first recording of the noun dates from the year 1583. Therefore this is an attested case of verb to noun conversion. Besides, some authors (Adams 1973, 104; Quirk *et al*. 1985: 1559) distinguish total from partial conversion, considering the former a type of derivation with semantic differentiation in the resulting word, as would be noun to verb cases such as *to milk* 'to draw milk from a cow or other domesticated animal' (*OED*), and *to beggar* 'to make someone a beggar', and verb to noun cases such as *embrace* (Collins, *CEPhiT* 1717: 108) 'action of folding in the arms', (*OED*) and *fall* (Bonnycastle, *CETA* 1786: 28) 'dropping down from a relatively high position' (*OED*). For Adams, partial conversion is related to syntax, and can be paralleled with ellipsis, as are the following adjective-to-noun cases *Primary* (< *primary planet*) (Ferguson, *CETA* 1756: 159), *ecliptic(k)* (< *ecliptic line*) (Whiston, *CETA* 1715: 14), and *iambick* (< *iambic verse*) (Greene, *CEPhiT* 1727: 7), though the use of these in the plural may also signify that partial conversion is just the first step towards full conversion, as can be seen in the following examples in the converted words in italics.

(15) Then by the Canon of *Sines*, *Tangents*, and *Secants*, the Radius it ſelf may eaſily be found (Whiston, *CETA* 1715: 8)

> You may eaſily diſtinguiſh them to be *Erraticks* or Planets (Harris, *CETA* 1719: 51)

In conclusion, after viewing conversion from several standpoints and given the controversial nature of this process I will deal with it with caution. As I have repeatedly stated in this chapter my analysis intends to deal primarily with affixation, which has in itself enough controversial elements, as we have seen in previous sections. Therefore, my attention to converted nouns will be slightly less comprehensive, but they will not be totally discarded, because I still regard conversion as a significant process —affixing or not— in the coining of new words in English. It might be very interesting to compare the productivity rates achieved in different periods of history, particularly the eighteenth century, by the three main word-formation processes, namely, affixation, compounding and conversion, and also evaluate the possibility that the rise and fall in one of them may somehow affect the others. In sum, nouns that may be considered as converted will be computed, listed regarding its pattern, coinage dates and origin, and studied separately from affixed nouns.

3.5. Other processes of word formation

This section deals with situations in word formation that cannot be included in any of the three major processes. They cannot be compared to affixation, compounding or conversion in terms of numbers of coined elements, but they

are relevant enough to be mentioned in a work that intends to give full coverage to word formation in English.

3.5.1. Word shortening

This process includes back formation, reanalysis or popular etymology, blending, clipping and abbreviations, processes which share a basic forming method that consists of the graphic and phonic reduction of any element in either the original word —either simple or complex— or phrase.

Back formation

Trask (1993) defines back formation as the process by which a word is formed by removing one of its morphemes, taking into account that this is similar to another different morpheme. For instance, *spectate* is formed from *spectator* (Gordon, *CETA* 1726: 111), *edit*$_{\text{Verb}}$ is formed from *editor*$_{\text{Noun}}$, *peddle*$_{\text{Verb}}$ from *pedlar*$_{\text{Noun}}$, *sculpt*$_{\text{Verb}}$ from *sculptor*$_{\text{Noun}}$, and *typewrite*$_{\text{Verb}}$ from *typewriter*$_{\text{Noun}}$ due to an incorrect interpretation of the agentive derivative morpheme, by applying the pattern by which *writer*$_{\text{Noun}}$ is derived from *write*$_{\text{Verb}}$. The same applies to *televise*$_{\text{Verb}}$, back-formed from *television*$_{\text{Noun}}$ mirroring the pattern *revise*$_{\text{Verb}}$ → *revision*$_{\text{Noun}}$. Another example found in the corpus is *wrinkle*$_{\text{Noun}}$, back-formed from *wrinkled*$_{\text{Past/Past participle}}$. In early grammars the only reference to back-formation —without mentioning the term at all— can be extracted from Maittaire's (1712) and Elphinston's (1765) definition of 'primitive'. They

intend to find an explanation for *belove*$_{\text{Verb}}$ as some sort of reversed derivative from *beloved*$_{\text{Adjective}}$, which had been interpreted as a further derivation of the compound *be* + *love* + *-ed*. Etymological research using the *OED* shows, however, that the *belove*$_{\text{Verb}}$ was already used around 1225, whereas *beloved*$_{\text{Adjective}}$ was used for the first time in 1398, invalidating therefore their interpretation. Besides, their attempt could also be inserted in the following section dealing with popular etymology.

Reanalysis or popular etymology

Some authors (Adams, 1973; Millward, 1996) have included this process in the group of back formation. Reanalysis consists of the reinterpretation of a compound word (generally a verb) and the subsequent removal of one of the affixes that is not an immediate constituent. This is the case of *sky-dive*$_{\text{Verb}}$, derived from *sky-diving*, which was originally *sky* + *diving,* but it is reinterpreted as *sky* + *dive* + *-ing.* The same happens in the case of *pied-pipe*, derived from *pied-piper*, which was in origin *pied* + *piper* and it is interpreted as *pied* + *pipe* + *-er.*

Blending

Blending is an unorthodox process by which some parts of two words are combined in order to make a new one. Bauer (2004: 22) claims that blending shares with compounding the presence of two lexemes in the base, and defines

a blend or a 'portmanteau word'[21] as "a word constructed from the beginning of one word and the end of another". This process is characterised by its arbitrariness and unpredictability, and it seems as if the two bases overlapped each other. Typical examples of blending are *smog* (< *smoke* + *fog*), *brunch* (< *breakfast* + *lunch*), *motel* (< *motor* + *hotel*), *chunnel* (< *channel* + *tunnel*). Lately new creative additions have been added, such as *nakations* (< *naked* + *vacations*) (*The New York Times Online* 27/04/2008), and *celebgeny* (< *celebrity* + *progeny*) (*The Irish Independent*: 28/07/2007). Blending has been studied since at least 1674, included in the section devoted to word formation by Wallis in his *Grammatica linguae Anglicanae*, but he and other early grammarians (Greenwood 1711) consider blends formations that the *OED* does not. Some examples provided by Wallis and Greenwood are *gruff* (< *grave* + *rough*), *sprout* (< *spring* + *out*), *trudge* (*tread*/*trot* + *drugde*), *greedy* (< *gripe* + *needy*) and *shatter* (< *shake* + *scatter*), which the *OED* defines as "'obscure', 'uncertain' or 'entirely a matter of conjecture'" (Sundby 1995: 25). In my corpus examples are almost inexistent, being *chickweed* (< *chicken* + *weed*) the only case that might be considered as representative of the process, maybe due to register constraints. In my example the resulting word is a blended compound in which only the first element loses some linguistic material.

[21] Bauer calls blends 'portmanteau words' because "there are two meanings packed up into one word" (Bauer 1988: 39).

Clipping

Bauer (1983: 233) defines clipping as "the process where a lexeme (simple or complex) is shortened, while still retaining the same meaning and still being a member of the same form class". In short, it is a process by means of which a word is derived from a longer one (containing two or more syllables) after removing a part of the original word without any fixed rules. The resultant clipping does not produce a change in meaning, category, or function of the derived word. Nevertheless it changes the style, since the resulting word is used generally in less formal environments than its longer equivalent, because of the familiarity of the speaker with the term in question or with the listener. This process, as happens with blending, is unpredictable and we cannot establish phonological or morphological patterns that may be applied to a choice of examples and foresee how other words would be clipped. Therefore, in the following examples the stressed syllable in the original word is kept in the derived one:

(16) *incenser* > *censer* (Curson, *CETA* 1702: 355)

fantasy > *fancy* (Morden, *CETA* 1702: 6)

withdrawing-room > *drawing-room* (Harris, *CETA* 1719: 2)

distress > *stress* (Balguy, *CEPhiT* 1733: 27)

adventure > *venture* (Collins, *CEPhiT* 1717: 72)

hydropsy > *dropsy* (Campbell, *CEPhiT* 1776: 53)

Whereas in the next set of examples, the stressed syllable is not kept in the derived one: *bi* (< *bisexual*), *deli* (< *delicatessen*), *exam* (< *examination*), *gym* (< *gymnasium*), *lab* (< *laboratory*, with BrE pronunciation), *porn* (< *pornography*) and *prof* (< *professor*). In some other cases, the derived word does contain the stressed syllable, but there is a tendency to germanise the stress pattern, so the stress becomes shifted from the second to the first syllable, such as in *advert* (< *advertisement*), *binocs* (< *binoculars*) and *photo* (< *photograph*).

In the previous samples we can observe that derivative words come from trisyllable or polysyllable words and are generally formed by their first — and sometimes second— syllables; however, the following examples show that at times it is the final part of the original that is kept: *bus* (< *omnibus*), *cello* (< *violoncello*), *copter* (< *helicopter*), *phone* (< *telephone*), *plane* (< *aeroplane*), '*Fro* (< *Afro*), *loid* (< *celluloid*), *Yard* (< *Montagnard*). On some occasions, it is the middle part of the word what remains, as in *flu* (< *influenza*), *fridge* (< *refrigerator*), *jams* (< *pyjamas*), *script* (< *prescription*).

There are also some irregular cases where the type of clipping only affects a few examples that do not adjust to any type of scheme, as in *Maths* (< *Mathematics*) and *specs* (< *spectacles*), which retain the final -*s* present in the original words; *turps* (< *turpentine*), which adds a final -*s* not present in the original; *mike* (< *microphone*) and *bike* (< *bicycle*) present a phonological connection that seems clear in the first case, but not in the first one; *pram* (<

perambulator), which affects only the meaning "hand-carriage, with three or four wheels, for one or two young children, pushed from behind" (*OED*), but not others, and as a result provides a word already existing in English, with the meaning of "flat-bottomed boat"; *Aussie* (< *Australian*), *commie* (< *communist*), *hanky* (< *handkerchief*) and *telly* (< *television*) share the suffix *-y* or *-ie*.

As with other processes, phonological and morphological changes are accompanied by semantic changes. Therefore, words such as *pants* (< *pantaloons*) or *lunch* (< *luncheon*) have supplanted the original long version, which has been abandoned. As a direct consequence, the first ones have stopped being considered as clippings and have become "original" themselves. Once more, and to conclude, we should consider whether the definitions given by Bauer (1983: 233), based on morphological and semantic criteria, and Adams (1973: 135) (based on syntactic criteria) could clash with one another, when we witness examples such as *fan* (< *fanatic*) and *curio* (< *curiosity*) provided by Adams (1973: 136) as instances of semantic change within the group of clippings. From a morphological point of view *fan* does not suffer a category change, but from a semantic perspective it seems unquestionable that the negative connotations (even religious) of *fanatic* have been softened to such a degree in fan that the derivative could stop being considered a clipping altogether. As regards *curio*, it does not undergo a category change either, but its meaning goes from the abstract property of the

original long version to a very concrete object with a specific origin, specifically "from China, Japan and the Far East" (*OED*) in the short version.

Clipping was widely criticised by early grammarians. Swift (1710) and Addison (1711) initiated this wave of criticism, and other authors joined in in subsequent years. Sundby (1995: 30) quotes a passage from Greenwood (1722: 200) in which the general view of the process is clearly expounded:

> But whatever may be allowed for our Forefathers in shortening the Words they borrowed from other Languages; I cannot but find fault with the humour of so miserably courtailing some of our Words in familiar Writings and Conversations, they often lose all but their first Syllables, as in *Mob*, *rep*, *pos*, *incog*, and the like; and as all ridiculous words make their first Entry into a Language by familiar Phrases, I dare not answer for these, that they will not in time be looked upon as a Part of our Tongue.

Acronyming

Acronyming is a relatively modern phenomenon, practically inexistent before the contemporary English period, for the simple reason that it relies too much on the degree of literacy of speakers, who must know the alphabet and the initial letters of the words. Therefore we are not going to deal with it because it falls out the temporal scope of our work. Acronyming has suffered a dramatic increase in the twentieth century, received a strong impulse during both world wars, and consolidated as a relatively productive word-formation process in the

second half of the past century. Defining acronym formation involves a considerable complexity due to the vast amount of possibilities to coin new words, and also to the ambiguity implied in the terminology we are to make use of. Acronyms could be simply defined as "words formed from the initial letters of a phrase" (Adams 1973: 136), a definition that Trask (1993: 5) specifies somewhat more by indicating that these initial letters belong to the main words in the phrase, hence, in general, articles and prepositions are not taken into account when forming acronyms. Some authors establish a division within this group of words, and consider, on the one hand, those formed by initials pronounced as a letter sequence, such as *BBC* (= 'British Broadcasting Company') o *UN* (= 'United Nations') cannot be considered as acronyms proper, but as abbreviations (Bauer 1983: 237), initialisms (Millward 1996: 336) or alphabetisms (López-Rúa 1999: 628). On the other hand, those pronounced as if they were words with regular phonological features, such as *laser* (= 'light amplification by the stimulated emission of radiation') or *NATO* (= 'North-Atlantic Treaty Organisation') do belong to the group of acronyms. In some instances, a specific element is given the appropriate name so that the acronym can be a word with meaning, as it happens with the computer language *BASIC* (= Beginners' All-purpose Symbolic Instruction Code). Finally, Bauer (1988: 40) speaks of another kind of acronyms that "tend to merge into blends when more than one letter is taken from each of the words of the title, as in the German *Ge*(*heime*) *Sta*(*ats*) *Po*(*lizei*), 'secret state police' or

Gestapo." He goes on affirming that even more complex forms could be found in other languages such as Indonesian, but these will be not dealt with in this section.

3.5.2. Proper names as a source of new words

Millward (1996: 206) resorts to Middle English to locate the emergence of new nouns derived from proper names, mainly from people's names, such as *jay* (< Lat. *Gaius*) or *jacket* (< Fr. *Jacques*), and geographic places, such as *damson* (< *Damascus*) or *magnet* (< *Magnesia*). The following periods are even more prolific when it comes to providing new examples, as in *grog* (< *Grogram*), *praline* (< *Pralin*) or *galvanic* (< *Galvan*); *tangerine* (< *Tanger*), *dunce* (< Duns) (Campbell, *CEPhiT* 1776: 50), *epicure* (< Epicurus) (Wollstonecraft, *CEPhiT* 1792: 55) and *vernier* (< *Vernier*) (Vince *CETA* 1790: 10); and also to increase the number of fields, including classic literature and mythology, as in *flora* (< Lat. *Flora*) or *hector* (< Lat. *Hector*); discoveries in botany after the colonisation of America and Africa, together with the suffix *-ia* give shape to words such as *fuchsia* (< *Fuchs*), *camellia* (< *Kamel*) or *magnolia* (< *Magnol*); and the names of some well-known commercial brands, which have become so popular that they have replaced the original nouns, such as *kleenex* (= 'paper tissue'), *victrola* (= 'gramophone'). Examples such as *Band-Aid*, trademark registered by *Johnson & Johnson Limited* have been

converted from a name into a noun, as in the following sentences found in the *OED*:

(17) *Band-Aid... 'a protective dressing for cuts and wounds'*

He was as adhesive as a band-aid.

And they can be subsequently derived either into adjectives and verbs by means of conversion (*vid*. section 3.4), as can be seen in the following examples.

(18) Band-aid [= 'provisional] *measures like tinkering with traffic will not revitalize the downtown area.* (Adj.)

The courts of Victoria…will probably be bandaided [= "aplied as a provisional repair"] *through their present time and space crisis*.

The examples mentioned above and the ones found in the corpus do not present affixes of any kind, and most cases are either conversions or clippings, therefore they will not be dealt with in this work. They are computed in our general list of nouns, but our study will not progress any further in this direction.

3.5.3. Reduplication

Katamba (1993: 180) and Bauer (1988: 25) explain reduplication as a process in which part of the phonological material of a base is taken as an affix in order to use it later within the same word. Sapir (1921: 76) states that:

> Nothing is more natural than the prevalence of reduplication, in other words, the repetition of all or part of the radical element. The process is generally employed, with self-evident symbolism, to indicate such concepts as distribution, plurality, repetition, customary activity, increase in size, continuance.

Sapir's explanation, though probably applicable to some languages in Africa, Asia and Oceania, is not the most adequate to explain reduplication in English, since this linguistic device is not used in the latter language for the same purpose as it is in the former. Millward (1996: 292, 335) proposes a much more interesting diachronic approach by declaring that if we exclude onomatopoeias such as *ha-ha* to represent laugh, the origin of which has to be looked for in the Old English period, we need to wait until the early Modern English period to find reduplicated words, and even at this moment most of them are loans from other languages, such as *dodo* (< Portuguese *doudo* "tonto"), *grugru* (< American Spanish *gru-gru* "palm tree, grub of the palm tree insect") or *haha* (French *haha* "ditch"). Common reduplicated words such

as *mama* or *papa* (French) were not incorporated until the seventeenth century. Probably the only native reduplication of this period could be *so so*.

In contemporary English new reduplications have enriched the fields of children's speech or jargon, as in *yum-yum* (= "delicious"), *hush-hush* (= "secret") o *boo-boo* (= "foolish mistake"). The elaboration of some reduplicative words has suffered an alteration in the vowel of some of their elements; therefore some authors (Bauer 1983: 213) have named it ablaut reduplication. This process started during the period of early Modern English, originating examples such as *fiddle-faddle* (= "trifling talk or action"), *dilly-dally* (= "trifling hesitancy"), and consolidated in Contemporary English with expressions such as *ric-rac* (= "decorative zigzag braid") or *ping-pong* (= "table tennis").

Reduplicative rhymes, which appeared also in early Modern English, exemplify another variant, with examples such as *helter-skelter* (= "chaos") or *hodgepodge* (= "haricot"), and have flourished in Contemporary English with terms such as *hanky-panky* (= "trickery, double or underhand dealing") or *fuddy-duddy* (= "old-fashioned person") and, above all, with the new tendency to form reduplications in which both elements provide their semantic load, as in *brain drain* (= "loss of highly trained people by emigration"), *culture vulture* (= "voracious for culture") or *walkie-talkie* (= "portable radio transmitter and receiver").

4. Productivity

This final section is intended to initiate a short debate on one of the delicate issues that qualify or restrict derivational morphology and which cannot be included in the previous sections, since it does not belong to the groups of units or processes, but to some other linguistic standpoints from which the use of those units and processes can be explained. The definition of productivity has posed a great deal of controversy among linguists since the twentieth century. Kastovsky (1986) and Bauer (2001) have made considerable efforts to defining morphological productivity, its degrees, its diachronic variation, and how to measure it. Bauer (2001: 25) quotes Rainer (1987) to offer six different definitions of productivity present in linguistic literature, some of which can be considered qualitative, others quantitative, some are synchronic, some others diachronic, in terms of the following: (i) the frequency of the output words, (ii) the number of available bases, (iii) the proportion of words actually used to the number of words potentially created by a particular process, (iv) the possibility of forming new words, (v) the probability of new forms occurring, and (vi) the number of new forms occurring in a specified period of time. Meanwhile, Adams (2001: 7) states that "productivity is an essential concept in any discussion of word formation, though it is also complex, elusive and difficult to discuss with any precision". Finally, Plag (2003: 52) assumes that since productivity is merely "the possibility of creating a new word, it should in

principle be possible to estimate or quantify the probability of the occurrence of newly created words of a given morphological category". Notwithstanding the above-explained theories, I will rather use the terms "productive" and "productivity" in a rather loose way across the following chapters and, if any, I will follow the theory proposed by Fernández-Domínguez (2009: 101-2):

> One straightforward option for productivity measurement is equating the number of analysable words derived by a give word-formation process to its productivity, i.e. a text or corpus is chosen, the frequency of the words derived by the process summed up, and the figure obtained by this operation is considered to represent the productivity of the process.

Considering the quantitative approach to data in this work, theories depending on qualitative definitions will not be contemplated. As regards synchronic and diachronic approaches, however, I will try to find some balance between them, since it is my intention to provide both a synchronic review of word formation processes in the eighteenth century and a diachronic evolution for every thirty-year section of the century. Besides, I will pay attention to the peaks and valleys in productivity for the affixes involved in the coining of the words used in the scientific register of English in the late Modern Period, independently from their status as productive or unproductive.

On the other extreme of productivity we can find nonce words or nonce-formations, newly coined words made up for the occasion. These are

unique examples that may have an ephemeral existence, never becoming lexicalised or used, as can be seen in many compound words present in the corpus, which are not present as entries in the *OED*. Sometimes, nonetheless, some of these nonce words become famous for some reason and can be found in dictionaries, as is the case of *yahoo* (Reid, *CEPhiT* 1764: 28), coined by Jonathan Swift in 1726 for his beastly but human-like characters in *Gulliver's Travels*. The noun was converted into a verb without the intervention of any affix in 1868, and later in the twentieth century it became a worldwide phenomenon when it was the name given to the most famous Internet portal. Some linguists (Bauer 2001: 63; Baayen & Lieber 1991: 815) consider, however, that the coinage of simplex words is not related to morphological productivity at all, and has more to do with the theory of linguistic creativity put forward by Chomsky (1966). A discussion on this controversy would probably extend the length of this chapter *ad infinitum*, and it does not really add any relevant information to our study, given that it deals with complex words.

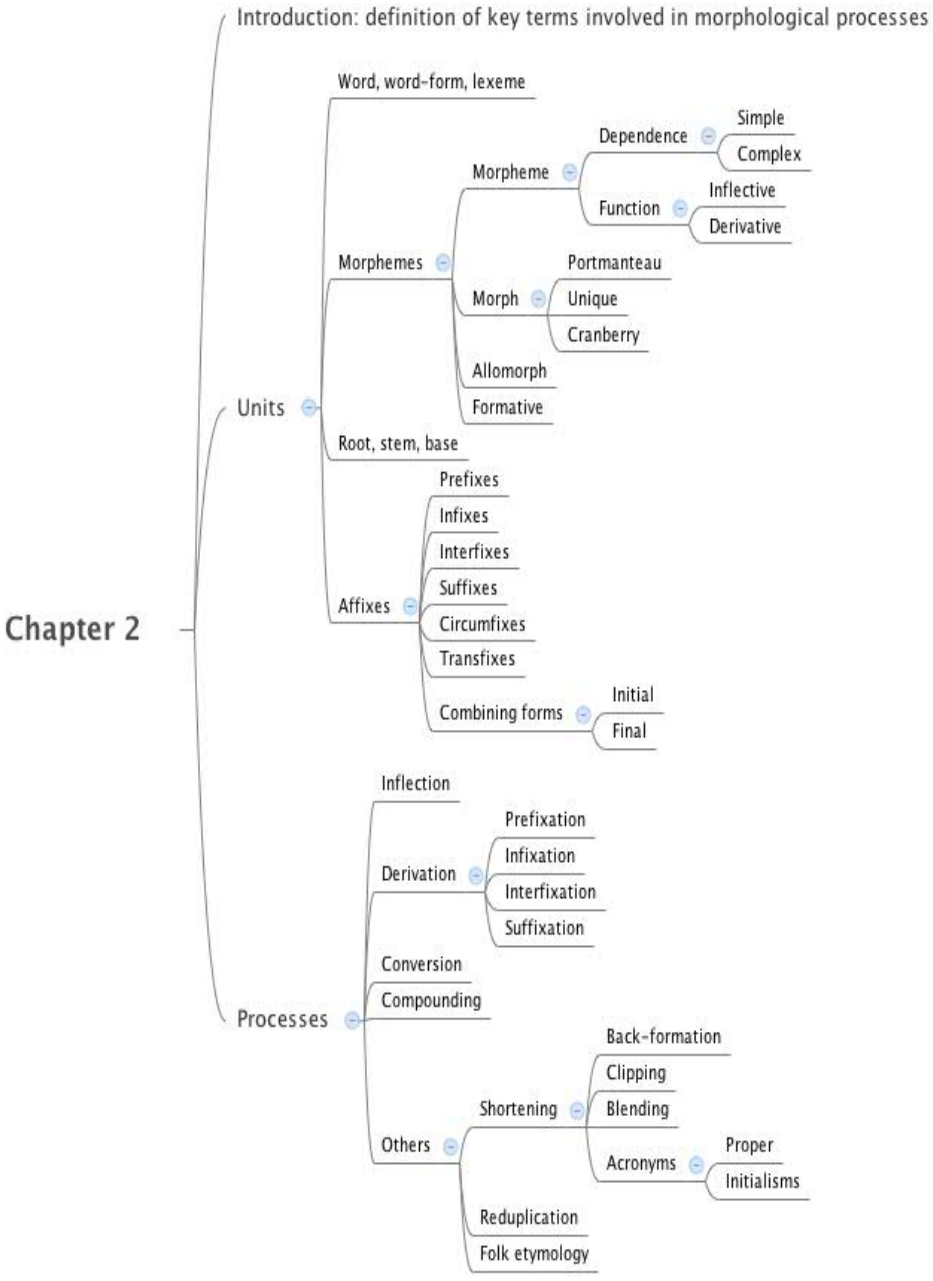


Figure 2.4: Structure of chapter 2.

3. Corpus and methodology

1. Introduction

This chapter provides some relevant information on the corpus used for my research on the scientific register of English. It explains the choice of the disciplines analysed, substantiates the use of the *OED* as a valid database for retrieving fairly accurate information on word formation and coinage dates, and justifies the focus on nouns as the word category used as a prototypical model for word formation in Late Modern English. Likewise it contains a detailed description of the methodology and the computational tools employed to analyse my data. My approach to scientific writing will be from the standpoint of computational morphology, which "deals with the processing of words in both their graphemic, i.e. written, and their phonemic, i.e. spoken form" (Trost 2003: 25), although I will only pay attention to their graphemic form, and corpus linguistics methodology will be used.

By the term 'corpus' I understand the definition that McEnery (2003: 449) applies exclusively to "a well-organized collection of data, collected within the boundaries of a sampling frame designed to allow the exploration of a certain linguistic feature (or set of features) via the data collected". McEnery

& Wilson (2001: 109) validate the use of corpus data by declaring that they "do have an important role to play in studying the frequencies of different morphological variants and the productivity of different morphemes." The criteria used for my study can be associated to quantitative analyses carried out by other scholars, like Caso[22] (1980) on semantic change, Culpeper & Clapham[23] (1996) on borrowing, Görlach[24] (2001) on the general acquisition of new words, and more predominantly, Camiña-Rioboo (2010, 2012)[25] on derivational morphology. Other linguists ratify quantitative studies with recourse to the *OED*, on which some of the previous analyses are based. For instance, Tieken-Boon van Ostade (2009: 57) considers the *OED* a "magnificent instrument for the purpose of studying the language of the lModE

[22] *Vid*. Introduction.

[23] Culpeper & Clapham (1996) investigated the effect of borrowing from Classical and Romance sources on the English lexicon. Their study, like Caso's on semantics and Camiña-Rioboo's on derivation, involved a series of computer searches, using in their case the electronic version of the *OED*. Their aim was to discover how many words and at what point in the history of the English language they were borrowed into English.

[24] Görlach studied the incorporation of new words into English in the twentieth century down to its early years at intervals of 20 years with recourse to dictionaries like the *Chronological English Dictionary*.

[25] The two studies by Camiña-Rioboo (2010, 2012) analyse morphological issues in philosophy and astronomy, respectively.

period, particularly since it is available as an online database which allows full-text searches as well as searches by date of occurrence", but she also acknowledges its limitations, by stating that "frequency of information should be interpreted with care due to the fact that different amounts of material are included for the eighteenth and nineteenth centuries. This makes comparisons hazardous". To this we might add that comparisons may be more misleading yet when dealing with older periods of the language, since sources tend to be scarcer. Finally, she warns about the potentially misleading feature 'Date chart', from which our current study has benefited greatly. However, I am conscious of this hazard and have considered the data obtained from date charts as primarily provisional. Future revisions of the *OED* and results obtained other research work will, no doubt, uncover earlier occurrences of many lexical items, and correct the misleading information in the dictionary.

As explained more comprehensively in chapter 1, I will examine astronomy and philosophy, two of the most important disciplines in history in the development of knowledge in general and science in particular. In the following sections computerised corpus analysis will be applied to my data in order to verify established theories that maintain that word formation in late Modern English was very limited, and consisted, by and large, of the combination of classical bases and affixes. In this respect, Bailey (2010: 186) states that,

> Many historians have alleged that the eighteenth century was a period with less innovation in vocabulary than the centuries before and after it. Such a conclusion is an artefact of statistical information drawn from the *Oxford English Dictionary*, though it was known from the very beginning of its publication that the eighteenth century was vastly underrepresented in the database upon which it was based. Revision of the dictionary now underway has added thousands of first uses in to the eighteenth century […] Many of these were words rarely used or restricted to technical fields."

Although I will review this point only marginally, I also intend to deal with it in my analysis, since preliminary research has already proved that some of the 'first occurrences' recorded in the *OED* date in fact from earlier periods. This can be seen in my samples, because some of the nouns included in them were only attested years, sometimes a century later. Besides, some nouns are not included as lexical entries in the dictionary at all.

Therefore, astronomy and philosophy are the disciplines selected as representatives of science to carry out this research work. This is not a random choice, and there are at least three reasons to encourage it. Firstly, a contrastive analysis of two clearly distinct disciplines —distinct from a contemporary standpoint, as already explained in detail in chapter 1 and as will be made clear in the UNESCO Classification below— will result in a wider linguistic scope. Consequently, more systematic findings will be available when all the data contained in the corpus have been examined in chapter 4. Secondly, both

astronomy and philosophy developed significantly in the time-span studied in this dissertation, so it is expected that those developments in the disciplines may have had a significant impact on language as well. And thirdly, the availability of the astronomy and philosophy texts compiled —only in their 2009 and 2010 beta versions, respectively— in *CETA* and *CEPhiT* at the time of writing provided a solid base on which this analysis of language could be firmly founded. We must bear in mind that it was the significance of astronomy and philosophy in the context of the scientific revolution and their divergent complementariness that made the members of the Research Group for Multidimensional Corpus-Based Studies in English (MuStE)[26] select these two disciplines as the first subsets to be included in the *Coruña Corpus of English Scientific Writing*. The availability of a representative, structured and balanced collection of samples in digital format, encoded to be used together with the information storage and retrieval tool designed specifically for this corpus, the *CCT*, made my analysis faster, more reliable and less prone to manual error. Manipulating such a large amount of data —over four hundred

[26] In this chapter I will make use of the pronouns "I" and "we" quite often. The first person singular refers especifically to my personal morphological analysis of these two subsets of the *Coruña Corpus*, whereas the plural will refer to compilation decisions made by MuStE, to which I have the honour to belong, as a whole.

eight thousand lexical items— without the aid of a computer and software tools would have rendered this analysis practically impossible.

Finally, I have chosen to study nouns based on the fact that nouns were regarded the key of languages in the seventeenth and eighteenth centuries. An example of which has been explained in chapter 1: Wilkins's attempt to create his analytical language focused only on providing new nouns to designate the elements of nature. Other parts of speech were, therefore, secondary to nouns. This situation remained unchanged in the eighteenth century, and Coote (1788: 9) provides an explanation to verbal derivation originating exclusively on substantive grounds, by stating that:

> From the invention of nouns or names, it was natural to proceed to the consideration of a word that might affirm something of the persons or things intimated by the substantive. It is, therefore, probable, that the introduction of what we now call a *verb* immediately succeeded the formation of the noun.

In sum, nouns have been chosen as the objective of this study because for the reason expounded above they make up the word category that increased the most since the Early Modern English period, either by borrowing (Nevalainen 2006a: 53-6), or affixing.

2. The corpus

The material has been extracted from the *Coruña Corpus: A Collection of Samples for the Historical Study of English Scientific Writing* (*Coruña Corpus*), a project currently under development by MuStE at Universidade da Coruña (Spain). The corpus has been designed to allow synchronic and diachronic studies in scientific discourse from most linguistic levels. It contributes to the study of the historical development of English for specific purposes, and provides linguistic material to study the scientific register and the style used in scientific texts at various points in the history of English. The compilation principles of the *Coruña Corpus* have already been explained at length by Crespo & Moskowich (2010), so they will not be dealt with comprehensively at this point, although it is worth noting that the *Coruña Corpus* has been designed to meet the standard guidelines on corpus linguistics. As a consequence, the selection of the texts for each discipline has followed the criteria put forward by Atkins *et al.* (1992) on discipline inventory, Leech (1992) on linguistic objectives, Biber *et al.* (1998) on the principles of representativeness, balance, and time-span, and Meyer (2002) and Crystal (2003) on the basic tenets for corpus compilation. The time span covered by these two subcorpora, as well as the *Coruña Corpus* itself, comprises the eighteenth and nineteenth centuries. When compiling the *Coruña Corpus* we have selected only one sample per author, in order to

increase idiolect variants (Lareo 2010), and two samples per decade, as a means of exploring diachronic evolution. In this period, authors used to publish works on several disciplines. We have decided to include only one work by every author, notwithstanding how productive he/she might have been, in one of the disciplines of the *Coruña Corpus*. This decision restricts the choice of authors greatly, and increases difficulty when trying to find sources for subsequent disciplines, but it helps preventing the proliferation of idiosyncrasies. Every sample contains roughly 10,000 words, excluding graphs, formulae, tables and figures. This editorial decision has been made in order to keep balance and facilitate comparative analyses. Also, quotations that do no represent the authors' speech have been left out, due to consistency. Moreover, it is our intention to avoid recurrent linguistic patterns; therefore, we have included different parts from texts in our samples, such as introductions, central chapters and conclusions. Short texts, however, have been included in full.

The samples selected for this research belong to the two first subcorpora compiled and encoded for that corpus, namely, the *Corpus of English Texts on Astronomy* (*CETA*) and the *Corpus of English Philosophy Texts* (*CEPhiT*). The *Coruña Corpus* has been compiled using the UNESCO Classification of Sciences as a starting point. By this classification astronomy falls into category I, containing natural sciences, that is to say, those making use of the scientific method to offer quantifiable, irrefutable data, and focusing

mainly on accuracy and objectivity; whereas philosophy can be included in Category VI, dealing with the human sciences, i.e. those not requiring mathematical treatment or experimental refutation[27], and which make use of the scientific method to study from an empirical point of view the human and divine aspects of the world. We must bear in mind that this classification of 'natural' and 'social' sciences can only be acknowledged from a contemporary perspective.

In spite of the wide time-span that the *Coruña Corpus* has to offer for linguistic research, for this work in particular I have only selected the samples corresponding to the eighteenth-century block of both *CETA* and *CEPhiT*. The reason behind this choice lies on the fact that our intention is to examine the scientific register of English in the late Modern Period, in order to assess the consolidation of the scientific discourse that emerged in the late seventeenth century (Banks 2005). I also intend to evaluate whether the degree of morphological innovation in the eighteenth century is really significant or, on the contrary, the vocabulary employed in these scientific texts is rather the consequence of innovations taking place in previous centuries. The use of these samples will, in theory, allow three sets of analyses, i.e. synchronic, diachronic and contrastive. Firstly, the inclusion of texts corresponding to two scientific

[27] Clark & Stafford (1982) propose a new division of knowledge in experimental and observational.

disciplines from one single century will offer a synchronic perspective of word-formation as a whole in the eighteenth century. Secondly, the recording of variables such as the origin, coinage dates of bases, affixes, and derived types will allow a double diachronic study: on the one hand, the processes to coin new lexical items and their productivity in the eighteenth century can be diachronically compared with those dating from previous periods in the history of English; on the other hand, we can carry out studies on morphological variability within the eighteenth century by setting up periods of thirty years. This can be compared with other works of similar characteristics by Kÿto *et al* (2000), which claim that "short-term change in diachrony can be safely studied over periods of thirty years" (Crespo & Moskowich 2009: 3). Thirdly, the contribution of astronomers and philosophers to lexical innovation will be scrutinised on a contrastive level to measure which of the two disciplines may have contributed to the expansion of scientific vocabulary to a greater extent.

All the texts encompassing the samples were written by different authors, as stated above, brought up and educated in different English-speaking countries[28], publishing original texts in English on astronomy and philosophy

[28] Samples belonging to the eighteenth century were written only by authors from the British Isles. We will need to wait until the nineteenth century to find North American authors in our corpus, being John Ewing's lecture on astronomy (1809) the first non-European sample in the *Coruña Corpus* so far.

between 1700 and 1800. These texts are not translations, although we must not forget the fact that some authors in this period also "read and wrote in Latin, and this may have exerted an influence on their use of the vernacular" (Crespo & Moskowich 2009: 3). To this we may add that some knowledge in Greek can also be attributed to them, which is also interesting from a morphological point of view, in case they made use of that knowledge to employ Greek affixes to coin new lexical elements in English. *CETA* includes twenty-one texts, whereas *CEPhiT* contains twenty. This slight difference in in number corresponds to the presence of two smaller astronomy samples in the 1770s, because we intended to include a journal article in the decade of the 1770s. So we were compelled to include a third short sample to make up the total 20,000-word count for the decade. The total word count[29] for our corpus is 408,981, from which 208,079 correspond to *CETA*, and 200,902 correspond to *CEPhiT*, as represented in Figure 3.1 below:

[29] The analyses of the subcorpora for this research were undertaken in 2008 and 2009 with beta versions of *CETA* and *CEPhiT* respectively, so the final count may vary slightly from that of the final versions when both corpora are published in the near future.

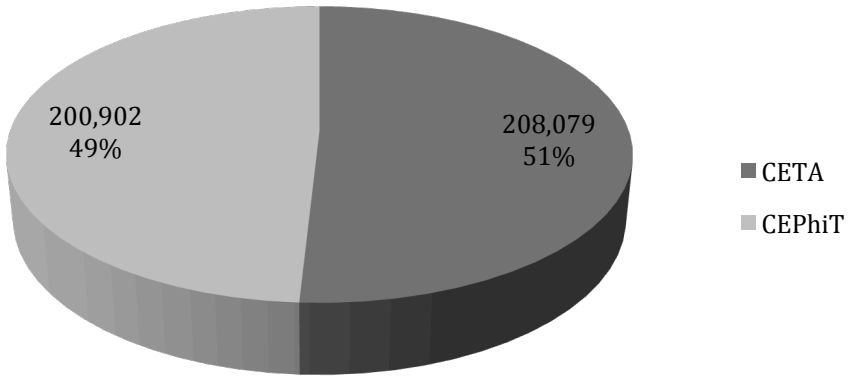


Figure 3.1: Total word count in *CETA* and *CEPhiT*.

CETA contains a slightly higher number of words than *CEPhiT*, but both disciplines make up roughly fifty per cent of the corpus each. Since my focus is on nouns, I have extracted them from the whole mass of words including the rest of the word categories, and I have computed 79,316 noun tokens. If we subtract the total number of noun tokens from the rest of the words in the corpus, we can establish that the percentage of nouns will be around 19%, which can be viewed graphically in Figure 3.2:

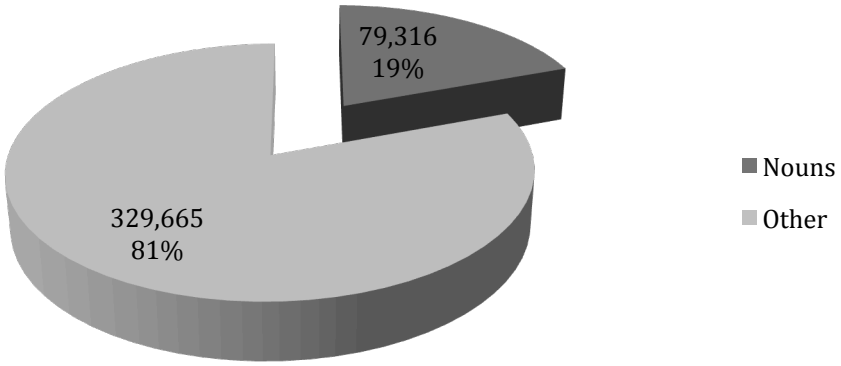


Figure 3.2: Number and percentage of nouns vs. other parts of speech in the corpus.

From these total 79,271 noun tokens present in the corpus, 38,117 belong to *CETA* and 41,154 belong to *CEPhiT*. In contrast with the overall word count in which *CETA* slightly outnumbers *CEPhiT*, philosophy texts contribute to my data with four per cent more nouns than astronomy. The following figure shows the distribution of the noun tokens in the two subcorpora analysed.

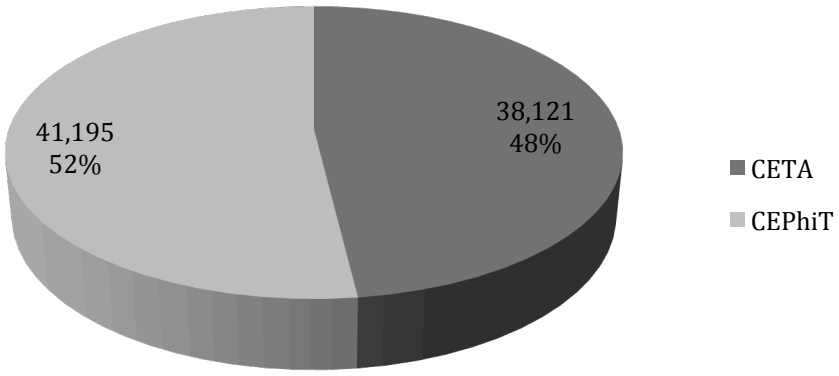


Figure 3.3: Distribution of noun tokens in *CETA* and *CEPhiT*.

As for the distribution of those 79,271 noun tokens, they belong to 4,530 noun types. Therefore, if we apply the following formula: $\frac{Tokens}{Types}$, the average number of tokens that each type produces in the samples can be obtained — also called *type-token ratio*. Thus, if:

$$\frac{tokens}{types} = \frac{79{,}210}{4{,}530} = 17.49$$

We can draw a broad preliminary conclusion: on average, each type is repeated almost eighteen times in the corpus. As we shall see later, when individual attention is paid to every type and both disciplines are compared, it will be made clear that some types produce many more tokens than others and, apart

from special cases such as nonce-formations, there are numerous cases known as *hapax legomena*, or "the occurrence of words having frequency 1" (Popescu *et al*. 2009: 227), which pose a very interesting line of research. In spite of the fact that these reasons can render the previous formula totally inadequate for our research, I have decided to include it now to compare the type-token ratios in the whole corpus with the results for each of the disciplines, in the hope that they will result in completely different figures. Also, when applied to individual samples this formula may provide interesting information on the style characteristic to authors and genres, as will be shown in tables for both subcorpora below.

2.1. *The Corpus of English Texts on Astronomy*

The following table presents the authors in chronological order, the dates of publication of the sampled works, their full title and the number of words contained in the samples:

Table 3.1: Authors included in *CETA*.

Author	*Year*	*Title*	*No. of Words*
Henry Curson	1702	*The theory of sciences illustrated, or the grounds and principles of the seven arts; grammar, logick, rhetorick, musick, arithmetick, geometry, astronomy. Accurately demonstrated and reduced to practice.*	10,247
Robert Morden	1702	*An Introduction to Astronomy, geography, navigation, and other mathematical sciences made easie by the description and uses of the coelestial and terrestrial Globes. In seven parts.*	10,154
William Whiston	1715	*Astronomical lectures, read in the publick schools at Cambridge.*	9,939
John Harris	1719	*Astronomical dialogues between a gentleman and a lady: wherein the doctrine of the sphere, uses of the globes, and the elements of Astronomy.*	9,907
George Gordon	1726	*An introduction to geography, astronomy, and dialling. Containing the most useful elements of the said sciences, adapted to the meanest capacity, by the description and uses of the terrestrial and celestial globes with an introduction to chronology.*	10,437
Isaac Watts	1726	*The knowledge of the heavens and the earth made easy: or, the first principles of astronomy and geography explain'd by the use of globes and maps.*	10,407
Samuel Fuller	1732	*Practical astronomy, in the description and use of both globes, orrery and telescopes wherein the most useful elements, and most valuable modern discoveries of the true astronomy are exhibited, after a very easy and expeditious manner, in an exact account of our solar system.*	10,232
Jasper Charlton	1735	*The ladies astronomy and chronology in four parts.*	10,358
Roger Long	1742	*Astronomy, in five books.*	10,474
James Hodgson	1749	*The theory of Jupiter's satellites: with the construction and use of the tables for computing their eclipses.*	11,106

John Hill	1754	*Urania: or, a compleat view of the heavens; containing the antient and modern astronomy, in form of a dictionary: illustrated with a great number of figures (Vol.I. Being the first of a compleat system of natural and philosophical knowledge).*	10,044
James Ferguson	1756	*Astronomy explained upon Sir Isaac Newton's principles and made easy to those who have not studied mathematics.*	10,519
Mathew Stewart	1761	*Tracts, physical, and mathematical. Containing an explication of several important points in physical astronomy; and a new method for ascer- taining the sun's distance from the earth, by the theory of gravity.*	12,180
George Costard	1767	*The history of astronomy, with its application to geography, history, and chronology; occasionally exemplified by the globes.*	10,315
Alexander Wilson	1773	*Observation on the solar spots.*	4,240
George Adams	1777	*A treatise describing the construction and explaining the use of celestial and terrestrial globes.*	10,566
John Lacy	1779	*The universal system: or mechanical cause of all the appearances and movements of the visible heavens: shewing the true powers which move the earth and planets in their central and annual rotations with a dissertation on comets, the Nature, cause, matter, and use of their tails, and the reasons of their long trajectories; likewise and attempt to prove what it is that moves the sun around its axis.*	5,908
William Nicholson	1782	*An introduction to natural philosophy.*	10,268
John Bonnycastle	1786	*An introduction to astronomy in a series of letters from a preceptor to his pupil.*	9,975
Samuel Vince	1790	*A treatise on practical astronomy.*	10,540
Margaret Bryan	1797	*A compendious system of astronomy.*	10,263

As can be observed in the table above, most samples in *CETA* contain approximately 10,000 words. It can be appreciated that James Hodgson's

(1749) and Matthew Stewart's samples (1761), however, go well over that limit, but these are special cases containing many numbers, variables and formulae embedded within sentences, which cannot be deleted without affecting the understanding of the text itself. Therefore we have decided to keep them, but stretching the final number of words, until a suitable number of appropriate material analysable under linguistic perspectives could reach the boundaries of 10,000 words. Likewise, as has already been explained above, there are three samples for the decade of 1770s. However, the final count for the decade is similar to other decades, that is to say, 20,000 words. From the 208,079 words contained in the samples of *CETA*, I have computed 38,117 nouns, which represent approximately 18.3 per cent of the total count, as seen in figure 3.4.

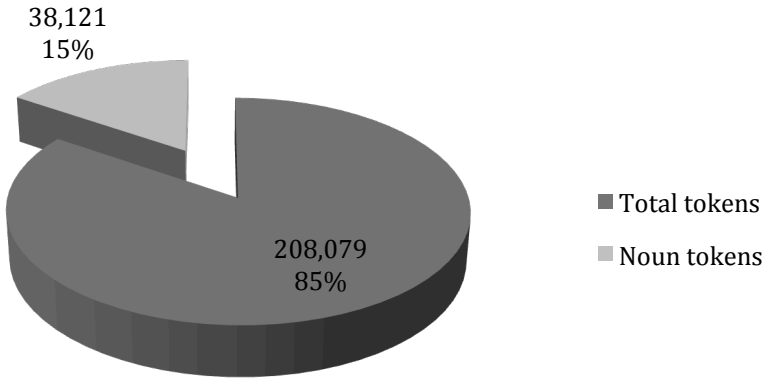


Figure 3.4. Words and noun tokens in *CETA*.

Table 3.2 below contains additional information about the number of noun tokens contained in every sample of *CETA*. Starting on the left end of the table, column 1 shows the name of the author and the year of publication; column 2 includes the total number of words for every sample; column 3 indicates the raw number for noun tokens; column 4 extracts the percentage that noun tokens represent in the total word count; column 5 lists the noun types present in the sample; and column 6 gives the results for the type-token ratio.

Table 3.2: Words, tokens and types in individual samples of *CETA*.

Sample	Total words	Tokens	%	Types	Type-token ratio
Curson (1702)	10,247	2,131	20.8	495	4.3
Morden (1702)	10,154	1,901	18.7	540	3.5
Whiston (1715)	9,939	1,689	17	365	4.6
Harris (1719)	9,907	1,602	16.2	454	3.5
Gordon (1726)	10,437	1,781	17.1	295	6
Watts (1726)	10,407	2,170	20.9	295	7.4
Fuller (1732)	10,232	2,109	20.6	350	6
Charlton (1735)	10,358	1,784	17.2	269	6.6
Long (1742)	10,474	1,998	19.1	293	6.8
Hodgson (1749)	11,106	2,681	24.1	255	10.5
Hill (1754)	10,044	1,843	18.3	401	4.6
Ferguson (1756)	10,519	1,673	15.9	299	5.6
Stewart (1761)	12,180	1,693	13.9	62	27.3
Costard (1767)	10,315	1,876	18.2	400	4.7
Wilson (1773)	4,240	707	16.7	246	2.9
Adams (1777)	10,566	1,835	17.4	250	7.3

Lacy (1779)	5,908	1,080	18.3	240	4.5
Nicholson (1782)	10,268	1,805	17.6	379	4.8
Bonnycastle (1786)	9,975	1,778	17.8	536	3.3
Vince (1790)	10,540	2,057	19.5	220	9.4
Bryan (1797)	10,263	1,927	18.8	460	4.2
TOTAL	**208,079**	**38,121**	**18.3**	**2060**[30]	**6.6**

This chapter aims at describing my corpus, and I do not intend to draw any conclusions yet, but a simple glance at the figures in column 6 (type-token ratio) of Table 3.2 shows that some authors offer a higher variety of nouns. Wilson (1773), for instance, contributes an average of just under three noun tokens used per single type, whereas others present a much higher recurrence, as is the case of Stewart (1761), with an average of over twenty-seven tokens per type. This considerable difference between both authors may have several explanations, among which we can include their different degree of linguistic competence and the genre/text-type employed —the text written by Wilson falls into the category of journal articles, whereas the one published by Stewart is a textbook, and therefore he may have needed to be more didactic or recursive in his explanations. In chapter 4 I will compare the figures of other

[30] The total number of types does not result from adding all the figures in column 5, since authors coincide in the use of some of those types and, therefore, I need to resort to a different figure from the list of different noun types in my database.

textbooks to check whether this is a general characteristic of the genre or, on the contrary, it is only an idiosyncrasy of this author in particular.

The principles underlying the *Coruña Corpus* have been devised to supply a wide range of genres/text-types[31] to compare different linguistic features intrinsic to each of them. Genres/text-types may vary from discipline to discipline, and in astronomy and philosophy we have contemplated eight genres: Essay, Treatise, Textbook, Lecture, Letter, Dialogue, Article and 'Others'. We have used this last label used when texts present miscellaneous features that make them ineligible for any of the previous genres. In the eighteenth century section of *CETA* we have managed to incorporate a wide variety of samples from different genres, divided as follows: 1 lecture, 1 dialogue, 5 treatises, 10 textbooks, 1 essay, 1 journal article and 1 letter, and we have included a dictionary (Hill 1754) under the category 'Others'. The next figure shows the distribution of genres in *CETA*, and the number of words contained in the samples making up those genres.

[31] At the time of writing there was no agreement among authors about the label to define genres and text-types. For example, when Görlach (2004: 88) defines the genres used in the *Coruña Corpus* he considers them text-types. Conversely, Paltridge (1997) treats them as genres. Other authors give various reasons to distinguish them but concede that these terms are often used interchangeably (Adolphs 2008: 80). It is not my intention to add any more arguments to this controversy, so I will make use of both terms synonymously across this work.

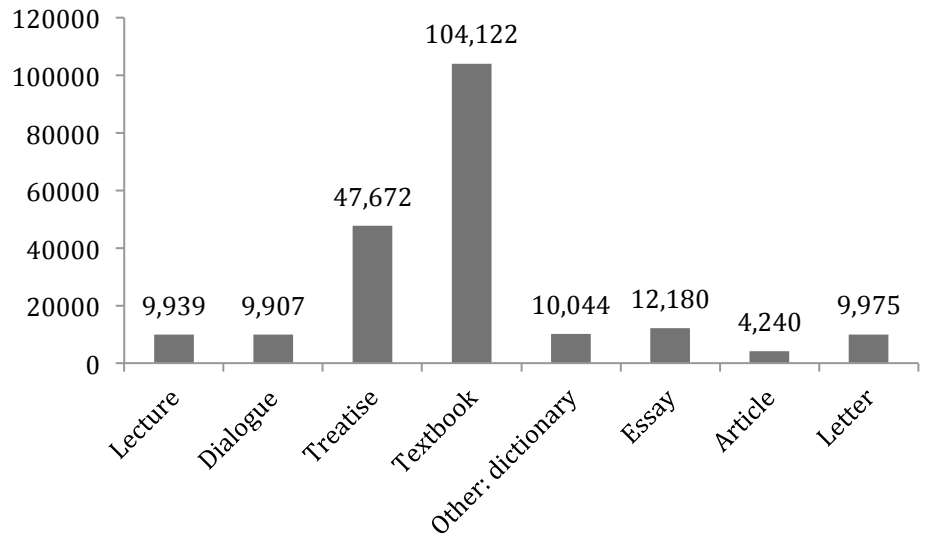


Figure 3.5. Tokens per genre/text-type in *CETA*.

Our *Coruña Corpus Tool* includes a metadata section that provides additional information about the authors and the texts selected for the *Coruña Corpus*. For example, we have contemplated the possibility of analysing the style of authors based on their place of education, in order to establish parallel word choice and word-formation processes among them. Therefore we have discovered that twelve astronomers were educated in various schools and universities in England, three in Scotland, and one in Ireland. Unfortunately, we are not sure of the place of education of five of our authors. If we suppose that also some of them were educated in England, this would raise the number to around 15-17 out of 21 astronomers educated in this country. It may sound a bit adventurous to suggest that the existence of the figure of the Astronomer

Royal[32], and the Royal Greenwich Observatory, had a notable significance in making of England the most appealing place for students on the British Isles to be educated as astronomers, but it is a very particular coincidence that most astronomy texts found in the 18th century have been written by authors educated in England. Figure 3.6 shows the geographical distribution of *CETA*.

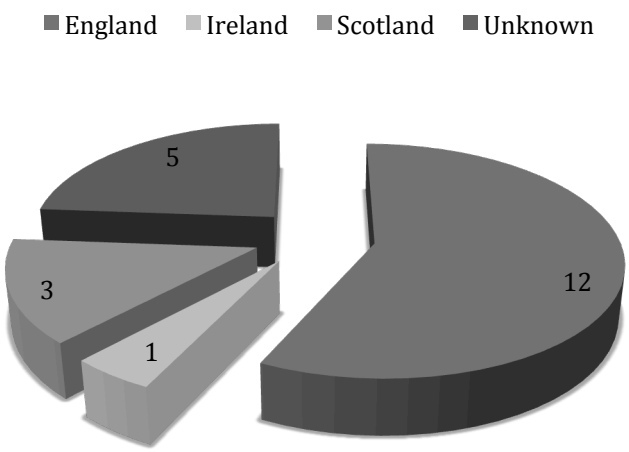


Figure 3.6: Places of education of authors in *CETA*.

For the compilation of the *Coruña Corpus* we have done our best to include texts written by women scientists where possible, because we consider that the variable 'sex of the author' poses another interesting line of study when

[32] John Flamsteed was appointed by king Charles II the first Astronomer Royal in 1675. One year later he became a fellow of the Royal Society.

comparing language use, let alone word-formation processes. But here it appears that we have committed ourselves to very challenging task. In the eighteenth century women had no access to academia[33] and, consequently, they would need to rely almost exclusively on autodidacticism. We presume that women wrote texts that were later published by their male relatives, brothers or husbands, and it is very likely that male astronomers have benefited from valuable data collected and tested by women that carried out their own experiments and established their own theories. As Herrero-López (2007: 75) puts it,

> […] existieron mujeres que lograron sobresalir, aunque muchas no obtuvieron el reconocimiento por parte de la sociedad en general pues sus logros se atribuían a los padres o esposos. Los problemas de identificación de autor se han complicado por la pérdida del apellido de algunas mujeres al casarse, o por la utilización de un pseudónimo masculino que garantizase que el trabajo fuese tomado en serio.

Most unfortunately, it was frown upon for women to be out in the fields watching the stars at night for eighteenth-century society. It may be interesting to mention that even the Royal Society, concerned as it has been throughout its

[33] Generally, universities in Europe did not accept female students, being Bologne in Italy one of the exceptions (Herrero-López 2007: 79).

history about the diffusion of science, did not consider the admission of women scientists until after World War II (Mason 1972). Therefore we have only managed to include one single sample from a woman astronomer, Margaret Bryan, who published her work towards the end of the century, as opposed to twenty samples written by men.

Finally, in order to provide as many variables as possible in our work, we have contemplated the age of the authors when their works were published, because it can also be viewed as a means to research on style and word formation. We are not completely sure of the age of six of the astronomers in our samples, so they will be left out of the computation. Besides, the sample written by Hodgson (1749) is in clear asynchrony with the rest, because he was 77 years of age when he published this work, so instead of calculating the average age of the authors in *CETA*, I have drawn the median, which results in 50 years. The distribution of age of the authors can be graphically observed in the following figure.

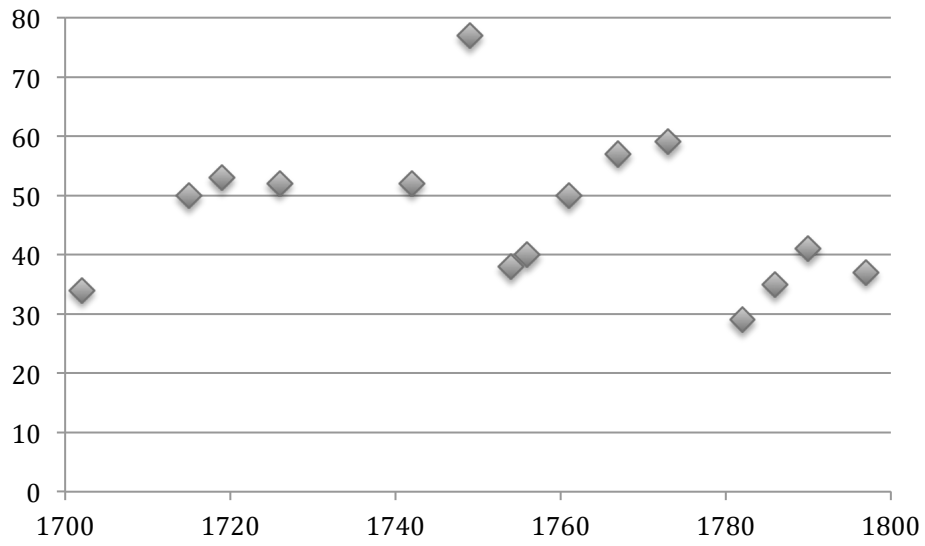


Figure 3.7: Age of the author when works were published in *CETA*.

2.2. *The Corpus of English Philosophy Texts*

The second subcorpus included in my analysis covers the human and divine aspects of philosophy. The importance of this discipline in the abstract concept of science and knowledge has already been explored in chapter 1. So I will focus now on our more concrete philosophical samples, in which I expect to find topics connected with religion, ethics, rhetoric, together with more mundane subjects such as marriage, feminism and politics. The following table shows the texts belonging to *CEPhiT*, their authors, dates of publication, full title of the works and number of words included in the samples:

Table 3.3: Authors included in *CEPhiT*.

Author	*Year*	*Title*	*No. of words*
Mary Astell	1700	*Some reflections upon marriage.*	10,076
George Cheyne	1705	*Philosophical principles of natural religion: containing the elements of natural philosophy, and the proofs for natural religion, arising from them.*	10,060
John Dunton	1710	*Athenianism: or, the new projects of Mr. John Dunton.*	10,059
Anthony Collins	1717	*A philosophical inquiry concerning human liberty*	10,071
Robert Greene	1727	*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.	10,122
Robert Kirkpatrick	1730	*The golden rule of divine philosophy*	10,045
Robert Balguy	1733	*The law of truth: or, the obligations of reason essential to all religion.*	10,040
Joseph Butler	1736	*The analogy of religion, natural and revealed, to the constitution and course of nature.*	10,049
George Turnbull	1740	*The Principles of Moral Philosophy. An Enquiry into the Wise and Good government of the moral world.*	10,030
David Hume	1748	*Philosophical essays concerning human understanding.*	10,019
Henry Bolingbroke	1754	*The philosophical works of the late right honorable Henry St. John, lord viscount Bolingbroke.*	10,023
Francis Hutcheson	1755	*A system of moral philosophy. Vol II. Book III.*	10,031
Thomas Reid	1764	*An inquiry into the human mind, on the principles of common sense.*	10,032
Adam Ferguson	1769	*Institutes of moral philosophy*	10,064
Edmund Burke	1770	*Thoughts on the cause of the present discontent*	10,017
George Campbell	1776	*The philosophy of rhetoric*	10,007
Catharine Macaulay	1783	*A treatise on the immutability of moral truth*	10,059
William Smellie	1790	*The philosophy of natural history*	9,993

Mary Wollstonecraft	1792	*Vindication of the rights of woman: with structures on political and moral subjects.*	10,058
Alexander Crombie	1793	*An essay on philosophical necessity*	10,047

Table 3.4 below mirrors the information provided for astronomy in table 3.2, but applied to *CEPhiT* on this occasion. Again, column 1 shows the name of the author and the year of publication; column 2 includes the total number of words for every sample; column 3 indicates the raw number for noun tokens; column 4 extracts the percentage that noun tokens represent in the total word count; column 5 lists the noun types present in the sample; and column 6 gives the results for the type-token ratio.

Table 4. Words, tokens and types in individual samples of *CEPhiT*.

Sample	Total words	Tokens	%	Types	Type-token ratio
Astell (1700)	10,076	1,724	17.1	609	2.8
Cheyne (1705)	10,060	1,888	18.8	447	4.2
Dunton (1710)	10,059	1,851	18.4	636	2.9
Collins (1717)	10,071	1,927	19.1	419	4.6
Greene (1727)	10,122	1,875	18.5	533	3.5
Kirkpatrick (1730)	10,045	1,627	16.2	594	2.7
Balguy (1733)	10,040	1,902	18.9	488	3.9
Butler (1736)	10,049	1,888	18.8	424	4.5
Turnbull (1740)	10,030	2,134	21.3	398	5.4
Hume (1748)	10,019	2,191	21.9	681	3.2
Bolingbroke (1754)	10,023	1,830	18.3	521	3.5
Hutcheson (1755)	10,031	2,283	22.8	644	3.5
Reid (1764)	10,032	2,000	19.9	635	3.1
Ferguson (1769)	10,064	2,704	26.9	641	4.2

Burke (1770)	10,017	2,157	21.5	739	2.9
Campbell (1776)	10,007	2,089	20.9	783	2.7
Macaulay (1783)	10,059	2,319	23.1	751	3.1
Smellie (1790)	9,993	2,486	24.9	646	3.8
Wollstonecraft (1792)	10,058	2,239	22.3	839	2.7
Crombie (1793)	10,047	2,081	20.7	464	4.5
TOTAL	**200,902**	**41,195**	**20.5**	**3777**[34]	**3.6**

We have managed to obtain a fairly regular number of words for every sample in *CEPhiT*, and only our utmost respect for sentence length has brought numbers over the customary 10,000-word limit, but very close to it, which makes comparison easier. Nevertheless, when I offer occurrence frequencies in the following chapter devoted to the analysis, I will use normalised figures. *CEPhiT* does not offer such a varied range of genres/text-types as *CETA*. Thirteen treatises, six essays and a textbook make up the eighteenth-century samples of the discipline. This can be considered a characteristic of the discipline, at least in the eighteenth century. Figure 3.8 shows the distribution of genres in *CEPhiT*, and the number of words contained in the samples making up those genres.

[34] *Vid*. note 7.

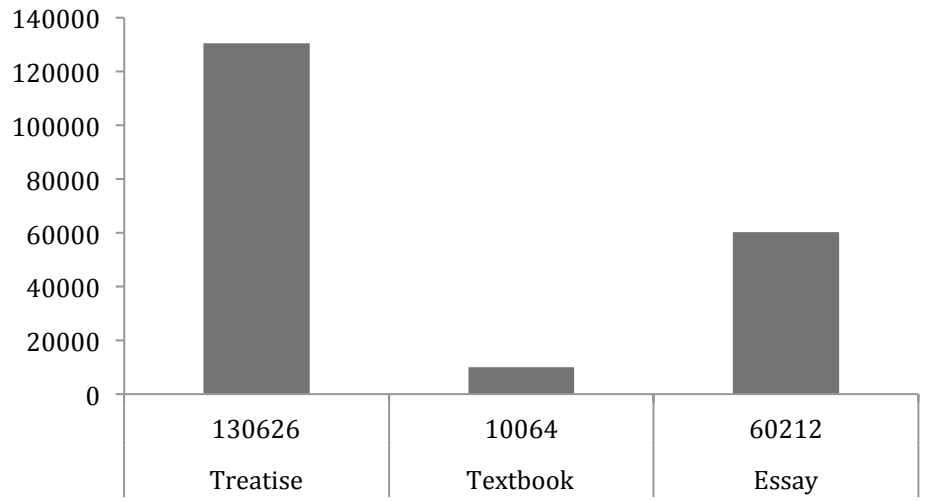


Figure 3.8: Genres/Text-types in *CEPhiT*.

Regarding the variable 'place of education' in *CEPhiT*, the situation is a bit different from the one shown for *CETA*. Practically half the philosophers in our corpus were educated in Scotland, if we count in Hutcheson, who was born in Northern Ireland, and studied both in Ulster and Scotland. The axis formed by Edinburgh and Glasgow universities contributed to vertebrate a period characterised by an outburst of intellectual accomplishments in Scotland. Hutcheson himself, Hume, Ferguson, Campbell, Reid, and other outstanding Scottish philosophers included in my corpus can be considered among the most influential thinkers in Europe and, due to the effect of the Scottish diaspora to America, they were also of great inspiration for social developments across the Atlantic (Herman 2003). Figure 3.9 shows the geographical distribution of *CEPhiT*.

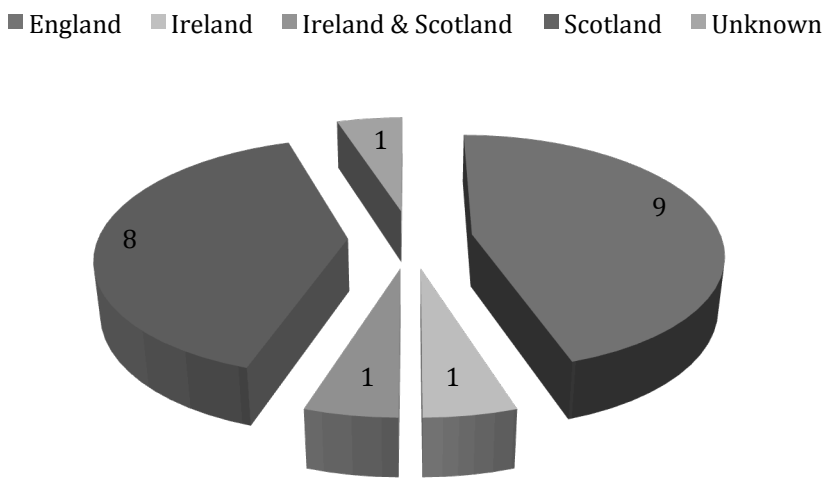


Figure 3.9: Places of education of authors in *CEPhiT*.

Women philosophers are more numerous than women astronomers in the eighteenth-century section of our corpus, mainly for the reasons explained above. The three outstanding women that have been included in *CEPhiT*, Mary Astell, Catherine Macaulay and Mary Wollstonecraft, exerted their influence as pioneers of feminism, and their works transcended frontiers, especially Macaulay's, whose ideas on equality were highly appreciated in North America. However, women still represent only 15 per cent of the total number of authors in the century. My samples are contained in works written by seventeen men compared to three written by women. Still, the use of normalised figures will allow comparative studies between women's and men's linguistic competence and style. Research already carried out for a pilot study

on this issue (Camiña-Rioboo, forthcoming) shows unpredicted results, considering the difficulties that women had to acquire an education. More information on this issue will be shown when the variable 'sex of the author' is developed in chapter 4.

The variable 'age of the author' has also been reviewed in *CEPhiT*. In this case we know all the dates of birth except for the 1730 text written by Kirkpatrick, whose sample has consequently been left out of our analysis for this variable in particular. As in the case of astronomy, one of which authors conditions the average result, also in philosophy Bolingbroke deviates sensibly from the age of the rest of philosophers —he was 77 years old when he published his work. Therefore we have again drawn the median, which results in 46 years. The distribution of age of the authors can be graphically observed in the following figure.

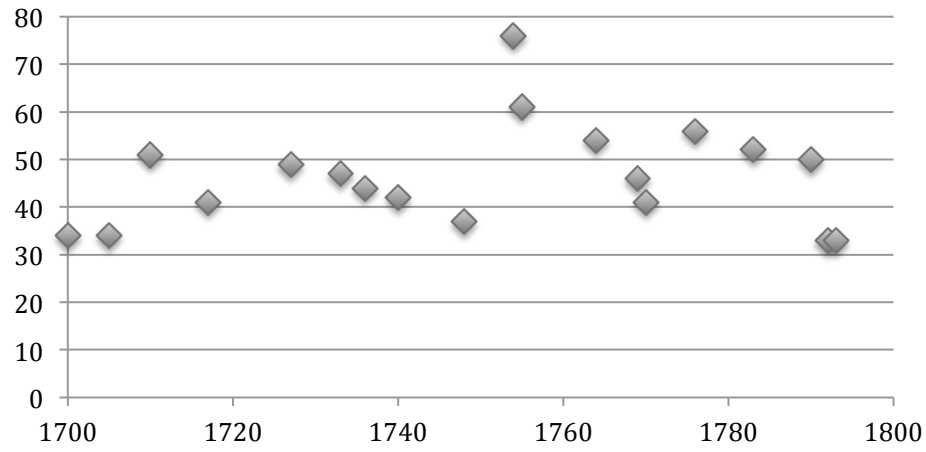


Figure 3.10: Age of the author when works were published in *CEPhiT*.

3. Methodology

The starting point of my study is the corpus formed by *CETA* and *CEPhiT* in their eighteenth-century versions and, as I stated above, I aim at analysing affixed nouns in scientific texts between 1700 and 1800. The process that I have proposed to carry out this research can be summarised as follows. The forty-one samples will first be segmented —tokenised—, and then several filters —disambiguation, degree of complexity of words, derivational processes and so on— will be applied to restrict the scope of our study further. Therefore, noun types will be extracted from the rest of data, and their number of tokens recorded. The noun types obtained will be divided into simple and complex, and the simple ones discarded. While these can still be used for further investigation on vocabulary choice, for example, and in contrasting linguistic performance by author, genre or discipline, only those nouns that have undergone word-formation processes are considered relevant to this study. The various noun-forming processes, i.e. affixation, compounding, conversion and other minor processes, will then be identified, and nouns will be classified according to each process. Finally the empirical data obtained and classified will be asserted as a way of measuring the evolution of these processes in the history of the English language, and in order to quantify their degree of productivity in lModE. Compounding, conversion and other processes will only be revised briefly in my work, whereas affixation and all

the elements involved in the process will be examined in depth. This whole procedure is laid out in figure 3.11.

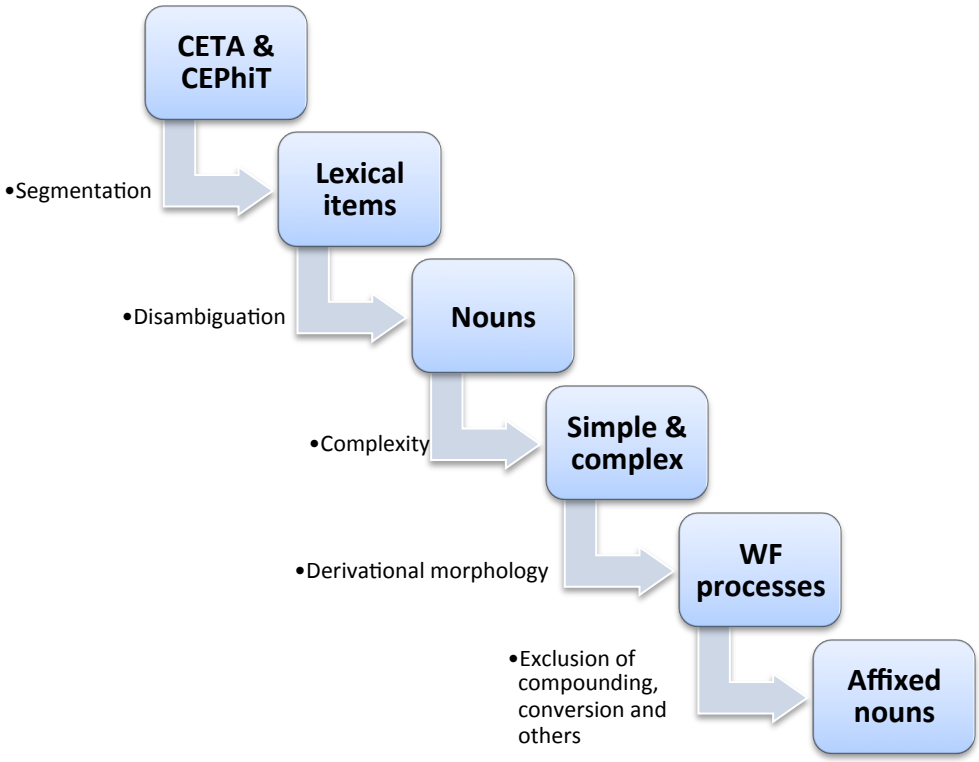


Figure 3.11: Initial process for analysis.

Any corpus-based analysis requires some previous steps to prepare the linguistic data for further processing, the first steps being the digitalisation of the original texts and their subsequent encoding to meet the needs required by computerised tools, so that these can recognise and process the data. In the case of the *Coruña Corpus* we have resorted to the eXtended Mark-Up Language (XML), given that it is a free cross-platform language that can be read,

exported and processed by most modern computer software, and more restrictively to the Text Encoding Initiative (TEI) subset to encode our corpus. Unfortunately, at the time of writing this piece of research the *Coruña Corpus* is not yet annotated to fulfil specific morphological studies, but it can nevertheless be a valuable aiding instrument in morphological analysis by tokenising the texts, that is to say, segmenting word tokens and counting them. As Mikheev (2003: 202) claims "tokenization and sentence splitting can be described as 'low-level' text segmentation which is performed at the initial stages of text processing." The following figure shows an example of tokenisation and generation of a frequency list using the *CCT*.

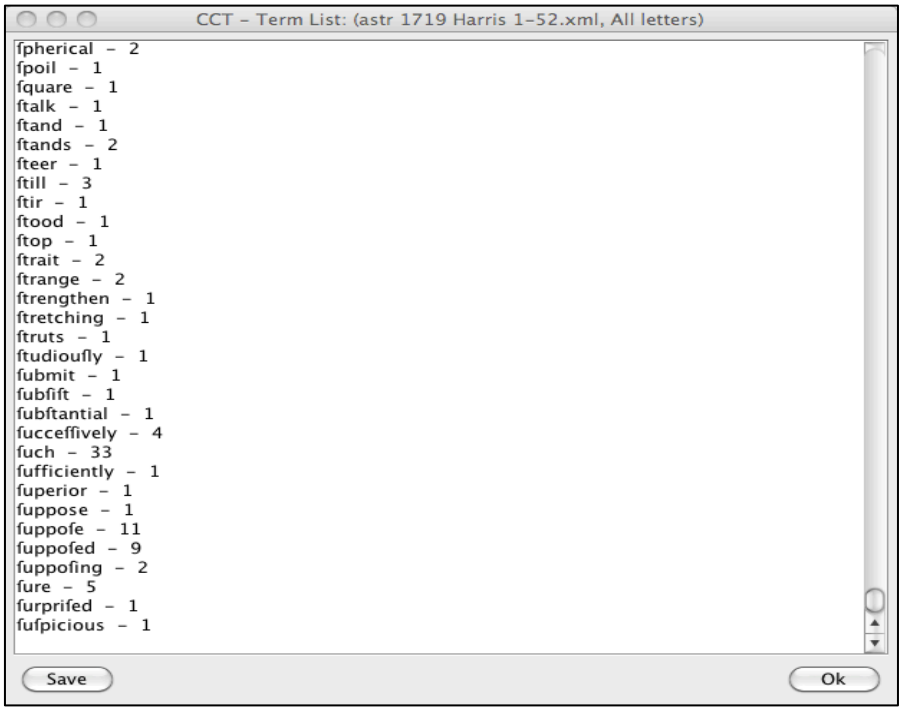


Figure 3.12: Tokenisation and frequency list generated with the *CCT*.

This process was performed individually on every sample for the two disciplines of the corpus, in order to obtain the full list of word tokens. Later, the word tokens were grouped into word types by combining all the individual word lists in a single matrix on a spreadsheet program. Only nouns were selected, and they were lemmatised. Since I am dealing with nouns, only singular and plural tokens needed unifying. Also, words with graphic variants were unified under a single noun type, as is the case of *dialling* (Morden 1702: 12), which also includes *dyaling* (Curson 1702: 365) and *dyalling* (Curson 1702: 364).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	TYPES	TOKENS	TokAstr	TokPhil	a02a	a02b	p00	p05	a15	a19	p10	p17	a26a	a26b
2682	nail	1	0	1										
2683	name	249	176	73	1	10	6	1		5	6		2	4
2684	narration	15	0	15										
2685	narrative	1	0	1										
2686	nation	60	12	48			1			1				1
2687	native	2	1	1	1									
2688	nativity	1	1	0	1									
2689	naturalist	3	1	2										
2690	nature	692	82	610	7	3	10	16	3	8	10	11	1	
2691	navigation	8	6	2					1					1
2692	navigator	10	9	1	1	1				2				
2693	navy	1	1	0										
2694	nearness	7	7	0										
2695	nebula	1	1	0									1	
2696	nebulosa	4	4	0		4								
2697	necessarian	3	0	3										
2698	necessity	72	4	68			2	8	2		4	22		1
2699	neck	7	5	2	2	2					1			
2700	need	10	3	7	1		2		2		1	1		
2701	needle	9	8	1	1	1				2			1	
2702	negation	1	0	1										
2703	negative	1	0	1										
2704	neglect	9	3	6			1			2		1		1
2705	negligence	1	0	1										
2706	negro	1	0	1										
2707	neighbour	7	2	5			3							

Figure 3.13: Type grouping after tokenisation.

As mentioned above, the *CCT* does not at present distinguish among parts of speech. At this point, consequently, I needed to combine computerised tools with manual disambiguation in order to extract nouns from the unwanted mass. This may seem a tedious task at first, but it was not so extreme. In fairness, numbers, formulae, grammatical words, inflected verbs, most adverbs and adjectives can be discarded without hesitation from the frequency lists generated above. There have been cases, however, for which careful individual supervision was required, as for example many instances of *-ing* word-forms, absolute uses of adjectives that might have been converted into nouns with use —lexicalised—, and conversion in general, since the controverted zero morph that I explained in chapter 2 cannot be observed in writing. For this task the *Coruña Corpus Tool* provided an invaluable help once more, since it works as a concordance program that performs individual searches, which can show all the instances of the word(s) in question in one window, together with left and right background —the so-called Key Word In Context (KWIC)—, so disambiguation can be performed relatively easily, as the next figure shows:

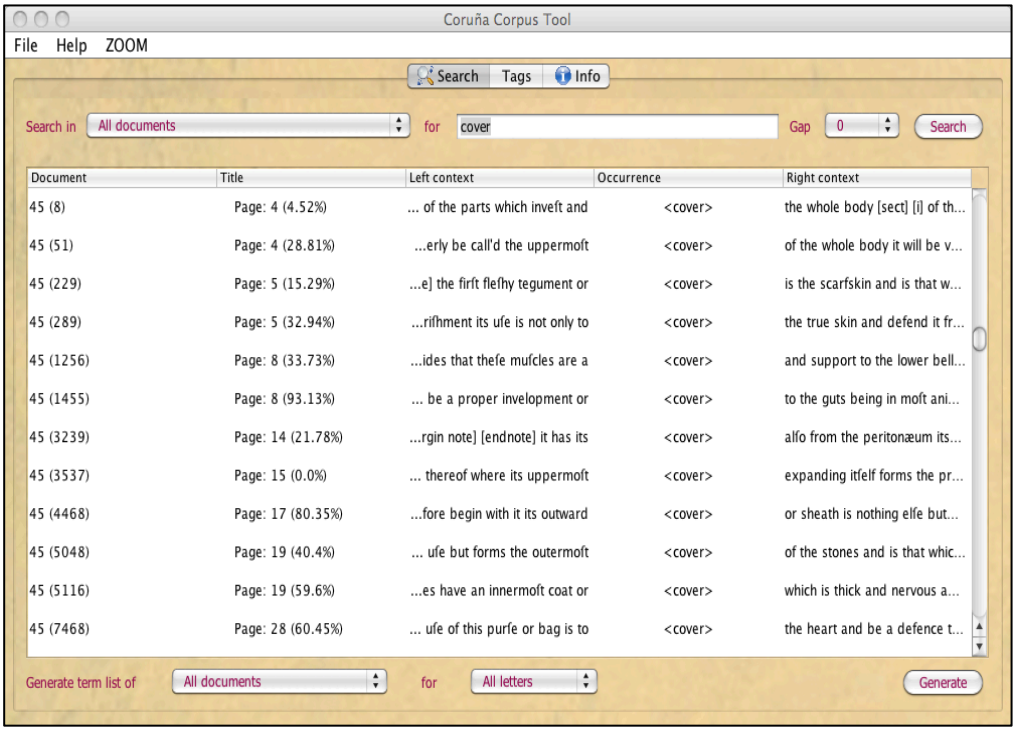


Figure 3.14: Results for *cover* in the search window of the *CCT*.

As Adolphs (2006: 5) points out, "this output format facilitates the analysis of lexical and grammatical patterns in the immediate environment of the search term". In the infrequent case that the left and right backgrounds are not clarifying enough, the *Coruña Corpus Tool* has another helping feature to offer: by clicking on the desired occurrence a pop-up window will show the fragment of the text in which the word in question will appear highlighted, for better identification, as shown below in figure 3.15:

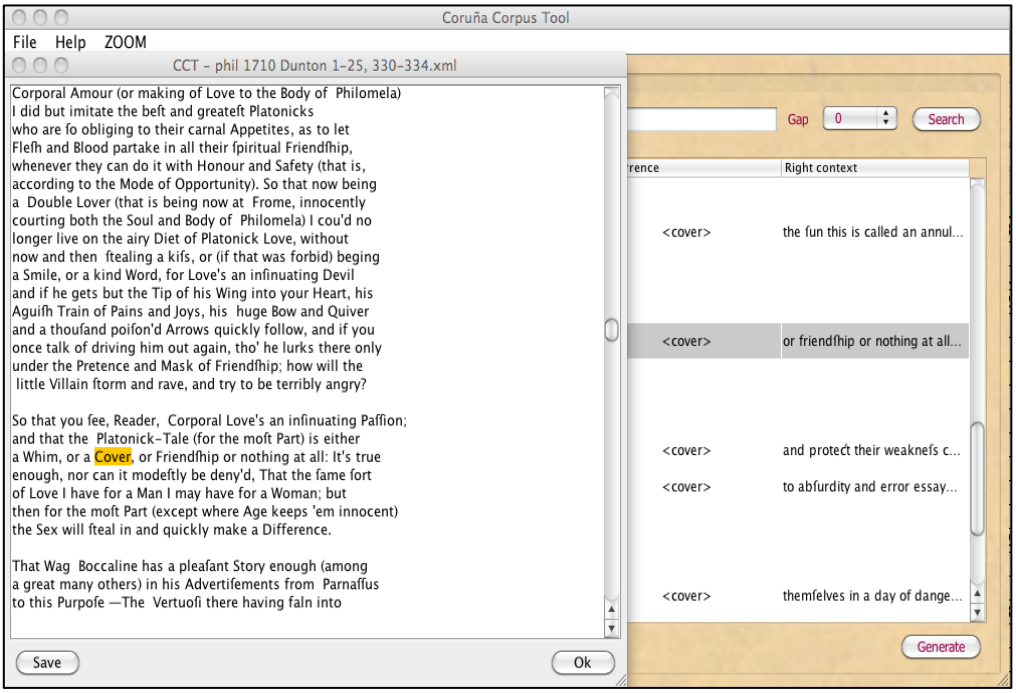


Figure 3.15: Disambiguating with the aid of the full-text option in the *CCT*.

3.3. The database

In order to process the data I have made up a database structured in three sections, the first of which focuses on the types and tokens present in the corpus. The second section deals with the processes, units and their origin, graphic variants and coinage dates according to the *OED*. The third section deals with the formulae necessary to process the relevant information from the previous two sections and establish principles, derive figures, and so on. The most significant fields contained in the first section are:

Types: Column A lists the different noun lexemes in the corpus after the disambiguation process mentioned above.

Tokens: column B shows figures for the total number of tokens included in both disciplines.

TokAstr and **TokPhil**: the following two columns show figures for the total number of tokens of every discipline. Column C provides figures for *CETA*, and Column D for *CEPhiT*.

The following forty-one columns reflect the tokens present in every individual sample in the corpus on an ascending timeline. I could have opted for splitting the database in two sections, one for *CETA* and one for *CEPhiT*, but I have decided to keep them together in order to observe side-by-side results for decades. Also for clarity, I have only included figures for those types that are indeed present in the samples, obviating zero results. Blank cells, therefore, indicate that the type in question has zero occurrences in that particular sample. The following figure depicts one very small fragment of the first section in my database:

	A	B	C	D	E	F	G	H	I	J	K
1	TYPES	TOKENS	TokAstr	TokPhil	a02a	a02b	p00	p05	a15	a19	p10
147	anatomist	4		4							
148	anatomy	9		9							2
149	ancestor	6		6				1			1
150	ancient	34	27	7		6			3	4	1
151	anecdote	1		1							
152	angel	36	2	34	1						13
153	anger	6		6							
154	angle	501	499	2	5	3		1	19		
155	angling	1		1							1

Figure 3.16: First section of the database dealing with types and tokens.

Contrary to the numerical and statistical nature of the first section, the second section in the database contains abundant and significant information about the processes, bases and affixes, their origin and their coinage dates, in order to establish patterns of noun formation and determine the period in the history of the English language that those new coinages were produced. I have adopted the taxonomy of the *OED* as regards affixes, i.e. *-meter*[1], *-meter*[2], and origins (Anglo-Norman, Old French, Ancient Greek, Scholastic Latin, and so on). Besides, I have recorded both the etymological origin of bases and affixes, and the date of first occurrence shown in the quotations of the different entries in the *OED*. The fields contained in this section are:

Process: I have identified all processes involved in the formation of new nouns in the corpus and divided them into five categories, (i) simple, (ii) affixation,

(iii) compounding, (iv) conversion, and (v) other. As can be observed in the following figure, when nouns are simple I do not research any further and the rest of the cells are left blank. Besides, affixation has been further subdivided into four different classes for reasons that will be explained in the following chapter.

Prefix/ICF: prefixes and initial combining forms have been grouped together, because the differences between them are not always clear, and different authors consider them differently, as explained in chapter 2.

PrefOrig: I am interested in attesting whether affixes and bases from different origins attach normally, so this column reflects the origin of the prefixes and initial combining forms.

Base: Bases used for further derivation are listed in the next column.

BCat: In order to attest which part of speech is used most frequently to coin new nouns I have included information on the base category.

BOrig: This column adds information on the origin of bases, which can be combined with PrefOrig and SufOrig to determine the mixed origins of many nouns. The terminology employed for this field has been adopted from that of the *OED*. Therefore, as regards Latin, for example, we may find labels such as Classical Latin, Middle Latin, Late Latin or Post-Classical Latin, among others.

Complex: If the base was not a simple word, the process by which it achieved its present state is noted. In the figure below, *foreknowledge* (Kirkpatrick 1730: 21) is formed by attaching the prefix *fore-* to the compound base *knowledge*.

BDate: This field provides information on the date in which the base was recorded for the first time.

Suffix/FCF: Suffixes and final combining forms are grouped following the same criterion used for prefixes and initial combining forms.

SufOrig: The origin of suffixes and final combining forms is recorded to check for cross-origin coinages.

Coinage: first-occurrence dates as recorded in the *OED* are offered on this field to review several topics. For example, I expect to attest whether the nouns were coined in the Late Modern Period, which falls directly under the scope of this study. Besides, I also intend to establish patterns that might show how long it takes for a base to start producing new derivatives. Finally, but on a lesser level, I will be capable to check whether some of the coinage dates offered by the *OED* correspond truthfully to the first use of words in history. It seems obvious that the higher the number of texts analysed in depth, the more inconsistencies can be found in the dictionary, which needs revising and many dates need to be taken back in time, for reasons explained by Tieken-Boon van Ostade (2009) and Bailey (2010) in the introductory section to this chapter.

Comments: This field adds extra linguistic information, such as graphic variants, processes different from the main ones explained above, presence or

absence of the type in the *OED*, and so forth. The next figure summarises what has been said so far about the second section in our database.

TYPES	Process	Prefix/ICF	PrefOrig	Base	BCat	BOrig	Complex	BDate	Suffix/FCF	SufOrig	Coinage	Comment
footing	deriv1			*foot*	v	OE		1400	ing_1	OE	1398	
footman	compound				n+n						1297	other: foot-man
footstep	compound				n+n						1220	
forbearance	deriv1			*forbear*	v	OE		888	ance	F	1591	
force	simple											
fore-father	deriv1	fore	OE	*father*	n	OE		825			1300	
fore-horse	deriv1	fore	OE	*horse*	n	OE		825			1480	
fore-knowledge	deriv1	fore	OE	*knowledge*	n	ME	comp	1300			1535	
fore-ordination	deriv1			*foreordain*	v	ME	deriv	1420	ation	L	1628	
foregoing	deriv1			*forego*	v	OE	deriv	825	ing_1	OE	1581	
forehead	deriv1	fore	OE	*head*	n	OE		825			1000	other: fore-head
forenoon	deriv1	fore	OE	*noon*	n	eOE		666			1506	
forepart	deriv1	fore	OE	*part*	n	CL		666			1400	
foresight	deriv1	fore	OE	*sight*	n	OE		950			1300	
forest	simple											
foretast	deriv1	fore	OE	*taste*	n	OF		1330			1435	

Figure 3.17: Second section of the database: processes, elements, origins and dates.

Finally, the third section contains all the formulae and the total results for every sample and process, namely, derivation, compounding, conversion and others, the type/token ratio per author, the use of simple and complex words, the choice of affixes and bases, and so on, all of which can be quantified to establish patterns of word formation together with other sociolinguistic variables such as place of education, sex of the author, age of the author when the work was published et cetera. All these formulae and results will be shown in depth in chapter 4. Figure 3.18 below shows total

results for word-formation processes in the corpus, as well as other significant sociolinguistic features of the samples

	A	E	F	G	H	I	J	K	L
1	**TYPES**	a02a	a02b	p00	p05	a15	a19	p10	p17
4536	TYPES	495	540	607	447	365	454	634	417
4537	TOKENS	2131	1901	1722	1888	1689	1602	1848	1924
4538	DERIVATION	233	249	325	234	193	208	325	227
4539	COMPOUNDING	15	20	14	12	3	21	21	11
4540	ZERO DERIVATION	16	24	44	23	24	19	39	26
4541	OTHER	1	2	5	0	0	4	3	1
4542	Token/type rate	4,31	3,52	2,84	4,22	4,63	3,53	2,91	4,61
4543	Sex of the author	Man	Man	Woman	Man	Man	Man	Man	Man
4544	Genre/Text-type	textbook	textbook	essay	treatise	lecture	dialogue	treatise	treatise
4545	Place of education	Unknown	England	England	Scotland	England	England	England	England
4546	Age when published	Unknown	34	34	34	50	53	51	41

Figure 3.18. Fragment of section three of the database: processes and other sociolinguistic characteristics of the texts.

The data contained in the three sections of the database interact by means of mathematical formulae, from which I intend to extract statistical information. Therefore these formulae connect the different processes with the affixes and bases used to coin new nouns, their origins and coinage dates, and results for every author and/or group of authors, decade, genre, discipline and so forth, can be drawn to establish patterns of word-formation across the history of the English language.

Summing up, I am presenting here a research work that contains an approach to language that resorts to various tools characteristic of other non-

linguistic disciplines. It analyses experimental data, brings together, compares and contrasts two distinct but complementary disciplines belonging to the natural and social sciences, and dissects their language by means of informatics and linguistics. The results obtained by means of my quantitative analysis will be shown in the next chapter, and I will elaborate morphological theories derived from those results.

4. Corpus material and analysis of data

1. Introduction

In this chapter I intend to present the data extracted from my corpus, and describe various morphological analyses that I have carried out on the nouns included in the eighteenth-century sections of *CETA* and *CEPhiT*. This material has been stored in a sizeable database containing over 325,000 cells of information, which have been structured to focus on a good number of variables, such as types and tokens, morphological processes, occurrences by sample, origins and coinage dates of affixes and bases, and variation in graphical representation, among others. I have applied mathematical formulae that compare and contrast single or multiple ranges of data, in order to generate patterns that may shed some light on the status of morphology in the late Modern English period. As stated in previous chapters, my goal is to determine the most productive periods, processes, affixes and bases, if any, in the history of the English language, as seen in eighteenth-century scientific texts, paying special attention to those taking place in the eighteenth century.

As I explained in chapter 3, the *Coruña Corpus* had not yet been encoded either from morphological or syntactical standpoints at the time of

writing this research work. Thus, the nature of the greatest part of my analysis could not be carried out by means of automated processes, and must be undertaken by hand. Indisputably, the *Coruña Corpus Tool* contributed to the process of building up my database to a great extent by segmenting the texts and generating frequency lists. Also, it provided crucial aid in disambiguating nominal from non-nominal types —as is the case of *-ing* words, which will be dealt with more in depth later. However, the remainder of my work implied manual classification, separation of mono- and multimorphemic types, and one-by-one input of the data recorded in each of the variables under scrutiny.

Due to the size and diversity of the data and variables processed for this research work, I have relied on already-existing nomenclatures regarding base and affix origins, and I have adopted those provided by the *Oxford English Dictionary*. Nevertheless, its high specification when referring to source languages, with as many as seven labels for Latin alone —Latin, Classical, Post-classical, Middle, Modern, Late and Scholastic— might render seemingly random results and, as a consequence preclude the creation of patterns. In order to prevent this, and for this work in particular, I have simplified the multiplicity of subdivisions of every language into one, i.e. Latin. Also, Anglo-French and Anglo-Norman examples have been merged into one because, in my opinion, this inclusive approach will not affect the final results presented. As a consequence, I have established six major etymological groups, namely, 'Greek', 'Latin', 'Anglo-Norman', 'French', 'English', and I have applied the

label 'Other' to a lesser group of examples coming from other languages, and to those whose origin is obscure or simply unknown[35].

Besides, I have also benefited from the coinage dates suggested by the *OED* —or rather, approximate dates, as they rely on the sample fragments attached to every entry of the dictionary, which I have revised one by one— in order to determine the first use of every noun contained in the corpus. As we shall see in the following chapter, I have found a few mismatches in connection with some of the coinage dates provided by the *OED*, since some nouns present in my corpus have been documented earlier than the dates recorded in the *OED*. Although, this does not contravene the general application of dates provided by the *Dictionary*, mainly because they are used merely for guidance. In this respect I will provide a list of noun types dated incorrectly in the *OED* in my conclusions in chapter 5, together with the new earlier date and the sample where they can be found. I hope that my findings may contribute to improving its contents.

[35] The following language labels may seem very generic at first sight, and indeed they are, but all the different stages of the source languages have been contemplated, i.e., Ancient Greek, early Old English, Old Norse, Classical Latin, Middle French, Early Scandinavian, among others. The simplification here offered is merely intended for ease of computation purposes. A wide range of variables would render the establishment of patterns rather complicated.

Three sorts of frequencies —raw, percentage and normalised— will be offered to represent different kinds of data and used for different purposes in this analysis. Raw frequencies and percentages will be given in the second section of this chapter in order to present concluding findings on issues such as types belonging to different processes, affixation classes, their origins and coinage dates, and some others. Given that the size of a sample may affect the level of statistical significance, the common base for normalisation must be comparable to the sizes of the corpora (or corpus segments) under consideration (McEnery *et al*. 2006: 53). Therefore, in subsequent sections normalised frequencies will be circumscribed to the direct comparisons of subcorpora, authors, samples, text-types or any other sets of data that may contain a clear disparity in their total number of words. These counts will be normalised to a rate per 10,000.

The second section of this chapter deals with the analysis of nouns and their classification, consisting of separating those nouns that have not undergone any kind of process from those which have. I have applied several intralinguistic variables to the latter, which have been examined on the basis of the morphological process that produced them. I have established four groups of processes, i.e., 'affixation', 'compounding', 'conversion' and 'other'. The first three groups meet the characteristics of their namesake processes extensively defined in chapter 2. 'Other' includes minor processes of word shortening, words coined by wrong interpretations of other words and plain

inventions. Affixed nouns were then decomposed into their forming elements, *viz*. bases and affixes, and those elements have been subcategorised regarding their origins, coinage dates, category and further decomposability in the case of bases.

Section three and its ensuing subsections are devoted to comparing the productivity of the processes and linguistic materials observed in the corpus by means of five extralinguistic variables. The first variable consists of comparing both subcorpora individually, as a means to characterise both disciplines from a linguistic point of view. The second variable deals with genres/text-types, and explores differences among them. The third variable studies language from the standpoint of the sex of the authors. By means of this variable I intend to determine a potential variation in the linguistic competence of men and women scientists at a moment in history in which women had no access to universities. The fourth variable contemplates the possibility that the location where scientists were educated might influence their use of the language in writing. Finally, the fifth variable considers the age of the authors as a hypothetical factor in the versatility and complexity of language use. I can anticipate that my forty-one samples may be adequate to present reasonably reliable patterns on the first three variables. However, their number might be low for the last two. Therefore, patterns for these must be considered purely experimental.

2. Nouns and processes

Following the procedure shown in figure 3.11 in the previous chapter, firstly, noun tokens have been classified by types, then divided by disciplines, and finally the resulting noun types have undergone a manual method of dissection into their constituent parts, if applicable. Simple nouns, that is to say, monomorphemic types that have not undergone any kind of formation processes (or 'de-formation', such as shortening and other analogous practices, as we shall see below) have been excluded at this stage and, though they are still taken into consideration and computed for the study of noun variation and *hapax legomena*, I will make no further mention to them in this section. They will be nonetheless reused in my analyses on extralinguistic variables in section 3. At a later stage complex nouns were differentiated from compounds, conversions and other nouns obtained by means of minor processes. I kept the latter for computation purposes and in order to mention those processes succinctly and give examples in their due sections, whilst the former were fully examined as regards its component parts, origins, coinage dates, graphical representations, etcetera.

After carrying out the previous procedures on the corpus, I have obtained the figures shown in chapter 3, that is, 79,271 noun tokens, 38,121 in *CETA* and 41,195 in *CEPhiT*. These tokens correspond to 4,530 noun types, from which 2,058 belong to *CETA* and 3,777 belong to *CEPhiT*. Those types

were subsequently checked by hand, divided into five groups, and computed. The results are as follows: 1,370 noun types that have not undergone any process (simple), 2,510 noun types that have experienced some sort of affixation, excluding zero-derivation (complex), 274 types formed by the combination of two or more free bases (compound), 376 types in which the base and the resulting noun share the same graphical form, but which contain a zero-morpheme (zero-derived), and 28 types that do not fall into any of the previous categories for various reasons. The following figure shows the percentage numbers that illustrate the division explained above, making reference to the processes affecting nouns in the corpus.

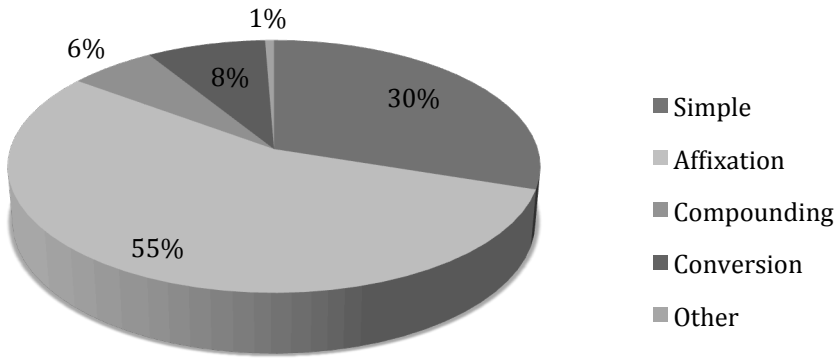


Figure 4.1. Morphological processes: types.

These results confirm the extensive use of complex nouns in scientific texts, which represent more than half the total number of types. If we add other non-simple processes to the equation, simple nouns represent less than one third of the total. Although, comparing these figures involving noun types with those obtained from noun tokens may be interesting. As figure 4.2 below illustrates, the percentage of simple noun tokens is higher than the rest, and thus it indicates a more recursive use of monomorphemic nouns. We may wish to restrict this test to simple and complex nouns and calculate their type-token ratios as follows:

$$\frac{simple\ tokens}{simple\ types} = \frac{39{,}305}{1{,}370} = 29.1$$

$$\frac{complex\ tokens}{complex\ types} = \frac{35{,}818}{2{,}510} = 14.3$$

Then, we shall realise that, on average, each simple noun is reused over twenty-nine times in the corpus, whereas repetition happens less than half those times as regards complex nouns, reused around fourteen times each. This can be graphically observed in figure 4.3 below.

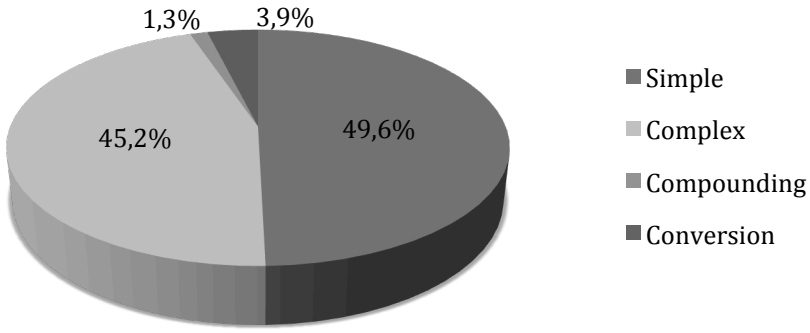


Figure 4.2. Morphological processes: tokens.

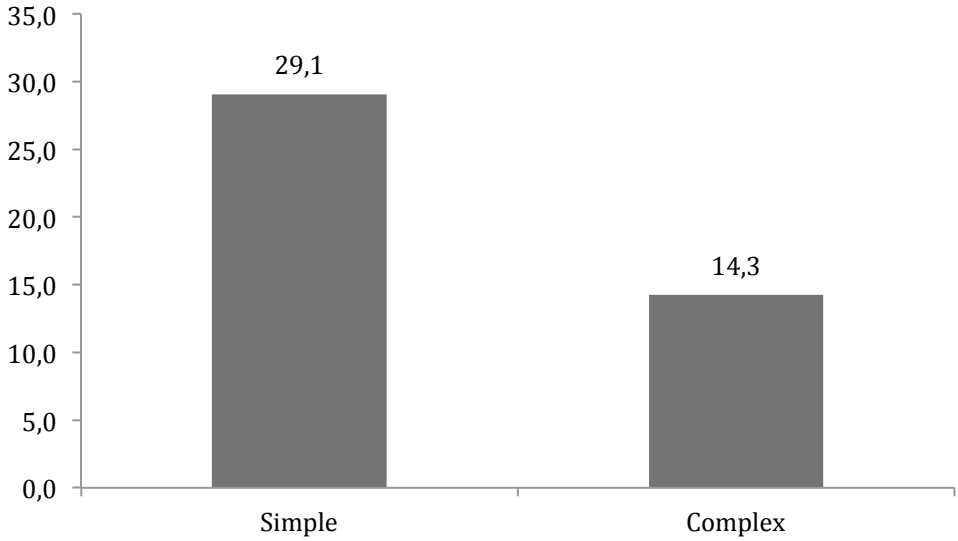


Figure 4.3. Type-token ratio of simple and complex nouns.

As we have observed above, simple nouns tend to be substantially more repeated than complex nouns across my samples. One final test on this matter can be performed on several levels to attest this principle: the presence of simple and complex *hapax legomena*, though it is not so clear that words that happen only once in a corpus contribute significantly to a supposed richness in the language (Popescu *et al*. 2009: 36-37). The first level comprises examining each sample individually, in order to assess the linguistic variation characteristic of every scientist. The resulting table with the data[36] will be used later in the comparison and contrast among authors by means of the variables 'sex', 'place of education' and 'age'. The second level involves each discipline individually, and it can be used to typify the features inherent to writings in astronomy and philosophy. These features will be outlined in section 3 of this chapter, and expounded in full in chapter 5. Lastly, the third level encompasses the whole corpus, and it is my hope that it may help drawing conclusions on the scientific register of English in general. It also represents the first stage towards further comparisons between scientific and non-scientific corpora. On this level I have computed 411 simple and 853 *hapax legomena*. These results support the principle detailed above, since most of the nouns used only once in the corpus are complex.

[36] For the details of every sample individually see table 4.20 in section 3.3.

The characteristics of affixation have been explained in chapter 2, so I will not review them here to prevent an unnecessary repetition of literature in a chapter devoted to data analysis. In brief, this section will examine those nouns containing at least one bound form. That is to say, nouns made up with a free form that acts as a base, and a bound form or affix, or those made up with two bound forms. Nouns containing more than one free form will be included in the group of compounds in section 2.2, and those presenting zero-morphemes will be studied in section 2.3, which deals with conversions. Those nouns not included in any of the previous groups will be paid due attention in section 2.4.

A preliminary outlook reveals a rather irregular pattern in the productivity rates of affixation throughout the Middle and Modern English periods. As we can see graphically in figure 4.4, alternating peaks and valleys reflect major differences among periods. My findings indicate that most noun types used in scientific writing, as seen in *CETA* and *CEPhiT*, were coined in the 1300s and 1500s, whereas the eighteenth century was manifestly poorer as regards noun formation. In this chapter and the next I will analyse probable causes for this variance.

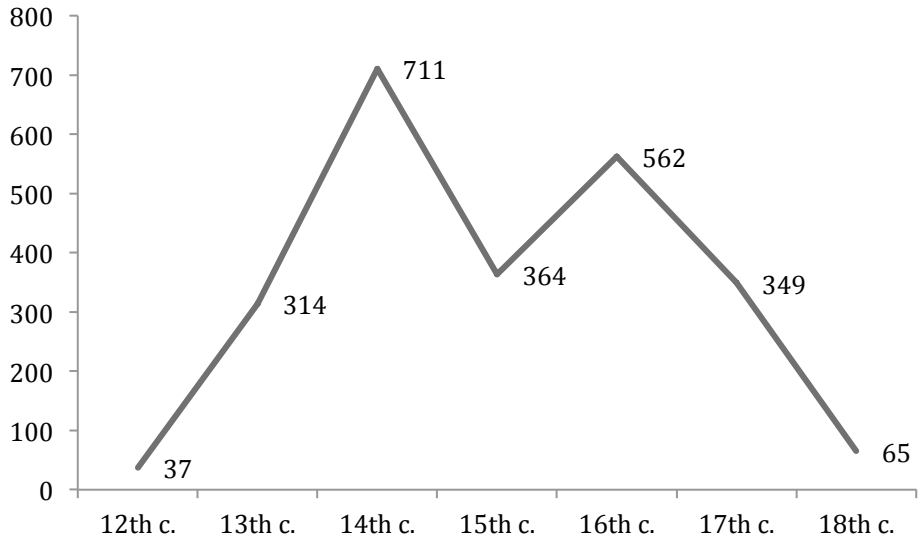


Figure 4.4. Affixation in English since the twelfth century.

The following paragraphs attempt to justify the proposed division of affixation in different classes. I will explain their characteristics, constituting elements, and their productivity on the different stages of the history of the language. Also, an exhaustive analysis of prefixes and suffixes will be carried out to determine their origin, productivity and the dates in which they contributed to coining new nouns in English. Special attention will be paid to the most productive affixes from various origin components, such as *-ion*$_1$ (Latin), *-ment* (Anglo-Norman), *ence-* (French) and *-ing*$_1$ (Old English).

2.1.1. Affixation classes

In chapter 1 I described some language internal and external reasons that affect the evolution of the English language and substantiate the heterogeneity of its

morphology. Apart from this, my corpus contains linguistic material coming from thirteen languages. Undeniably Greek, Latin, Anglo-Norman, French and Old English provide the highest number of elements to derive new nouns in English. But some linguistic material comes also from other Romance languages, such as Italian and Spanish, from Germanic languages, among which we can find German, Dutch, Norse, Swedish and early Scandinavian, and even from Turkish. Besides, some of the nouns found come from uncertain or unknown origins. Furthermore, many affixes in English have been imported from Greek, Latin and French, therefore obscuring the distinction between borrowed words and words coined by means of genuine derivation processes in English. The noun-forming suffix *-eer*, for example, was adapted or evolved in the seventeenth century from French *-ier*, but are we supposed to consider it English or foreign? This is a hard choice. Finally, numerous bases and affixes regarded as French were in origin Latin or Greek, and French acts only as the intermediary for these materials to permeate English (Durkin 2008). Because of all these reasons, attempting an analysis of complex nouns from a single angle seems rather unattainable.

In order to overcome these problematic circumstances, I have established a further division of complex words into four groups, which I denominate 'affixation classes I, II, III' and 'IV', and these classes have been studied separately. This division has already put forward in other works (Camiña-Rioboo 2010, 2012). These four classes group nouns that can be

differentiated with regard to their bases, either free or bound, their affixes, neutral or non-neutral (Aronoff 1983), and their coinage dates. The criteria to distinguish among them will be explained more comprehensively in the following paragraphs but, in short, nouns encompassed by classes I and II contain at least one free form and an affix and those belonging to classes III and IV are made up with bound forms. The differences between classes I and II are rather subtle. On the one hand, nouns contained in class I present a free base with strong boundaries (Katamba 1993) and a neutral affix. On the other hand, class II incorporates nouns with free bases with either strong or weak boundaries and neutral and non-neutral affixes alike. Their subdivision is therefore more linked to the recording —or not recording— of the bases at the time of coining the complex word, and to the uncertainty of whether they are adaptations of foreign words or derivations, as explained above. Classes III and IV are not as closely related as the previous two. Class III is formed by nouns containing a bound base and a non-neutral affix, and Class IV incorporates a construction of affixes/combining forms, most of which originated on classical grounds.

After applying these criteria to the 2,510 complex noun tokens present in the corpus, I have subsequently divided them into four groups, and then incorporated them into their respective classes in the following manner: 740 types fulfil the characteristics of class I, 630 types can be included in class II; 1,106 types fall into class III; and finally, 34 types belong to class IV. The

following figure illustrates with percentages the distribution of nouns as described above.

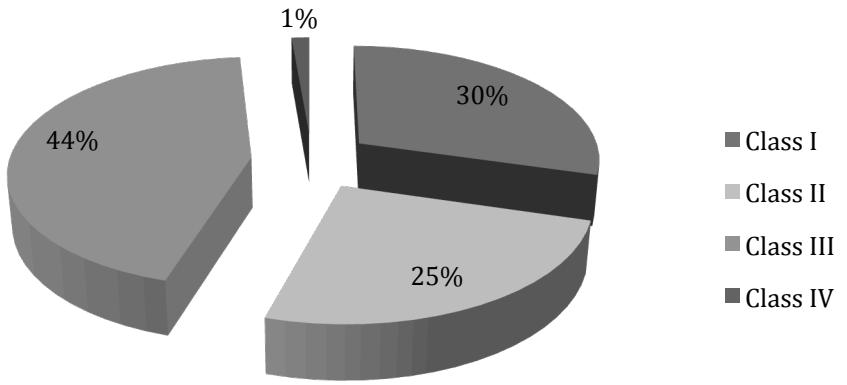


Figure 4.5. Affixation classes.

As we can see, affixation class III is by far the most productive of them all. However, affixation involving free bases represented by classes I and II outnumbers that of bound bases.

Another formula that may act as a distinguishing element among these classes is the type-token ratio. Types are, on average, reused to a different degree depending on the affixation classes to which they belong. For example, every type included in class I produces 5.3 tokens, that is to say, every noun is repeated in the corpus slightly over five times. In contrast, types belonging to class II produce 11.2 tokens. Those contained in class III are employed 22.1

times in texts. Finally, types in class IV can be observed 12.6 times in the samples. As we shall see later, these ratios show some degree of variation when my astronomy and philosophy samples are compared, especially as it comes to class IV, but this issue will be dealt with in more depth later in this chapter. Figure 4.6 shows the disparity of type-token ratios characterising the affixation classes in the corpus.

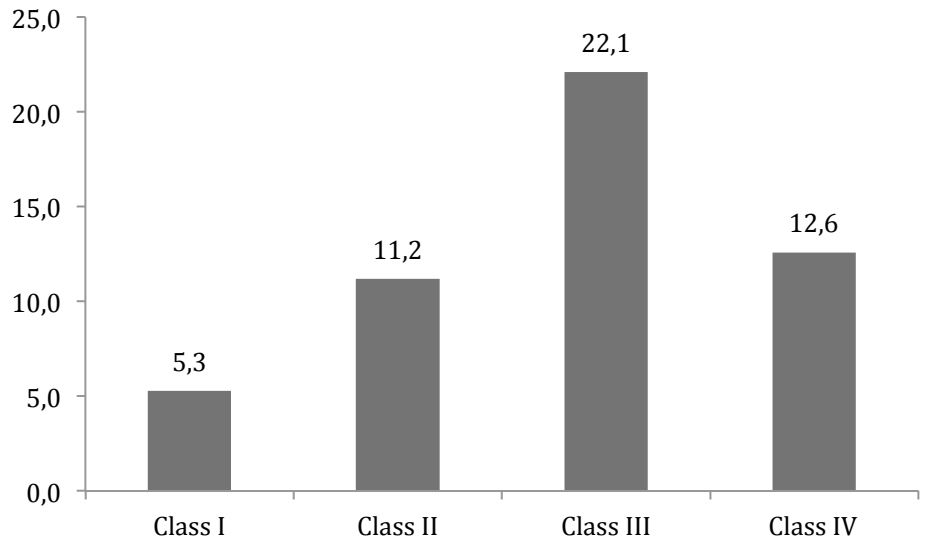


Figure 4.6. Type-token ratio in affixation classes I-IV.

Furthermore, these four classes have followed very different lines from a diachronic point of view. My findings show, for instance, that nouns coined in the twelfth century were almost exclusively from classes I and III. But this situation was very different in the sixteenth century, when numbers concerning

nouns from class II were equivalent to those in classes I and III. Conversely, in the eighteenth century most examples are coinings from class I, that is to say, combinations of free bases and neutral affixes. More importantly still, this class seems to follow a steady line to dominate noun formation from the 1500s onwards. The following table shows the absolute numbers for affixation classes between the twelfth and the eighteenth centuries, and figure 4.7 displays evolutionary lines based on the percentage that each class epitomises in the same period.

Table 4.1. Types in affixation classes per century.

	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
Class I	18	72	107	75	192	170	49
Class II	2	38	187	118	177	92	9
Class III	17	202	411	168	185	76	3
Class IV	0	2	6	3	8	11	4

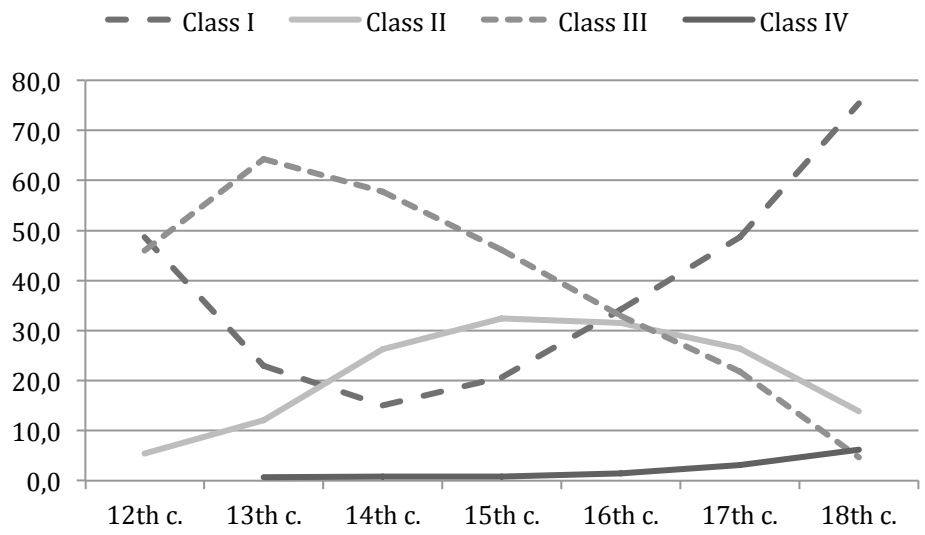


Figure 4.7. Percentage of affixation classes per century.

a. Affixation class I.

This class categorises straightforward combinations of free bases, which behaved as independent lexical items and attached to neutral affixes in order to coin new nouns. These bases had been recorded previously to their use to make up derived nouns. Examples including verbal bases such as *develop* (1592) → *development* (1756), nominal, as *connection* (1680) → *disconnection* (1735), and adjectival, as *impracticable* (1656) → *impracticability* (1747) can be observed in the corpus.

I have found a total of 740 complex noun types classifiable under the premises defining this subsection, of which 251 belong to *CETA* and 598 belong to *CEPhiT*. I need not say that the total number of types is smaller than

the sum resulting from the addition of the figures of both subcorpora because *CETA* and *CEPhiT* share a good number of those types (109).

The English component in class-I complex nouns is remarkably the most frequent in affixation class I. English bases contribute to providing 276 new nouns, whereas French bases are used to make up 233 nouns. Latin and Anglo-Norman provide 151 and 56 bases, respectively. The use of Greek bases in this class of affixation is very marginal, with only 3 instances. Figure 4.8 shows the percentage figures of base origin in this class:

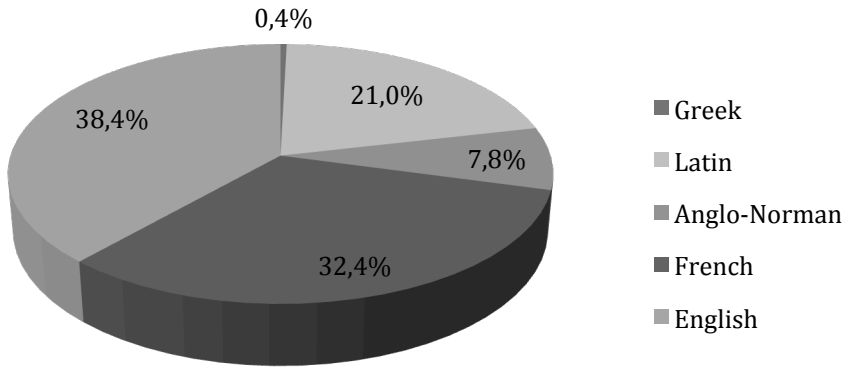


Figure 4.8. Base origin in affixation class I.

b. Affixation class II

Finding out the coinage date of a complex noun may at times involve a relatively arbitrary task and, generally, the further back in time we research, the

more uncertain our results will become. To make things worse, the presence of Romance affixes in a great number of English nouns increases the complexity in determining whether the noun in question is the outcome of an adaptation of the original or, rather, a derived type genuinely obtained by means of a combination of linguistic material (i.e., bases and affixes) which already existed in English at that particular moment in time.

This is the reason to include this class as a separate group in my research. As happens with class I, it includes nouns made up by free bases liable to be used in combination with neutral affixes. Conversely, however, class II groups those nouns that the *OED* regards as instances borrowed from other languages, but given that their bases had been present in the English language before they were coined, they were also likely to have been coined by means of already-existing base + affix combinations. Such are the cases shown in the following examples:

(19) … wanting the advantage of ſuch a diſpoſition and **arrangement**, the ſcience was ſo far defective (Hill, 1754: 9)

(20) They turn from barren to fertile earth, which indicates ſomething analogous to a **ſelection** of food (Smellie, 1790: 7)

The two nouns in question, *selection* and *arrangement*, were seemingly coined in 1646 and 1727 respectively, and they are regarded as borrowings from Latin

and French, but the verb *select* had been in use since 1567. As for the verb *arrange*, it had been used in English since as early as 1375. Besides, both affixes *-ion* and *-ment* had been producing new lexical items for centuries. So there is not much evidence that they needed to be borrowed from other languages when the elements that make them up had been present in English for years, centuries even. My theory does not prove either way, but at least it raises some doubt on the truthfulness of some origins proposed by etymological dictionaries. As a matter of fact, I have found instances of affixed nouns included in class II whose coinage date precedes that of their base, as is the case of *assert* (apparently coined in 1638), *assertion* (1420), and *assertor* (1566). It would also account for anaphoric examples such as the following:

(21) the reſt are found by Addition or **Subſtraction** of the mean Motion for common Julian Years (Hodgson, 1749: 89)

This construction would rely on the Latin term **substractionem*, which in fact did not exist in Latin. The verb *substract*, however, was coined some fifty years earlier than the derived noun, which made use of the productive suffix *-ion* to emulate other similar constructions such as *association*, *conviction* and *detection*, to number but a few. More cases of anaphoric derivations will be offered in section 2.4.

I have found 630 complex noun types amenable to fall in this subsection, of which 241 belong to *CETA* and 587 belong to *CEPhiT*. On the one hand, the Romance component is clearly predominant: French provides the highest number of bases (342), followed by Latin (176) and Anglo-Norman (79). On the other hand, the presence of the Germanic component is very low, and English-originated material contributes with only 23 bases. Additionally, the presence of Greek is merely testimonial, with 3 bases. Figure 4.9 illustrates the percentages of bases coming from the most productive languages. It is worth noting the apparent dissimilarity of base origins in classes I and II, which may in itself justify an independent study of both.

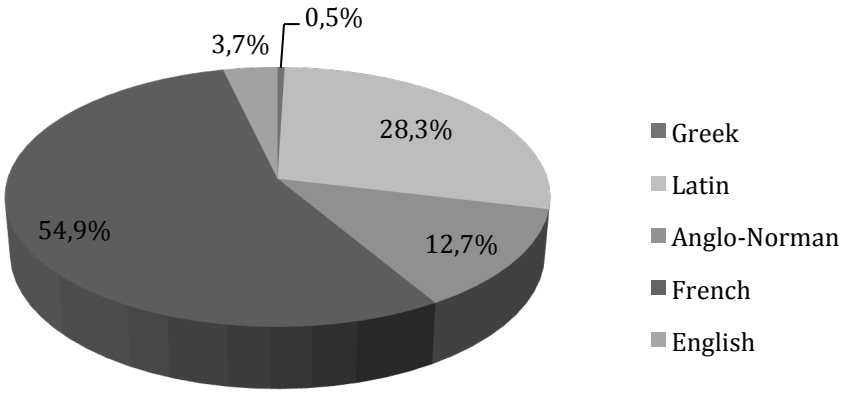


Figure 4.9. Base origin in class II.

Word category of free bases

The syntactic categories to which bases belong may be also of some interest. In the following paragraphs I shall deal with this as it may help complete our overview of derivational processes in the eighteenth-century scientific register of English. Because class III is made up with bound bases, and prefix/initial combining form + suffix/final combining form combinations will not be included in this analysis. Only the categories included in classes I and II will be considered. I have computed 1,370 different bases that contribute to the coining of complex nouns. Of those, 685 are verbs, 335 adjectives, 291 nouns, 24 proper nouns, and the remaining 35 instances have been derived from other parts of speech, but many of them have verbal foundations, such as participial stems and participial adjectives. No significant divergence can be observed in the use of word categories between classes I and II in order to coin new nouns. Therefore percentages of their results can be presented together, as illustrated in the following figure. Examples are shown in table 4.10 below.

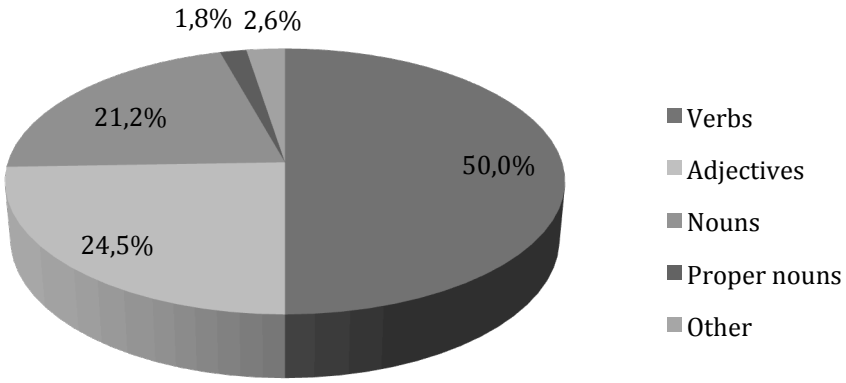


Figure 4.10. Base category in affixation classes I and II.

Table 4.2. Base category in derivation classes I and II with examples.

Category	**Types**	**Examples**
Verb	685	*assemblage, broacher, contrivance, development, impregnation*
Adjective	335	*ductility, criticism, distemperature, inconclusiveness, irascibility*
Noun	291	*colatitude, courtship, legislature, metaphysician, prepossession*
Proper noun	24	*Arabian, epicurean, Greenlander, Laplander, Platonism*
Other	41	*correction, directory, drunkenness, effrontery, interestedness*

Complexity of free bases

In chapter 2 we saw examples provided by Hockett (1958: 79) that prove that affixation is a finite process. Therefore we cannot keep adding affixes to bases *ad infinitum* to further derive new words. The information stored in my

database allows determining the ratio of new coinages derived from simple or complex bases in scientific writing.

The immense majority of the bases used to coin new nouns are simple (1,074), and the remaining (174) can be classified as derivations, compounds, zero-derivations and other. The line between a few derived and compound bases is blurred, and some cases deserve a deeper study. Such are the cases of *awful* (> *awfulness*) and *overflow* (> *overflowing*), which the *OED* considers derivations, but depending on our perspective could be also considered compounds (see chapter 2, section 2.3). Notwithstanding this, the remaining 'non-simple' bases have been divided as follows:

a) 100 are the product of previous derivations, such as *behave* (> *behaviour*, Kirkpatrick 1730: 30), *changeable* (> *changeableness*, Watts 1726: 7) and *debase* (> *debasement*, Campbell 1776, 48).

b) 10 are compounds, such as *housewife* (> *houswifery*, sic. Harris 1719: 24) and *knight-errant* (> *knight-errantry*, Macaulay 1783: 30).

c) 54 are zero-derived bases, of which many come from participles in *-ate,* such as *animate* (> *animation*, Crombie 1793: 42) and *articulate* (> *articulations*, Smellie 1790: 22). Other examples include the verbs *apart* (> *apartment*, Hume 1748: 32) and *class* (> *classing*, Ferguson 1769: 64)

d) 10 include other constructions, such as i) shortenings: *ready* (> *readiness*, Vince 1790: 29); ii) backformations: *sigh* (> *sighing*, Dunton 1710: 13); iii) errors, such as *whiggamore* (> *whig*, Burke 1770: 5); iv) onomatopoeias: *dash* (> *dashing*, Greene 1727: 8); and v) other alterations: *astonish* (> *astonishment*, Bonnycastle 1783: 43).

c. Affixation class III

In this class I have included all those nouns that have been considered borrowings or adaptations from foreign words, but whose constituent parts are clearly inferred from the complex word, although there is no evidence that those parts had been already recorded in English. Examples of class III are *morality*, *observance*, *passage* and *participation*, among others. This group also contains nouns, with independence of their origin, whose base/affix boundaries cannot be plainly deduced from the resulting complex noun without the presence of a non-neutral affix. Such is the case with *assumption* (<**assump+tion*, Latin), *omission* (<**omiss+ion*, Anglo-Norman), *amplitude* (<**ampli+tude*, French) and *strength* (<**streng+th*, Old English). As I have already mentioned, they have generally been considered as borrowings, but here I will tackle a more inclusive approach and review them from the perspective of word formation. Therefore they will be regarded as complex nouns, in line with the theories established in other corpus-based studies

(Dalton-Puffer 1996). Accordingly, I have listed 1,106 noun types belonging to class III, 566 in *CETA* and 929 in *CEPhiT*, most of which were coined in the Middle English period. The majority of bases come from Romance origin, 561 from French, 202 from Anglo-Norman and 300 from Latin, in clear contrast with the Germanic origin, from which Old and Middle English contribute only with 34 bases. All this information is illustrated in percentages in figure 4.11.

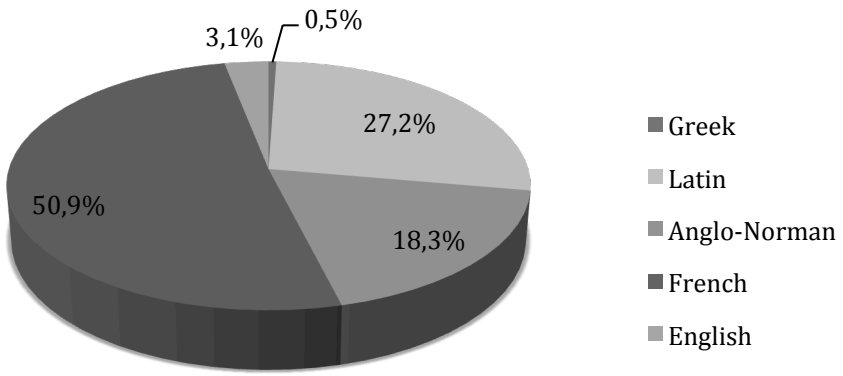


Figure 4.11. Base origin in affixation class III.

It may be clarifying to establish now the diachronic evolution of the use all these bases, free and bound, across the history of English. Only 0.5% of bases are of Greek origin (13), and their influence was seen at the end of the Middle English period, but it cannot be compared to other languages, or to the importance that Greek affixes enjoyed.

Latin contributes with 699 bases that make up to 26.2% of the total. Its impact on English has been significant since the thirteenth century, and became the second most productive contributor two centuries later. The use of linguistic material coming from Latin reached its peak in the seventeenth century, in which the scientific revolution took place, and remained at the top until the end of the time span covered in this study.

Anglo-Norman provides 13.6% of bases to the corpus (363). Its diachronic figures parallel those of French, which may indicate the prevalence of the Romance over the Germanic component in this hybrid group. Anglo-Norman bases had their peak about two hundred years after the Conquest, representing a quarter of the total bases found, and then their impact on new formations gradually decreased to minimal levels, computing below 5% in the seventeenth and eighteenth centuries.

French represents 46.9% of the total number of bases (1,251). The predominance of French increases from one third of the total noun formations in the 1200s, second only to Old English, to reach over fifty per cent during most of the Middle English period. In Modern English, French and Latin provided most of the bases for new coinages.

Old English contributes 12.8% of bases to my corpus (341). We have seen that these figures fluctuate greatly from one affixation class to another — thirty-one per cent in class I as opposed to three per cent in classes II and III— and from period to period. For instance, its predominance in the twelfth

century is almost absolute, contributing up to two thirds of coinages. Conversely, in the seventeenth century the presence of Germanic bases in new nouns is practically inexistent, due to various reasons —lack of prestige, among others— already mentioned in chapter 1.

So far, I have provided an assortment of raw figures and percentages to explain the influence of bases coming from five different origins. But a graphic representation of this diachronic evolution will indeed show this in a clearer manner, as seen in figure 4.12.

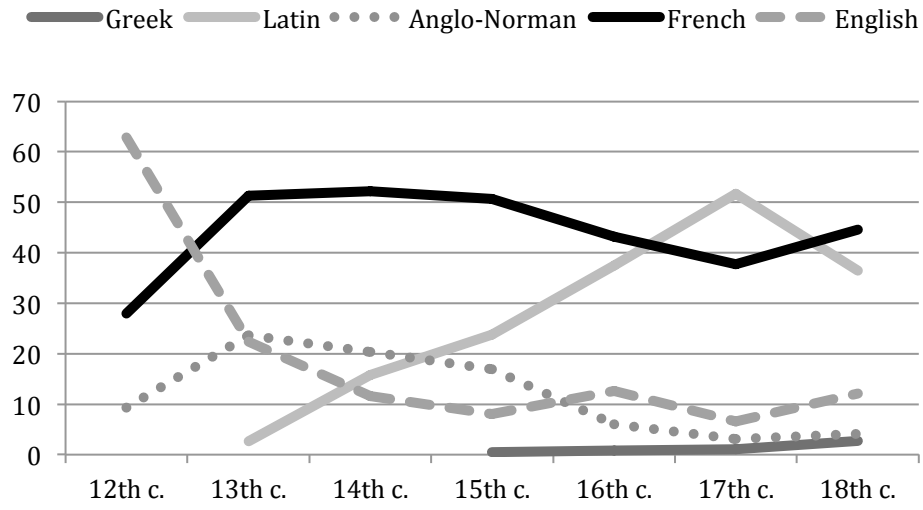


Figure 4.12. Evolution of base origins across the Middle and Modern English periods.

d. Affixation class IV

Finally, I have arranged all prefix (or initial combining form) + suffix (or final combining form) combinations into this group. The forming elements used to

coin new nouns that have been included here come almost exclusively from Greek or Latin, and some authors (Bauer, 1983; Adams, 2001) define this word-formation process as a variety of compounding, and label it 'neoclassical compounding'. However, if we take into account the true nature of compounding, not of derivation, which consists basically of the combination of two free morphemes, we can concur that the process here depicted shows different implications, since basically the two elements making up class IV nouns are bound. Aside from this, there is another reason to treat these types as a subcategory of affixation. Some of these classical combining forms have become free morphemes across the history of the language. Such is the case of *scope*, coined in 1534, which is present in the corpus in three instances: *telescope* (1648), *microscope* (1651), and *helioscope* (1675). Also, the association of *scope* is not restricted to Latin or Greek any longer, and it attaches to both Romance and Germanic bases to coin new nouns, in examples —not extracted from my corpus— such as *reflectoscope*, *sniperscope*, *vectorscope* and *volumescope*. The process to become a completely free base does not seem to have concluded yet, since after checking the three hundred and seventy-five entries of the word *scope* included in the *OED*, none of them would be incorporated to my affixation class I group.

I have found 34 instances of nouns liable to belong to affixation class IV. Besides, six possible combinations can be observed regarding initial and final combining forms: Greek + Greek (18 types), Greek + Latin (4), Greek +

French (5), Latin + Greek (1), Latin + Latin (2) and Latin + French (2). Table 4.3 illustrates class IV with examples that describe the origin of the two components of the nouns included here.

Table 4.3. Nouns in class IV.

Origins	**Examples**
Greek + Greek	*apology, biography, metropolis, ontosophy, zoophyte*
Greek + Latin	*barometer, dialect, geometer, perimeter*
Greek + French	*geometry, micrometer, thermometer, trigonometry, diameter*
Latin + Greek	*trigon*
Latin + Latin	*matrimony, patrimony*
Latin + French	*quadruped, solicism*

2.1.2. Affixes involved in derivational processes

In the previous section I have revised various matters concerning bases. I will proceed now to review units other than bases. The limits between neutral and non-neutral affixes seem straightforward. The former attach to bases with strong boundaries, while the latter attach to those with weak boundaries. Combining forms, however, share some of the features of both neutral and non-neutral affixes. Apart from attaching to other combining forms to coin nouns studied in affixation class IV above, they can also attach to free and bound bases to fall within the other three classes established in this chapter. Examples abound in the corpus and nouns are coined using this material on a regular basis. Later formations can be found without difficulty. The Greek initial combining form *geo-*, for instance, contributes with other final

combining forms to making up nouns such as *geoblast* and *geophyte* (coined in 1860 and 1900, respectively), and behaves as a prefix in cases such as *geomagnetism* and *geomathematics* (recorded in in 1938 and 1963, respectively). The same applies to final combining forms such as Greek -*logy*, found in nineteenth-century combinations such as *petrology* (1811). It also gives birth to other nouns, some of them nonce formations, whose bases are English, as is the case of *undergroundology* (1820) and *hatology* (1837), provided by the *OED*.

Notwithstanding all this, to attempt distinctions between prefixes and initial combining forms on the one hand, and suffixes and final combining forms on the other, may not be convenient at this point. And therefore, as I suggested in chapter 3, affixes and combining forms will be studied together. From this point onwards I shall refer to prefixes and initial combining forms as 'prefixes', and 'suffixes' will comprise both suffixes and final combining forms.

Tables listing the prefixes and suffixes found in the corpus are provided below. These tables classify those affixes found in the corpus in five groups regarding their origin, namely, Greek, Latin, Anglo-Norman, French and Old English. Each table shows four columns, the first of which lists the affixes under study; the second, the total number of types found in the corpus; the third and fourth indicate how many types have been found in *CETA* and *CEPhiT*, and the fifth contains examples taken from the corpus. Besides, at the bottom

of section three sets of dual figures specify the number of affixes and the total number of types (in brackets) per origin. It must be also observed that I have not evaluated the phonological qualities of affixes. Consequently, allomorphs such as *in*$_2$-, *im*$_2$- and *ir*$_2$- will be analysed individually.

It is noteworthy that similar bases could attach to different affixes from different origins depending on the period of the English history in which new nouns were coined. This is the case of the doublets -*ance* (French)/-*ancy* (Latin) and -*ence* (French)/-*ency* (Latin) forming abstract nouns. The data collected for my corpus under the variable 'coinage dates' suggest a higher frequency in the use of the French versions, and the appearance of the Latin versions at a later period, as can be observed in figure 4.13.

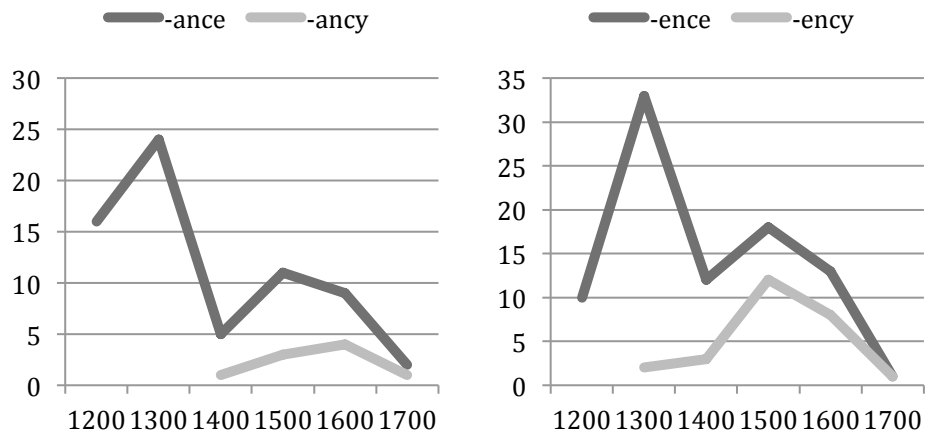


Figure 4.13. Doublets -*ance*/-*ancy* and -*ence*/-*ency*.

a. Prefixes

According to my data, the prefixes used in the scientific register of English come primarily from classical sources. Twenty-nine instances are derived from Greek, and produce 63 types; Latin-based prefixes add 27 elements to obtain 107 new coinages; Anglo-Norman and French provide only 2 and 3 prefixes each, which attach to other affixes/bases to coin 9 and 5 new noun respectively; and Old English contributes with 8 prefixes that make up 33 coinages. Figure 4.14 shows the absolute numbers for prefix origins in the corpus.

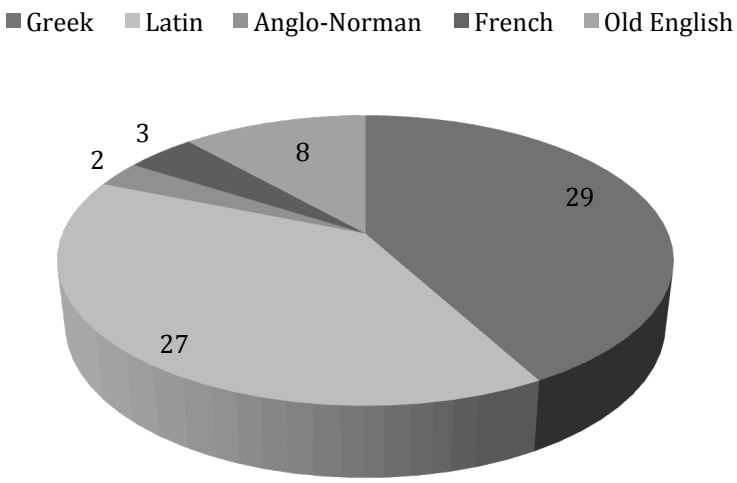


Figure 4.14. Prefixes and their origin.

Despite the fact that Greek provides English with a higher number of prefixes, those from Latin seem to be the most productive. Figure 4.15 below

displays the percentages of the types produced by the combination of those affixes. As regards noun tokens, Greek and Latin also predominate, summing up 85% of the nouns under scrutiny. This seems to emphasise the prevalence of classical sources in prefixation in scientific English.

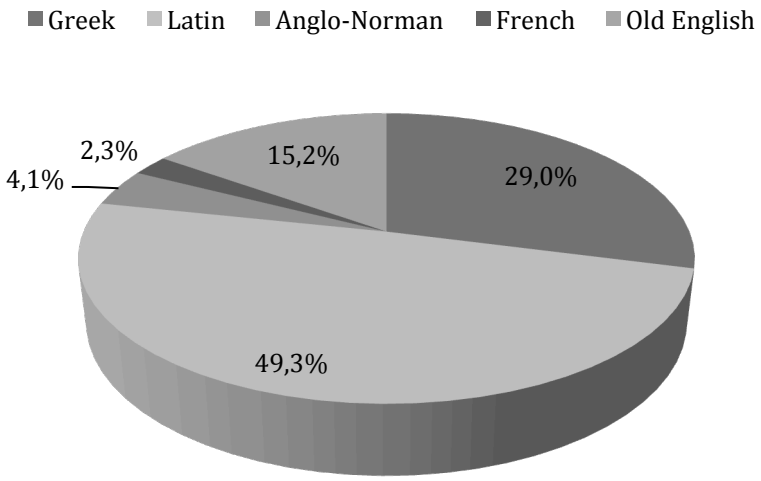


Figure 4.15. Percentage of prefixed nouns and their origins: types.

I have examined the diachronic influence of the above-mentioned origins in prefixation from the twelfth century onwards, and my findings show that if percentage figures within each century are considered, Latin prefixes have been producing new nouns steadily throughout the time span studied. Its predominance is such that in the eighteenth century three quarters of all the nouns coined by prefixation make use of Latin-based prefixes. Greek follows Latin in order of importance, and both French and Anglo-Norman influence is

almost inexistent. These percentages are best illustrated in figure 4.16, and table 4.4 contains a classification of prefixes by origin, The data included in this table —raw frequencies in the corpus and both subcorpora, and examples of each prefix found in the samples— are listed in descending order of frequency for use of interpretation.

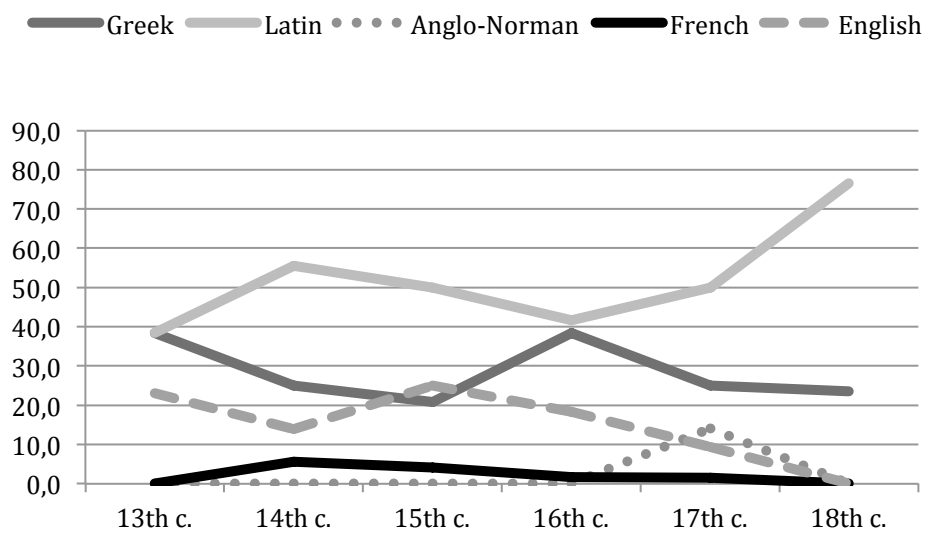


Figure 4.16. Diachrony: percentage of components in prefixed nouns.

Table 4.4. Prefixes.

Greek origin				
Prefixes	Corpus	*CETA*	*CEPhiT*	Examples
para-	7	2	6	*parallax, paronomasia*
dia-	6	3	5	*diameter, diaphragm*
peri-	5	4	1	*perigeo(-ee), perihelion*

hypo-	4	1	4	*hypocrisy*
anti-	3	1	3	*antidote, antithesis*
apo-	3	3	1	*apogee, apology*
arch(i)-	3	-	3	*archbishop, archetype*
astro-	3	3	2	*astronomy, astro-theology*
geo-	3	1	3	*geography, geometer*
hyper-	2	-	2	*hyperbola*
meta-	2	1	2	*metaphysic*
micro-	2	1	1	*microscope*
onto-	2	-	2	*ontology*
poly-	2	-	2	*polytheism*
sym-	2	-	2	*symmetry*
baro-	1	1	-	*barometer*
bio-	1	-	1	*biography*
cosmo-	1	1	-	*cosmography*
di$_2$-	1	1	-	*distich*
epi-	1	1	-	*epicycle*
gymno-	1	-	1	*gymnosophist*
helio-	1	1	-	*helioscope*
hemi-	1	1	-	*hemisphere*
metro-	1	1	-	*metropolis*
syn$_1$-	1	1	1	*synthesis*
tele-	1	1	-	*telescope*
thermo-	1	-	1	*thermometer*
trigono-	1	1	-	*trigonometry*
zoo-	1	-	1	*zoophyte*
Total	29 (63)	20 (30)	20 (44)	

Latin origin

Prefixes	Corpus	*CETA*	*CEPhiT*	Examples
dis-	22	2	22	*disadvantage, disconnection, disgrace*
in$_3$-	19	4	17	*inaction, inadvertence, incoherence*

re-	11	6	9	*reappearance, recollection, research*
semi-	7	7	1	*semi-orbit, semicircle, semi-tychonic*
co-	5	5	-	*codeclination, cotangent*
in$_2$-	4	-	4	*incumbrance, influx*
inter	4	1	4	*intercourse, intermarriage*
sub-	4	3	3	*subdivision, subzenith*
de-	3	1	3	*decrease, demerit*
equi-	3	2	2	*equilibrium, equinox*
pre-	3	1	3	*prepossession, prerogative*
super-	3	2	3	*superstructure*
im$_1$-	2	-	2	*impropriety*
matri-	2	-	2	*matrimony*
quadru-	2	1	1	*quadruped*
tri-	2	2	1	*triangle*
circum-	1	1	-	*circumjovial*
im$_2$-	1	-	1	*impatience*
ir$_2$-	1	-	1	*irreligion*
juxta-	1	-	1	*juxtaposition*
omni-	1	-	1	*omniscience*
patri-	1	-	1	*patrimony*
pene-	1	1	-	*penumbra*
pro$_1$-	1	-	1	*proverb*
subter-	1	1	-	*subterfuge*
uni-	1	1	-	*unicorn*
vice-	1	1	-	*vice-president*
Total	27 (107)	18 (42)	21 (83)	

Anglo-Norman origin

Prefixes	Corpus	*CETA*	*CEPhiT*	Examples
non-	6	1	5	*non-intelligence, nonsense*
counter-	3	-	3	*counter-motion, counterpart*
Total	2 (9)	1 (1)	2 (8)	

French origin

Prefixes	Corpus	*CETA*	*CEPhiT*	Examples
sur-	3	2	3	*surplus*
demi-	1	-	1	*demi-god*
mis$_2$-	1	-	1	*mischief*
Total	3 (5)	1 (2)	3 (5)	

Old English origin

Prefixes	Corpus	*CETA*	*CEPhiT*	Examples
mis$_1$-	9	3	7	*mis-spelling, miscarriage, misfortune*
fore-	8	4	5	*fore-father, fore-knowledge, foretast*
un$_1$-	6	1	6	*unbeliever, uncertainty, unfitness*
out-	5	1	5	*outcry, outrage*
over-	2	1	1	*over-measure*
off-	1	-	1	*off-set*
step-	1	-	1	*step-mother*
twi-	1	1	-	*twilight*
Total	8 (33)	6 (11)	7 (26)	

b. Suffixes

I have attested a total of 120 suffixes that produce 2,326 noun types. These can be considered from the perspective of their origin again, and classified as follows: Greek, 18 suffixes and 27 types; Latin, 32 suffixes and 933 types; Anglo-Norman, 7 suffixes and 162 types; French, 40 suffixes and 620 types; and Old English[37], 23 suffixes and 584 types. These absolute numbers can be graphically viewed in the following figure:

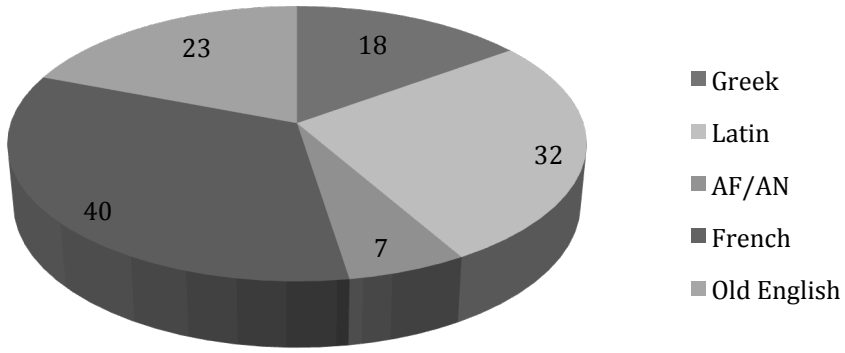


Figure 4.17. Suffixes and their origin.

The outlook of suffixation reveals some differentiating features from those typifying prefixation. In the first place, the former is much more frequent

[37] *Vid*. note attached to table 4.5.

than the latter. Indeed, if we exclude cases encompassed in affixation class IV (32 instances), the probability of coining a new noun by means of a suffix is over twelve times as high as it is when a prefix is involved. Besides, prefixes only represent one fourth of my total number of affixes. Additionally, the average productivity of a prefix is lower, contributing to four new types each, whereas suffixes form more than nineteen types each. Lastly, type-token ratios prove that nouns coined with a base + suffix pattern are present up to almost fifteen times in the samples (14.7), whereas those coined with a prefix + base pattern can be seen under nine times (8.96).

In the second place, there is a significant difference in the pattern of source languages used between prefixation and suffixation. If we take into consideration the findings disclosed above, as shown in figure 4.15, Greek-based materials and the amount of types formed with them are not predominant any longer. The presence of the Old English, French and Latin components is notable, and the number of types coined with these stand out, especially with the latter.

A diachronic study of the influence of source languages, illustrated in figure 4.16 below, shows that Latin-based suffixes experiment a steady increase from the twelfth century onwards. Their outbreak takes place between the fourteenth and fifteenth centuries and stays as a major source until the end of the late Modern period. French-based suffixes overcome Old English elements in the thirteenth century, about two hundred years after the Conquest,

and remain productive across the Middle and Modern English periods. Old English suffixes, though numerous and fairly productive throughout the whole history of the language, decrease in opposite relation to the rise of French and Latin. Anglo-Norman shows a uniform pattern, ranging between five and ten per cent in the time-span studied. Greek suffixes are not as productive as prefixes, contributing to between 0.3 and 3.9 per cent of the types coined, and experiencing a revival in eighteenth-century scientific texts. The following figure shows the diachronic lines of source languages in Middle and Modern English. Moreover, table 4.5 contains a classification of suffixes by origin, providing a list in descending order of frequency, raw frequencies in the corpus and subcorpora, and examples of each suffix contained in the samples.

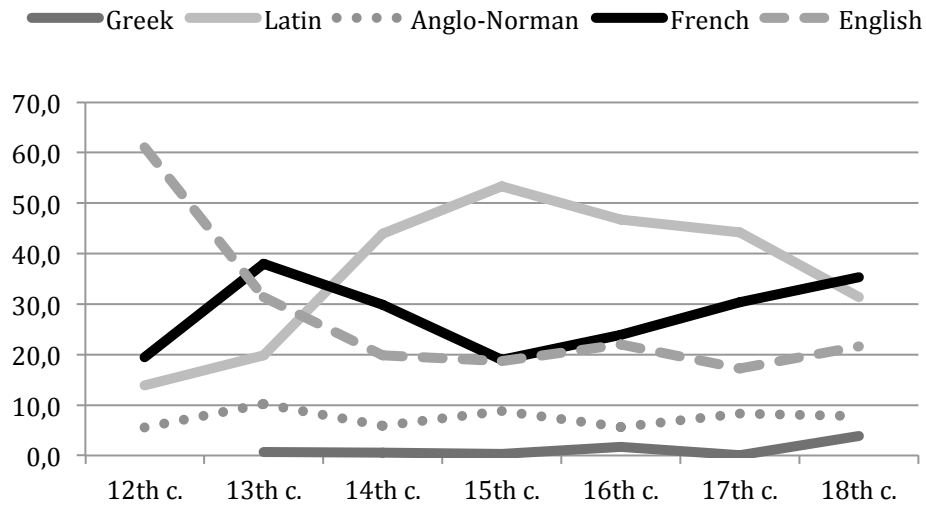


Figure 4.18. Diachrony: percentage of components in suffixed nouns.

Table 4.5. Suffixes.

Greek origin

Suffixes	Corpus	*CETA*	*CEPhiT*	Examples
-logy	5	2	5	*phraseology, ontology*
-sis	4	3	2	*analysis, apsis*
-graphy	3	2	2	*biography*
-ad$_1$	1	1	-	*Olympiad*
-cosm	1	-	1	*diacosm*
-enchyma	1	-	1	*parenchyma*
-ene	1	-	1	*gangrene*
-gon	1	1	-	*trigon*
-gram	1	1	-	*diagram*
-logue	1	1	1	*dialogue*
-nomy	1	1	1	*astronomy*
-ose$_2$	1	-	1	*tuberose*
-pathy	1	-	1	*sympathy*
-phragm(a)	1	-	1	*diaphragm*
-phyte	1	-	1	*zoophyte*
-polis	1	1	-	*metropolis*
-sophy	1	-	1	*ontosophy*
-sphere	1	1	1	*atmosphere*
Total	18 (27)	10 (14)	14 (20)	

Latin origin

Suffixes	Corpus	*CETA*	*CEPhiT*	Examples
-ion$_1$	413	211	359	*alleviation, cohesion, cultivation*
-ity	163	56	149	*ductility, electricity, irritability*
-ation	86	19	82	*colonization, pulsation, versification*
-or	64	33	51	*assertor, interlocutor, reflector*
-ency	26	5	26	*complacency, subserviency, tendency*
-an	24	15	15	*elean, plebeian, ptolomean*
-al$_1$	22	10	16	*disposal, revival, trial*

-tude	20	9	17	*attitude, infinitude, vicissitude*
-ate$_1$	18	7	16	*automate, disparate, novitiate*
-ian	17	6	16	*academician, necessarian, rhodian*
-ary$_1$	15	7	12	*boundary, emissary, religionary*
-acy	13	3	12	*accuracy, celibacy, intimacy*
-ancy	9	3	8	*ascendancy, infancy, preponderancy*
-ine$_1$	8	1	8	*intestine, machine, routine*
-mony	6	3	6	*acrimony, parsimony*
-ary$_2$	3	1	2	*luminary, salary*
-ile	3	2	1	*bissextile, projectile*
-meter$_1$	3	2	1	*barometer, perimeter*
-scope	3	2	1	*microscope, telescope*
-ar$_1$	2	2	1	*scholar*
-arian	2	1	1	*libertarian*
-ean	2	1	2	*epicurean*
-ite$_2$	2	-	2	*appetite*
-and	1	1	1	*command*
-ane	1	-	1	*membrane*
-ar$_2$	1	1	1	*pillar*
-ary$_3$	1	-	1	*janissary*
-et$_3$	1	1	1	*tenet*
-ication	1	-	1	*ramification*
-lect	1	-	1	*dialect*
-oid	1	1	-	*spheroid*
-ule	1	1	-	*globule*
Total	32 (933)	27 (404)	30 (811)	

Anglo-Norman origin

Suffixes	Corpus	*CETA*	*CEPhiT*	Examples
-ment	95	41	84	*agreement, enjoyment, firmament*
-our	23	10	21	*behaviour, governour, odour*
-y$_5$	16	10	14	*city, enquiry, majesty*
-ory	11	4	7	*observatory, repository, territory*
-y$_4$	10	5	8	*adultery, ministry, scrutiny*

-ee	5	-	5	*debauchee, repartee*
-er$_4$	2	1	1	*remainder*
Total	7 (162)	6 (71)	7 (140)	

French origin

Suffixes	Corpus	*CETA*	*CEPhiT*	Examples
-ence	87	38	83	*coincidence, deference, interference*
-y$_3$	76	31	70	*epilepsy, philanthropy, syzygy*
-ance	67	27	63	*hindrance, luxuriance, reluctance*
-ure	52	29	46	*conjuncture, failure, legislature*
-age	28	11	26	*marriage, patronage, usage*
-er$_2$	28	22	21	*commissioner, mariner, passenger*
-ist	27	4	25	*dramatist, machinist, optimist*
-tion	24	14	20	*causation, deviation, reproduction*
-ery	23	3	22	*debauchery, effrontery, machinery*
-ism	21	6	17	*athenianism, favouritism, patriotism*
-ice$_1$	20	8	18	*auspice, interstice, precipice*
-ent	18	12	15	*constituent, incident, percipient*
-et$_1$	17	8	11	*bullet, cabinet, puppet*
-ant	16	9	12	*disputant, octant, stimulant*
-cule	16	6	16	*animalcule, corpuscle, radicle*
-ic	13	2	12	*logic, republic, topic*
-el$_2$	9	7	6	*damsel, sequel, vessel*
-ard	8	5	6	*dotard, reward, standard*
-ess$_1$	8	1	7	*adulteress, enchantress, poetess*
-ition	8	4	7	*coalition, competition, transposition*
-ry	8	2	7	*bravery, pageantry, pleasantry*
-ise$_2$	6	2	6	*promise, treatise*
-ine$_4$	5	2	5	*concubine, jessamine*
-ive	5	2	4	*negative, perspective*
-cide	3	-	3	*homicide*
-ician	3	-	3	*arithmetician, dialectician*
-ish$_2$	3	1	2	*anguish, relish*
-ite$_1$	3	2	2	*areopagite, proselyte*

-meter$_2$	3	2	2	*micrometer, thermometer*
-metry	3	1	2	*trigonometry, symmetry*
-ade	2	-	2	*henriade*
-ier	2	-	2	*courtier, frontier*
-eer$_1$	1	-	1	*scrutineer*
-ine$_5$	1	-	1	*glycine*
-ison	1	1	1	*comparison*
-istic	1	-	1	*characteristic*
-ot	1	-	1	*zealot*
-ped	1	-	1	*quadruped*
-ture	1	-	1	*shelter*
-yer	1	-	1	*lawyer*
Total	40 (620)	29 (262)	40 (551)	

Old English origin

Suffixes	Corpus	*CETA*	*CEPhiT*	Examples
-ing$_1$	166	66	124	*blending, dialling, twirling*
-er$_1$	126	47	100	*biographer, improver, recliner*
-ness	117	46	93	*distinctness, fixedness, loftiness*
-ty$_1$	65	32	61	*certainty, frailty, poverty*
-le	37	25	27	*disciple, female, needle*
-th$_1$	19	15	18	*breadth, growth, mirth*
*-ics**	11	4	11	*economics, metaphysics, optics*
-ship	9	3	8	*courtship, rivalship, sonship*
-dom	5	3	5	*freedom, thraldom*
-y$_1$	5	1	5	*ivy, lofty*
-hood	4	1	3	*falsehood, neighbourhood*
-el$_1$	3	3	1	*feuel*
-ish$_1$	3	1	3	*foolish, rubbish*
-ar$_3$	2	-	2	*beggar*
-ful	2	1	2	*handful*
-ing$_3$	2	1	2	*farthing*
-red	2	1	2	*hatred*
-er$_3$	1	1	1	*manner*
-ess$_2$	1	1	1	*riches*

-ling	1	-	1	*starveling*
-ster	1	-	1	*lobster*
-y₂	1	1	1	*assembly*
-y₆	1	-	1	*fally*
Total	23 (584)	19 (253)	23 (473)	

* *-ics* is not an OE suffix. I have included it here because it is an English creation dating from the 16th century, applied originally to *mathematics* and *economics*, and then to most sciences that had so far taken endings in *-ic*

Deeper studies on each individual suffix may leave doors open to additional sub-classifications and the application of new variables. These new lines of analysis would, perhaps, increase the number of frequencies derived from them, and more specific theories would be then put forward. However, the limited extension of this work recommends the opposite. Consequently, I have focused only on the most productive suffixes coming from each component expounded above. Due to its low productivity, Greek has been excluded from the analysis, so this will cover Latin *-ion₁*, Anglo-Norman *-ment*, French *-ence*, and Old English *-ing₁*.

-ion₁

This is by far the most productive affix in the corpus, used to form 413 types. Stein (2007: 87) has defined it succinctly as a "suffix added to verbs to form nouns". In fact, most bases to which this affix attaches in the corpus are verbs, such as *regulation* (< *regulate*) and *impregnation* (< *impregnate*), but there are

also a few instances derived from adjectival bases, such as *aversion* (< *averse*) and *imperfection* (< *imperfect*), and very few nominal bases, such as *fruition* (< *fruit*).

The suffix *-ion* attaches exclusively to Latin, Anglo-Norman and French bases and the majority of its coinages fall into affixation class II. This circumstance might attest their condition as adaptations of foreign words. However, *-ion* is also the affix that presents the highest number of anaphoric formations, coined in the seventeenth century, and imitating patterns of Latin, though the bases from which they are derived did not exist in that classical language. The following examples display anaphoric constructions and their coinage dates:

(22) *Condescension* (< **condescensionem*), coined in 1642. Found in Hutcheson (1755:184), Astell (1700: 70) and Wollstonecraft (1792: 65).

(23) *emersion* (< **emersionem*), coined in 1633. Found in Hodgson (1749: 103, 104, 110, 111*).

(24) *exertion* (< **exertionem*), coined in 1668. Found in Balguy (1733: 40), Macaulay (1783: 2, 6, 37) and Bonnycastle (1786: 49), among others.

(25) *exhaustion* (< **exhaustionem*), coined in 1646. Found in Greene (1727: 3).

(26) *retrogression* (< **retrogressionem*), coined in 1646. Found in Adams (1777: 10, 26).

Coinages ending in *-ion* reached their peak in the 1300s and remained very productive until the eighteenth century in which, for uncertain reasons, the corpus contains no examples. Figure 4.19 shows the diachronic evolution in the productivity of this suffix.

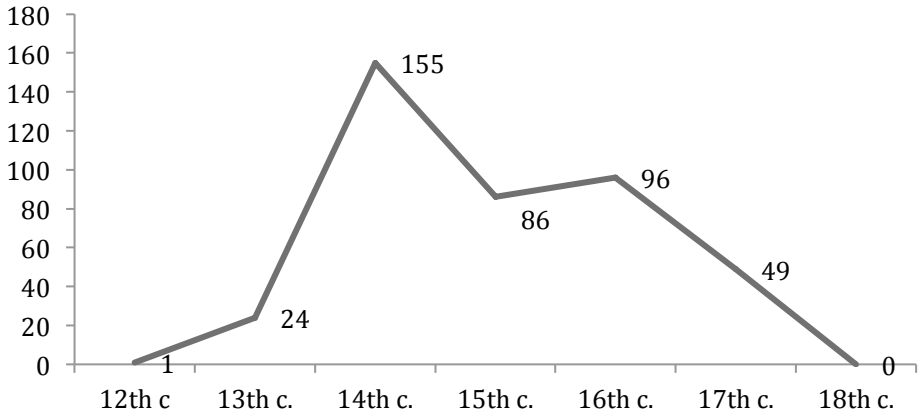


Figure 4.19. Diachronic productivity of *-ion*.

-ment

According to the *OED*, a suffix "Forming abstract nouns from verbs and (less commonly) from adjectives", as seen in *accompanyment* (< *accompany*), *development* (< *develop*) and *measurement* (< *measure*), all of them coined in

the eighteenth century. Only one instance of *-ment* attaching to an adjective is present in the corpus: *rudiment* (< *rude*).

This suffix attaches primarily to Anglo-Norman, French and Latin bases, as in *punishment* (< *punish*), *amendment* (< *amend*) and *inducement* (< *induce*), respectively. Exceptionally it attaches to some Old English verbs, as in and *acknowledgement* (< *acknowledge*) and *amazement* (< *amaze*), which represent two of the earliest instances of Romance affixes combining with Germanic bases —coined in 1567 and 1595. Two previous corpus studies on the productivity of this suffix derive in two contradictory theories: a) there are very few examples between 1150 and 1350 —Ciszek (2008: 123) found no instances of nouns in *-ment* between 1150 and 1350 in the *Helsinki Corpus*—, and b) its most productive peak is established between 1250 and 1350 (Dalton-Puffer 1996: 108). My corpus contains some evidence that may challenge both theories. Firstly, 16 instances coined between 1175 and 1330, as seen in table 4.6:

Table 4.6. Coinages with *-ment* between 1150-1350.

Type	Coinage date
sacrament	1175
judg(e)ment	1225
ornament	1230
com(m)andment	1250
commencement	1250
firmament	1250
instrument	1290

torment	1290
advancement	1297
amendment	1297
element	1300
parliament	1300
testament	1300
monument	1325
parchment	1325
concealment	1330
nourishment	1330

Secondly, this suffix peaked in the 1500s and continued to be moderately productive throughout the Modern English period. Figure 4.20 below shows the diachronic productivity of *-ment*. In the figure, the sixteenth century coincides with the highest peak in its use, increasing its impact on the language since the 1300s. Most coinages from the year 1500 onwards involve free bases —and 100% of the cases after 1600. The valley represented in the eighteenth century will be explained in the concluding remarks in chapter 5, in association with the general decrease in all the domains of affixation.

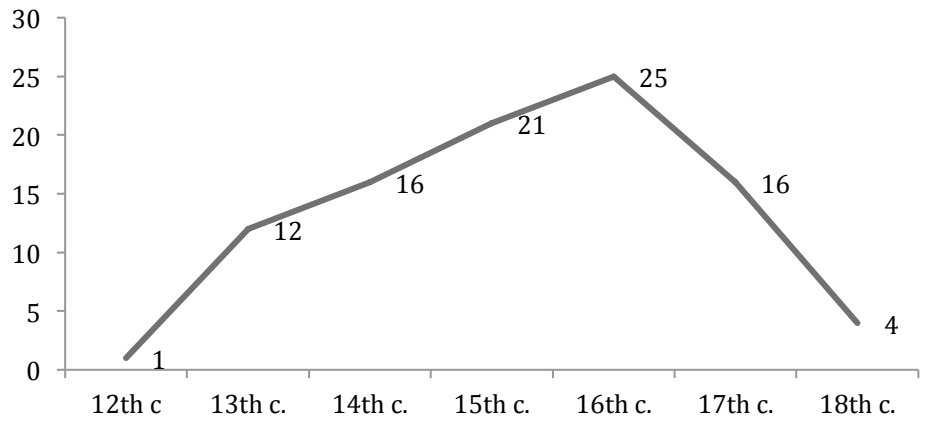


Figure 4.20. Diachronic productivity of *-ment*.

-ence

Stein (2007: 44) states that this suffix is "added to adjectives ending in *-ent* or *-escent* to form nouns", and indicates "the state, condition or quality denoted by the adjective". This means that it attaches only to Latinate bases, as attested in the samples. In my corpus *-ence* is mainly used in combination with deverbal adjectives, as in *abhorrence* (< *abhorrent*), *inconsistence* (< *inconsistent*), *recurrence* (< *recurrent*) and *subservience* (< *subservient*), all of them coined in the seventeenth century. Such a productive pattern may have produced nouns directly, without the intervention of an adjective, as in seventeenth-century *acquiescence* (< *acquiesce*) and *emergence* (< *emerge*) and eighteenth-century *interference* (< *interfere*). The case of *antecedence*, derived in theory from the nominal base *antecedent* poses a different problem, because we might be tempted to interpret the base as a conversion from the adjective.

Nevertheless, the situation seems to be the contrary, since the first written evidence of the noun was attested in 1393, whereas the first use of the adjective seems to have occurred in 1543 (*OED*).

The diachronic productivity of -*ence* can be seen in the following figure. No instances of its use have been attested in the corpus in the twelfth century, though it became fairly productive in the 1300s. A noticeable decrease is observed after the fourteenth century, which remains relatively steady until the end of the Modern English period. The cause for this decrease may spring from the competence of the Latin suffix -*ency*, introduced in the language in the 1400s, which I already mentioned in section 2.1.2 above.

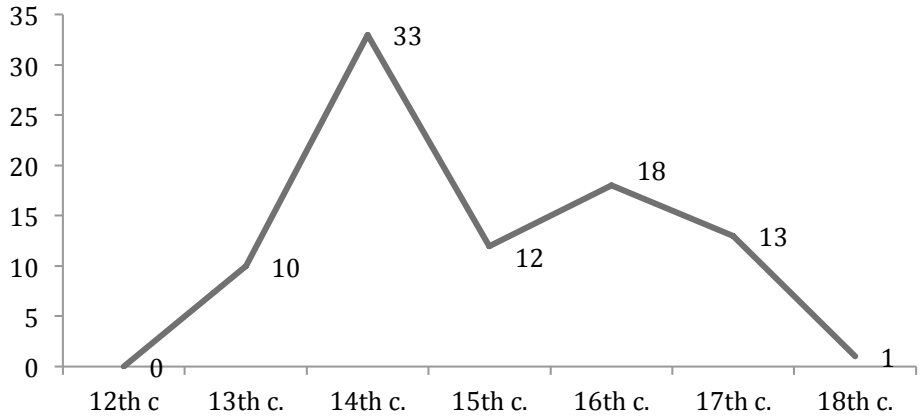


Figure 4.21. Diachronic productivity of -*ence*.

-ing

According to Beard (1995: 196) *-ing* is "the most productive suffix in English, marking both inflectional and derivational derivations". Thus, it can form syntactic and lexical derivations alike by attaching to any verb, producing present participles, adjectives and nouns. This presents great difficulty for research works using untagged corpora containing several hundred thousands words because, as some authors acknowledge (Banks 2012), a lot of time-consuming disambiguating must be carried out by hand. In my case, instances were excluded for two chief reasons: 1) because *-ing* was not an affix in some words with that ending, as is the case of *ring, thing* and *wing* for example, and 2) because the resulting word was not a noun.

A considerable amount of literature has been published since the 1960s on gerundive nominalisations to identify their typology, and the similarities and distinctions observed in their potential classification. Consider, for example, Lees (1969), Chomsky (1970), Jackendoff (1977), Abney (1987), Blevins (1994) and Baker (2003), to name but a few. More recently there have been attempts to systematise their nominal properties from functional and semantic perspectives (Heyvaert 2003, 2008, 2010). But it is not my intention to contribute to this theoretical discussion, because my focus is on the general status of morphology in the eighteenth century, rather than on an extensive study of particular affixes and their idiosyncrasies. Therefore I will contemplate Chomsky's approach to the differences between gerundive and

derived nominal in examples such as *John's being eager to please* and *John's eagerness to please*. He (1970: 187) points out that,

> Many differences have been noted between these two types of nominalization. The most striking differences have to do with the productivity of the process in question, the generality of the relation between the nominal and the associated proposition, and the internal structure of the nominal phrase.
>
> Gerundive nominals can be formed fairly freely from propositions of subject-predicate form, and the relation of meaning between the nominal and the proposition is quite regular. Furthermore the nominal does not have the internal structure of a noun phrase; thus we cannot replace John's by any determiner […], nor can we insert adjectives into the gerundive nominal.

And I will adopt a compromise decision regarding the constructions to be included here, in order to restrict my study to only those *-ing* nouns susceptible to be substituted by other nouns, including zero-derived and affixed nouns (see examples below and proposed substitutes in square brackets). Besides, these coincide with the ones listed as independent entries in the *OED*. Some may regard this solution simplistic, because it excludes examples in which the *-ing* form functions as head of noun phrase (27) or prepositional phrase (28)

followed by a complement, be it another noun phrase or an object pronoun, such as:

(27) and that juſtifies my **calling** [**call*] this Project, The double Courtſhip (Dunton 1710: 4).

(28) A man that is in danger of **confounding** [**confusion*] them, is indeed to be pitied (Reid 1764: 55).

Apart from this, I have found no instances of the pattern N(-*ing*) + V + Complement in the samples, such as *Annotating is indispensable* and *Disambiguating needs time*, so there is no need for further comments on their inclusion or exclusion at this point. However, they would be excluded for the reasons explained above. Notwithstanding the loss of some admittedly interesting cases for discussion, my compromise solution still leaves 1,119 tokens manually catalogued belonging to 166 types, which seem reasonable numbers that may illustrate the significance of the suffix. I will undertake a deeper study of other N(-*ing*) syntactic constructions in future research works.

The following examples show instances of -*ing* nouns in the corpus. As we can see, highlighted nouns can be replaced with those in square brackets, without affecting the syntactic structure of the sentences.

(29) As the Height of the Sun at Noon is called its Meridian Altitude, or its **Culminating** [*culmination*], ſo the Height of the Sun in the Eaſt or Weſt is call'd its Vertical Altitude (Watts 1726: 37).

(30) If a fact be certain, there is no **reaſoning** [*reason*] againſt it (Turnbull 1740: 16).

(31) The ſubject of morality has been greatly perplexed by the **blending** [*blend*] of theſe queſtions together (Ferguson 1769: 109).

To conclude this discussion, the following example illustrates the common practice of capitalising nouns in the seventeenth and eighteenth centuries. The author acknowledges the different syntactic functions of *duckings* and *crossing*, and writes the noun with a capital letter.

(32) and I have read of ſtrange Ceremonies and **Duckings** [*ducks*], which they make young Navigators undergo, at the firſt Time of their **croſſing** [**cross*] the Equator (Harris 1719: 35).

In contrast with the previous three suffixes, *-ing* attaches to Latinate or Germanic bases alike. The following examples illustrate formations coined using materials with Latin, Anglo-Norman, French and English components, respectively.

(33) But when through the fault of either ſide the eſſential ends of this relation are defeated, *viz*. the **procreating** and **educating**[38] of offspring, and a friendly ſociety for life (Hutcheson 1755: 177).

(34) If there was any real Repugnance, or Inconſiſtence between them; ſuch a **Proceeding** might naturally be expected (Balguy 1733: 2).

(35) For when we determine a thing to be probably true, ſuppoſe that aEvent has or will come to paſs, 'tis from the Mind's **remarking** in it a Likeneſs to ſome other Event, which we have obſerved has come to paſs (Butler 1736: ij).

(36) then let the Angles ACF, AFC, made by the **meeting** of the Semidiameters of the Earth with the Line connecting the Tops of the Mountains… (Whiston 1715: 8).

The diachronic productivity of *-ing* varies significantly from century to century in the time-span studied. As happens to the other three suffixes outlined here, with the exception of *-ence*, it peaked in the fourteenth and sixteenth centuries, and declined after them. Its diagram, depicted in figure 4.22, follows a trend that encompasses all affixation in general, rather than describing a tendency inherent to this suffix individually.

[38] These two nouns are not listed in the *OED*, so they might be nonce-formations.

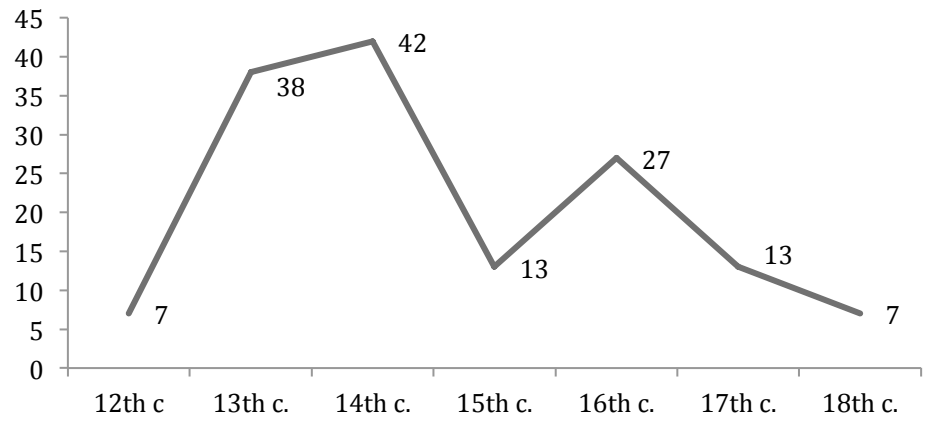


Figure 4.22. Diachronic productivity of *-ing*.

2.1.3. Cross-origin affixation

To conclude this review of affixation I will provide figures on the interconnecting relationships among linguistic elements and their origin components. In the following table the first row shows the origin of affixes, first column on the left indicates the origin of bases, and the rest of the cells contain quantitative information on how often affixes and bases from one origin combine with another. As we can see, certain patterns are much more productive than others.

Table 4.7. Cross-origin associations.

Base/Affix	Greek	Latin	AN	French	English
Greek	5	8	-	3	-
Latin	17	384	108	434	15
Anglo-Norman	2	53	21	106	10
French	15	180	124	339	29
English	1	62	61	178	290

The pattern Latin base + French affix is the most frequent in the corpus, with 434 noun types. In fact, nominal combinations of these two components, Latin (base) + French (base) of the type *animalcule*, *quadrature*, and French (base) + Latin (affix) such as *reprizal*, *spheroid,* make up the majority of affixed nouns in English (54.7 per cent). English bases combine with Latinate affixes on a regular basis, producing examples such as *behaviour*, *burial* and *hindrance*, and vice versa, as in *measuring*, *nobleness*, *recorder* and *rivalship*. Greek affixes attach to Greek, Latin and French bases, but no instances have been found of Greek + Anglo-Norman or Greek + English combinations.

My calculations allow measuring the probability of a particular component in cross-origin affixation. That is to say, how likely a component is to combinations with other components as opposed to combinations with itself. Figure 4.23 illustrates those percentages, and clearly reveals that English materials are very prone to combine among them to produce new nouns. On the other side of the spectrum, Anglo-Norman and Greek elements tend to

attach to materials from other origins. The rates of Latin and French are more balanced.

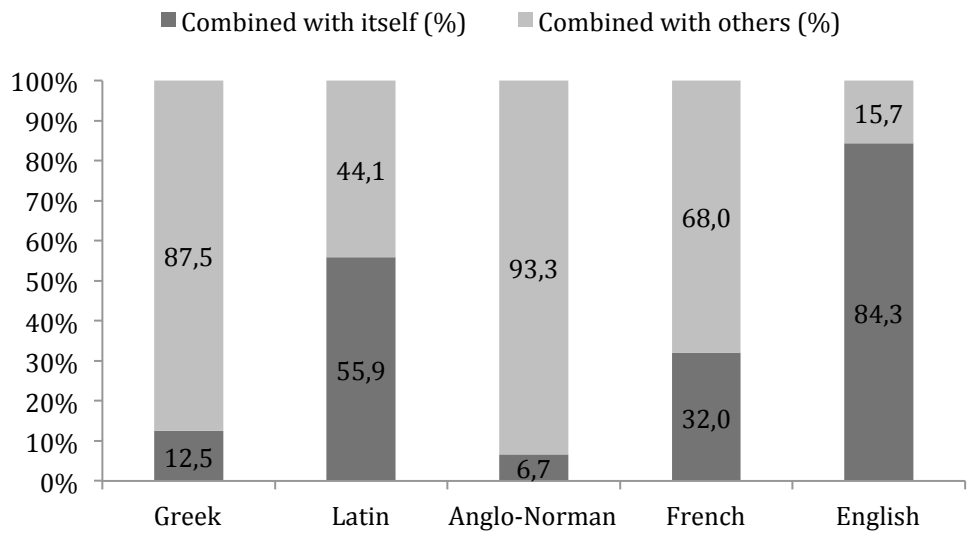


Figure 4.23. Cross-origin affixation.

2.2. Compounding

Compounding has been a productive device to coin new nouns since Old English times, and according to my data it has increased steadily ever since. This study was not intended to deal with compounds in depth, but considering that compounds "occur most abundantly as tools of terminology in scientific fields which explore tentative and mysterious phenomena, such as astronomy and particle physics" (Raad 1989: 133), I will give a brief account of those found in our corpus. We must not forget, however, what I stated above

concerning the samples included in the *Coruña Corpus*: they have not been formatted or analysed for ease of retrieval. Consequently I have only taken into account those compounds that can be implied from the token lists generated by the *Coruña Corpus Tool*, namely, those which show what Mikheev (2003: 208) calls 'true-hyphens', that is to say, lexical hyphens linking two free bases. Any other analysis is only possible if we assume the adjectival nature of many nouns listed here, which would be modifying other nouns. Apart from this, some authors included in my analysis display singular interpretations in the use of compounds, which suggest a concept of compound somewhat slacker than in our days. Examples that include -*'s* as part of the compound, such as such *New-Year's-Day* (Fuller 1732: 17)*, sun's-beam*[39] (Bryan 1797: 107) and *woman's-man* (Harris 1719: 31), add even more controversy to my analysis.

Bearing these issues in mind, 34 out of the 65 examples dating from the eighteenth century can be considered nonce formations, since they are not listed as entries in the *OED*. A total of 271 compound nouns have been found, and their historical distribution can be observed in Figure 4.24 below. Regarding base categories, Millward (1996) and Nevalainen (2006a) agree the most productive type of late Modern English compound noun was noun + noun, as was the case in ME. This can be observed in our data too, with 84

[39] *Sun-beam* was coined in the eleventh century, but there are no instances of *sun's-beam* as a compound noun in the *OED*.

instances. Another eight different combinations used to create new nouns have been identified: adjective + noun (33 types), noun + verb (3), preposition + noun (2), verb + noun (1), proper noun or name + noun (1), noun + adjective (1), adverb + noun (1), and adverb + verb (1). Table 9 provides a list of compounds regarding the categories of their bases, with figures on their productivity and some examples. Meanwhile, Figure 4.24 shows the evolution of compounding in the corpus.

Table 4.8. Base categories involved in compounding.

Base category	Types	Tokens	Examples
adjective + noun	79	196	*back-stairs, prime-vertical, self-approbation*
adverb + noun	4	10	*insight, welfare, well-doing*
adverb + verb	2	54	*offspring, inlet*
noun + adjective	3	4	*sum-total, captain-general, aurora borealis*
noun + noun	176	757	*help-mate, marriage-contract, slave-trade*
name + noun	1	1	*London-measure*
prep. + noun	3	19	*behalf, by-stander*
pronoun + noun	1	2	*she-wit*
verb + noun	2	4	*brimstone, pastime*

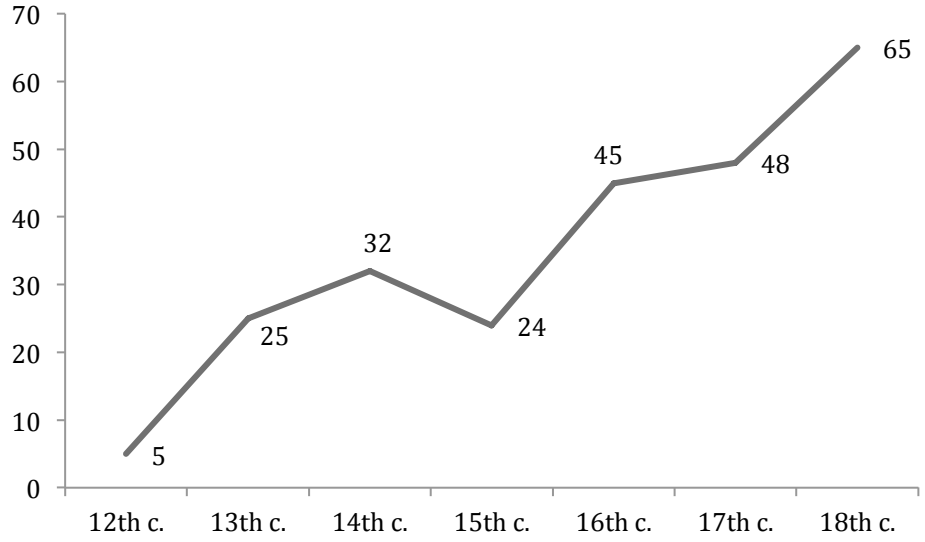


Figure 4.24. Evolution of compounding as seen in the corpus.

2.3. Conversion

In the present study, where affixes have been at the heart of my interest, those productive word-formation processes that do not include them have been analysed and quantified, but not included or computed within affixation.

Conversion is one of the most productive means of word formation in present-day English, and several parts of speech can be shifted from one to another rather freely. In the Middle and Modern English periods, however, conversion was more restricted, and notwithstanding that nouns and verbs were regularly used as sources of new converted elements, other word categories were not as free as they seem nowadays. Marchand (1969) claims that noun-to-verb and verb-to-noun conversions dominated the period, though adjectives

were used at times, especially in the coining of new verbs. Millward (1996) sees a direct connection between the increasing use of conversion and the loss of inflectional endings in Middle English. Regarding adjective-to-noun conversions, Kastovsky (2005: 33) suggests the need to explore them from a syntactic rather than from a morphological point of view, by stating:

> Many of these instances can be treated as ellipses of a noun […], and some of these have undergone lexicalisation, but they usually do not adopt the properties of genuine derivatives. Consequently, they are indeed better regarded as representing a basically syntactic, and not a morphological phenomenon. If such formations become entrenched in the lexicon, it is a consequence of their having become lexicalised/ institutionalised, but not the result of any word-formation process.

However, it is highly complex to measure the degree of lexicalisation of some of these conversions retrospectively. Besides, some of them can be found in the plural (*curve*, *perpendicular*, *representative*, *universal*, etcetera). Also, some dictionaries lack consistency in regarding them as nouns or not. So I have considered them as partial conversions, as explained in chapter 2, and computed them in my analysis, given that their figures do not affect the main body of affixation.

I have found 376 converted nouns in the corpus, of which 239 are adjective-to-noun and 124 verb-to-noun conversions. The rest of instances

come from adverbs (4), participial adjectives (4), proper nouns (3), conjunctions (1) and interjections (1). Forty of these constructions were coined in the eighteenth century, and most of them have adjectival (34) vs. verbal (4) bases, results that seem to contradict Marchand's words above. The following table illustrates this morphological process with examples extracted from the corpus:

Table 4.9. Base categories involved in conversion.

Base category	Types	Examples
adjective	239	*fluid, tangent, trine, tychonick, universal, exponent*
adverb	4	*hereafter, wherewithal*
conjunction	1	*but*
interjection	1	*check*
name	3	*dunce, epicure, vernier*
particip. adjective	4	*attribute, damned, vanquished, governed*
verb	124	*amount, attack, contest, dispute, laugh, struggle*

Graph 4.25 shows the evolution of conversion in Middle and Modern English. My findings confirm that conversion increased its productivity in the modern period with respect to Middle English in absolute numbers, and reaches its highest peak in the sixteenth century. We have seen in previous graphs that most processes begin a constant decrease in the 1600s that continues in the 1700s.

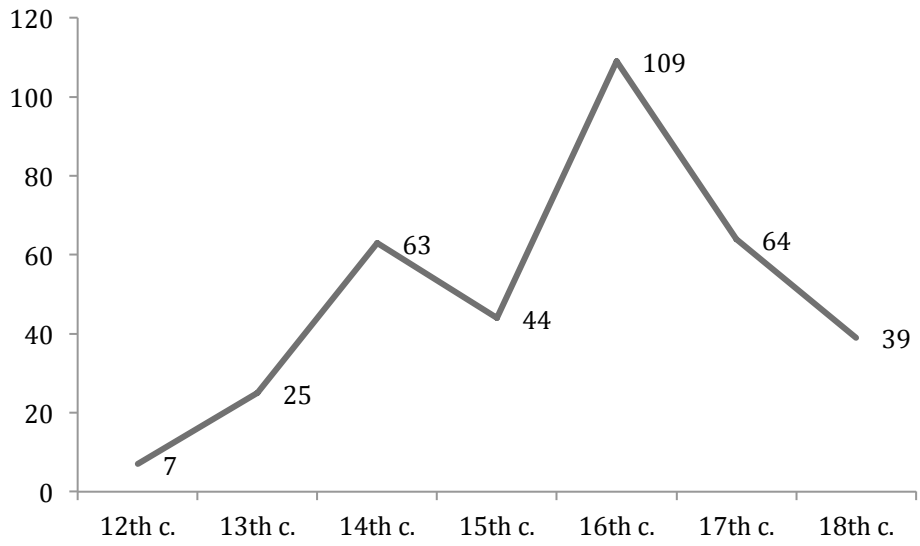


Figure 4.25. Evolution of conversion as seen in the corpus.

As far as bases are concerned, the significance of Old English materials in conversion suggests a significant difference to other affixing processes, where Romance bases are much more numerous than Germanic ones. Here, combined numbers for Old English and other Germanic languages —labelled 'other'— are comparable to that of affixation class I, as can be observed in figure 4.26.

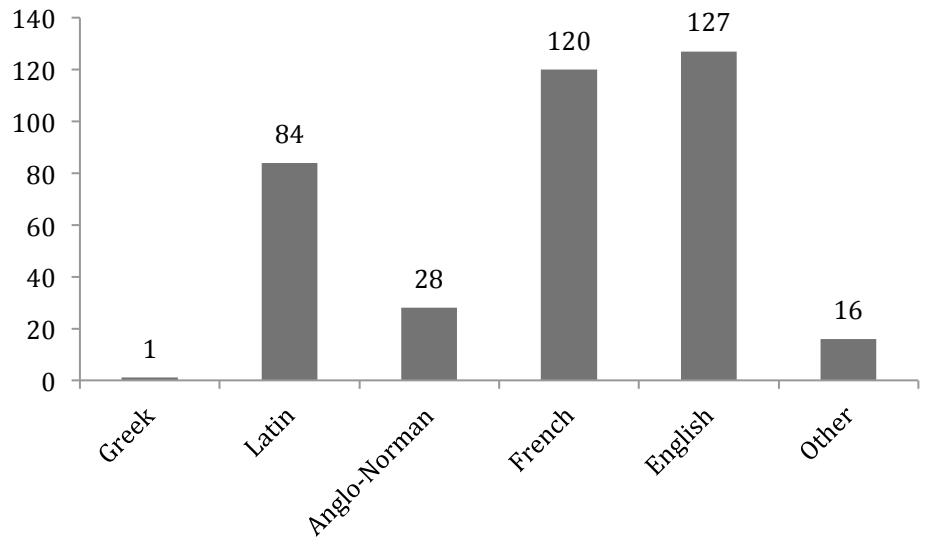


Figure 4.26. Components in conversion.

2.4. Other processes

In my data twenty-eight instances cannot be classified under any of the morphological processes described in the previous sections. Alongside onomatopoeic endeavours to coin new nouns I have identified some others that result from errors in the interpretation of foreign words, backformations, inventions and analogical formations. Moreover, I have found cases in which linguistic material, not only morphological but also phonological, is either transposed —metathesis— or removed at the beginning of the noun —aphaeresis or back clipping—, in the middle —syncope or middle clipping—, or at the end —apocopation or front clipping— of words for several reasons.

These can appear in combination with compounding as well. The following nouns exemplify these processes.

a. Onomatopoeia: *rattle*.
b. Error: *junto* (< Spanish *junta*).
c. Backformations: *fog* (< *foggy*, adj.), *wrinkle* (< *wrinkled*, adj.).
d. Analogy: abode (< *bode*, by association with the verb abide).
e. Metathesis: *dirt* (< *drit*).
f. Clipping:
    i. Aphaeresis: *censer* (< *incenser*), *chymist* (< *alchymist*), *crew* (< *accrewe = accrue*), *dropsy* (< *hydropsy*), *stress* (< distress) *venture* (< *adventure*), *drawing-room* (< *withdrawing + room*).
    ii. Syncope: *fancy* (< *fantasy*), *governess* (< *governeress*), *Whig* (< whiggamore).
    iii. Apocopation: *maid* (< *maiden*), pun (< *punctilio*, n), *chickweed* (< *chicken* + *weed*).

However, all the devices and examples taken together represent only a marginal percentage of all the instances of word formation in the history of English in general and late Modern English in particular.

2.5. Language changes throughout the eighteenth century

In one of her chapters on *CETA* Moskowich (2012) argues that the compilation criteria underlying the *Coruña Corpus* are such that they allow the analysis of language in 30-year periods. This is so because the texts incorporated to the corpus are mostly first editions, therefore the date of publication is not far from that of writing the texts —though this is arguable, especially in the case of collections of lectures, letters, etcetera. Obviously, a demonstrable proximity between writing and publishing dates is required. She continues by saying that "where first editions were not possible, we have chosen editions published within thirty years of the work's initial publication" were selected. These criteria are founded on Kytö, Rudanko and Smittenberg's theory (2000: 92) that language change can be observed within 30-year periods.

Following this assumption, I have decided to exploit this feature embedded in the corpus as a means to quantify processes and coinages throughout the eighteenth century. I will establish three blocks or sub-periods of roughly thirty years each and examine them one by one. The distribution of the texts in *CETA* and *CEPhiT* persuaded me to set the dividing lines between the years 1732-1733 and 1765-1766, that is, thirty-three years in the first and second sub-periods, and thirty-two in the third —because the latest sample in the century dates from 1797— in order to include a balanced number of words and samples in every block. It must be noted that some coinage dates extracted

from the *OED* were replaced by my own dates, because I have found earlier uses of those nouns in my corpus. My measurements are shown in table 4.10.

Table 4.10. Processes by 33-year periods.

	1700-1732	1733-1765	1766-1797
Affixation	32	20	14
Compounding	34	17	14
Conversion	17	5	19

There are several points of interest within the findings themselves. As we can see, a total sixty-six new nouns were coined by means of affixation in the eighteenth century,[40] which follows the decreasing tendency of the previous century, falling from thirty-two coinages in the first period to seventeen in the third. Compounding virtually matches the numbers given for affixation in overall numbers for the whole century. It must be said that nonce-formations play a very significant role in compounding, and half the cases found are not recorded in the *OED*. Zero derived nouns outnumber both affixations and compounds in the third sub-period, and revert the decreasing trend of the other

[40] This number differs with that of figure 4.4 above by one noun. This is no mistake, but the product of two different criteria. The *Coruña Corpus* considers the year 1700 as part of the eighteenth century, so computation for the eighteenth century includes this year also and, as a consequence, one more type as well. However, for all other computations the customary division between centuries is kept (1101-1200, 1201-1300, 1301-1400, etcetera).

two processes after the first third of the century. Focusing only on affixation, the following table contains a frequency list, produced in alphabetical form, of the noun-forming affixes in each of the sub-periods described above. I have quantified the types that each affix produce and offer examples of coinages.

Table 4.11. Affixation in the eighteenth century.

	Prefixation	Suffixation	1700-32	1733-65	1766-97	Examples
-acy		x	-	1	-	| - | *inaccuracy* | - |
-ade		x	-	1	-	| - | *henriade* | - |
-age		x	1	-	-	*assemblage* | - | - |
-an		x	-	1	-	| - | *elean* | - |
-ance		x	2	-	-	*disappearance* | - | - |
-ancy		x	1	-	-	*ascendancy* | - | - |
-ant		x	1	-	-	*stimulant* | - | - |
-arian		x	-	-	1	| - | - | *libertarian*
astro-	x		1	-	-	*astro-theology* | - | - |
-ation		x	1	-	1	*civilization* | - | *colonization*
co-	x		2	-	1	*codeclination* | - | *colatitude*
-cosm	x	x	1	-	-	*diacosm* | - | - |
dia-	x		1	-	-	*diacosm* | - | - |
dis-	x		-	1	-	| - | *disconnection* | - |
-ence		x	-	1	-	| - | *interference* | - |
-ency		x	-	-	1	| - | - | *ascendency*
-er$_1$		x	1	-	-	*biographer* | - | - |
-ery		x	2	1	-	*prudery* | - |
-ian		x	-	1	1	| - | *academician* | *necessarian*
-ics		x	-	-	1	| - | - | *empirics*
in-$_3$	x		2	-	1	*inaction* | - | *inebriety*
-ine$_5$		x	-	-	1	| - | - | *glycine*

-ing$_1$		x	2	4	1	*meeting* | *classing* | *blending*
-ion$_1$		x	1	-	-	*cultivation*
-ism		x	4	1	-	*patriotism* | *favouritism* | - |
-ist		x	1	-	1	*machinist* | - | *optimist*
-ity		x	1	3	1	*intrepidity* | *irascibility* | *respectability*
-ment		x	1	3	-	*arrangement* | *development* | - |
-meter$_2$		x	1	-	-	*micrometer* | - | - |
micro-	x		1	-	-	*micrometer* | - | - |
-ness		x	1	-	-	*interestedness* | - | - |
-onto		x	1	-	-	*ontosophy* | - | - |
-or		x	2	-	-	*vociferator* | - | - |
semi-	x		1	2	2	*semiduration* | *semiaxis* | *semi-orbit*
-sophy		x	1	-	-	*ontosophy* | - | - |
sub-	x		1	-	-	*subzenith* | - | - |
-y$_6$		x	-	-	1	| - | - | *fally*
Total affixes			26	12	13	
Total types			32*	20	14	

* In fact, this sum equals 35, but six affixes combine in pairs and only occur once, so *diacosm*, *micrometer* and *ontosophy* are doubled.

As it stands, then, only 37 affixes take part in new formations throughout the century, 26 in the first sub-period, which produce 32 new types, 12 in the second, forming 20 types, and 13 in the third, amounting to 14 types. The most productive affixes are *-ing* (7 types), *semi-*, *-ity* and *-ism* (5 types each), and *-ment* (4 types). Comparing these figures with those from previous centuries, the lower productivity of *-ion* stands out (only 1 type, in the first sub-period).

3. Other variables under study

As explained in the introduction, this section describes an evaluation of the forty-one samples in the eighteenth-century sections of *CETA* and *CEPhiT* by means of five extralinguistic variables. By analysing the data under these parameters I attempt to observe patterns on the following topics: a) in the first place, astronomy and philosophy will be compared and contrasted, in order to determine, if possible, unique characteristics inherent to each discipline; b) the second variable focuses on the different genres/text-types integrating the corpus, so as to find distinctive peculiarities in language use among them; c) the third variable deals with a contrastive review of the sex of the authors, and attempts at finding similarities and/or differences in the writings of men and women scientists; d) comparing the samples from the fourth and fifth standpoints will provide preliminary conclusions on linguistic use depending on the places where authors were educated, and e) on their ages at the time of writing these texts. These analyses will tackle the elements explained in section 2, namely, variety of types, types-per-token ratios, word-formation processes, affixation classes, base/affix origin and categories, etcetera. All the patterns established below are exclusively related to the data obtained from this corpus, and consequently arguable outside this context. Their validity must be checked against other scientific and non-scientific corpora, and corpora dating from different periods in the history of the English language.

3.1. Disciplines and morphological processes

In chapter 2 we saw that the eighteenth-century sections of *CETA* and *CEPhiT* are made up by a similar number of words (208,079 and 200,902, respectively). Also, at the first stage of the analysis absolute figures representing the total noun tokens in both disciplines are still amenable to direct comparison (38,121 in *CETA* vs. 41,195 in *CEPhiT*). From these figures we can infer that the number of nouns in the philosophy subcorpus is 3.8 per cent higher than that of astronomy. Although, we may concur that this preliminary conclusion does not seem overtly significant at this point, except for the sake of declaring that philosophers make use of a marginally higher number of nouns than astronomers. Categorising these tokens within their respective noun types may nonetheless clarify things further. By doing so, the 79,316 noun tokens contained in the whole corpus were classified within 4,558 different noun types. The subsequent division and quantification of the types occurring in each discipline individually show the following results: 2,060 belong to *CETA* and 3,777 to *CEPhiT*, while both subcorpora share 1,278 of those noun types. These results are rather surprising. Such a substantial contrast between both subcorpora could not be easily devised before all the data were quantified. The distribution of tokens and types in both disciplines is illustrated in figure 4.27.

As we shall see later, when more elements of comparison and contrast are taken into consideration, the divergence between *CETA* and *CEPhiT* will rise, especially when dealing with affixed nouns.

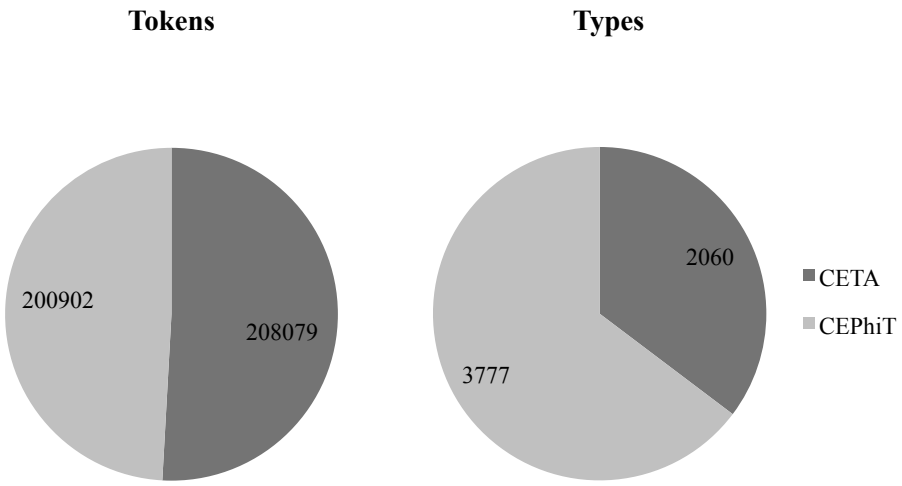


Figure 4.27. Noun tokens and types in *CETA* and *CEPhiT*.

The first of those elements of comparison and contrast involves quantifying the word-formation processes in both disciplines. I have carried out the same classification used in section 2, which establishes five groups of nouns: i) simple, ii) complex, iii) compound, iv) zero-derivation, and v) others not belonging to any of the previous groups. Details concerning noun-forming processes used in both disciplines are given in table 4.12.

Table 4.12. Processes in *CETA* and *CEPhiT*.

CETA					
	Simple	Complex	Compound	Zero	Other
Tokens	21,764	14,606	366	1,374	11
Types	737	1,074	127	114	8
% Tokens	57%	38%	1%	4%	< 1%
% Types	36%	52%	6%	6%	< 1%
Type-token ratio	29.5	13.6	2.9	12.1	1.4
Tokens/1,000 words	104.6	70.2	1.8	6.6	0.1
Hapax legomena	115	197	67	19	1
Hapax/1,000 words	0.6	0.9	0.3	0.1	< 0.1
CEPhiT					
	Simple	Complex	Compound	Zero	Other
Tokens	17,541	21,212	692	1,691	59
Types	1,120	2,138	165	328	26
% Tokens	43%	51%	2%	4%	< 1%
% Types	30%	57%	4%	9%	1%
Type-token ratio	15.7	9.9	4.2	5.2	2.3
Tokens/1,000 words	87.3	105.6	3.4	8.4	0.3
Hapax legomena	296	656	99	156	14
Hapax/1,000 words	1.5	3.3	0.5	0.8	0.1

These results reveal a series of interesting issues that require explanations at this point. In the first place, the relationship between simple and complex tokens is reversed in both disciplines. That is to say, *CETA* shows a higher number of simple than complex tokens (104.6 vs 70.2 per 1,000 words), whereas *CEPhiT* contains more complex than simple tokens (105.6 vs 87.3 per 1,000 words). However, this is not applicable when analysing types. Here, the

relationship in both disciplines is more balanced, i.e. complex types are predominant in the astronomy and philosophy samples. Although percentages differ by a significant margin, being slightly higher in *CEPhiT* (58%) than in the *CETA* (52%).

A general overview suggests that the philosophy subcorpus outnumbers the astronomy subcorpus without exception as far as nouns in all processes are concerned. Nevertheless, a closer look shows that differences in the use of types between *CETA* and *CEPhiT* vary appreciably from one process to another. For example, percentages show that simple and compound types have a greater influence in astronomy than in philosophy, whereas philosophers prefer complex, zero-derived and 'other' nouns to a greater degree than astronomers. It is in the case of simple types that the greatest difference is found (≈6%), whereas those labelled 'other' show the narrowest margin (< 1%). Figure 4.28 illustrates the differential coefficients of types in *CETA* and *CEPhiT*. Positive numbers stand for higher percentages in the astronomy subcorpus and negative numbers indicate higher percentages in the philosophy corpus.

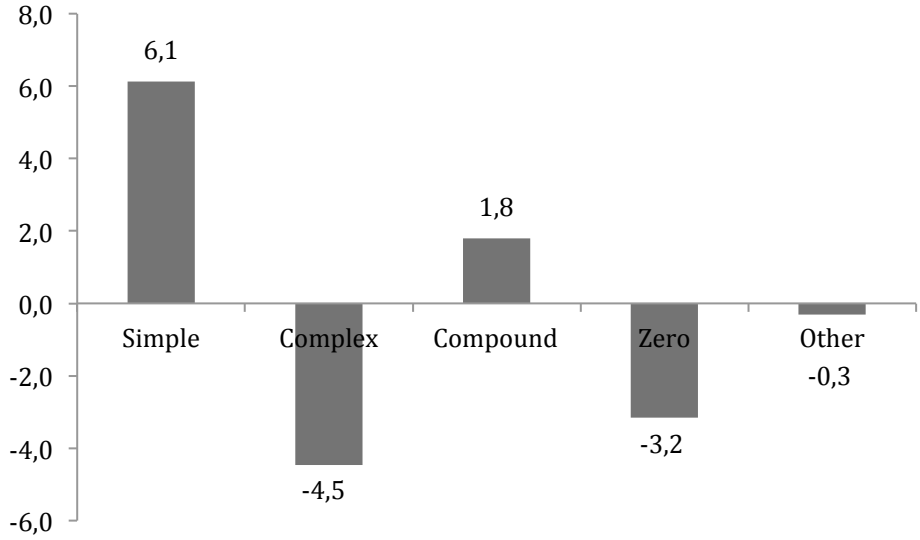


Figure 4.28. Processes per discipline. Differential coefficient (%).

In terms of frequency astronomers reuse their nouns to a higher extent than philosophers. This is seen in the overall type-token ratio where *CETA* uses each type on average 18.5 times, whereas *CEPhiT* does so 10.9 times. Type-token ratios by process suggest that simple nouns tend to occur more often than any other group in both disciplines, though the ratio in astronomy is the highest of all, with 29.5 times per 1,000 words. Compound nouns present the only exception to the general tendency, since their type-token ratio in the astronomy subcorpus is lower than in the philosophy subcorpus. Figure 4.29 below illustrates the differences of type-token ratios per noun-forming processes in *CETA* and *CEPhiT*. Positive numbers refer to higher percentages in the former and negative numbers indicate higher percentages in the latter.

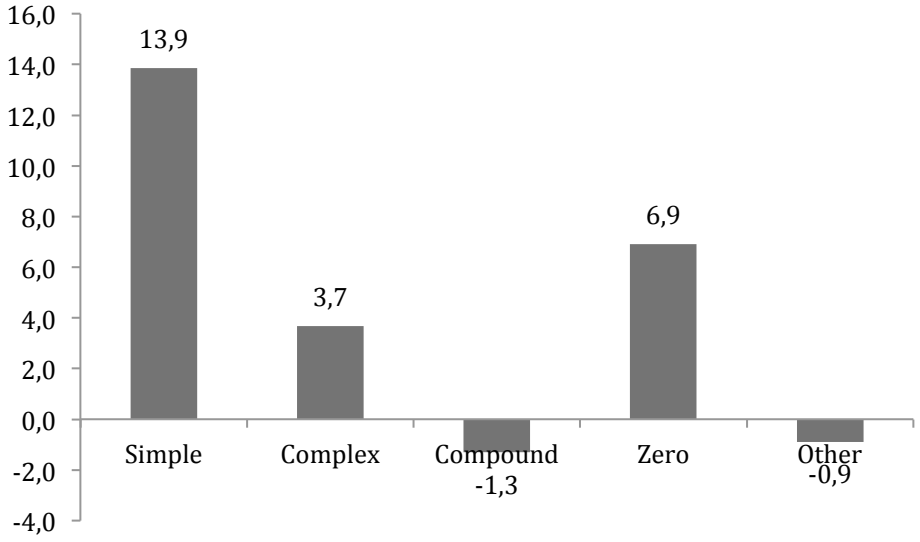


Figure 4.29. Type-token ratio in *CETA* and *CEPhiT*.

Also, *hapax legomena* are more frequent in philosophy than in astronomy. General figures show a total 6.1 nouns per 1,000 words in *CEPhiT* versus 1.9 per 1,000 words in *CETA*. If we look into processes individually the highest numbers of nouns occurring only once in the corpus are complex, whereas the greatest difference between the two subcorpora can be observed in zero-derivation, *CEPhiT* showing eight times as many instances as *CETA*.

Narrowing my focus onto affixation, I have subdivided the group of complex nouns included above into four affixation classes, which basically mirror those explained in section 2.1. Details for both astronomy and philosophy are given in Table 4.13.

Table 4.13. Affixation classes in *CETA* and *CEPhiT*.

CETA				
	Class I	Class II	Class III	Class IV
Tokens	1275	2082	10899	350
Types	251	241	566	16
% Tokens	9%	14%	75%	2%
% Types	23%	22%	53%	1%
Type-token ratio	5.1	8.6	19.3	0.7
Tokens/1,000 words	6.1	10.0	52.4	0.1
CEPhiT				
	Class I	Class II	Class III	Class IV
Tokens	2623	4973	13538	59
Types	598	587	929	24
% Tokens	12%	23%	64%	< 1%
% Types	28%	27%	43%	1%
Type-token ratio	4.4	8.5	14.6	2.5
Tokens/1,000 words	13.1	24.8	67.4	0.3

Apart from documenting the higher number of affixed tokens and types contained in the philosophy subcorpus, these results related to affixation classes show other similarities and divergences in the linguistic choice of astronomers and philosophers. Perhaps the most relevant aspect that can be extracted from these figures is the fact that most philosophers use coinages from classes I and II to a greater extent than astronomers, whereas the latter resort to nouns from classes III and IV more frequently than the former. This means that the majority of *CEPhiT* nouns are made up with free bases, whereas bound bases make up most *CETA* nouns. In total, the philosophy samples show

55% of nouns containing free bases, whereas the astronomy samples show only 45%. More importantly, the difference in class IV is especially meaningful, because a more extended use of prefix + suffix constructs might be taken as a fundamental characteristic of the discipline. The distribution of affixation classes is illustrated in figure 4.30.

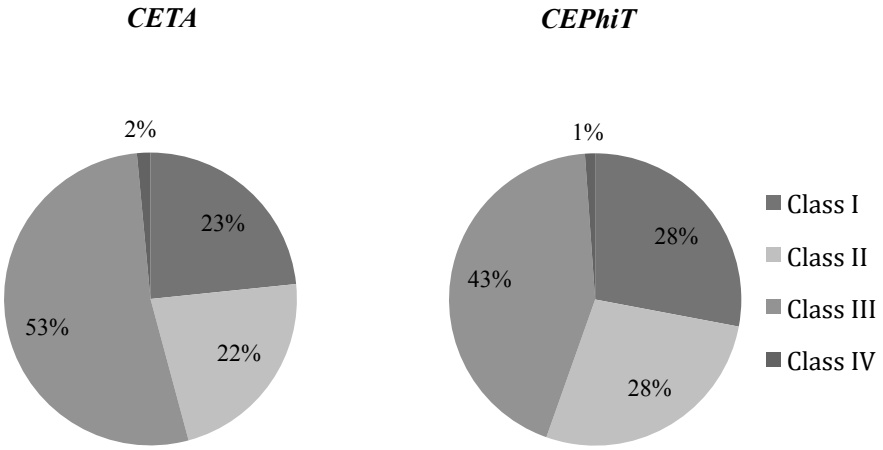


Figure 4.30. Affixation classes in *CETA* and *CEPhiT*.

Regarding similarities, tokens included in affixation class III are the most numerous in both subcorpora, reaching 74% of the total in *CETA* and 64% in *CEPhiT*. Additionally, in spite of the different values for type-token ratios explained above that both disciplines show, their distribution across processes

is very similar in both, becoming practically equivalent in relation to class II, as can be observed in figure 4.31.

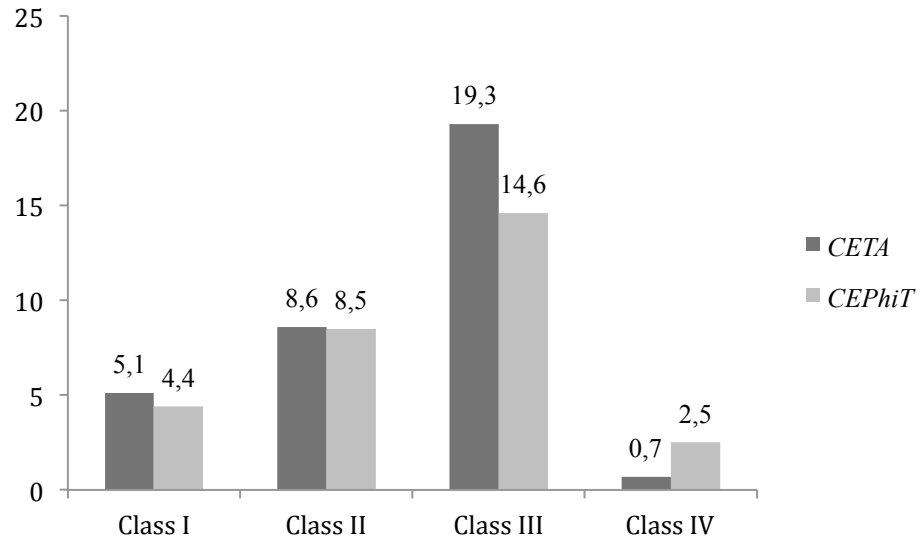


Figure 4.31. Distribution of type-token ratios by affixation classes in *CETA* and *CEPhiT*.

Coinage dates might represent another differentiating component between both subcorpora. Section 2.1 in this chapter revealed that affixation reached its peak in the fourteenth and sixteenth centuries, if the data extracted from the corpus *in toto* is taken into account. Now I intend to determine whether *CETA* and *CEPhiT* may follow divergent time-related patterns regarding the complex nouns they incorporate, so I will examine their frequencies separately. My findings are laid out in table 4.14, which

inventories the figures of frequencies per disciplines and centuries throughout the Middle and Modern English periods.

Table 4.14. Coinage dates of complex nouns in *CETA* and *CEPhiT*.

	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
CETA types	17	170	339	142	198	121	22
% CETA	2%	17%	34%	14%	20%	12%	2%
CEPhiT types	31	277	628	313	473	278	47
% *CEPhiT*	2%	14%	31%	15%	23%	14%	2%

These findings confirm the existence of convergent and divergent tendencies in writing astronomy and philosophy texts. On the one hand, both disciplines follow frequencies similar to the ones characterising the whole corpus, showing two peaks in the fourteenth and sixteenth centuries followed by two valleys in the fourteenth and eighteenth. The fact that *CEPhiT* uses considerably more complex noun types that *CETA* in absolute numbers is unsurprising given the differences between the two subcorpora in this respect. On the other hand, the distribution of complex nouns along the timeline varies from one subcorpus to the other. If we pay attention to percentages complex nouns coined in the 1200s and 1300s have a higher frequency in *CETA* (17% and 34%) than in *CEPhiT* (14% and 31%), whereas those nouns coined in the 1400-1600 period show a higher frequency in *CEPhiT* (15%, 23% and 14%) than in *CETA* (14%, 20% and 12%). These numbers may arguably suggest that philosophers make use of a more modern vocabulary. The following figure

illustrates the differential coefficient regarding the coinage of complex nouns in both subcorpora.

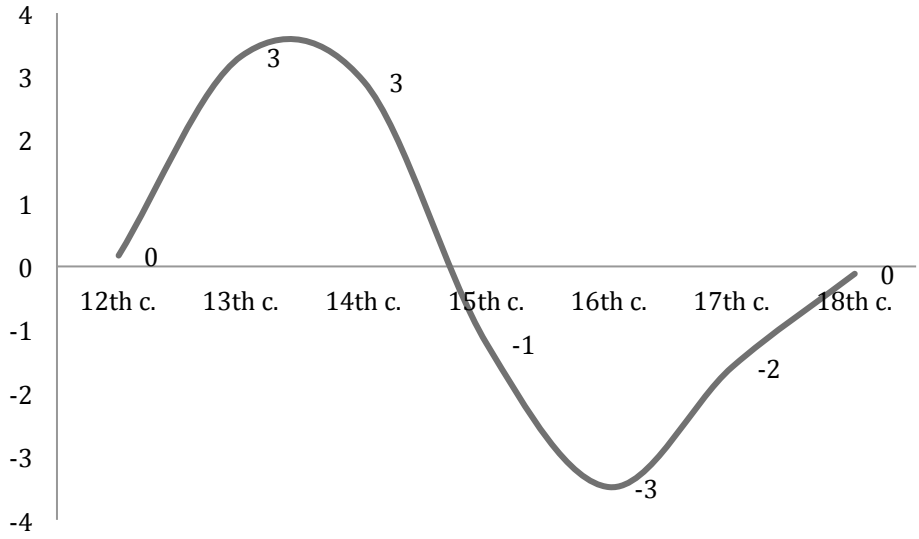


Figure 4.32. Coinage dates and lineal tendencies in *CETA* and *CEPhiT* (%).

Frequencies concerning prefixation and suffixation for both subcorpora have been presented in section 2.1.1. Given that *CEPhiT* contains a number of complex noun types considerably higher than that of *CETA* (2,138 versus 1,074 types, respectively), it must make use of a higher number of affixes and bases as well. Certainly, 167 affixes can be found in philosophy texts (53 prefixes and 114 suffixes) versus 137 in astronomy (46 prefixes and 91 suffixes). Differences also arise when reviewing the number of types that each affix can produce. The ratios for both disciplines, prefixed and affixed nouns give the following results: a) each prefix in *CETA* produces, on average, 1.7 nouns,

while 2.9 nouns are coined using the prefixes included in *CEPhiT*; and b) suffixes in the astronomy samples produce, on average, 10.9 types versus the 17.4 nouns obtained by combining suffixes in the philosophy samples.

Lastly, we can also search for differentiating features in origin components. This means that I intend to draw a contrasting line between the two disciplines by paying attention to the origin of all elements involved in affixation. To achieve this I have quantified all bases, prefixes and suffixes, classified them by discipline and later I have applied the sub-classification used in section 2 regarding the origin of these materials, i.e. Greek, Latin, Anglo-Norman, French and English. Finally I have calculated the percentage of each origin component by means of the following formula:

$$\frac{(Prefixes + Bases + Suffixes\ from\ one\ component) \times 100}{Total\ elements\ used\ in\ affixation} = Origin\ component$$

For example, *CETA* uses 30 prefixes, 6 bases and 14 suffixes from Greek origin. The total number of elements involved in affixation in the astronomy samples amounts to 2,140, as compared to 4,246 in the philosophy samples. Therefore the percentage frequency of the Greek component in *CETA* is obtained as follows:

$$Greek\ component = \frac{(30 + 6 + 14) \times 100}{2{,}140} = 2.3$$

By applying this formula to all components in both subcorpora I have obtained the results for their percentage frequencies, which I have laid out in table 4.15.

Table 4.15. Component origin in *CETA* and *CEPhiT*: percentage frequency.

	Greek	Latin	AN	French	English
CETA	2%	34%	11%	34%	19%
CEPhiT	2%	33%	11%	37%	18%

These results reveal a great similarity between *CETA* and *CEPhiT*. As we can see in the table, French- and Latin-based affixes and bases are the most productive in both subcorpora, whereas Greek-based affixes and bases produce the least coinages. Besides, the use of Greek- and English-based materials is equivalent in *CETA* and *CEPhiT*, since only insignificant decimals —unprinted in the table— separate them. There are also minimal differences in the Latin and English components (1% in each), and especially in the French component (3%). However, I do not regard this evidence conclusive enough to formulate a hypothesis that may define any characteristics inherent to any discipline in particular.

3.2. Comparison by genres/text-types

I mentioned in chapter 2 that the *Coruña Corpus* incorporates a wide variety of genres/text-types in order to allow researchers to identify, compare and

contrast linguistic features essential to them. The corpus that I have used for my research contains samples from eight different genres/text-types in the following proportion: a) one lecture, b) one dialogue, c) eighteen treatises, d) eleven textbooks, e) seven essays, f) one journal article, g) one letter, and h) one text classified under the label 'others', which is in this case the dictionary published by John Hill in 1754. The total number of words included in each genre/text-type is graphically illustrated in figure 4.33.

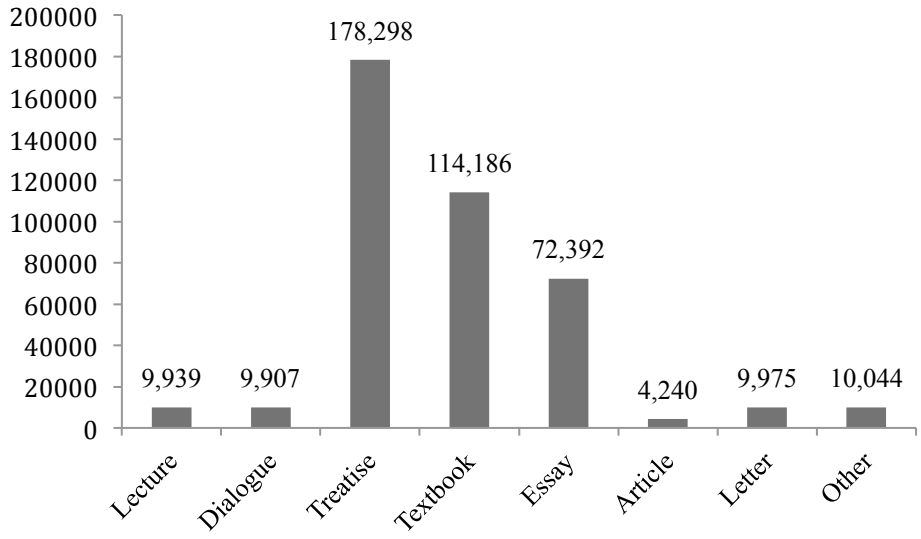


Figure 4.33. Total number of tokens per genre/text-type in the corpus.

Given that the general word count of samples in the corpus is around 10,000, we can see that the figure shows a great disparity in the distribution of tokens belonging to each genre/text-type. Normalised figures will be thereby used in all the evaluations that follow, which will deal with type and token frequencies,

ratios and percentages, noun forming processes, affixation classes and coinage dates. The details for types and tokens by genres/text-types are given in table 4.16:

Table 4.16. Analysis by genre/text-type.

	Tokens	Types	Type-token ratio	Tokens/ 1,000 words	Types/ 1,000 words
Lecture	1,689	365	4.6	170	36.7
Dialogue	1,602	454	3.5	162	45.8
Treatise	35,070	3,380	10.4	197	19.0
Textbook	23,117	1,818	12.7	**202**	15.9
Essay	13,510	13,510	7.0	187	26.6
Article	707	246	**2.9**	167	**58.0**
Letter	1,778	536	3.3	178	53.7
Other	1,843	401	4.6	183	39.9

These figures demonstrate that journal articles make use of the highest number of noun types, 58.0 per 1,000 words, than other genres by a wide margin. At the opposite end of the spectrum, textbooks only supply 15.9 noun types per 1,000 words. If we pay attention to the frequency of tokens shown by the same pair of genres, those fifty-eight types found in journal articles occur 167 times per 1,000 words, whereas the nineteen types obtained from textbooks produce 202 tokens. As a consequence, the type-token ratios of both genres stand out as clearly unalike, i.e. journal articles reuse each type only 2.9 times, whilst textbooks reuse them 12.7 times. Hence we can infer that recurrence seems an

archetypal characteristic of textbooks. It may correlate a purpose to fix definite patterns by means of repetition.

A closer look at the remaining genres seems to confirm that their type-token ratios are inversely proportional to the degree of didacticism present in them. That is to say, those intended to disseminate science in a more formal manner —highly specialised articles and university lectures, for example— show lower ratios, while those aiming at the popularisation of science — textbooks, treatises and essays, in descending order— resort to repetition to a higher extent. Genres that communicate science by imitating conventional conversations, such as dialogues and letters remain halfway between the other two groups. Meanwhile, defining general characteristics for highly heterogeneous genres such as 'other' is not easy. In the future, when supplementary samples are added to this category in the *Coruña Corpus*, more reliable data and more comprehensive information will be available. The type-token ratios are represented in Figure 4.34.

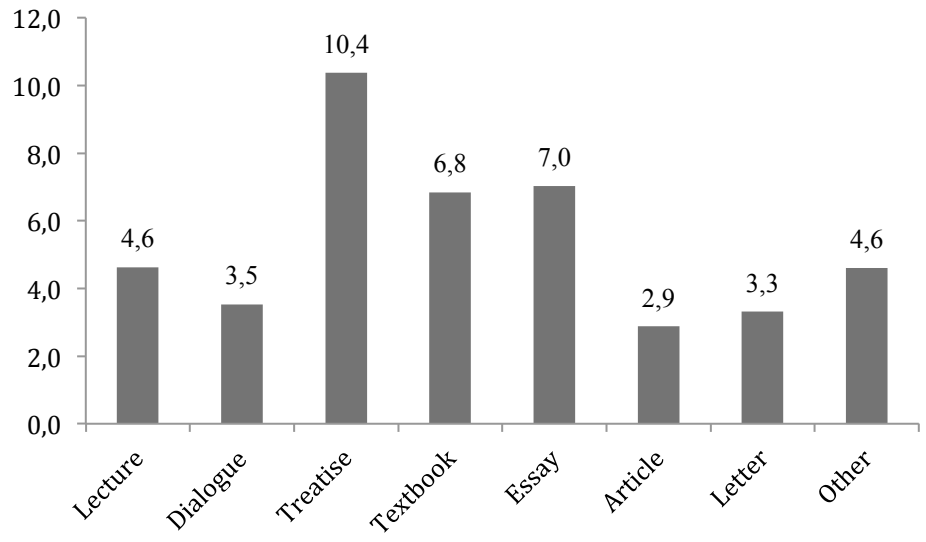


Figure 4.34. Type-token ratios by genre/text-type.

The following step will be studying these genres/text-types from the perspective of the five word-formation processes described above. But comparing eight categories by means of types, tokens, percentages, ratios and frequencies per 1,000 words will undoubtedly result in a very large table containing a multitude of figures that may hinder their interpretation. Besides, the further division of complex nouns into affixation classes I-IV will definitely add more confusion to the study. Therefore, for the sake of simplicity I will only state those results related to types per 1,000 words and their percentages. Also, I have not included the data referring to 'other' minor processes, because numbers rarely exceed 0.1 types per 1,000 words by genre. Therefore, only results for simple, complex, compound and zero-derived nouns

will be given. The details on noun-forming processes by genre/text-type are shown in table 4.17.

Table 4.17. Analysis of genres/text-types by word-formation process.

Types per 1,000 words				
	Simple	Complex	Compound	Zero
Lecture	14.7	19.3	0.3	2.4
Dialogue	20.5	20.9	**2.1**	1.9
Treatise	6.0	10.5	0.9	1.4
Textbook	5.6	8.4	0.9	1.0
Essay	7.7	15.9	0.5	2.3
Article	**22.2**	**32.8**	0.5	**2.6**
Letter	19.7	31.2	0.5	2.3
Other	18.4	17.8	1.1	2.4
Percentages				
	Simple	Complex	Compound	Zero
Lecture	40%	53%	1%	7%
Dialogue	45%	46%	5%	4%
Treatise	32%	55%	5%	8%
Textbook	35%	53%	**6%**	6%
Essay	29%	**60%**	2%	**9%**
Article	38%	57%	1%	4%
Letter	37%	58%	1%	4%
Other	**46%**	45%	3%	6%

As we can see in the table, each genre/text-type shows a singular distribution regarding word-formation processes. Complex words predominate, with different margins, in all genres with the exception of 'others', where simple

nouns occur more often than any other. Considering that the only sample included in the category 'other' is a dictionary, we may assume that the definitions of entries in dictionaries contain a higher number of simple than complex nouns. Table 4.16 above proved that journal articles present the highest amount of noun types and the lowest type-token ratio. Hence it is unsurprising that the same genre shows the highest rates of types per 1,000 words in three of the four processes studied, i.e. simple or no-process, affixation and conversion, while dialogues show a higher frequency in compounding.

If we pay attention to the second section of the table, all genres show higher rates of complex than simple nouns, with the exception of dictionaries, where this relationship is reversed by a narrow margin (46% vs. 45%). The highest percentage of affixed nouns, either by means of an affix (60%) or a zero-affix (9%), corresponds to essays, and compounds are more generally used in textbooks (6%). Figure 4.35 represents the percentage relationship between simple and complex nouns in the corpus by genres/text-types. Essays show the widest margin (31%), whereas in dialogues and dictionaries numbers are equivalent.

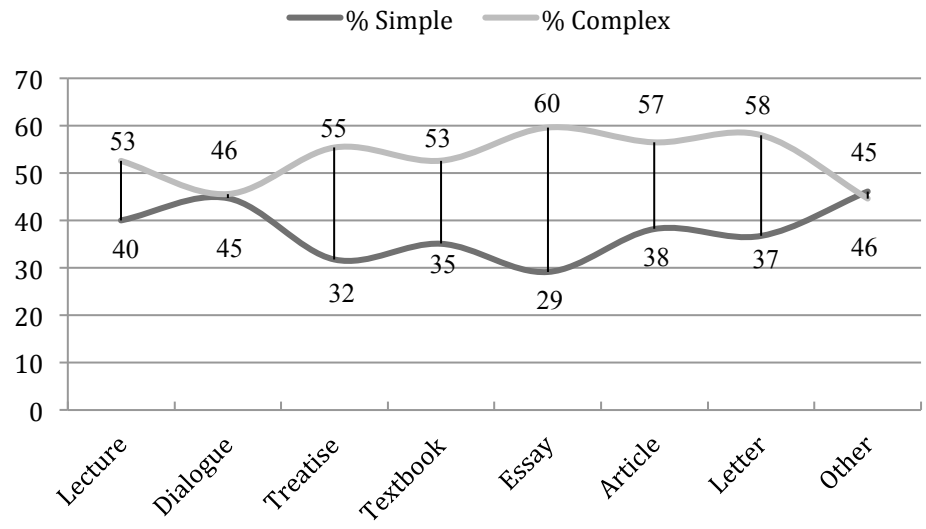


Figure 4.35. Relationship simple vs complex nouns by genre/text-type (%).

When the subdivision of complex nouns in affixation classes is carried out, the data concerning each genre/text-type gives the results shown in table 4.18 below. As seen in the first section of the table, the number of types per 1,000 words indicates that class III predominates across all genres. Besides, letters present the highest frequencies in classes I and III (5.0 and 18.9 noun types per 1,000 words respectively), while the highest numbers regarding classes II and IV correspond to journal articles (9.7 and 0.7 types per 1,000 words). Textbooks show the lowest numbers for al affixation classes, sharing those for class IV with treatises.

Table 4.18. Analysis of genres/text-types by affixation class.

Types per 1,000 words				
	Class I	Class II	Class III	Class IV
Lecture	3.1	4.2	11.6	0.4
Dialogue	4.3	3.9	12.0	0.6
Treatise	2.7	2.8	4.8	0.1
Textbook	1.8	2.1	4.4	0.1
Essay	3.6	4.6	7.4	0.2
Article	4.7	**9.7**	17.7	**0.7**
Letter	**5.0**	6.9	**18.9**	0.3
Other	3.0	3.6	10.9	0.4
Percentages				
	Class I	Class II	Class III	Class IV
Lecture	16%	22%	60%	2%
Dialogue	21%	19%	57%	**3%**
Treatise	**26%**	27%	46%	1%
Textbook	22%	25%	52%	1%
Essay	23%	**29%**	47%	1%
Article	14%	**29%**	54%	2%
Letter	16%	22%	**61%**	1%
Other	17%	20%	**61%**	2%

The second half of the table reveals that the addition of percentages affixation classes involving bound bases (classes III and IV) generally exceeds that of free bases (classes I and II). Treatises and essays are exceptions to this rule by narrow margins. At the other end, dictionaries and letters show the widest margin between the use of bound and free bases (with 26% and 25% difference, respectively). It is in class III that we can see higher differences

among genres, treatises showing rates up to fifteen per cent lower than letters and dictionaries, for example. It seems that the less formal genres, or those that are not addressed to a close audience, that is to say, those that represent orality more truthfully contain more class III nouns. This is the case of letters and lectures, for example, which are generally addressed to a familiar audience, and also dialogues, which are simulations of real face-to-face conversations. At the other end, treatises, essays, textbooks and articles offer a more literary discourse and, therefore, contain more class III nouns. Figure 4.36 illustrates differences in affixation classes by genre/text-type in the corpus.

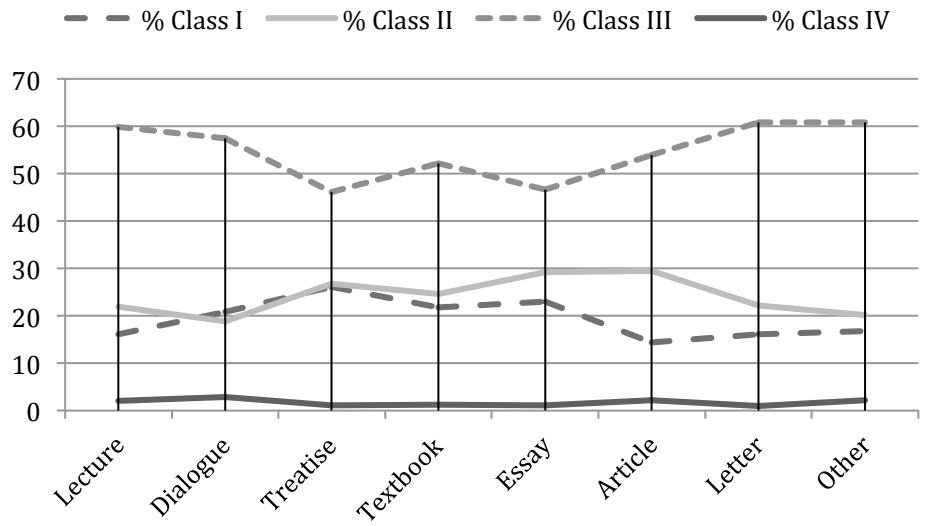


Figure 4.36. Differences in affixation classes by genre/text type (%).

Ultimately, the variable of genre/text type will be filtered by coinage dates. It is my intention to assess whether or not the choice of nouns can be

connected with a specific genre regarding the period in the history of English in which those nouns were coined. The details are provided in table 4.19 below.

There is no need to mention that all genres show the highest frequencies in the 1300s and 1500s, because this trend in noun-forming processes has been already mentioned in previous sections of this chapter. But other facts observable in the results shown in the table have more relevance. For example, journal articles concentrate their noun choice in five of the seven centuries recorded, showing no types from the twelfth or eighteenth centuries. That also means that their numbers for other centuries are higher than in other genres, especially in the 1300s (14.9 types per 1,000 words).

Within each genre, patterns related to the percentage of nouns coined in every century are varied. In the twelfth century, for example, the highest rates correspond to dialogues and dictionaries (3%); dialogues show the highest share in the thirteenth century, with 29%; as I mentioned earlier, articles contain a high number of nouns coined in the fourteenth century, which represent 40% of their total. In the fifteenth and sixteenth centuries essays show higher percentages than other genres (22% and 13%), though in the latter century this position is shared with that of treatises. In the seventeenth century textbooks and treatises also share the same percentage (13%). Finally, in the eighteenth century percentage margins are low, as the inclusion of nouns coined in the 1700s is low as well, for reasons anticipated in chapter 2, and

which will be fully explained in the next chapter. Treatises and textbooks show the highest rates (4%). These two text-types were generally addressed to students, so authors may have included the latest coined nouns intentionally to make the most modern terminology available for their readers.

Table 4.19. Analysis of genres/text-types by coinage date.

Types per 1,000 words							
	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
Lecture	0.3	4.9	8.4	2.9	4.3	1.9	0.2
Dialogue	**0.8**	7.5	9.3	2.1	4.4	1.7	0.3
Treatise	0.3	2.0	3.9	1.9	2.9	0.1	0.5
Textbook	0.2	1.8	3.2	1.3	2.1	0.2	0.4
Essay	0.4	3.0	5.8	2.9	4.2	0.3	**0.6**
Article	—	**9.2**	**14.9**	4.0	5.0	**3.1**	—
Letter	**0.8**	6.7	12.6	**4.6**	**6.4**	1.9	**0.6**
Other	**0.8**	5.3	8.2	2.2	3.9	1.8	0.3
Percentages							
	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
Lecture	1%	22%	37%	13%	19%	7%	1%
Dialogue	**3%**	**29%**	36%	8%	17%	5%	1%
Treatise	2%	15%	29%	14%	**22%**	**13%**	**4%**
Textbook	2%	17%	31%	13%	20%	**13%**	**4%**
Essay	2%	16%	30%	**15%**	**22%**	12%	3%
Article	—	25%	**40%**	11%	13%	11%	—
Letter	2%	19%	36%	13%	18%	8%	2%
Other	**3%**	24%	37%	10%	18%	6%	1%

Numbers for coinages in the eighteenth century are so low that they render the analysis of noun-forming processes by 30-year periods rather impractical. As a consequence I have omitted this study from a genre/text-type perspective. The same can be said for origin components, as this analysis does not offer any contrastive results. They are similar to those comparing disciplines, as shown in section 3.2. Therefore I have omitted this step as well.

3.3. Sex of the author

In this section I attempt to determine characteristics inherent to texts written by men and women scientists. My goal is to assess whether or not writers from either sex make a different use of nouns in their writings. In line with my previous studies in section 3, I will look for differences in frequencies, type-token ratios, processes, affixation classes and periods in the history of the English language in which those nouns were coined.

However, before I start analysing the two elements of this variable separately, I would like to pay due attention to every individual sample in the text, because we may observe some features in some authors that will condition the overall outcome of the study. This is especially significant in the case of women writers, because findings in each of the four samples written by them will account for twenty-five per cent of the total numbers in the analysis of the sex of the author. Conversely, samples written by men will not reach three per

cent of the total. The details for every author are laid out in table 4.20 below. I will bring up the information given in this table when I deal with the following two variables, which study linguistic characteristics regarding the places of education and ages of the authors in sections 3.4 and 3.5.

The table contains three sets of rows that provide information about the sample in question, identified with the first letter of the subcorpus to which it belongs —'a' for astronomy or *CETA* and 'p' for philosophy or *CEPhiT*— followed by the publication date. In case that the texts belong to the same discipline and year, as is the case of the astronomy samples written by Curson and Morden in 1702, and Gordon and Watts in 1726, the publication date will be followed by '(a)' or '(b)' respectively. Finally, a gender symbol representing man/woman (♂/♀) will identify the sex of the author. Additionally, samples written by women have been additionally highlighted for ease of identification.

Most of the texts contain around 10,000 words, except for two astronomy samples published in 1773 and 1779. This implies that the normalised and percentage figures are exponentially analogous (the division of the second by the first will be equal to 10 in most of the cases), so the latter have been omitted.

Table 4.20. Details of all samples included in the corpus.

	Tokens	Types	Type-token ratio	Tokens/10,000 words	Types/10,000 words
p1700 ♀	1724	609	2.8	1711.0	604.4
a1702(a) ♂	2131	495	4.3	2079.6	483.1
a1702(b) ♂	1901	540	3.5	1872.2	531.8
p1705 ♂	1888	447	4.2	1876.7	444.3
p1710 ♂	1851	636	2.9	1840.1	632.3
a1715 ♂	1689	365	4.6	1699.4	367.2
p1717 ♂	1927	419	4.6	1913.4	416.0
a1719 ♂	1602	454	3.5	1617.0	458.3
a1726(a) ♂	1781	295	6.0	1706.4	282.6
a1726(b) ♂	2170	295	7.4	2085.1	283.5
p1727 ♂	1875	533	3.5	1852.4	526.6
p1730 ♂	1627	594	2.7	1619.7	591.3
a1732 ♂	2109	350	6.0	2061.2	342.1
p1733 ♂	1902	488	3.9	1894.4	486.1
a1735 ♂	1784	269	6.6	1722.3	259.7
p1736 ♂	1888	424	4.5	1878.8	421.9
p1740 ♂	2134	398	5.4	2127.6	396.8
a1742 ♂	1998	293	6.8	1907.6	279.7
p1748 ♂	2191	681	3.2	2186.8	679.7
a1749 ♂	2681	255	10.5	2414.0	229.6
a1754 ♂	1843	401	4.6	1834.9	399.2
p1754 ♂	1830	521	3.5	1825.8	519.8
p1755 ♂	2283	644	3.5	2275.9	642.0
a1756 ♂	1673	299	5.6	1590.5	284.2
a1761 ♂	1693	62	**27.3**	1390.0	50.9
p1764 ♂	2000	635	3.1	1993.6	633.0
a1767 ♂	1876	400	4.7	1818.7	387.8
p1769 ♂	**2704**	641	4.2	**2686.8**	636.9
p1770 ♂	2157	739	2.9	2153.3	737.7
a1773 ♂	707	246	2.9	1667.5	580.2
p1776 ♂	2089	783	2.7	2087.5	782.5
a1777 ♂	1835	250	7.3	1736.7	236.6
a1779 ♂	1080	240	4.5	1828.0	406.2
a1782 ♂	1805	379	4.8	1757.9	369.1
p1783 ♀	2319	751	3.1	2305.4	746.6
a1786 ♂	1778	536	3.3	1782.5	537.3
a1790 ♂	2057	220	9.4	1951.6	208.7
p1790 ♂	2486	647	3.8	2487.7	647.5
p1792 ♀	2239	**839**	**2.7**	2226.1	**834.2**
p1793 ♂	2081	464	4.5	2071.3	461.8
a1797 ♀	1928	460	4.2	1878.6	448.2

The text written by Stewart (*CETA* 1761) stands out by its noticeable scarcity of noun types. As far as tokens are concerned there are five other samples that show fewer frequencies than this one, but it seems that Stewart reuses every noun type 27.6 times on average, the maximum in the corpus by a wide margin, remaining its closest follower around 17 units lower. Hence it contributes to the general type count with only 5.1 types per 1,000 words. Inversely, the sample by Macaulay (*CEPhiT* 1783) uses up to 222.6 types per 1,000 words, so she only repeats each of her nouns 2.7 times. Another two authors, Kirkpatrick and Campbell (*CEPhiT* 1730 and 1776, respectively) show the same type-token ratio as Macaulay, though their absolute numbers remain lower than those seen in her text (574 and 783 in theirs versus 839 types in hers).

Another fact that can be observed in the table is the increase of frequencies in texts written in the second half of the century. With the honourable exception of Hume's work (*CEPhiT* 1748), the eight samples containing the highest frequency of types per 1,000 words were published after 1750.

Given that my corpus contains thirty-seven samples written by men and only four by women scientists, the total word count for each sex is rather imbalanced. Indeed, texts written by men contain 368,525 words, whereas those written by women have 40,456 words. This section will therefore have a

constant recourse to normalised figures (per 1,000 words) to account for that difference in numbers.

The details of noun types and tokens from the perspective of the sex of the author are given in table 4.21. The first two columns on the left express the total noun count and their types. The subdivision 'men' has 71,106 tokens and 4,254 types, whereas the same count for the subdivision 'women' contains 8,210 tokens and 1,729 types. These figures are not very relevant so far, but they become meaningful as soon as they are used as bases for normalising frequencies.

Table 4.21. Tokens and types in men and women scientists.

	Tokens	Types	Type-token ratio	Tokens/ 1,000 words	Types/ 1,000 words	*hapax leg.*/1,000 words
Men	71106	4254	16.7	193	11.5	4.3
Women	8210	1729	4.7	203	42.7	19.7

The rest of the columns show exceptionally interesting results regarding convergent and divergent characteristics of both subdivisions. As for the first, both men and women use a similar number of nouns in their texts. Women scientists incorporate ten more nouns per thousand words, but this difference does not seem significant enough to establish a gender-based feature. It is in the other three sets of data that clear divergences in language use are found. Indeed, in their texts women incorporate 42.7 types per 1,000 words, as

opposed to men, who provide 11.5 types per 1,000 words in theirs. As seen in the type-token ratio, women reuse their nouns 4.7 times, whereas men repeat theirs 16.7 times. Also, women show higher frequencies in the single use of nouns or *hapax legomena*, with 19.7 types per 1,000 words, versus those achieved by men, 4.3 types. In chapter 5 I will delineate these features further in order to explain differences in language use by men and women scientists.

Regarding noun-forming processes, I have performed the same subdivision applied to the previous variables in this research work due to consistence. Therefore, information about simple, complex, compound and zero-derived nouns will be discussed in the following lines. The details of frequencies for nouns formed by means of other processes have been given along with the rest in table 4.22 below, but they will be excluded from my argument because they represent less than 1 per cent of the total.

A first look at the table confirms a higher frequency of noun types per 1,000 words in texts written by women in all processes. Complex nouns set the widest gap between sexes (6.4 types per 1,000 words in men versus 24.3 types in women). Besides, percentage frequencies of tokens reveal that not only do women use simple nouns less regularly than men do (43% vs. 50%, respectively), but also that complex nouns are the most frequent in their texts (51%). Conversely, men resort to simple nouns more often than they do regarding complex nouns (50% and 45%, respectively). Also, male scientists use compounds more regularly (6%) than female scientists do (2%). The

outlook becomes more balanced when dealing with percentage frequencies of types, since simple types show a 3% margin between both sexes (30% in men vs. 33% in women), and complex show a still narrower margin of 2% (55% in men vs. 57% in women).

Type-token ratios add more contrast still. Figures show a virtual equivalence as far as compounding, conversion and minor processes are concerned, but there is a clearer disparity in the simple and complex noun columns. These indicate that men reuse their simple nouns 27.6 times versus women, who repeat their simple nouns only 6.2 times. Moreover, complex nouns are reused 13.5 times by men, whereas they are repeated 4.3 times by women. This may be due to the fact that women were more concerned than men about the use of a more polished and varied language, given that they were more in need to be seen as valid scientists in a male-dominated world.

Table 4.22. Morphological processes by sex of the author.

Men					
	Simple	Complex	Compound	Zero	Other
Tokens	35,769	31,642	926	2,709	59
Types	1,295	2,343	256	337	23
% Tokens	50%	45%	1%	4%	< 1%
% Types	30%	55%	6%	8%	< 1%
Type-token ratio	27.6	13.5	3.6	8.0	2.6
Tokens/1,000 words	**97.1**	85.9	2.5	7.4	0.2
Types/1,000 words	3.5	6.4	0.7	0.9	0.1

Women					
	Simple	Complex	Compound	Zero	Other
Tokens	3,536	4,174	132	356	11
Types	567	982	35	137	8
% Tokens	43%	51%	2%	4%	< 1%
% Types	33%	57%	2%	8%	< 1%
Type-token ratio	6.2	4.3	3.8	2.6	1.4
Tokens/1,000 words	87.4	**103.2**	**3.3**	**8.8**	**0.3**
Types/1,000 words	**14.0**	**24.3**	**0.9**	**3.4**	**0.2**

Further subdividing complex nouns into their affixation classes supports the principles discussed above, which confer different characteristics to texts written by scientists of either sex. The details are presented in table 4.23.

From the perspective of token and type frequencies my results put in evidence that women include more tokens per 1,000 words and many more types per 1,000 words. Differences range from 3.1 to 7.5 times concerning tokens, and 0.5 to 45.1 times as regards types. Frequencies found in men only exceed those for women in class IV tokens (1.1 versus 0.7 per 1,000 words):

Table 4.23. Affixation classes by sex of the author.

Men				
	Class I	Class II	Class III	Class IV
Tokens	3,259	6,244	21,739	69
Types	656	597	1057	45
% Tokens	10%	20%	69%	1%
% Types	28%	25%	45%	1%
Type-token ratio	5.0	10.5	20.6	12.1
Tokens/1,000 words	8.8	16.9	59.0	1.1
Types/1,000 words	9.2	8.4	14.9	0.5
Women				
	Class I	Class II	Class III	Class IV
Tokens	639	809	2,698	65
Types	230	251	493	50
% Tokens	15%	19%	65%	1%
% Types	23%	26%	50%	1%
Type-token ratio	2.8	3.2	5.5	3.5
Tokens/1,000 words	**15.8**	**20.0**	**66.7**	0.7
Types/1,000 words	**28.0**	**30.6**	**60.0**	**1.0**

Type-token ratios do not contradict any of the information given above, since figures for women are lower in each affixation class. Notwithstanding diversity in numbers, both genders share similar relationships from the standpoint of ratios with respect to affixation classes, as shown in figure 4.37 below.

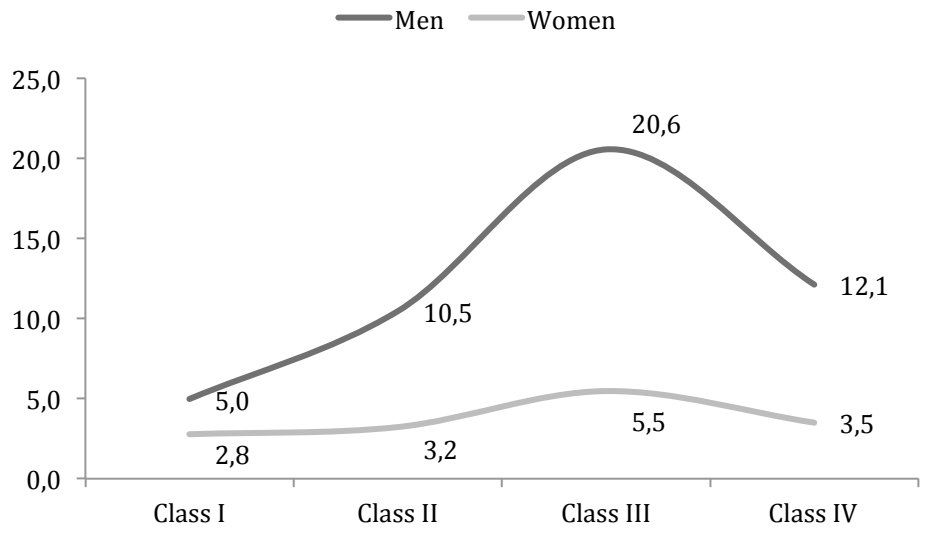


Figure 4.37. Type-token ratios in affixation classes by sex of the author.

Also, percentages show similarities between both groups. The two of them make a more repeated use of nouns belonging to class III, followed by those belonging to classes II and I in descending order, either in token or type rates. Men, however, tend to use noun types made up with free bases slightly more frequently than women do, with a combined result for class I + class II marginally superior to that of class III + class IV (53% vs. 47%). Conversely, texts written by women present a higher frequency of nouns made up with bound bases (51% vs. 49%).

Applying the filter by coinage dates to the variable 'sex of the author' also shows some interesting differences between texts written by men and women scientists. Table 4.24 includes the particulars of the data filtered in this

way. In the first section of the table there is nothing particularly relevant, as women writings show a consistent higher frequency of types per 1,000 words across the seven centuries included in my study. The most productive periods in language as regards the coinage of nouns show a wider margin between both sexes, especially in the fourteenth century (2.2 types per 1,000 words in texts written by men vs. 9.6 types per 1,000 words in those written by women).

The second section of the table displays the percentages of the nouns used in writing science by authors of both sexes from a diachronic point of view. As we can infer from these rates, both men and women use the same amount of nouns coined in the twelfth (2%) and fifteenth centuries (14%), but their progression fluctuates from a more intensive use in the thirteenth and fourteenth centuries in texts written by women, and then those written by men have a heavier load after the fifteenth century. This might imply that men prefer nouns coined more recently, whereas women favour vocabulary consolidated in previous centuries. Labov (2001: 293) states that women "conform more closely than men to sociolinguistic forms that are overly prescribed but conform less than men when they are not". So this might explain why women scientists in my corpus resort to well-established nouns. For Nevalainen (2006b: 574) Labov's words also mean that "women typically emerge as the leaders of linguistic change" but the materials provided by my corpus and the data obtained do not prove this leadership. The differential coefficients in the percentage rates are shown in figure 4.38 below. Positive

numbers indicate higher rates in texts written by men, whereas negative numbers show higher rates in texts written by women.

Table 4.24. Analysis of genres/text-types by coinage dates.

Types per 1,000 words							
	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
Men	0.2	1.1	2.2	1.1	1.9	1.2	0.4
Women	0.6	5.2	9.6	4.0	5.7	3.5	0.5
Percentages							
	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
Men	2%	14%	28%	14%	**24%**	**15%**	**5%**
Women	2%	**18%**	**34%**	14%	20%	12%	2%

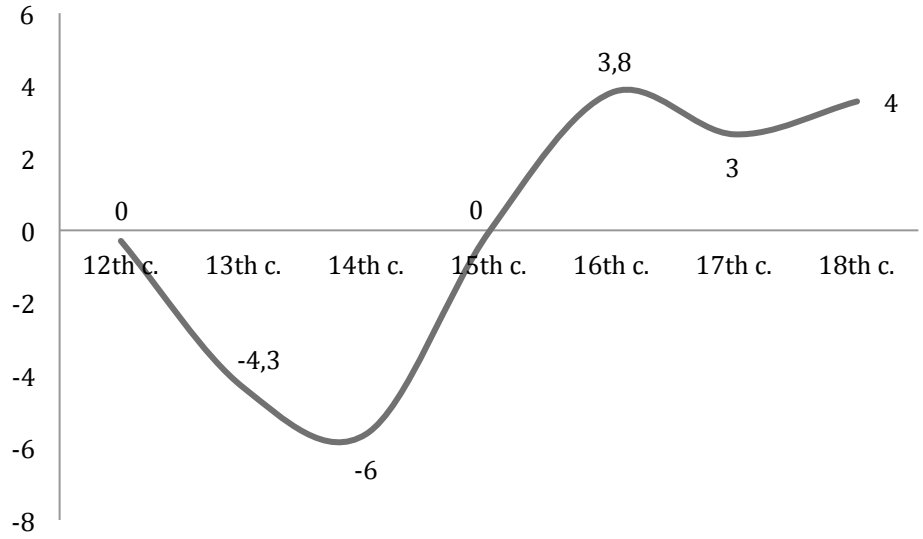


Figure 4.38. Differential coefficients in men and women regarding coinage dates (%).

3.4. Place of education

The third variable under study will assess to what extent the place of education may have had an influence in the linguistic habits acquired by scientists. By means of this variable and the filters used in the previous two variables in sections 3.1 and 3.2, I intend to establish patterns of noun formation and noun use in my corpus. The records extracted from the metadata section in the *Coruña Corpus Tool* shows that we lack information about the places of education of six of our authors —Jasper Charlton, Henry Curson, George Gordon, James Hodgson, Robert Kirkpatrick and John Lacy. But the remaining thirty-five authors provide a total of 350,880 words, which clearly represents enough material to carry out this analysis. Thus, I have classified the texts into four groups as follows:

a) England: twenty-one samples and 213,409 words.
b) Ireland: two samples and 20,249 words.
c) Ireland/Scotland: one sample and 10,031 words.
d) Scotland: eleven samples and 107,191 words.

The particulars concerning types and tokens are displayed in table 4.25. The first two columns on the left side refer to absolute numbers. Given the differences in sums across groups I will not deal with them now, and I will concentrate on the normalised figures per 1,000 words displayed in the

remaining four columns. We can see that the group representing authors educated in Ireland/Scotland show the lowest type-token ratio, since they reuse their noun types only 3.5 times. Also, their texts contain the highest amount of nouns, as observed in their frequencies concerning tokens (227.6 per 1,000 words). Besides, their frequencies of noun types and *hapax legomena* also are at the top of the range (64.2 and 29.4 times per 1,000 words, respectively), which speaks for itself as regards language competence.

Conversely, it is authors educated in England that reuse their nouns the most, 12.3 times, and contribute the lowest numbers of tokens, types and *hapax legomena* to the corpus (187.6, 15.2 and 6.0 times per 1,000 words, respectively). Between these two extremes, the group of authors educated in Ireland show similar results to those of Ireland/Scotland.

Table 4.25. Types and tokens by place of education.

	Tokens	Types	Type-token ratio	Tokens/ 1,000 words	Types/ 1,000 words	*hapax leg.*/1,000 words
England	40,037	3,250	12.3	187.6	15.2	6.0
Ireland	4,266	988	4.3	210.7	48.8	23.8
Ireland/Scotland	2,283	644	3.5	227.6	64.2	29.4
Scotland	21,646	2,368	9.1	201.9	22.1	8.7

Filtering these results by processes will allow us to deepen in the study of the linguistic habits contained in the four groups mentioned earlier. My results are laid out in table 4.26, which is rather dense, so the lowest values

relating to type-token ratios and the highest relating to frequencies per 1,000 words in each category have been printed in bold characters to make recognition easier.

The data displayed in the table confirms that the group Ireland/Scotland shows lower type-token ratios in four of the five processes contemplated —no process (4.8 times), affixation (3.0 times), conversion (1.9 times) and other processes (1.0 times)—, whereas authors in the group Ireland reuse their compounds the least (1.7 times).

Frequencies corroborate the linguistic proficiency of authors in the two groups described above. On the one hand, Ireland stands out in terms of noun use, chiefly in simple, zero-derived and 'other' nouns, as well as in the inclusion of a higher number of compound types. On the other hand, Ireland/Scotland predominates in the use of different simple and complex noun types, with 18.9 and 39.4 times per 1,000 words, respectively, especially if we compare these figures with those of Scotland and England, which are much lower.

Table 4.26. Morphological processes by place of education.

England					
	Simple	Complex	Compound	Zero	Other
Tokens	19,999	18,146	512	1,349	31
Types	1,026	1,809	150	247	18
% Tokens	50%	45%	1%	3%	< 1%
% Types	32%	56%	5%	8%	1%

Type-token ratio	19.5	10.0	3.4	5.5	1.7
Tokens/1,000 words	93.7	85.0	2.4	6.3	0.1
Types/1,000 words	4.8	8.5	0.7	1.2	0.1

Ireland

	Simple	Complex	Compound	Zero	Other
Tokens	2,027	1,958	84	190	7
Types	331	540	49	64	4
% Tokens	48%	46%	2%	4%	< 1%
% Types	34%	55%	5%	6%	< 1%
Type-token ratio	6.1	3.6	**1.7**	3.0	1.8
Tokens/1,000 words	**100.1**	96.7	4.1	**9.4**	**0.3**
Types/1,000 words	16.3	26.7	**2.4**	3.2	**0.2**

Ireland/Scotland

	Simple	Complex	Compound	Zero	Other
Tokens	907	1,201	102	72	1
Types	190	395	21	37	1
% Tokens	40%	53%	4%	3%	< 1%
% Types	30%	61%	3%	6%	< 1%
Type-token ratio	**4.8**	**3.0**	4.9	**1.9**	**1.0**
Tokens/1,000 words	90.4	**119.7**	**10.2**	7.2	0.1
Types/1,000 words	**18.9**	**39.4**	2.1	**3.7**	0.1

Scotland

	Simple	Complex	Compound	Zero	Other
Tokens	10,135	10,354	282	848	27
Types	755	1,337	86	181	9
% Tokens	47%	48%	1%	4%	< 1%
% Types	32%	56%	4%	8%	< 1%
Type-token ratio	13.4	7.7	3.3	4.7	3.0
Tokens/1,000 words	94.6	96.6	2.6	7.9	**0.3**
Types/1,000 words	7.0	12.5	0.8	1.7	0.1

Percentages also point out the differences between these places of education as regards noun choice. The proportions of simple, complex, compound and zero-derived nouns are dissimilar in all four groups, as can be best observed in the following graph, containing the distribution of types in each group in connection with the morphological processes. Due to its marginally low numbers, 'other' nouns have been omitted in the figure.

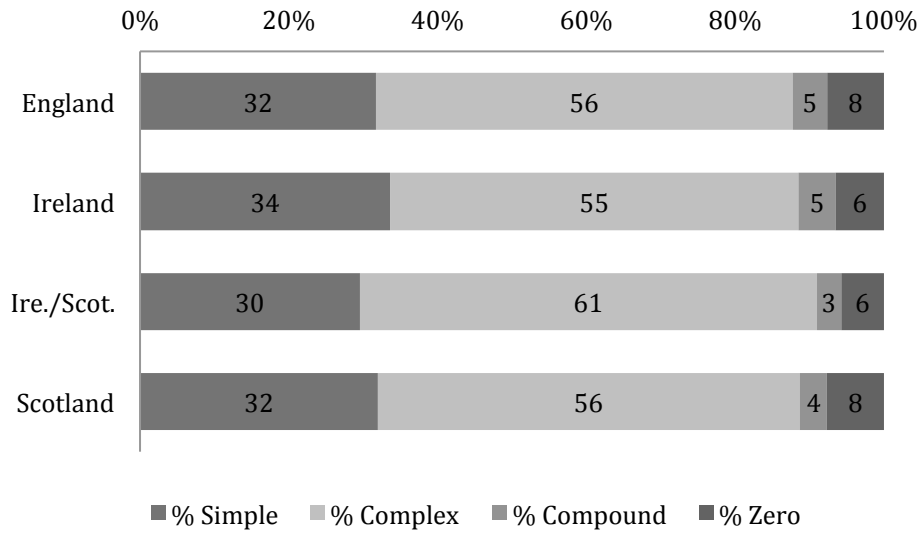


Figure 4.39. Types by process and place of education (%).

As we can see in the figure, samples in the group Ireland/Scotland contain the minimum and maximum levels of simple and complex nouns respectively, showing 30% and 61% of their total number. However, authors in this group do not seem to favour compounds as much as other authors do, representing only 3% of all their noun types. Besides, they show a low

percentage of zero-derived instances. The levels of simple types are the highest in samples written by those authors educated in Ireland. English and Irish writers seem to prefer compounds to a higher extent than other authors do. Lastly, zero-derivations present higher rates in samples belonging to the England and Scotland groups.

A narrower filter that focuses on complex nouns further characterises the writings in these four groups. The details of this subdivision are given in table 4.27. Again, the group Ireland/Scotland shows the highest frequencies in practically all affixation classes.

More remarkable results can be observed in the second section of the table including percentages. Writers educated in England and Scotland make a more regular use of nouns made up with free bases, that is, the combination of nouns included in classes I and II is higher than that of classes III and IV, whereas on the contrary those educated in Ireland and Ireland/Scotland favour nouns coined by means of bound bases. Furthermore, the group England stands out due to its predominance in the use of class I nouns. As figure 4.7 above shows, these nouns are on average the most recent coinages in the language; therefore English scientists use a higher percentage of newer nouns in their writings. Though my data do not provide conclusive results that will allow me to provide a demonstrable pattern to account for this, I might speculate that English authors had access to the latest publications in London that included the latest coinages and the most modern linguistic fashions. As a consequence,

their subsequent writings would reflect more truthfully the new developments of the capital. Also, I might venture another hypothesis that would relate this situation to the results obtained by individual disciplines. Given that most astronomers are English and most philosophers are Scottish, and the nouns used in astronomy tend to be more recent, then it is more likely that English authors tended to use a higher percentage of class I nouns. If we take the eighteenth century, for example, 90 per cent of the nouns coined in this century are present in *CETA*, vs. 70.2 per cent in *CEPhiT*. In spite of these partial explanations, I am aware that these conjectures are by no means definitive; however, I believe that they may not be too far from to the truth.

Table 4.27. Affixation classes by place of education.

Types per 1,000 words				
	Class I	Class II	Class III	Class IV
England	2.3	2.0	4.0	0.1
Ireland	5.1	7.1	14.3	**0.2**
Ireland/Scotland	**7.5**	**11.5**	**20.4**	0.0
Scotland	2.7	3.6	6.0	**0.2**
Percentages				
	Class I	Class II	Class III	Class IV
England	28%	24%	47%	1%
Ireland	19%	27%	54%	1%
Ireland/Scotland	19%	29%	52%	< 1%
Scotland	22%	29%	48%	1%

Diachrony also plays a significant role in the distinction among these four groups. Following the same procedure that I have used in combination with other variables, I have computed the noun types coined across the Middle and Modern English periods from the standpoint of the place of education, and my results are given in table 4.28. The figures concerning types per 1,000 words serve to emphasise the predominance of the group Ireland/Scotland in every single century under study. However, the gap between this group and its counterparts seems to narrow after reaching its peak in the fourteenth century, to the point of becoming only marginally valuable in the 1600s and 1700s (with a difference of 0.1 times per 1,000 words).

Table 4.28. Analysis of places of education by coinage dates.

Types per 1,000 words							
	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
England	0.2	1.7	3.2	1.5	2.2	1.4	0.4
Ireland	0.9	6.7	11.0	4.3	5.7	3.6	0.9
Ireland/Scotland	**1.4**	**10.2**	**15.6**	**6.9**	**7.6**	**3.7**	**1.0**
Scotland	0.3	2.1	4.5	2.2	3.5	2.2	0.7
Percentages							
	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
England	2%	16%	30%	14%	21%	13%	3%
Ireland	3%	20%	33%	13%	17%	11%	3%
Ireland/Scotland	3%	22%	34%	15%	16%	8%	2%
Scotland	2%	14%	29%	14%	22%	14%	5%

In the second section of the table percentage numbers circumscribe the highest rates to the fourteenth century. In contrast with other variables studied in this research work, the margins between nouns coined in the fourteenth and sixteenth centuries are relatively wider. Moreover, the number of nouns used for the first time in the thirteenth century exceeds those whose coinages are dated in the sixteenth century, especially when dealing with figures related to Ireland and Ireland/Scotland. The language used by authors included in these latter two groups can be considered more traditional not only because they include nouns coined in earlier centuries, but also because they present the lower frequencies of nouns belonging to class I, which are the most recently coined, as seen in table 4.27 above. In order to establish a pattern in noun use that qualifies the latter group, I have compared its percentages with an average sum of the other three, all of which is shown in figure 4.40. In the figure positive numbers indicate higher rates in Ireland/Scotland, and negative numbers denote higher rates in the combination of the other three groups.

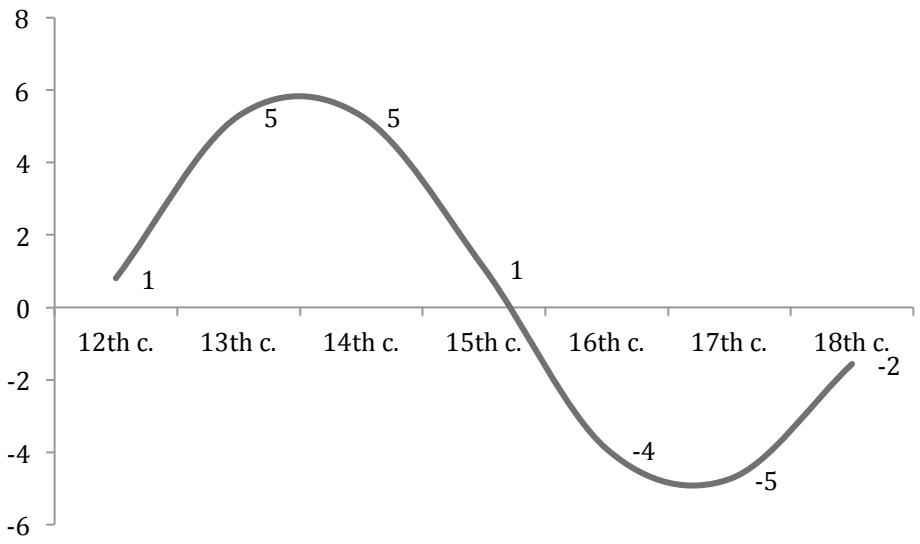


Figure 4.40. Differential coefficients in coinage dates by places of education (%).

As seen in the figure, samples in the group Ireland/Scotland show positive numbers in the first four centuries and negative in the last three. This implies that authors in this group favour nouns coined in the Middle English period, whereas on average, the rest of authors in the other three groups select nouns used for the first time in the Modern English period.

3.5. Age of the author

The last of the variables studied in this research work deals with the ages of the authors when their works were published. Evidently, this analysis makes little sense if we take every different age individually, because results would be biased by idiosyncratic uses of the language by every author. So it needs some

kind of arrangement because I intend to establish significant patterns that encompass several authors. As is acknowledged in the metadata section in the *Coruña Corpus Tool*, the ages of seven of our authors in the eighteenth-century blocks of *CETA* and *CEPhiT* are unknown to us —those of George Adams, Jasper Charlton, Henry Curson, Samuel Fuller, George Gordon, Robert Kirkpatrick and John Lacy. But these aside, there are still twenty-four different ages associated with the remaining thirty-four authors in my corpus, which renders single-age studies rather impractical.

Therefore, I have attempted to establish four groups of ages that contain a balanced number of samples —bearing in mind that most of our authors were in their fifties when they published their works— and, at the same time, that may be representative of different stages in the writing activity of these scientists. These four groups include the texts written by the authors whose ages are: a) under 36 years old, b) between 36 and 45 years old, c) between 46 and 55 years old, and d) over 55 years old. This division allocates seven samples (70,638 words) to the first group, nine samples (91,552 words) to the second, twelve samples (123,276 words) to the third, and six samples (55,722 words) to the fourth. Table 4.29 shows the details of the four groups of ages as far as types and tokens are concerned.

Table 4.29. Types and tokens by age of the author.

	Tokens	Types	Type-token ratio	Tokens/ 1,000 words	Types/ 1,000 words	*hapax leg.*/1,000 words
Ages < 36	13,416	2,119	**6.3**	189.9	30.0	13.2
Ages 36 – 45	17,798	1,990	8.9	194.4	21.7	8.4
Ages 46 – 55	24,289	2,703	9.0	197.0	21.9	9.1
Ages > 55	11,466	1,765	6.5	**205.8**	**31.7**	**13.7**

Authors over 55 years of age make use of a higher frequency in three of the categories represented in the table. In their writings they include more tokens (205.8 per 1,000 words), more types (31.7 per 1,000 words) and more *hapax legomena* (13.7 per 1,000 words) than any of the other three groups. The last two records may be interpreted in connection with their higher mastery of the language, which they may have achieved over more years than other authors. Also, they are only second to authors under 36 years of age as regards the type-token ratio. Significantly enough, it is the latter authors that follow the frequencies shown by the group of seniors. Furthermore, these authors show the lowest type-token ratio (6.3) of all groups.

The high frequencies shown by authors in the junior group prevent the establishment of meaningful patterns for most elements in the table. Except for the increasing nominalisation of texts that can be observed in the column related to noun tokens per 1,000 words, which grow gradually in every age

group, the other columns only follow rising patterns in the last three age groups.

If additional elements are added to the equation there will be more information that may help to characterise the writings of authors in relation with their ages. In order to do this, now I will proceed to filter the nouns used in their texts by morphological processes. The details are given in table 4.30 and the highest values have been printed in bold characters for ease of interpretation.

Table 4.30. Morphological processes by age of the author.

Ages < 36					
	Simple	Complex	Compound	Zero	Other
Tokens	6,300	6,429	180	495	12
Types	717	1,181	62	151	8
% Tokens	47%	48%	1%	4%	< 1%
% Types	34%	56%	3%	7%	< 1%
Type-token ratio	**8.8**	5.4	2.9	**3.3**	**1.5**
Tokens/1,000 words	89.2	91.0	2.5	7.0	0.2
Types/1,000 words	**10.2**	16.7	0.7	2.1	0.1
Ages 36 - 45					
	Simple	Complex	Compound	Zero	Other
Tokens	8,345	8,695	196	550	12
Types	650	1,136	62	141	7
% Tokens	47%	49%	1%	3%	< 1%
% Types	33%	57%	3%	7%	< 1%
Type-token ratio	12.8	7.7	3.5	3.9	1.7
Tokens/1,000 words	91.2	**95.0**	2.1	6.0	0.1
Types/1,000 words	7.1	12.4	0.6	1.5	0.1

Ages 46 - 55					
	Simple	Complex	Compound	Zero	Other
Tokens	5,530	5,041	235	644	16
Types	891	1,472	128	197	15
% Tokens	50%	45%	1%	4%	< 1%
% Types	33%	54%	5%	7%	1%
Type-token ratio	13.7	7.4	**2.5**	4.8	1.7
Tokens/1,000 words	98.9	87.8	2.5	7.6	0.2
Types/1,000 words	7.2	11.9	**1.0**	1.6	0.1
Ages > 55					
	Simple	Complex	Compound	Zero	Other
Tokens	12,188	10,822	314	939	26
Types	520	1,055	49	134	7
% Tokens	48%	44%	2%	6%	< 1%
% Types	29%	60%	3%	8%	<1%
Type-token ratio	10.6	**4.8**	4.8	4.8	2.3
Tokens/1,000 words	**99.2**	90.5	**4.2**	**11.6**	**0.3**
Types/1,000 words	9.3	**18.9**	0.9	**2.4**	0.1

I will not pay attention to absolute figures at this point, because the differences in the total number of words for the four groups will consequently result in higher numbers in those containing more numerous samples. Authors in the first group show the following characteristics: their type-token ratios are the lowest in three of the five processes studied —simple (8.8 times), zero-derived (3.3 times) and 'other' nouns (1.5 times)— and the highest frequency of simple noun types (10.2 per 1,000 words). Authors in the second group use complex nouns the most (95.0 tokens per 1,000 words) and reuse them the most as well (7.7 times each). Those in the third group show the higher frequencies in the

use of compounds (1.0 types per 1,000 words) and, accordingly, the lowest type-token ratio in nouns obtained by means of the same process (2.5 times). Authors over 55 years present the highest frequency of simple noun tokens (99.2 per 1,000 words), compounds (4.2 per 1,000 words), conversions 11.6 and 'other' (0.3 per 1,000 words). Also, they present the highest frequencies of zero-derived noun types (2.4 per 1,000 words) and, more importantly complex nouns (18.9 types per 1,000 words). This is especially significant for the next stage of my analysis, or the subdivision in complex words by affixation classes.

As seen in the table, noun complexity plays a capital part in the writings of senior writers. The percentages involving types in all groups suggest that the relationship between simple and complex types varies depending on age, i.e. simple nouns experiment a decrease and complex nouns show an increasing tendency with age, as shown in the following figure.

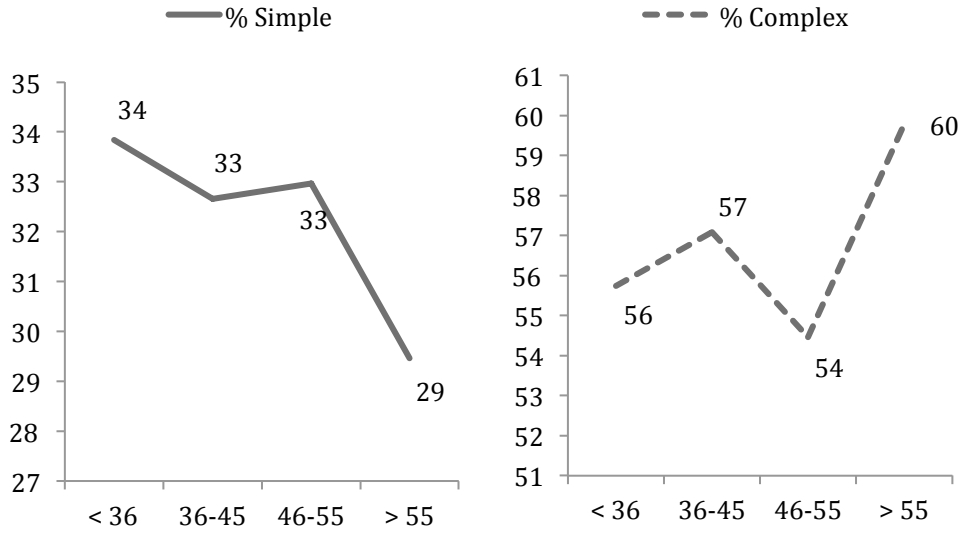


Figure 4.41. Simple and complex nouns by age of the author (%).

Further filtering does not seem to add much more information to the comparison and contrast of this variable. The details of affixation classes provided in table 4.31 below demonstrate that authors under 36 years old show the highest frequency in nouns belonging to class I (3.9 per 1,000 words), whereas authors over 55 years old seem to prefer those in class III (9.8 noun types per 1,000 words) and, to a lesser extent, in class II (5.3 per 1,000 words). Frequencies for affixation Class IV are rather low and show similar figures across the four groups of ages.

In the second section of the table, percentages show that authors whose ages range between forty-six and fifty-five years use nouns with free bases (classes I and II) more frequently than those with bound bases (classes III and

IV), in contradistinction with authors from other groups. However, differences are minimal.

Table 4.31. Affixation classes by age of the author.

Types per 1,000 words				
	Class I	Class II	Class III	Class IV
Ages < 36	**3.9**	4.3	8.3	0.1
Ages 36 – 45	2.7	3.3	6.3	0.2
Ages 46 – 55	3.0	3.1	5.7	0.2
Ages > 55	3.7	**5.3**	**9.8**	0.2
Percentages				
	Class I	Class II	Class III	Class IV
Ages < 36	24%	26%	50%	< 1%
Ages 36 – 45	22%	26%	51%	1%
Ages 46 – 55	**25%**	26%	47%	1%
Ages > 55	19%	**28%**	**52%**	1%

Finally, as can be inferred from table 4.32 below, which includes the details of diachrony in the corpus by age of the author, there are no observable tendencies in any of the groups.

Table 4.32. Analysis of coinage dates by ages of the authors.

Types per 1,000 words							
	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
Ages < 36	0.4	3.4	6.4	3.0	4.0	2.5	0.5
Ages 36 – 45	0.3	2.6	4.6	2.1	3.1	1.8	0.4
Ages 46 – 55	0.3	2.5	4.6	2.1	3.2	1.8	0.6

Ages > 55	**0.5**	**3.5**	**7.0**	**3.3**	**4.9**	**2.7**	**0.7**
Percentages							
	12th c.	13th c.	14th c.	15th c.	16th c.	17th c.	18th c.
Ages < 36	2%	17%	31%	15%	20%	12%	2%
Ages 36 – 45	2%	17%	31%	14%	21%	12%	2%
Ages 46 – 55	2%	16%	30%	14%	21%	12%	3%
Ages > 55	2%	16%	31%	15%	22%	12%	2%

On the one hand, frequencies only emphasise the characteristics explained above, the most relevant of which determines that authors over 55 years of age make use of the highest number of types per 1,000 words consistently throughout the Middle and Modern English periods. On the other hand, percentages are practically equivalent in the fourth age group, and margins do not generally exceed one per cent.

Summing up, in this chapter I have described a comprehensive analysis on my corpus. Besides, I have scrutinised each relevant piece of data with care so that I can provide meaningful patterns of word formation. Finally, I have dealt with intra- and extralinguistic variables that may clarify to some extent the evolution of noun formation in the scientific register of the English language across several centuries of its history, putting special emphasis on the eighteenth century. In the next chapter I will present my concluding remarks.

5. Concluding remarks

The preceding chapters have dealt broadly with the causes and consequences of language change, with a more detailed focus on the Scientific Revolution and the adoption of English as a means to disseminate scientific knowledge in English-speaking areas in the eighteenth century (chapter 1). In them I have described the morphological processes involved in the coining of new words in English, and how new nouns can be formed to enrich the language (chapter 2). Also, I have reviewed the computer tools used and the section of the *Coruña Corpus* selected for my research (chapter 3). And ultimately, by means of those morphological processes and those computer tools I have decomposed the corpus into minor units, i.e. nouns, bases and affixes. I have classified the linguistic data contained in the corpus so that I can establish patterns of noun formation in order to present a series of seemingly reliable results, in the hope that the methodology employed may also be used in other linguistic studies (chapter 4).

This method has also helped to determine a number of extralinguistic variables, such as the comparison of a) the two disciplines, astronomy and philosophy, b) genres/text types, c) sex of the authors, d) place of education,

and e) age of the author when the work was published. The conclusions I have reached by doing so will be presented in the following pages.

First of all, linguistic theories have so far tried to explain the contrast in the production of new vocabulary between the eighteenth and nineteenth centuries, being lower in the former than in the latter. Linguists have related this contrast to the lack of studies focused on the 1700s. However, my analysis shows a steep deceleration in the production of new vocabulary in the eighteenth century, which contradicts what I expected to find in my corpus, if we take into consideration the changes and innovation brought about by the scientific revolution. Beal (2004: 16-17) justifies this decrease in innovation by stating that the eighteenth century follows Shakespeare's times, a period of "exuberant lexical innovation" when a "large injection of vocabulary" had already taken place, therefore the "gap in the vocabulary of English had been filled". She continues by saying that even the most reputed writers such as Swift and Addison were advocates of the linguistic conservatism characteristic of the period, and campaigned against the inclusion of new words (especially monosyllables obtained from clippings in the case of Swift). Another plausible reason might be a conscious attempt to 'freeze' the English language, so that it could acquire a more prestigious status as a conveyor of science. We must bear in mind that at the time language change was viewed as language corruption, so any attempts to put English on the same level with Greek and Latin would involve bringing the evolution of the language to a halt.

Second, an extensive use of complex nouns can be observed across all the samples in the corpus, which nearly doubles the use of simple nouns (2,510 vs. 1,370), let alone compounds, conversions and others. This in itself may be acknowledged as a characteristic of scientific writing. Forthcoming research comparing language complexity in scientific and non-scientific texts might shed some light on the differences, if they exist, and might therefore result in an interesting study. In spite of the fact that more complex than simple nouns are contained in my corpus, the type-token ratio is higher in the latter than in the former. This characteristic could probably be extrapolated to other registers of the English language. It may seem too obvious that simple nouns are reused more often than complex ones, and my analysis shows quantitatively that the former are reused almost twice as many times as the latter in the scientific register. However, it does not explain why this is so. The fact is that texts require a suitable balance between clarity and complexity in order to accomplish the dissemination of science. Accordingly, the language employed must be a combination of elements familiar to readers and new ones that will keep their interest high. Simple nouns represent those familiar elements, and they are used to convey clarity and make the reader feel comfortable on safe grounds, while complex nouns embody the challenge to acquire new knowledge and. Readers interpret simple nouns as ‘ordinary’, so they are not so aware of their repetition. For this reason simple nouns are reused more frequently. Complex nouns are considered more ‘special’, and must therefore

be used sparingly, which is why they are less reused than simple ones. This also implies a higher need of vocabulary richness to prevent the feeling of recurrence.

The analysis carried out on the eighteenth century section of *CETA* and *CEPhiT* indicates an average of slightly over 343.1 nouns were coined per century. As we saw in more detail in figure 4.4 in the previous chapter, the fourteenth and sixteenth centuries were extremely productive, doubling average figures, whereas the thirteenth and the fifteenth centuries show considerably lower figures, almost half. The situation portrayed in figure 4.5 in the years following the Scientific Revolution is somewhat different, with a general decrease in the coinage of new nouns across the eighteenth century. Besides, the significance of the main processes involved in new coinages seems to have been reverted, that is to say, affixation has lost its hegemony, whereas zero derivation and, especially, compounding show a significant contrast in numbers, as observed in figures 4.25 and 4.26. Seemingly, zero derivation in the 1700s is comparable to that of previous centuries, but its numbers are still high in relation to the little affixation used, whereas the increase in compounding is considerable. As we saw in section 2.2 in the same chapter, the fact that many compounds coined in the period may be considered nonce formations indicates that compounding was a rather unconstrained process at the time and, as a consequence, writers of science attempted free creations at will. In any case, it may be necessary to contemplate lexical

innovation in the eighteenth century by analysing texts from ensuing centuries, in order to have a wider perspective. All in all, those theories putting forward a steep decrease in the production of new words in the century, due to the various reasons expounded succinctly above, and more in depth in chapter 1, seem to be on the right track if we consider the results of my analysis.

My study has shown that, if we pay attention to the word category of the free bases involved in the coinage of new nouns, verbs represent the highest number (50%), followed by adjectives (24%), and other nouns (21%). Proper nouns and other categories make up the remaining 5 per cent. The wide difference between verbal bases and those belonging to other word categories may be justified by the need for scientists to explain natural processes, both in the physical world and in the realms of the mind. Therefore, the presence of a high frequency of verbs and deverbal nouns in the corpus seems fairly understandable in both disciplines. In fact, the differences between astronomy and philosophy concerning the word category of the bases used is around ±2 per cent.

With regard to the subprocesses of affixation, suffixation is more common than prefixation because not only does it create new types, but also implies a change in the category of the resulting word. Thus, by providing new types belonging to different word categories, suffixation helps enlarging word paradigms and, consequently, contributes to language plasticity and communication to a greater extent. Besides, it becomes a key process in the

explanation and dissemination of new theories, discoveries and inventions, since it adds more accuracy to the scientific register of English, in which precision is more important than linguistic rhetoric. I have found 69 different prefixes in the corpus that produce 217 noun types, versus 120 suffixes that help coining 2,326 new noun types. Most prefixes have Greek (29) and Latin (27) origins, while most suffixes are have French (40), Latin (32) and English (23) origins.

My division of affixed nouns in four classes, and the subsequent individualised study of every class, has provided some interesting results. For example, more than 50 per cent of the complex nouns in English have been coined using the pattern free base + neutral affix (the one shown in affixation classes I and II), and from the sixteenth century onwards this has been the most regular pattern. If we pay attention to the situation in the century covered by the samples in my corpus, roughly 90 per cent of the nouns coined between 1700 and 1799 contain free bases. Complex nouns formed by a bound base + non-neutral affix (affixation class III) reached its peak in the fourteenth century (58%) and from this moment they suffered a steep descent that continued during the eighteenth century (4%). Also, nouns included in affixation class III still show the highest type-token ratio. My speculation is that the presence of bound bases makes the border between the base and the affix rather diffuse. Consequently, this circumstance may contribute to segregate nouns belonging

to this class still further from those in affixation classes I and II, and at the same time brings them closer to simple words.

The linguistic materials involved in noun formation have also been analysed with regard to their origin. To do so, I have divided them in six groups under the labels of Anglo-French, French, English, Greek and Latin, and computed their occurrences. I have also subdivided these occurrences into bases and affixes and, ultimately, affixes were in turn subdivided into prefixes and suffixes. Each of these subdivisions was accounted for and examined in depth. As regards noun bases, French provides the highest number in all three base + affix subprocesses, that is, affixation classes I, II and III, especially in the latter two, whereas native English bases are on a similar —but lower— level in affixation class I. Nevertheless, the Romance component of materials regarding bases remains predominant in the three subprocesses, given that Latin and Anglo-French also contribute a moderately high number of instances. The above said does not apply to affixation class IV, which is composed exclusively by Greek and Latin bound elements. However, as we have seen, some of those bound elements have become free bases of their own, as is the case of *scope*. In this respect it must be acknowledged that my reputable etymological source has been the *OED*, and I have respected the information contained in it. Notwithstanding this, it must be acknowledged that in the eighteenth century authors had a significant knowledge of Latin, and a good number of them still published works in Latin as well, so they were more likely

to resort to Latin linguistic materials in order to create new words than to those of French. It might be, then, that materials labelled as French by the *OED* are ultimately Latin, French being merely the vehicle that conveyed them through history.

If we take the complexity of bases into consideration, new nouns are predominantly coined by the addition of simple bases. These coinages represent 86 per cent of the total, while the remaining 14 per cent contains complex bases. These have originated from derivations (57%), zero derivations (31%) and compounds (6%), while bases obtained by means of other processes make up for the remaining 6 per cent.

Regarding affixes, the situation varies greatly between prefixes and suffixes. While most prefixes —roughly 81 per cent— are of Greek and Latin origin, and almost half the prefixed nouns were coined using Latin prefixes in absolute numbers, most suffixes —slightly over 33 per cent— come from French. Notwithstanding this, it is English materials that present the highest productivity per prefix. In fact, every native prefix produces over four new nouns on average. The same can be said about suffixes. It must be noted here that *-ion* was by far the most productive affix for centuries, and that this suffix alone is present in 413 nouns in the corpus. But leaving *-ion* aside, native English materials also produce the highest number of nouns per suffix, over 25 on average. This feature implies that even though a lot of foreign linguistic elements entered the language across the centuries, scientists have always

resorted to native materials to make up complex words in English, in spite of the fact that it was not easy for English to compete with Latin as the vehicle to disseminate scientific knowledge.

The productivity of *-ion*, in decline since the fifteenth century, reached its lowest record in the eighteenth century, when only one instance has been found: *cultivation* (Bryan 1797: 122). The most productive suffixes in the latest century under study are *-ing* (7 instances), *-ity* (5), *-ism* (5) and *-ment* (4), while the most productive prefix is *semi-* (7). We might explain the decline of suffixes such as *-ion* from the prism of neutrality or non-neutrality explained in chapter 2. It generally attaches to Latin past participles as a non-neutral suffix, and on many occasions we may find that the distinction between *-ation*, *-tion*, and *-ion* is not very straightforward. Contrarily, the most productive suffixes in the eighteenth century are all neutral, so the segmentation between the base and the affix is clearly seen, so readers will find it easier to relate the derived word to the original one. Once again, we must remember that men of science in this period intended to address their writings to the widest audience possible, so simplicity of language became a must. Using linguistic elements closer to readers seems a cleverly valid way to achieve this goal.

Focusing on cross-origin creations, most contributing languages can combine with others to a certain extent. Greek, Anglo-French and French bases and affixes become attached to elements coming from other origins more frequently than with their own, which confers them with a high degree of

versatility. English bases and affixes combine more or less freely with virtually all materials from other origins, with the exception of Greek. I have found neither instances of Greek prefix + English base, English prefix + Greek base nor English base + Greek suffix in the corpus, and only one instance of English base + Greek suffix. Notwithstanding the above, English materials tend to combine among themselves much more often than with others in scientific writing. On the one hand, this may be due to the fact that authors writing in their native English were more conscious of the smaller units that make up words and their potential combinations. Writing in Greek, on the other hand, may have implied the implicit use of words as full units.

Fourth, comparing and contrasting the morphological features of the two disciplines have provided some unanticipated results. The two disciplines include a similar number of noun tokens, showing *CETA* slightly higher figures than *CEPhiT* —208,079 vs. 200,902 noun types respectively. But contrary to my expectations, philosophy contains significantly more types than astronomy. Out of the 4,458 noun types present in the corpus, 2,060 belong to *CETA* samples and 3,777 belong to *CEPhiT*, while the two of them share 1,279 types. When developing the analysis further, we can also observe that my astronomy samples contain a higher percentage of simple noun types (36% vs. 30%) and compounds (6% vs. 4%), whereas the philosophy section contains a higher percentage of complex noun types (57% vs. 50%) and zero-derived (9% vs. 6%). As a consequence of all of this, the type-token ratio must be clearly

higher in *CETA*. In fact, astronomers reuse each simple noun type 29.5 times, and every complex noun type 13.6 times. Conversely, figures for philosophers are much lower and more balanced. In their samples simple nouns are reused 10.9 times and complex nouns 9.9 times. Furthermore, the number of *hapax legomena* in philosophy texts reaches almost 30 instances per 1,000 words, versus 10 instances per 1,000 words in astronomy texts. These unforeseen results in astronomy and philosophy may be explained by a higher verbosity in the latter discipline and, as a consequence, a wider set of nouns is required to prevent reiteration. We must not forget that attempts at explaining facts that do not belong to tangible nature, so complexity becomes also one of its characteristics. In contraposition, observational sciences like astronomy can convey meanings with a less elaborate discourse, since readers can relate nouns and physical objects in a more straightforward manner.

I have also observed differences between the two disciplines at the levels of affixation classes and the period when these nouns were first coined. On the one hand *CETA* presents more affixed nouns belonging to classes III and IV than *CEPhiT*, that is to say, the complex types of the former subcorpus were primarily coined using bound elements, either bases or affixes. On the other hand *CEPhiT* shows more affixed nouns belonging to classes I and II than *CETA*, which means that nouns in the astronomy subcorpus were primarily made up with free elements. My expectations to find differences between the two disciplines were not fulfilled with regard to other variables.

For example, the origins of the linguistic materials used in coinages in astronomy texts practically mirror those in philosophy.

As stated above, minor differences also arise when paying attention to the period in the history of the English language from which both disciplines extract their nouns. Whereas *CETA* retrieves its types more often from the earlier centuries, i.e. the thirteenth or fourteenth, *CEPhiT* shows a more recurrent use of nouns coined in the later three centuries. Both show even numbers in the eighteenth century. The reason for this difference may need to be ascribed to the nature of the disciplines themselves. Astronomy as science has dealt with celestial objects throughout the centuries, and therefore its vocabulary has been established and fixed in early times. As I have already explained in chapter 1, philosophy suffered major changes and restricted its scope further —for example, cosmology was excluded from its study—, so it focused mainly on epistemological issues. Its evolution has been affected by several schools of thought and changing reality and values, and has therefore diversified its scope, which now covers topics related to knowledge and the mind, reason, social values and so on. This multiplicity of topics can also be observed in the philosophy samples in the corpus if we compare *CEPhiT* with *CETA*.

Fifth, if we follow the classification of samples proposed by the *Coruña Corpus* with respect to the variable "genre/text-type", they fall into eight different categories: lectures, dialogues, treatises, textbooks, essays, journal

articles, letters, and 'others', which contains a dictionary in this particular case. My analysis shows that journal articles provide the highest number of types per 1,000 words (58) and, as a consequence, the lowest type-token ratio (2.9). Texts falling into this category are generally addressed to a learned audience, and therefore repetition may not be as necessary as that of other genres. On the contrary, textbooks show the lowest number of types per 1,000 words (15.9), the highest number of noun tokens per 1,000 words (202) and the highest type-token ratio (12.7). Textbooks are addressed to students without extensive knowledge of the subject matter, and besides, repetition is still a common resource in didactics nowadays. Therefore the figures presented by the samples of this text type are not too surprising.

Most genres/text-types contain more complex than simple nouns, with the exception of the dictionary in 'others' (Hill 1754). Complex nouns are used more often than simple nouns by only one percentage point in dialogues. On the other side of the spectrum, essays show the widest margin between simple and complex nouns (29% vs. 61%, respectively). The remaining genres/text-types, however, show margins in the range of twenty percentage points.

Treatises and essays are, on the one hand, the only genres/text-types that include more nouns showing coinages with free bases than nouns with bound bases, that is to say, percentages for complex nouns belonging to affixation classes I and II combined represent over fifty per cent of the total.

On the other hand, 63 per cent of the nouns in 'others', and 62 per cent in letters have been coined using bound elements.

Unfortunately, no relevant results have been obtained with the last three analyses on genres/text-types involving noun choice: i) period in the history in which nouns were coined, ii) origin of the linguistic components, and iii) processes by 30-year periods in the eighteenth century. The first one shows that nouns coined in the fourteenth century are preponderant in all genres/text-types, so no contrasting data can be shown. As regards the second one, my review on the origin of components gives the same results as the comparison between disciplines explained above, so I will not repeat it here. Finally, the production of nouns in the eighteenth century is so low that it precludes subdividing the century further.

Sixth, the variable "sex of the author", included here in order to assess the differences —if any— between texts written by men and women scientists, has also produced relevant results. Women do not seem to lack linguistic ability compared to their male counterparts, in spite of being banned from high education, but rather on the contrary. As a matter of fact, the average type-token ratio is only 4.7 among women versus an average 16.7 among men. Also, the text written by Mary Wollstonecraft (1792) shows the lowest type-token ratio in the whole corpus, which can be interpreted as a proof of her excellent proficiency in language. Furthermore, as we saw in the previous chapter women make use of more complex nouns, they write nearly four times as many

different noun types, as well as almost five times as many *hapax legomena* as men. All indicators prove that women were at the least on the same level as men —higher, rather— as far as linguistic competence is concerned. All in all, the general characteristic that we can draw from these figures is that men scientists make use of more simple and compound nouns, whereas women scientists employ more affixed nouns in their writings.

If we look to more differences between men and women by paying attention to the period when the nouns used were coined, we will come to the conclusion that men use more "modern" vocabulary, because they use more nouns occurring for the first time in the sixteenth, seventeenth and eighteenth centuries, whereas women's figures surpass men's in the thirteenth and fourteenth centuries.

There is no need to say that women were subject to great difficulties to research and publish their works in the 1700s. If they had the intention to be taken seriously as scientists in a world monopolised by men they needed to prove their excellence in every respect. Therefore their use of a higher number of complex nouns may have followed an intentional strategy to show their mastery of the language to their male contemporaries. Regarding their use of nouns coined in earlier centuries, the explanation may be twofold. First, we

must bear in mind that they had no access to universities or their libraries[41], so their autodidactic education had to rely on private collections belonging to relatives, who were not always likely to own the latest scientific works published. This might justify their use of a more traditional vocabulary when compared to men, since women may have had to resort to older sources, whereas men could access to the most recent materials. Second, being a woman and a scientist was intrepid enough in the eighteenth century. Thus, attempting to be an innovator of the language as well may have been too revolutionary at the time. Consequently women may have decided to explain innovation in science using well-established linguistic elements in their writings, in order to remain on the safe side and attract as little negative criticism as possible.

Seventh, when applying the variable "place of education" to the authors reviewed in this dissertation, we must bear in mind that it may not be a coincidence that most astronomers in the eighteenth century were educated in England. We cannot ignore the significance that the creation of the Greenwich

[41] In Spain, for example, women had to wait until Antonia Gutiérrez Bueno obtained the right to make use of the Biblioteca Nacional de España in 1837, which had been prohibited until that moment. By then she had already published a historical and biographical volume on outstanding women, and a collection of medical articles that she had translated from French (García-Ejarque 1997: 528).

Observatory had on astronomy studies on the Isles. Likewise, the nomination of John Flamsteed as first Astronomer Royal was to become a milestone in the social recognition of astronomers and mathematicians, who had until then been relegated to secondary roles in academic environments, as I have already mentioned in chapter 1. The same can be said about the higher number of philosophers educated in Scotland, given that the universities of Edinburgh, Glasgow and Aberdeen were key in the emergence of the Scottish Enlightenment. This needs be said as a partial disclaimer, because the figures obtained by studying this variable may be too similar to those of the disciplines themselves.

After examining the four groups established for the analysis, that is, 'England, Ireland, Ireland/Scotland' and 'Scotland', I have realised that authors educated in 'Ireland/Scotland' show the lowest type-token ratio. Noun types in this group are reused 3.5 times. Besides, these samples contain the highest number of nouns tokens, 227.6 per 1,000 words. They also show the highest frequencies of noun types (64.2 per 1,000 words) and *hapax legomena* (29.4 times per 1,000 words).

Conversely, authors educated in England reuse their nouns 12.3 times. Also, samples in this group contain the lowest numbers of noun tokens (187.6 per 1,000 words), noun types (15.2 per 1,000 words) and *hapax legomena* (6.0 per 1,000 words) in the corpus. Between these two groups, we can find authors educated in Ireland, which show similar numbers to those from

Ireland/Scotland, and authors educated in Scotland on a step lower to the previous.

The highest use of complex nouns can be found in texts from the group 'Scotland' (61%), while the highest number of simple nouns belongs to the group 'Ireland'. Thus, not only do authors educated in Ireland/Scotland make use of the highest number of noun types, but also their nouns are principally complex. A closer look on complex nouns divided by affixation classes will show that samples in this group present the highest frequencies in every class, except for class IV.

Samples included in the group 'England' contain the highest percentage of nouns made up with free bases (52%), especially significant is the fact that most of their nouns belong to affixation class I (28%), almost ten percentage points more than samples in the other groups.

When I analysed this variable from the standpoint of diachrony, I observed that nouns coined in earlier centuries predominate in the texts written by authors educated in Ireland/Scotland. Their percentages are higher than the other groups in the thirteenth, fourteenth and fifteenth centuries. In fact, the combined percentages of these centuries make up 71 per cent of the total nouns included in these samples. This situation is reverted in the following three centuries, when this group shows lower numbers than the other three. In the last three hundred years under study, contrariwise, samples belonging to authors educated in Scotland include more modern coinages, and the combined

percentages of the sixteenth, seventeenth and eighteenth centuries reach 41 per cent.

It must be noted that only one scientist, Francis Hutcheson, forms the group of authors educated in Ireland/Scotland. His relevance as one of the key figures of the Scottish Enlightenment and the influence that he had on other prominent philosophers such as Hume can explain his mastery of the language and therefore the figures shown above. However, this does not apply to his preference for words coined in earlier centuries. It may have just been his personal choice.

Eighth, the age of the author is the last variable that I have paid attention to in this dissertation. And again this variable has rendered some relevant results. Similarly to what I did in the section above, I divided authors in four groups. As already explained, the first group is formed by authors under 36 years of age; the second consists of authors between 36 and 45; the third includes authors between 46 and 55; and the fourth is made up with those older than 55. When computing the number of types, tokens, type-token ratio and hapax *legomena* it seems clear that those authors in the fourth group have a higher mastery of the language. It is understandable that at a later age, writers will have improved their writing techniques and, at the same time, they are expected to know and make use of a richer vocabulary.

My data shows that these authors include a higher frequency of nouns in three of the above categories in normalised figures: tokens (205.8), types

(31.7) and *hapax legomena* (13.7). Besides, their type-token ratio is the second highest. However, I cannot establish a definite pattern for a variable so far, due to an unexpected feature. It is true that a linear progression can be observed in all figures ranging from the second to the third group, and from the third to the fourth as well; however, the first group —formed by authors under 36— contradicts the above by showing the lowest type-token ratio (6.3) and the second highest frequencies regarding tokens, types and *hapax legomena*. Four astronomers and three philosophers under 36 years of age, five English and two Scottish, form this heterogeneous group without any apparent connecting bonds. The key to this puzzle, though, may be found in the city of London, to which the seven authors are closely related. It is not clear where Curson and Morden were born, but both of them worked and published their writings in London; Nicholson and Wollstonecraft were born in the city; Bonnycastle studied there; finally, Cheyne and Crombie moved from their homeland in Scotland to London in their youth. The importance of the capital in terms of publications may have had an impact in the number of texts available for the younger generation of scientists, who had access to more materials written by more numerous authors, and could therefore expand their scientific vocabulary.

Texts written by authors over 55 years of age show the highest normalised frequencies of complex (18.9) and zero-derived nouns (2.4). Differences regarding compounding and other processes are not significant. Further filtering does not seem to add much more information to the

comparison and contrast of this variable. The details of affixation classes provided in table 4.31 below demonstrate that authors under 36 years old show the highest frequency in nouns belonging to class I (3.9 per 1,000 words), whereas authors over 55 years old seem to prefer those in class III (9.8 noun types per 1,000 words) and, to a lesser extent, in class II (5.3 per 1,000 words). Frequencies for affixation Class IV are rather low and show similar figures across the four groups of ages.

Percentages in my tables show that authors whose ages range between forty-six and fifty-five years use nouns with free bases (classes I and II) more frequently than those with bound bases (classes III and IV), in contradistinction with authors from other groups. However, differences are minimal. The same happens when resorting to percentages in affixation classes. In spite of the fact that texts written by authors between forty-six and fifty-five make a more frequent use of nouns with free bases (classes I and II) than those with bound bases (classes III and IV), differences in numbers are not significant.

In any case, and for contrasting purposes, it is true that nouns belonging to affixation class I predominate among the youngest authors (3.9 per 1,000 words), whereas nouns in class III (9.8 noun types per 1,000 words), and —to a lesser extent— those in affixation class II (5.3 per 1,000 words), seem to be preferred by the elder authors and. Frequencies for affixation Class IV are too low and uniform among all writers in the corpus to deserve further recognition here.

Finally, checking every word in my corpus against the terminological database of the *OED* has proven that a few of the first coinage dates of the matching entries in the dictionary have been recorded incorrectly. This is inevitable since its compilers cannot review each and every written document in existence. Although, there can be no doubt that research works like the one that I have carried out, increasingly more numerous, will help improving the accuracy of coinage dates in the dictionary. In order to do so, I have included below a list of inaccuracies containing information about the subcorpus, author and sample where I found them, which also shows the *OED* date of first use and the earlier date recorded. All of this can be seen the Appendix.

Also, we can suggest the inclusion of coinages that have been left out for whichever reasons either as entries or in the entries of their base nouns, as is the case of *sub-zenith* (*CETA*, Fuller 1732: 20) and *semi-orbit* (*CETA*, Costard 1767: 281). Other words not listed are *almantar* (*CETA*, Hill 1754: 7) *antaeci* (CETA, Fuller 1732: 25), *ascendency* (*CEPhiT,* Campbell 1776: 34), *atriplex* (*CEPhiT*, Smellie 1790: 10), *ayenia* (*CEPhiT*, Smellie 1790: 10), *blending* (*CEPhiT*, Ferguson 1769: 109), *carystian* (CEPhiT, Reid 1764: 26), *diacosm* (*CEPhiT*, Greene 1727: 13), *Elean* (*CEPhiT*, Reid 1764: 25), *empirics* (*CEPhiT*, Campbell 1776: 6), *henriade* (*CEPhiT*, Hume 1748: 38), *oviparous*[42]

[42] We can consider *oviparous* as an elliptical adjective, but the *OED* has already accepted *viviparous* as a noun. I reckon that *oviparous* should therefore obtain the same recognition.

(*CEPhiT*, Smellie 1790: 47), *procreating* (*CEPhiT*, Hutcheson 1755: 177), *solicism* (*CEPhiT*, Dunton 1710: 4), *tychonick* (*CETA*, Morden 1702: 6). I have excluded from this list all the compounds not recorded in the *OED*, because authors tended to be very creative in the eighteenth century, and most of their nonce-formations did not survive.

I hope that this doctoral dissertation may have shed some light on the idiosyncrasy of noun formation in the eighteenth century in the scientific register of English. I also hope that the conclusions that I have reached after analysing my data and the methodology used may help future research works on morphology in the late Modern English period, and may serve as the subject matter for contrastive studies with other periods of the history of the English language.

Bibliography

Primary sources

The Corpus of English Texts on Astronomy

Adams, George (1777). *A Treatise describing the construction and explaining the use of celestial and terrestrial globes. Designed to illustrate, in the most easy and natural manner the phoenomena of the Earth and Heavens, and to shew that correspondence of the two spheres. With a great variety of astronomical and geographical problems.* London: printed and sold by the author.

Bonnycastle, John (1786). *An Introduction to Astronomy in a Series of Letters.* London: J. Johnson.

Bryan, Margaret (1797). *A compendious system of astronomy in a course of familiar lectures; in which the principles of that science are clearly elucidated, so as to be intelligible for those who have not studied the Mathematics. Also trigonometrical and celestial problems, with a key to the ephemeris, and a vocabulary of the terms of science used in the lectures, which latter are explained agreeably to their application in them.* London: M. Bryan.

Charlton, Jasper (1735). *The Ladies Astronomy and Chronology, in Four Parts*. London: T. Gardner.

Costard, George (1767). *The history of astronomy, with its application to geography, history, and chronology; occasionally exemplified by the globes.* London: J. Lister.

Curson, Henry (1702). *The theory of sciences illustrated; or, the grounds and principles of the seven liberal arts: grammar, logick, rhetorick, musick, arithmetick, geometry, astronomy*. London: R. Smith.

Ferguson, James (1756). *Astronomy explained upon Isaac Newton's principles and made easy to those who have not studied mathematics.* London: printed for, and sold by the author.

Fuller, Samuel (1732). *Practical Astronomy, in the description and use of both globes, orrery and telescopes.* Dublin: Samuel Fuller.

Gordon, George (1726). *An introduction to geography, astronomy, and dialling. Containing the most useful elements of the said sciences, adapted to the meanest capacity, by the description and uses of the terrestrial and celestial globes. With an introduction to chronology*. London: J. Senex.

Harris, John (1719). *Astronomical Dialogues between a Gentleman and a Lady*. London: T. Wood.

Hill, John (1754). *Urania: or, a compleat view of the heavens; containing the antient and modern astronomy, in form of a dictionary: Illustrated with a great number of figures.* London: T. Gardner.

Hodgson, James (1749). *The theory of Jupiter's satellites, with the Construction and use of the tables for computing their eclipses.* London: W. and J. Mount & T. Page.

Lacy, John (1779). *The universal system: or mechanical cause of all the appearances and movements of the visible heavens: shewing the true powers which move the earth and planets in their central and annual rotations. With a dissertation on comets, the nature, cause, matter, and use of their tails, and the reasons of their long trajectories: likewise an attempt to prove what it is that moves the sun round its axis.* London: J. Buckland.

Long, Roger (1742). *Astronomy, in five books* (I). Cambridge: R. Long.

Morden, Robert (1702). *An Introduction to astronomy, geography navigation, and other mathematical sciences made easie by the description and uses of the cœlestial and terrestrial Globes.* London: R. Morden & R. Smith.

Nicholson, William (1782). *An Introduction to Natural Philosophy* (I). London: J. Johnson.

Stewart, Matthew (1761). *Tracts, physical and mathematical: containing, an explication of several important points in physical astronomy and a*

new method for ascertaining the sun's distance from the earth. London: A. Millar & J. Nourse.

Vince, Samuel (1790). *A Treatise on Practical Astronomy.* Cambridge: J. and J. Merrill, J. Nicholson & W. Lunn.

Watts, Isaac (1726). *The knowledge of the heavens and the earth made easy: or, the first principles of astronomy and geography explain'd by the use of globes and maps: with a solution of the common problems by a plain scale and Compasses as well as by the globe.* London: J. Clark, R. Hett & R. Ford.

Whiston, William (1715). *Astronomical Lectures* (I). London: R. Senex & W. Taylor.

Wilson, Alexander (1773). Observations on the Solar Spots. *Philosophical Transactions (1683-1775)*, *XLIV*, 1-19.

The Corpus of English Philosophy Texts

Astell, Mary (1700). *Some Reflections Upon Marriage.* London: John Nutt.

Balguy, John (1733). *The law of truth: or, the obligations of reason essential to all religion.* London: J. Pemberton.

Bolingbroke, Henry (1754). *The Philosophical Works of the late Right Honorable Henry St. John, Lord Viscount Bolingbroke.* London: D. Mallet.

Burke, Edmund (1770). *Thoughts on the cause of the present discontents.* London: J. Dodsley.

Butler, Joseph (1736). *The analogy of religion, natural and revealed, to the constitution and course of nature.* London: Knapton.

Campbell, George (1776). *The Philosophy of Rhetoric* (I). London: W. Strahan & T. Cadell.

Cheyne, George (1705). *Philosophical Principles of Natural Religion: Containing the Elements of Natural Philosophy, and the Proofs for Natural Religion, arising from them.* London: G. Strahan.

Collins, Anthony (1717). *A Philosophical Inquiry Concerning Human Liberty.* London: R. Robinson.

Crombie, Alexander (1793). *An Essay on Philosophical Necessity.* London: J. Johnson.

Dunton, John (1710). *Athenianism: or, the new projects of Mr. John Dunton.* London: Tho. Darrack.

Ferguson, Adam (1769). *Institutes of Moral Philosophy.* Edinburgh: A. Kincaid & J. Bell.

Greene, Robert (1727). *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* (I). Cambridge: Cambridge University Press.

Hume, David (1748). *Philosophical Essays Concerning Human Understanding.* London: A. Millar.

Hutcheson, Francis (1755). *A System of Moral Philosophy* (II [book III]). Glasgow: R. and A. Foulis.

Kirkpatrick, Robert (1730). *The Golden Rule of Divine Philosophy.* London: R. Kirkpatrick.

Macaulay Graham, Catharine (1783). *A Treatise on the Immutability of Moral Truth.* London: A. Hamilton.

Reid, Thomas (1764). *An Inquiry into the Human Mind, On the Principles of Common Sense.* Edinburgh: A. Millar.

Smellie, William (1790). *The Philosophy of Natural History* (I). Dublin: William Porter.

Turnbull, George (1740). *The Principles of Moral Philosophy. An Enquiry into the Wise and Good Government of the Moral World.* London: John Noon.

Wollstonecraft, Mary (1792). *Vindication of the Rights of Woman: With Structures on Political and Moral Subjects.* London: J. Johnson.

Works cited

Abney, Steven (1987). *The English Noun Phrase in its Sentential Aspect.* Ph.D. Thesis. Massachussetts Institute of Technology, Cambridge, Ma.

Adams, Valerie (1973). *An Introduction to Modern English Word Formation (English Language Series)*. London: Longman.

Adams, Valerie (2001). *Complex Words in English, English Language Series.* Harlow: Pearson.

Addison, Joseph (1711). *The Spectator*, *135*.

Adolphs, Svenja (2006). *Introducing Electronic Text Analysis.* London: Routledge.

Adolphs, Svenja (2008). *Corpus and context: investigating pragmatic functions in spoken discourse.* Amsterdam: John Benjamins.

Agassi, Joseph (1973). Continuity and discontinuity in the history of science. *Journal of the History of Ideas*, *34*(4), 609-626.

Aronoff, Mark (1983). A decade of morphology and word formation. *Annual Review of Anthropology*, *12*, 355-375.

Aronoff, Mark & Fudeman, Kirsten A. (2005). *What is morphology?* Oxford: Blackwell.

Ash, John (1775). *The New and Complete Dictionary of the English Language.* London: printed for Edward and Charles Dilly.

Atkins, Sue, Clear, Jeremy, & Ostler, Nicholas (1992). Corpus design criteria. *Literary and Linguistics Computing*, *7*(1), 1-16.

Atkinson, Dwight (1999). *Scientific discourse in sociohistorical context. The philosophical transactions of the Royal Society of London 1675-1975.* New Jersey: Lawrence Erlbaum Associates.

Baayen, R. Harald & Lieber, Rochelle (1991). Productivity and English derivation: a corpus-based study. *Linguistics*, *29*, 801-843.

Bacon, Francis, Jardine, Lisa & Silverthorne, Michael (2000). *The New Organon, ed.* Cambridge: Cambridge University Press.

Bailey, Richard W. (2007). The ideology of English in the long eighteenth century. In Marina Dossena & Charles Jones (Eds.), *Insights into Late Modern English* (pp. 21-44). Bern: Peter Lang.

Bailey, Richard W. (2010). Variation and change in eighteenth-century England. In Raymond Hickey (Ed.), *Eighteenth-Century England: Ideology and Change* (pp. 182-199). Cambridge: Cambridge University Press.

Baker, Mark (2003). *Lexical Categories: Verbs, Nouns and Adjectives.* Cambridge: Cambridge University Press.

Balteiro, Isabel (2007). *A Contribution to the Study of Conversion in English.* Münster: Waxmann Verlag Gmbh.

Banks, David (2005). Emerging scientific discourse in the late seventeenth century: A comparison of Newton's *Opticks* and Huygen's *Traité de la lumière*. *Functions of language*, *12*(1), 65-86.

Banks, David (2012). Thematic structure in eighteenth century astronomical texts: a study of a small sample of articles from the Corpus of English Texts on Astronomy. In Isabel Moskowich & Begoña Crespo (Eds.),

Astronomy 'playne and simple': The Writing of Science between 1700 and 1900 (pp. 221-238). Amsterdam: John Benjamins.

Barber, Charles (1976). *Early Modem English.* Edinburgh: Edinburgh University Press.

Baskerville, John (2010). *A Vocabulary, Or Pocket Dictionary: To which is Prefixed, a Compendious Grammar of the English Language, 1765.* Charleston, SC: BiblioLife.

Bauer, Laurie (1983). *English Word Formation.* Cambridge: Cambridge University Press.

Bauer, Laurie (1988). *Introducing English Morphology*. Edinburgh: Edinburgh University Press.

Bauer, Laurie (2001). *Morphological Productivity.* Cambridge: Cambridge University Press.

Bauer, Laurie (2004). *A Glossary of Morphology.* Edinburgh: Edinburgh University Press.

Beal, Joan C. (2004). *English in modern times, 1700-1945*. London: Arnold.

Beal, Joan C. (2010). Prescriptivism and the suppression of variation. In Raymond Hickey (Ed.), *Eighteenth-Century English: Ideology and Change* (pp. 21-37). Cambridge: Cambridge University Press.

Beard, Robert (1995). *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation.* Albany, NY: State University of New York Press.

Bevis, John (1765). *A Pocket Dictionary; Or, Complete English Expositor.* London: J. Newbery.

Biber, Douglas, Conrad, Susan & Reppen, Randi (1998). *Corpus Linguistics.* Cambridge: Cambridge University Press.

Bicknell, Alexander (1790). *The Grammatical Wreath; Or a Complete System of English Grammar: Being a Selection of the Most Instructive Rules from All the Principal English Grammars, Etc.* London: printed for the author.

Birch, Thomas (1756). *The History of the Royal Society.* London: A. Millar.

Blevins, James P. (1994). A lexicalist analysis of gerundiven nominals in English. *Australian Journal of Linguistics*, *14*, 1-38.

Bloomfield, Leonard (1933). *Language.* London: George Allen & Unwin.

Booij, Geert (2007). *The Grammar of Words: An Introduction to Linguistic Morphology.* Oxford: Oxford University Press.

Borges, Jorge L. (1993). El idioma analítico de John Wilkins. In *Otras inquisiciones.* Madrid: Alianza.

Botvina, Renata (2005). Francis Bacon's natural philosophy as a universal language. *Studies in Logic, Grammar and Rhetoric*, *8*(21), 89-99.

Boyle, Robert (1661). A Proëmial Essay, wherein, with some considerations touching experimental essays in general, is interwoven such an introduction to all those written by the author, as is necessary to be perus'd for the better understanding of them. In *Certain Physiological*

Essays Written at Distant Times, and on Several Occasions. London: Henry Herringman.

Browne, Richard (1969). *The English School Reformed, 1700.* Menston, England: The Scolar Press.

Burridge, Kate (2004). *Blooming English.* Cambridge: Cambridge University Press.

Butterfield, Herbert (1965). *The Origins of Modern Science*. New York: The Free Press.

Camiña-Rioboo, Gonzalo (2005). Tmesis and in(ter)fixation: the unity of morphemes and/or words in question. In Isabel Moskowich (Ed.), Reinterpretations of English. *Essays on Language, Linguistics and Philology (II)* (pp. 71-81). A Coruña: Universidade da Coruña.

Camiña, Gonzalo (2010). New nouns for new ideas. In M. Lluisa Gea-Valor, Isabel García, & M. José Esteve (Eds.), *Linguistic and Translation Studies in Scientific Communications* (pp. 156-176). Bern/Berlin: Peter Lang.

Camiña-Rioboo, Gonzalo (2012). Accounting for observations of the heavens in the 18th century: New nouns to explain old phenomena. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The Writing of Science between 1700 and 1900* (pp. 93-122). Amsterdam: John Benjamins.

Carstairs-McCarthy, Andrew (2002). *An Introduction to English Morphology: Words and their Structure.* Edinburgh: Edinburgh University Press.

Caso, Arthur L. (1980). The production of new scientific terms. *American Speech*, *55*(2), 101-111.

Chomsky, Noam (1966). *Topics in the Theory of Generative Grammar.* The Hague: Mouton De Gruyter.

Chomsky, Noam (1970). Remarks on nominalization. In Roderick A. Jacobs & Rosenbaum (Eds.), *Transformational Grammar* (pp. 184-221). Waltham, Mass., Toronto & London: Ginn and Company.

Ciszek, E. (2008). *Word Derivation in Early Middle English.* Frankfurt am Main: Peter Lang.

Clark, Geoffrey & Stafford, Russel (1982). Quantification in American archaeology: a historical perspective. *World Archaeology*, *14*(1), 98-119.

Collyer, J. (1735). *The General Principles of Grammar, 1735.* Menston, England: The Scolar Press Limited.

Coote, Charles (1788). *Elements of the Grammar of the English Tongue.* London: printed for the author.

Crespo, Begoña & Moskowich, Isabel (2009). The limits of my language are the limits of my world: the scientific lexicon from 1350 to 1640. *Skase Journal of Theoretical Linguistics*, *6*(1), 45-48.

Crespo, Begoña & Moskowich, Isabel (2010). *CETA* in the context of the *Coruña Corpus*. *Literary and Linguistics Computing*, *25*(2), 153-164.

Crombie, Alistair C. (1953). *Robert Grossteste and the Origins of Experimental Philosophy, 1100-1700.* Oxford: Clarendon Press.

Crombie, Alistair C. (1969). *Augustine to Galileo: The History of Science A.D. 400-1650.* London: Penguin.

Cronin, Vincent (1992). *The Flowering of the Renaissance.* New York: Harper Collins.

Crystal, David (2003). *A Dictionary of Linguistics & Phonetics.* Oxford: Blackwell.

Culpeper, Jonathan & Clapham, Phoebe (1996). The borrowing of Classical and Romance words into English: a study based on the electronic Oxford English Dictionary. *International Journal of Corpus Linguistics*, *1*(2), 199-218.

Dalgarno, George, Cram, David, & Maat, Jaap (2001). *George Dalgarno on Universal Language: The Art of Signs (1661), The Deaf and Dumb Man's Tutor (1680), and the Unpublished Papers.* Oxford: Oxford University Press.

Dalton-Puffer, Christiane (1996). Middle English is a Creole and its opposite: On the value of plausible speculation. In Jacek Fisiak (Ed.), *Linguistic Change under Contact Conditions* (pp. 35-50). Berlin: Mouton De Gruyter.

Dilworth, Thomas (1767). *A New Guide to the English Tongue, 1751.* Menston, England: The Scolar Press.

Dressler, Wolfgang U. (1979). *On a polycentristic theory of word formation*. Proceedings from 12th International Congress of linguists, Vienna.

Duhem, Pierre M. M. (1991). *The Aim and Structure of Physical Theory.* Princeton: Princeton University Press.

Durkin, Philip (2008). Latin loanwords of the early modern period: How often did French act as an intermediary? In Richard Dury, Maurizio Gotti, & Marina Dossena (Eds.), *English Historical Linguistics 2006* (pp. 185-202). Amsterdam/New York: John Benjamins.

Elphinston, James (1765). *The Principles of the English Language Digested; Or, English Grammar Reduced to Analogy.* London: James Bettenham.

Fenning, Daniel (1967). *A New Grammar of the English Language, 1771.* Menston, England: The Scolar Press.

Fernández-Domínguez, Jesús (2009). *Productivity in English Word-formation. An approach to N+ N compounding* (341). Bern: Peter Lang.

Firth, John R. (1930). *Speech.* London: Benn's Sixpence.

Fogg, Peter W. (1792). *Dissertations, Grammatical and Philological.* London: Peter W. Fogg.

Fogg, Peter W. (1796). *Elementa Anglicana, Or, The English Grammar.* Stockport: Printed for the author.

Fowler, Henry W. & Crystal, David (2009). *A dictionary of modern English usage.* Oxford: Oxford University Press.

García-Ejarque, Luís. (1997). *La Real Biblioteca de S.M. y su personal (1712-1836).* Madrid: Tabapress.

Gill, Alexander, *et al*. (1972). *Logonomia Anglica, 1619.* Stockholm: Almqvist & Wiksell International.

Görlach, Manfred (1991). *Introduction to Early Modern English.* Cambridge: Cambridge University Press.

Görlach, Manfred (1994). *The Linguistic History of Engish.* London: Macmillan.

Görlach, Manfred (2001). *Eighteenth-century English.* Heidelberg: Universitätsverlag.

Görlach, Manfred (2004). *Text Types and the History of English.* Berlin: Mouton de Gruyter.

Gotti, Maurizio (1996). *Robert Boyle and the Science of Language.* Milano: Angelo Guerini.

Gotti, Maurizio (2001). The Experimental Essay in Early Modern English. *European Journal of English Studies*, *5*(2), 221-239.

Gotti, Maurizio (2003). *Specialized discourse: Linguistic features and changing conventions.* Bern: Peter Lang.

Gotti, Maurizio (2005). *Investigating Specialized Discourse.* Bern/Berlin: Peter Lang.

Gough, James (1969). *A Practical Grammar of the English Tongue, 1754.* Menston, England: The Scolar Press.

Gower, Barry (2002). *Scientific method: An historical and philosophical introduction.* London: Routledge.

Grant, Edward (1996). *The Foundations of Modern Science in the Middle Ages: Their Religious, Institutional, and Intellectual Contexts.* Cambridge: Cambridge University Press.

Grant, Edward (2007). *A History of Natural Philosophy: From the Ancient World to the Nineteenth Century.* Cambridge: Cambridge University Press.

Greenwood, James (1722). *An Essay towards a Practical English Grammar, Describing the Genius and Nature of the English Tongue.* (2nd ed., with additions). London: printed for John Clark.

Greenwood, James (1968). *An Essay towards a Practical English Grammar, 1711.* Menston, England: The Scolar Press.

Hall, Marie B. (1975). The Royal Society's Role in the Diffusion of Information in the Seventeenth Century (1). *Notes and Records of the Royal Society of London*, *29*(2), 173.

Halliday, Michael A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning.* Baltimore: University Park Press.

Halliday, Michael A. K., & Webster, J. J. (2004). *The Language of Science.* London/New York: Continuum.

Hård, Mikael & Jamison, Andrew (2005). *Hubris and Hybrids: A Cultural History of Technology and Science.* London: Routledge.

Harrington, Jonathan (2010). *Phonetic Analysis of Speech Corpora.* Oxford: Wiley-Blackwell.

Haugen, Einar (1966). *Linguistics in language planning*. Proceedings from the UCLA Sociolinguistics Conference 1964. The Hague: Mouton, 50-71.

Henry, John (2002). *The Scientific Revolution and the Origins of Modern Science.* New York: Palgrave.

Herman, Arthur (2003). *The Scottish Enlightenment: the Scots' invention of the modern world.* London: 4th Estate.

Herrero-López, Concepción (2007). Las mujeres en la investigación científica. *Criterios*, *8*, 75-96.

Herschel, John F. W. (1830). *A Preliminary Discourse on the Study of Natural Philosophy.* London: Printed for Longman, Rees, Orme, Brown, Green & J. Taylor.

Heyvaert, Liesbet (2003). *A Cognitive-Functional Approach to Nominalization in English.* Berlin: Mouton de Gruyter.

Heyvaert, Liesbet (2008). On the constructional semantics of gerundive nominalizations. *Folia Linguistica*, *42*(1-2), 39-82.

Heyvaert, Liesbet (2010). A cognitive-functional perpective on deverbal nominalization in English. Descriptive findings and theoretical ramifications. In Monika Rathert & Artemis Alexadiou (Eds.), *Semantics of Nominalizations acros Languages and Frameworks* (pp. 51-82). Berlin: Mouton de Gruyter.

Hickey, R. (2004). How and why supraregional varieties arise. In M. Dossena & C. Jones (Eds.), *Insights into Late Modern English* (pp. 351-373). Bern: Peter Lang.

Hickey, Raymond (2004). How and why supraregional varieties arise. In Marina Dossena & Charles Jones (Eds.), *Insights into Late Modern English* (pp. 351-373). Bern: Peter Lang.

Hockett, Charles F. (1958). *A Course in Modern Linguistics*. New York: Macmillan.

Hope, Jonathan (2000). Rats, bats, sparrows and dogs: biology, linguistics and the nature of Standard English. In Laura Wright (Ed.), *The Development of Standard English, 1300-1800. Theories, Descriptions, Conflicts.* (pp. 49-56). Cambridge: Cambridge University Press.

Hunter, Michael (2007). Robert Boyle and the early Royal Society: a reciprocal exchange in the making of Baconian science. *The British Journal for the History of Science*, *40*(1), 1-23.

Information Retrieval Lab (2012). *The Coruña Corpus Tool* [computer software]. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy*

'playne and simple': The Writing of Science between 1700 and 1900. Amsterdam: John Benjamins.

Jackendoff, Ray. (1977). *X-Syntax: A Study of Phrase Structure.* Cambridge, Ma: Massachussetts Institute of Technology Press.

Jackendoff, Ray. (2010). *Meaning and the Lexicon: The Parallel Architecture, 1975-2010.* Oxford: Oxford University Press.

Jacob, Margaret C. (1988). *The Cultural Meaning of the Scientific Revolution.* Philadelphia: Temple University Press.

Jespersen, Otto (1933). *Essentials of English Grammar.* Allen & Unwin: London.

Jespersen, Otto (1956). *A modern English grammar on historical principles: Morphology*. London: George Allen & Unwin Ltd.

Johnson, Sameul (1755). *A Dictionary of the English Language.* London: W. Strahan.

Jones, John (1701). *Practical phonography.* London: Richard Smith.

Kadublowski, David (2008). Pampered 'Nakations'. *The New York Times* (Online*)*. Retrieved from www.nytimes.com. [Acessed: 27 April 2008].

Kastovsky, Dieter (1986). The problem of productivity in word formation. *Linguistics*, *24*(3), 585-600.

Kastovsky, Dieter (1989). Typological changes in the History of English Word-Formation. In Heiz-Joachim Müllenbrock & Renate Noll-

Wiemann (Eds.), *Anglistentag 1988 Göttingen.* (pp. 281-293). Tübingen: Max Niemeyer.

Kastovsky, Dieter (1996). Verbal derivation in English: a historical survey. Or: Much ado about nothing. In Derek Britton (Ed.), *English Historical Linguistics 1994* (pp. 93-118). Amsterdam: John Benjamins.

Kastovsky, Dieter (2005). Conversion and/or zero: word-formation theory, historical linguistics, and typology. In Laurie Bauer & Salvador Valera-Hernández (Eds.), *Approaches to Conversion/Zero-Derivation* (pp. 31-50). Münster: Waxmann.

Kastovsky, Dieter. (2006). Vocabulary. In Richard M. Hogg & David Denison (Eds.), *A History of the English Language* (pp. 199-270). Cambridge: Cambridge University Press.

Katamba, Francis (1993). *Morphology.* Basinkstoke: Macmillan Press.

Katamba, Francis (2005). *English Words: Structure, History, Usage.* London: Routledge.

Kearney, Hugh F. (1964). *Origins of the Scientific Revolution.* London: Longmans.

Keene, Derek. (2000). Metropolitan values: migration, mobility and cultural norms. In Laura Wright (Ed.), *The Development of Standard English 1300-1800. 1300-1800. Theories, Descriptions, Conflicts.* (pp. 93-114). Cambridge: Cambridge University Press.

Kuhn, Thomas S. (1996). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kytö, Merjia, Rudanko, Juhani & Smitterberg, Erik (2000). Building a bridge between the present and the past: A corpus of 19th-century English. *ICAME journal*, *24*, 85-98.

Labov, William (1972). *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.

Labov, William (2001). Principles of Language Change: Social Factors. Chichester: Wiley.

Lane, A. (1769). *A Key to the Art of Letters, 1700.* Menston, England: The Scolar Press.

Lareo, Inés (2010). New trends exploring the language of science: The *Corpus of English Texts on Astronomy* (*CETA*) and its tool (*CCT*) in the context of the *Coruña Corpus.* In M. Lluisa Gea-Valor, Isabel García-Izquierdo, & M. José Esteve (Eds.), *Linguistic and Translation Studies in Scientific Communication* (pp. 131-156). Bern: Peter Lang.

Lareo, Inés & Montoya-Reyes, Ana (2007). Scientific Writing: Following Robert Boyle's Principles in Experimental Essays–1704 and 19981. *Revista Alicantina de Estudios Ingleses*, *20*, 119-137.

Lass, Roger (1999). *The Cambridge History of the English Language: 1476-1776* (3). Cambridge: Cambridge University Press.

Leech, Geoffrey (1992). Corpora and theories of linguistic performance. In Jan Svartvik (Ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium, 4-8 August 1991* (pp. 105-122). Berlin: Mouton de Gruyter.

Lees, Robert (1969). *The Grammar of English Nominalizations.* The Hague: Mouton.

Lenski, Daniel (2000). The Status of Etymology in the Synchronic Morphology of English. *myxo. css. msu. edu/danimal/academic,* 5. Retrieved from http://myxo.css.msu.edu/danimal/academic/

Lieber, Rochelle (2010). *What is morphology?* Cambridge University Press: Cambridge.

López-Rúa, Paula (1999). On the motivation of initialisms. In F. Toda *et al.* (Eds.), *Actas del XXI Congreso Internacional AEDEAN* (pp. 627-633). Seville: University of Seville.

Lyons, John (1977). *Semantics.* Cambridge University Press: Cambridge.

Lowth, Robert (1778). *Isaiah: A new translation.* London: J. Nichols.

Mackintosh, Duncan (1797). *A Plain, Rational Essay on English Grammar.* Boston: Manning & Loring.

Maittaire, Michael (1712). *The English Grammar: Or, an Essay on the Art of Grammar, Applied to and Exemplified in the English Tongue.* London: printed by W. B. for H. Clements.

Marchand, Hans (1953). Notes on English suffixation. *Neuphilologische Mitteilungen*, *54*, 246-272.

Marchand, Hans (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach.* Munich: C. H. Beck.

Martin, Benjamin (1970). *Institutions of Language, 1748.* Menston, England: The Scolar Press.

Mason, Joan (1992). The admission of the first women to the Royal Society of London. *Notes and Records of the Royal Society of London*, *46*(2), 279-300.

Matthews, Peter H. (1974). *Morphology: an introduction to the theory of word-structure.* Cambridge: Cambridge University Press.

Meyer, Charles F. (2002). *English Corpus Linguistics: An Introduction.* Cambridge: Cambridge University Press.

McEnery, Tony & Wilson, Andrew (2001). *Corpus Linguistics* (2nd ed.). Edinburgh: Edinburgh University Press.

McEnery, Tony (2003). Corpus linguistics. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics* (pp. 448-463). Oxford: Oxford University Press.

McEnery, Tony, Xiao, Richard & Tono, Yukio (2006). *Corpus Based Language Studies: An Advanced Resource Book.* London: Routledge.

Mikheev, Andrei (2003). Text segmentation. In Ruslan Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics* (pp. 201-218). Oxford: Oxford University Press.

Mill, John S. (1843). A System of Logic: Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation. Falta referencia

Millward, Celia M. (1996). *A Biography of the English Language: Principles and Applications* (2nd ed.). Boston: Thomson Learning.

Moessner, Lilo (2006). The Birth of the Experimental Essay. In Vijay K. Bhatia & Maurizio Gotti (Eds.), *Explorations in Specialized Genres.* (pp. 59-77). Bern/Berlin: Peter Lang.

Moskowich, Isabel (1995). *Los escandinavos en Inglaterra y el cambio léxico en inglés medieval.* A Coruña: Servicio de Publicacións. Universidade da Coruña.

Moskowich, Isabel (2001). Morfología flexiva del inglés moderno. In I. De la Cruz & J. Martín (Eds.), *Lingüística histórica inglesa* (pp. 624-654). Barcelona: Ariel.

Moskowich, Isabel (2002). El mito vikingo: el escandinavo como el *otro* en la Inglaterra medieval. *Cuadernos del CEMYR*, *10*, 55-79.

Moskowich, Isabel (2012). *CETA* as a tool for the study of modern astronomy in English. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy*

'playne and simple': The Writing of Science between 1700 and 1900 (pp. 35-56). Amsterdam: John Benjamins.

Moskowich, Isabel & Crespo, Begoña (2007). Presenting the *Coruña Corpus*: A collection of samples for the historical study of English Scientific writing. In Javier Pérez Guerra *et al.* (Eds.), '*Of Varying Language and Opposing Creed': New Insights into Late Modern English* (pp. 531-544). Bern: Peter Lang.

Moskowich, Isabel & Crespo, Begoña (Eds.) (2012). *Astronomy 'Playne and Simple': The Writing of Science Between 1700 and 1900.* Amsterdam: John Benjamins.

Moskowich, Isabel, Lareo, Inés, Camiña, Gonzalo & Crespo, Begoña (Compilers) (2012). *Corpus of English Texts on Astronomy* [CD]. In Isabel Moskowich & Begoña Crespo (Eds.), *Astronomy 'playne and simple': The Writing of Science between 1700 and 1900* (pp. 35-56). Amsterdam: John Benjamins.

Moskowich, Isabel & Montoya-Reyes, Ana (2003). Thirteen Paston letters in search of a standard. In *Revista Canaria de Estudios Ingleses, 46* (pp. 13-34).

Murray, Lindley (1795). *English Grammar.* Menston, England: The Scolar Press.

Myers, Scott. (1984). Zero-derivation and inflection. In *MIT working papers in linguistics: Papers from the January 1984 MIT workshop in morphology*, *7.* Cambridge University Press: Cambridge.

Nevalainen, Terttu (2006a). *An Introduction to Early Modern English.* Edinburgh: Edinburgh University Press.

Nevalainen, Terttu (2006b). Historical sociolinguitics and language change. In Ans van Kemenade & Bettelou Los (Eds.), *The Handbook of the History of English* (pp. 558-588). Malden, MA: Blackwell Publishing.

Nybakken, Oscar (1959). *Greek and Latin in Scientific Terminology.* Ames, IA: Iowa State University Press.

Ong, Walter (1982). Orality and technology: The technologizing of the word. London: Routledge.

Oxford English Dictionary Online (1989). 2nd edition. Oxford: Clarendon Press.

Pahta, Päivi (2001). Creating a new genre: Contextual dimensions in the production and transmission of early scientific writing. *European Journal of English Studies*, *5*(2), 205-220.

Paltridge, Brian (1997). *Genre, frames and writing in research settings.* Amsterdam: John Benjamins.

Patrick, James (Ed.). (2007). *Renaissance and Reformation.* New York: Marshall Cavendish.

Plag, Ingo (1999). *Morphological productivity: structural constraints in English derivation.* Berlin: Walter de Gruyter.

Plag, Ingo (2003). *Word-Formation in English.* Cambridge: Cambridge University Press.

Popescu, Ioan-Iovitz *et al*. (2009). *Word frequency studies.* Berlin: Walter de Gruyter.

Priestley, Joseph (1768). *The Rudiments of English Grammar, Adapted to the Use of Schools; With Notes and Observations, For the Use of those who Have Made some Proficiency in the Language.* London: printed for T. Becket and P.A. De Hondt, and J. Johnson.

Priestley, Joseph (1970). *A Course of Lectures on the Theory of Language and Universal Grammar, 1762.* Menston, England: The Scolar Press.

Quirk, Randolph, *et al*. (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Raad, B. L. (1989). Modern trends in scientific terminology: morphology and metaphor. *American Speech*, *64*(2), 128-136.

Rainer, Franz (1987). Productivitätsbegriffe in der Wortbildungstheorie. In Wolf H. G. Dietrich & Hans-Martin Gauger (Eds.), *Grammatik und Wortbildung romanicher Sprachen* (pp. 187-202).

Samuels, Michael L. (1963). Some applications of Middle English dialectology. *English Studies*, *44*, 81-94.

Sanders, Gerald (1988). Zero derivation and the overt analogue criterion. *Theoretical Morphology,* 155-175.

Sapir, Edward (1921). *Language. An Introduction to the Study of Speech.* New York: Hartcourt Brace.

Saussure, Ferdinand (1989). *Cours de linguistique générale.* Wiesbaden: Otto Harrassowitz.

Snyder, Laura J. (2009). Hypotheses in 19th Century British Philosophy of Science: Herschel, Whewell, Mill. In Michael Heidelberger & Gregor Schiemann (Eds.), *The Significance of the Hypothetical in the Natural Sciences* (pp. 59-76). Berlin: Mouton De Gruyter.

South, Robert (1720). *Thirty-Six Sermons and Discourses, on Several Subjects and Occasions.* Dublin: Patrick Dugan & Joseph Leathly.

Spencer, Andrew (1991). *Morphological Theory.* Oxford: Blackwell.

Sprat, Thomas (1667). *The History of the Royal Society of London for the Improving of Natural Knowledge.* London: Martyn & Allestry.

Staniford, Daniel (1800). *A Short but Comprehensive Grammar.* Charlestown: printed for the author.

Štekauer, Pavol (1992). On some issues of zero morpheme in English. *Linguistica Pragensia*, *2*, 73-87.

Štekauer, Pavol (2000). *English Word-Formation: A History of Research, 1960-1995.* Tübingen: Günter Narr Verlag.

Stein, Gabrielle (2007). *A Dictionary of English Affixes.* Munich: Lincom.

Stillman, Robert E. (1995). *The New Philosophy and Universal Languages in Seventeenth-Century England. Bacon, Hobbes and Wilkins.* New Jersey: Associated University Presses.

Stockwell, Robert P. & Minkova, Donka (2001). *English Words: History and Structure.* Cambridge: Cambridge University Press.

Stump, Gregory T. (1998). Inflection. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology* (pp. 13-43). Oxford: Blackwell.

Subbiondo, Joseph L. (2001). Educational Reform in Seventeenth-Century England and John Wilkins. *Language & Communication*, *21*(3), 273-284.

Subbiondo, Joseph L. (2007). Competing models for a 17th century universal language: A study of the dispute between George Dalgarno and John Wilkins. In Douglas E. Kibbee (Ed.), *History of linguistics 2005* (pp. 112-119). Amsterdam: John Benjamins.

Sundby, Bertil (1995). *English word-formation as described by English grammarians, 1600-1800.* Novus Forlag.

Swift, Jonathan (1710). The False Refinements in Our Style. *The Tatler*, *230*.

Taavitsainen, Irma (2000). Scientific language and spelling standardisation. In Laura Wright (Ed.), *The Development of Standard English, 1300-1800. Theories, Descriptions, Conflicts.* (pp. 131-154). Cambridge: Cambridge University Press.

Taavitsainen, Irma & Pahta, Päivi (1997). The corpus of Early English medical writing. *ICAME Journal*, *21*, 71-78.

Taavitsainen, Irma & Pahta, Päivi (Eds.). (2004). *Medical and scientific writing in late medieval English.* Cambridge: Cambridge University Press.

Taavitsainen, Irma, Pahta, Pahta, Mäkinen, Martti & Hickey, Raymond (2005). *Middle English Medical Texts.* Amsterdam: John Benjamins.

Taton, René (Ed.). (1964). *The Beginnings of Modem Science 1450-1800.* London: Thames and Hudson.

The Irish Independent (2007). Celebgeny, 29 August 2007, p. 28.

The Royal Society (1728). *The Charters and Statutes of the Royal Society of London, for Improving Natural Knowledge*. London: The Royal Society.

Tieken-Boon van Ostade, Ingrid (2009). *An Introduction to Late Modern English.* Edinburgh: Edinburgh University Press.

Trask, Robert L. (1993). *A Dictionary of Grammatical Terms in Linguistics.* London: Routledge.

Trost, Harald (2003). Morphology. In Ruslan Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics* (pp. 27-47). Oxford: Oxford University Press.

Tuite, Thomas (1726). *The Oxford Spelling-Book; Being a Complete Introduction to English Orthography. In a Method much more Clear*

and Intelligible than any Book of this Nature, hitherto Extant. London: printed for J. Hazard.

Valle, Ellen (1999). *A collective intelligence: The life sciences in the Royal Society as a scientific discourse community (1665-1965).* Turku: University of Turku.

Valle, Ellen (2006). Reporting the doings of the curious: authors and editors in the Philosophical Transactions of the Royal Society of London. *News discourse in early modern Britain: selected papers of CHINED 2004,* 71.

Van Dyke, Carolynn (1992). Old words for new worlds: modern scientific and technological word-formation. *American Speech*, *67*, 383-405.

Voigts, Linda E. (1979). Anglo-Saxon plant remedies and the Anglo-Saxons. *Isis,* 250-268.

Vos, Alvin (1979). "Good Matter and Good Utterance": The Character of English Ciceronianism. *Studies in English Literature, 1500-1900*, *19*(1), 3-18.

Ward, William (1967). *An Essay on Grammar, 1765.* Scolar Press: Menston, England.

Watts, Isaac (1753). *The Works of the Late Reverend and Learned Isaac Watts, DD.* London: T. and T. Longman.

Whewell, William (1840). *The Philosophy of the Inductive Sciences: Founded upon their History.* London: John W. Parker.

Whitehead, Alfred N. (1933). *Science and the Modern World.* Cambridge: Cambridge University Press.

Whitney, William D. (1889). *Sanskrit Grammar.* Cambridge, Mass.: Harvard University Press.

Wilkins, John (1668). *An Essay towards a Real Character, and a Philosophical Language.* London: Samuel Gellibrand and John Martin.

Wood, James (1777). *Grammatical Institutions, Or, A Practical English Grammar: On a Plan Entirely New.* Newcastle: T. Saint.

Wright, Laura (1994). On the writing of the history of Standard English. In Francisco Fernández, Miguel Fuster, & Juan José Calvo (Eds.), *English Historical Linguistics 1992* (pp. 105-115). Amsterdam: John Benjamins.

Wright, Laura (1996). About the evolution of standard English. In Elizabeth M. Tyler & J. Jane Toswell (Eds.), *Studies in English Language and Literature: 'Doubt Wisely', Papers in Honour of E.G. Stanley* (pp. 99-115). London: Routledge..

Wright, Laura (2001). The role of international and national trade in the standardisation of English. In Isabel Moskowich, Begoña Crespo, Emma Lezcano, & Begoña Simal (Eds.), *Re-interpretations of English. Essays on Language, Linguistics and Philology, I* (pp. 189-207). A Coruña: Universidade da Coruña.

Appendix

Nouns coined at an earlier date than recorded in the *OED*.

Noun	**Subcorpus**	**Year**	**Author**	***OED***
aegean	*CEPhiT*	1727	Robert Greene	1814
bellow	*CEPhiT*	1710	John Dunton	1779
co-altitude	*CETA*	1726	Isaac Watts	1833
co-declination	*CETA*	1726	Isaac Watts	1812
co-latitude	*CETA*	1726	Isaac Watts	1790
day-tide	*CETA*	1756	James Ferguson	1818
fiction-monger	*CEPhiT*	1730	Robert Kirkpatrick	1835
glossing	*CEPhiT*	1730	Robert Kirkpatrick	1875
hedge-parsley	*CEPhiT*	1790	William Smellie	1830
ill-breeding	*CEPhiT*	1700	Mary Astell	1800
inebriety	*CEPhiT*	1783	Catherine Macaulay	1786
interference	*CEPhiT*	1727	Robert Greene	1783
iron-wire	*CETA*	1719	John Harris	1841
glycine	*CEPhiT*	1790	William Smellie	1851
grand-niece	*CEPhiT*	1755	Francis Hutcheson	1830
ludicrous	*CEPhiT*	1776	George Campbell	1782
prime-vertical	*CETA*	1732	Samuel Fuller	1846
semi-tychonic	*CETA*	1786	John Bonnycastle	1794
shun	*CEPhiT*	1769	Adam Ferguson	1822
spinal-marrow	*CEPhiT*	1790	William Smellie	1794
stumbling-block	*CEPhiT*	1792	Mary Wollstonecraft	1845
two-thirds	*CETA*	1767	George Costard	1777
vector	*CETA*	1702	Robert Morden	1704
vociferator	*CETA*	1702	Robert Morden	1814
wheat-sheaf	*CETA*	1754	John Hill	1846