UNIVERSITAT DE BARCELONA

Voice in Scientific English: A Corpus-Based Approach to the Use of Active and Passive Voice in Medical Research Articles in English

Aaron Ventura Campos

UNIVERSITAT DE BARCELONA

Voice in Scientific English: A Corpus-Based Approach to the Use of Active and Passive Voice in Medical Research Articles in English

Doctoral thesis submitted by

Aaron Ventura Campos

For the degree of

Doctor of Philosophy

by the University of Barcelona

Doctoral programme: *Linguistic, literary, and cultural studies*

Line of research: *Lexis, communication, and cognition*

Doctoral thesis supervised by

Dr. Isabel Verdaguer Clavera

University of Barcelona

Dr. Suganthi John

University of Birmingham

VENTURA CAMPOS AARON - 53286400T

Firmado digitalmente por VENTURA CAMPOS AARON - 53286400T
Fecha: 2024.05.18 00:34:36 +02'00'

UNIVERSITAT DE BARCELONA

Voice in Scientific English: A Corpus-Based Approach to the Use of Active and Passive Voice in Medical Research Articles in English

Tesi doctoral presentada per

Aaron Ventura Campos

com a requeriment per a l'obtenció del títol de

Doctor per la Universitat de Barcelona

Programa de doctorat: *Estudis lingüístics, literaris i culturals*

Línia de recerca: *Lèxic, comunicació i cognició*

Tesi doctoral supervisada per

Dra. Isabel Verdaguer Clavera

Universitat de Barcelona

Dra. Suganthi John

University of Birmingham

Abstract

The notion of voice has attracted significant attention in academic writing (John 2005, Herrando-Rodrigo 2019, 2022, *inter alia*) as it is conceived as a compelling device to ascertain how specialised discourse writing is shaped. The interrelatedness shown between voice, authorial (in)visibility and (im)personality makes it a hard concept to delimit, as each of them may partially entail any of the others.

As impersonality, at least in medical writing, has traditionally been related to the extensive use of the passive voice, taking grammatical voice as a starting point to operationalise the notion of impersonality seems well-justified. The corpus-based study presented in this thesis aims at quantifying a set of active and passive structures recurrent in RCTs. They are quantified through corpus searches using SQS and CQPS. The high frequency of these structures, after scrutinising a set of real RAs, allows their inclusion within the restricted set of lexicogrammatical patterns typically occurring in medical register. Quantitative and qualitative analyses are carried out to determine the extent to which active and passive structures primarily relate to (im)personality, and secondarily, to authorial (in)visibility. This study concludes with a proposal to scientifically categorise the structures analysed in the corpus-based study through a Cartesian coordinate system, in which mainly two distinct parameters or axes are used, namely personality (x-axis) and grammatical voice (y-axis).

The corpus-based study is complemented with a close textual analysis, whose main purpose is to attempt to disentangle the complex and multi-layered dimension of impersonality by exploring the extent to which the strategies of (im)personalitsation described interact to textually characterise medical discourse as being considerably impersonal. Through a manual classification of the lexicogrammatical patterns analysed, I attempt to assess the degree of impersonality they carry with them by analysing the extent to which these instances relate to authorial presence (or absence) in the text. This study is also aimed at assessing the pragmatic functions these patterns perform within the

text as far as authorial (in)visibility and (im)personality are concerned. Through the exploration of the interplay between these patterns, an evaluation of how such interaction affects the transmission of knowledge and the degree of impersonality it transmits has also been done. The gradation of structures from more personal to more impersonal-like is represented through a continuum, so that the overall picture of impersonality and authorial visibility (also including information on grammatical voice) in RCTs is visually shown.

The results of both studies decipher the interrelatedness between the notions of impersonality, authorial (in)visibility and the use of grammatical voice in medical discourse using a mixed-method approach, in which quantitative data are subsequently qualitatively analysed. The effect other lexicogrammatical patterns exert over this impersonal perception is also seen as fundamental, as they may be used to either reinforce this impersonal reality or counteract it. To explore the correlation between these patterns and three distinctive analytical dichotomised parameters, namely *personality* vs. *impersonality*, *active voice* vs. *passive voice*, and *authorial visibility* vs. *authorial invisibility*, has been crucial to unravel the complex notion of impersonality, and both a Cartesian coordinate system and a cline have proved to be excellent methods to do so.

Abstract (in Catalan)

La noció de veu ha atret una atenció significativa en l'escriptura acadèmica (John 2005, Herrando-Rodrigo, 2019, 2022, *inter alia*) ja que es concep com un element determinant per a entendre com es conforma l'escriptura del discurs especialitzat. La interrelació que es mostra entre veu, (in)visibilitat autorial i (im)personalitat fa que sigui un concepte difícil de delimitar, ja que cadascun d'ells implica parcialment els altres.

Com la impersonalitat tradicionalment s'ha relacionat amb l'ús extensiu de la veu passiva, almenys en l'escriptura acadèmica, prendre la veu gramatical com a punt de referència per tal d'operacionalitzar la noció d'impersonalitat sembla justificat. L'estudi basat en corpus presentat en aquesta tesi té com a objectiu quantificar un conjunt d'estructures actives i passives recurrents als RCTs, mitjançant cerques de corpus utilitzant SQS i CQPS. L'elevada freqüència d'aquestes estructures, després d'analitzar detingudament un conjunt d'articles mèdics reals, permet la seva inclusió dins del conjunt restringit de patrons lexicogramaticals que típicament ocorren en el registre mèdic. S'han realitzat anàlisis quantitatives i qualitatives per determinar en quina mesura les estructures actives i passives es relacionen principalment amb la (im)personalitat, i secundàriament, amb la (in)visibilitat autorial. Aquest estudi conclou amb una proposta de categorització científica de les estructures analitzades en l'estudi basat en corpus a través d'un sistema de coordenades cartesianes, en el qual principalment s'empren dos paràmetres o eixos, concretament l'eix de personalitat (eix x) i l'eix corresponent a la veu gramatical (eix y).

L'estudi basat en corpus es complementa amb una anàlisi textual detallada, amb el principal objectiu d'intentar desxifrar la complexa i estratificada dimensió d'impersonalitat, explorant fins a quin punt les estratègies d'(im)personalització descrites interactuen per caracteritzar textualment el discurs mèdic com a considerablement impersonal. Mitjançant la classificació manual dels patrons lexicogramaticals analitzats, s'ha avaluat el grau d'impersonalitat que comporten, delimitant fins a quin punt aquestes ocurrències es relacionen amb la presència (o absència) autorial en el text. Aquest estudi

també té com a objectiu avaluar les funcions pragmàtiques que aquests patrons exerceixen dins del text pel que fa a la (in)visibilitat autorial i a la (im)personalitat. Mitjançant l'exploració de la interacció entre aquests patrons, també s'ha fet una avaluació de com aquesta interacció afecta la transmissió de coneixements i el grau d'impersonalitat que transmet. La gradació d'estructures, des de les més personals a les més impersonals, s'ha representat a través d'un continu, per tant que els dos conceptes (im)personalitat i (in)visibilitat autorial (també afegint-hi informació sobre la veu gramatical) en els RCTS es puguin mostrar visualment.

Els resultats d'aquests dos estudis desxifren la interrelació entre les nocions d'impersonalitat, (in)visibilitat autorial i l'ús de la veu gramatical en el discurs mèdic utilitzant un enfocament de mètodes mixtos, on les dades quantitatives han estat subseqüentment analitzades qualitativament. També s'ha considerat l'efecte que altres patrons lexicogramaticals exerceixen sobre aquesta percepció impersonal, ja que aquests poden ser emprats per a reforçar aquesta realitat impersonal o contrarestar-la. Explorar la correlació entre aquests patrons i tres paràmetres d'anàlisi dicotomitzats, concretament *personalitat* vs. *impersonalitat*, *veu activa* vs. *veu passiva*, i *visibilitat autorial* vs. *invisibilitat autorial*, ha estat crucial per desxifrar la noció complexa d'impersonalitat, i tant el sistema de coordenades cartesià com el continu han resultat ser excel·lents mètodes per a poder-ho fer.

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or part, for a degree at this or any other university.

Aaron Ventura Campos

Acknowledgements

The present doctoral dissertation would not have been possible without the help and support of many people. The following lines are dedicated to all of them.

First and foremost, I would especially like to thank my thesis supervisors, Dr. Isabel Verdaguer (University of Barcelona) and Dr. Suganthi John (University of Birmingham), for all their support. Both have been a constant source of inspiration. Their vocation, dedication, and expertise have contributed not only to the timely and proper completion of this doctoral thesis but also to my growth as a researcher and as a person. You have enabled me to achieve what I thought I would never accomplish, and without you, this thesis would probably never have seen the light of day. Thanks for believing in me! For all these reasons, I will always be grateful to you.

I would like to express my heartfelt gratitude to Dr. Anna Poch†, without whom none of this would have been possible. Thanks to your English syntax classes, I developed an interest in the fascinating world of Linguistics. Thank you for listening to me and giving me advice. You opened the doors for me to meet Dr. Isabel Verdaguer, and for that, I will always be very grateful!

I would also like to thank all those who made my research stays at the University of Birmingham unforgettable, especially those working at CARE (Centre for Advanced Research in English) and at CCR (Centre for Corpus Research). I would like to thank Dr. Suganthi John, former Director of CARE, for opening the doors of this institution to me and allowing me to connect with many renowned linguists with whom I could discuss, contrast ideas, and receive valuable advice, which ultimately materialised in this doctoral thesis. I am grateful for the conversations with Dr. Susan Hunston and Dr. Paul Thompson, former director of the CCR. It was a pleasure to share ideas about the thesis with them. I would also like to thank Dr. Neil Millar for his support, giving me access to

the RCTs Corpus. Many thanks to Michelle Devereux, who helped me with anything I needed during my various research stays in Birmingham.

I would also like to express my gratitude to Dr. Isabel Herrando-Rodrigo, Dr. Marzia Coltri, and Dr. Lobat Asadi. All three of you, in one way or another, made me feel very accompanied during those summer stays when I felt truly alone in Birmingham. Thank you for making the journey a bit easier.

I would like to dedicate a few words to my colleagues in the Department of Modern Languages and Literatures and English Studies at the University of Barcelona. With many of you, I have had the opportunity of sharing great joys within the research and the teaching innovation projects we have been part of. Thank you for integrating me into the world of university teaching and research. You have always counted on me, despite knowing that due to work commitments, I could not offer as much as I would have liked. This gratitude is especially dedicated to my department colleagues and friends, Dr. Mª Luz Celaya, Dr. Natàlia Judith Laso, Dr. Elisabet Comelles, Dr. Emilia Castaño, Dr. Júlia Barón, Dr. Montserrat Forcadell, and Dr. Joe Hilferthy. Thank you all for taking care of me, for helping me when I most needed it, and for being part of my academic and personal life. You are all a source of inspiration for me!

I would also like to thank Dr. Joan Carles Mora, Eva Cerviño, Dr. Cristina Aliaga, and Dr. Mireia Ortega. Thank you for always helping me and for making me feel part of the group that teaches English Phonetics and Phonology at the University of Barcelona. Especially, thank you, Eva, for your support during very difficult times. You have always taken a moment to listen to me and you are one of the people who has helped me the most to see the positive side of things. I would also like to thank Dr. Maria Grau, for the several conversations we have had. Your positivity has been a true balm during certain moments of confusion.

I would also like to thank Dr. Elisenda Bernal for her unwavering faith in me. Thank you for opening the doors for me to teach at Pompeu Fabra University. I am also grateful for the trust Dr. Salvador Climent placed in me by offering me the opportunity to teach at the

Open University of Catalonia, as well as Dr. Albert Morales, for helping me with everything related to the subjects I currently teach at this institution.

It is also a pleasure to spend many hours of the day with a great group of education professionals at Col·legi Santa Teresa – Ganduxer. First, I would like to thank the school for its support over these last years, especially for accommodating my extremely complicated schedules to facilitate my teaching commitments elsewhere. I would also like to thank my department colleagues for their constant support. Thank you, Maria Rosa Jordà, for your friendship, for taking care of me, and for making my life much easier at school. Thank you, Laura Garcia, for your friendship and support on all levels. Thanks also to Bàrbara Almeida, because you have helped me a lot in this final stretch of my thesis.

This thesis has also been made possible thanks to the unconditional support of my friends. Despite the physical distance with many of them, they have always been present in one way or another during difficult times. I would especially like to thank the friendship of Dr. Aislín Kearney, one of the people who has left the greatest mark on my life, as well as Pere Corominas, Laura Martí, Ester Magrinyà, Lucía Mérida, and Andrés Henao. I love you all very much, and without your advice and support, I would not be here today!

Per últim, agrair tot el suport moral a tots els membres de la meva estimada família. En especial, als meus pares, i en particular a la meva mare, que literalment s'ha deixat la pell perquè els seus dos fills poguessin tenir el millor en un context d'adversitats constants. Ella és la meva heroïna i el motor que ens ha fet tirar endavant a tots els que l'envoltem. Gràcies per sempre ser-hi! Agrair també el suport del meu germà Èric i de la Gibet. Gràcies per sempre ser-hi i per aportar la serenor que moltes vegades ens manca a la resta de membres de la família. Evidentment, voldria també agrair el suport del Jefferson, sense el qual aquesta tesi probablement mai s'hagués arribat a materialitzar. Jeff, estoy convencido que desde arriba nos han permitido cerrar esta etapa para poder iniciar otra que, seguramente, ponga nuestras vidas patas arriba. A pesar de ciertas dificultades en el camino, hemos mostrado ser un equipo unido y dispuesto a luchar por todo lo que queremos. Us estimo molt a tots i a cadascun de vosaltres!

No vull oblidar-me de totes aquelles persones que avui dia no estan presents entre nosaltres, però que han compartit una gran part de la meva vida i han estat també font insaciable d'inspiració. En especial voldria dedicar aquesta tesi doctoral als meus avis† i a la meva tieta Inma†, els quals no han pogut gaudir d'aquesta última etapa, però que m'han estat acompanyant de ben segur a cada passa que donava. Espero que allà on estigueu, us sentiu ben orgullosos; el meu agraïment etern a vosaltres. Us tinc molt presents i us agraeixo de tot cor l'estima i la dedicació que sempre heu tingut cap a mi. Us estimo, i sempre us estimaré.

Aaron Ventura

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

aIMRaD – Abstract, Introduction, Materials and Methods, Results, and Discussion

AMWA – American Medical Writers Association

AN – Annals of Internal Medicine

ANN INTERN MED – Annals of Internal Medicine

BM – The British Medical Journal

BMJ – The British Medical Journal

BRIT MED J – The British Medical Journal

CCR – Centre for Corpus Research (University of Birmingham)

CEQL – Common Elementary Query Language

CoP – Community of Practice

CQPweb – Corpus Query Processor Web

CQPS – Corpus Query Processor Syntax

DC – Discourse Community

DOI – Digital Object Identification

EAL – English as an Additional Language

ELF – English as a Lingua Franca

FDA – Food and Drug Administration (United States)

ICMJE – International Committee of Medical Journal Editors

IMRD – Introduction, Materials and Methods, Results, and Discussion

IMRaD – Introduction, Materials and Methods, Results, and Discussion

JA – The Journal of American Medical Association

J AM MED ASSOC – The Journal of American Medical Association

JAMA – The Journal of American Medical Association

KWIC – Key Words in Context

LLC – London-Lund Corpus of Spoken English

LLC:c – London-Lund Corpus of Spoken English (complete and most updated version)

LLC:o – London-Lund Corpus of Spoken English (original corpus)

LLC:s – London-Lund Corpus of Spoken English (supplement to the original corpus)

LA – The Lancet

LANCET – The Lancet

LNC – The Lancet

LOB – Lancaster-Oslo-Bergen Corpus

NE – The New England Journal of Medicine

NEJM – The New England Journal of Medicine

NEW ENG J MED – The New England Journal of Medicine

PLM – Plain Language Movement

POS – Part-of-Speech

RA – Research Article

RCT(s) – Randomised Controlled Trial(s)

RCT Corpus – Randomised Controlled Trials Corpus

SELF – Studying English as a Lingua Franca

SEU – Survey of English Usage

SQS – Simple Query Syntax

SSE – Survey of Spoken English

SVO – Subject, Verb, and Object

URL – Uniform Resource Locator

VOICE – Vienna-Oxford International Corpus of English

WrELFA – Written ELF in Academic Settings

According to family legend, Ferguson's grandfather departed on foot from his native city of Minsk with one hundred rubbles sewn into the lining of his jacket, travelled west to Hamburg through Warsaw and Berlin, and then booked passage on a ship called the *Empress of China*, which crossed the Atlantic in rough winter storms and sailed into New York Harbor on the first day of the twentieth century. While waiting to be interviewed by an immigration official at Ellis Island, he struck up a conversation with a fellow Russian Jew. The man said to him: *Forget the name Reznikoff. It won't do any good here. You need an American name for your new life in America, something with a good American ring to it.* Since English was still an alien tongue to Isaac Reznikoff in 1900, he asked his older, more experienced compatriot for a suggestion. *Tell them you're Rockefeller*, the man said. *You can't go wrong with that.* An hour passed, then another hour, and by the time the nineteen-year-old Reznikoff sat down to be questioned by the immigration official, he had forgotten the name the man had told him to give. *Your name?* the official asked. Slapping his head in frustration, the weary immigrant blurted out in Yiddish, *Ikh hob fargessen (I've forgotten)!* And so it was that Isaac Reznikoff began his new life in America as Ichabod Ferguson.

Extracted from "4321", by Paul Auster (2017)

Introduction

All things happen
by virtue of necessity.
Democritus

I) Introduction to the Thesis and Research Questions

The notion of voice has attracted significant attention in academic writing (Elbow, 1994; Charles, 1999; Tang, 2004; Dressen-Hammouda, 2013, 2014; Stapleton & Helms-Park, 2008; Matsuda, 2001; Matsuda & Tardy, 2007, 2008; Ivanič & Camps, 2010; Tang & John, 1999a; John, 2005; Herrando-Rodrigo, 2014, 2019, 2022; *inter alia*), as it is conceived as a compelling device to ascertain how discourse writing, and particularly, how specialised writing is shaped and constructed (Tardy, 2012a; Hewings, 2012; Silver, 2012). The various connotations attached to voice, such as its equivalence to stance (Pho, 2008), or as its conception as a superordinate of stance (Hyland, 2012; Thompson, 2012), to name a few, have allowed for a set of conceptualisations which may be interpreted and understood in many ways (Castelló et al., 2012). This is mainly one of the reasons why voice is conceived as a complex notion to delimit and account for.

Indeed, the fact that the notion of voice seems to be inherent to academic writing (Dressen-Hammouda, 2014) shows an interrelatedness between this notion and that of authorial identity, since authors, as individuals belonging to a specific community of practice, make use of their epistemological means to socially construct their selves and portray themselves towards their community. They do so by choosing noticeably conventionalised linguistic choices that shape and mark their discourse off other types of registers (Hyland, 2010b, 2011, 2012). These linguistic resources are strategically used to transmit not only relevant specialised content, but also essential information on the involvement of the writers in the research carried out. The notion of voice is then closely linked to the strategical ways in which writers represent and position themselves towards what they are describing and claiming, as well as towards the audience they are targeting.

The repertoire of linguistic strategies used in medical writing largely goes hand in hand with the visible, or invisible, representation of authors. In fact, the alternation of different types of strategies allows for several degrees of authorial manifestations in the text, which switch on and off at the authors' discretion. These linguistic choices range

from those asserting complete authorial visibility, when authors use the structure *we + active verb*, to those manifesting relatively complete authorial invisibility, when authors use an existential *there* pattern. A wide range of other intermediate categories, such as the agentless passive voice structure, are used to partially manifest their degree of involvement in the text. In addition, these linguistic choices also perform specific rhetorical functions that serve to adhere to conventionalised abstract features inherent in medical writing, such as neutrality, objectivity, rigour, and detachment. Indeed, these abstract concepts can be comprised within the umbrella term of impersonality.

Agency backgrounding or agency omission are key features of medical writing, as they are extremely frequent, and they are utterly motivated insofar as they let the audience focus on the data and empirical evidence rather than on the authors' experiences or opinions. Adopting an impersonal style is then seen as a rhetorical choice that is aligned with the way medical information is transmitted (and has been transmitted through time) and that is used to persuade other members of the community of the validity of the claims being made (Molino, 2010). Despite the perceivable impersonal reality of medical writing, personal-like structures are also used to counteract impersonality and favour an interpersonal dimension (Oliver del Olmo, 2004; Hyland, 1998) so that the messages given are clearly framed and thus perceived as trustworthy and convincing.

The original conception of this dissertation aimed at accounting for the various linguistic choices connected to the notion of grammatical voice existing in medical writing. First, the main interest primarily focused on the extensive use of the passive voice to transmit the aforementioned sense of impersonality. The study of the passive voice, however, somehow implied the study of its active counterpart, so active and passive voice structures were seen as the starting point for the analysis of the impersonal dimension of medical research articles through a corpus-based analysis.

The choice of the passive voice as a starting point was justified because it is undeniable that medical English shows an extensive inclination towards the use of these marked structures (Fromkin & Rodman, 1993). In fact, the passive voice is even portrayed as a staple of medical prose insofar as it serves as a discourse marker (Rundblad, 2007) and

allows for the topicalisation of relevant information. The way a passive sentence organises information is particularly useful and helpful for the empiric transmission of information, as it demotes the presence of the agent, either through backgrounding or omission, and promotes new research knowledge which relates to the objects under study (Swales, 1990; Solstad et al., 2004; Baratta, 2009; Banks, 1994; Givón, 1994, 2017; Ding, 2002). This reinforces the idea that the passive is an impersonal textual device serving as a guiding signpost which eases the readability of the complex amount of information being presented (Baratta, 2009), despite its various negative considerations as a pompous and verbose linguistic choice which theoretically obscures medical writing (Malmfors et al., 2004; Stuart, 2007; Matthews & Matthews, 2014; *inter alia*). As said, the study of the passive voice, however, implies the study of its active counterpart, the active voice. By using a personal active sentence, authors can promote themselves as the doers of the research activities carried out in the study and highlight their overall presence in the text.

As stated through the corpus-based study presented in this thesis, the high number of passive voice structures, alternated with active voice sentences, serves to partially represent the degree of authorial manifestation. Indeed, it is a partial representation because there exist other lexicogrammatical resources available to writers to manifest themselves in the text and extend this impersonalised character over medical writing. In fact, these other resources fail at being dichotomised within the active and the passive parameters.

This same issue appears when trying to delimit the extent of impersonalisation carried out by all these structures. For instance, an active sentence can be seen as personal if its subject is realised by the personal pronoun *we*, whereas it can be conceived as being more impersonal if its subject consists of an inanimate noun phrase, such as *this trial* or *this study*. In fact, if one reads a medical research article, it is possible to notice a lot of resources which flee from the dichotomisation *active* vs. *passive* and notably contribute to the perception of impersonality of medical texts.

As already noticed, the passive is frequently associated with the notion of impersonalisation in medical texts and it may seem that medical writing is conceived as impersonal due to the extensive use of this particular linguistic resource. However, this results in an oversimplification. The examples commented above illustrate and justify the blurring of the existing semantic correlation between the active voice and personality, on the one hand, and the passive voice and impersonality, on the other. Indeed, the dichotomised grammatical parameters *active* and *passive* fail to fit into the pragmatic and textual notions of authorial (in)visibility and (im)personality, as these concepts can be represented in a gradual manner. When noticing this, the thesis was redesigned by reconceptualising the parameters used to account for these nuances, both bearing in mind active and passive grammatical forms, on the one hand, and considering the notions of authorial (in)visibility and (im)personality, on the other.

Most authors have accounted for partial representations of (im)personality and authorial (in)visibility in academic writing. One of the most prolific aspects as regards personality is the study of self-mentions through first person pronouns and possessive adjectives (Brno, 2005; Dontcheva-Navrátilová, 2013; Harwood, 2005a, 2005c; Hewings & Coffin, 2007; Hyland, 2001b, 2002a; Ivanič, 1998; Martínez, 2005a; Tang & John, 1999b; Vergaro, 2011; John, 2009a; *inter alia*). Some others focus part of their research on the use of the passive voice in specialised registers (Millar et al., 2013; Swales, 1990; Solstad et al., 2004; Baratta, 2009; Banks, 1994; Givón, 1994, 2017; Ding, 2002; Tarone et al., 1998; Marín-Arrese et al., 2001, *inter alia*), as well as on the effects that nominalisations exert over specialised discourse (Liardet, 2016; Tyrikkö & Hiltunen, 2009; Hunston, 2002; Hyland, 2008, 2009; Ravelli, 1996; Hao, 2020; Guillén, 1998; Banks, 2005; Halliday, 1998, *inter alia*).

The study of abstract rhetors as impersonalising devices is also well documented (Johns, 2001; Master, 1991, 2001; Wiraszka, 2023; Seskauskiene, 2022; Khedri, 2016, 2020; McGrath, 2016; Hyland, 1998, *inter alia*), as well as the extensive use of reduced relative clauses to load the message with descriptive lexical information (Chomsky, 1957; Smith, 1961; Harwood, 2018; Wanner, 2009; Elsness, 2014; Hare et al., 2007; McKoon & Ratcliff, 2003; Merlo & Stevenson, 1998, *inter alia*). Other researchers,

from a systemic functionalist approach, also account for the analysis of thematic and rhematic structures and how the organisational packaging of information reveals authorial manifestation and impersonality in specific registers (Halliday, 1973; Halliday & Matthiessen, 1994; Eggins, 2004; Rodríguez-Vergara, 2017, *inter alia*). Lastly, not many researchers describe existential *there* patterns as totally depersonalising strategies (Sun & Cheng, 2015; Huddleston & Pullum, 2002; Basilico, 1997; Loock, 2013, *inter alia*).

Some of the previous authors have attempted to gather some of these linguistic strategies and have tried to gradually classify them according to their degree of (im)personality and/or authorial visibility (Martínez, 2001; Rundblad, 2007; Herrando-Rodrigo, 2014, 2019, 2022; Poulsen, 2015; Seoane & Hundt, 2017). This dissertation is aimed at collecting these various resources and extending the set of linguistic features devoted to portraying medical Research Articles as the archetypes of impersonalisation. The description of these wide array of lexicogrammatical patterns, the observation of their raw and relative frequencies, the analysis of their main functions, and their characterisation within a Cartesian coordinate system and within a continuum to represent the distinct layers of impersonalisation are aspects of novelty worth highlighting in this dissertation.

Indeed, the Cartesian coordinate system is aimed at including the most recurrent verbal constructions in medical discourse, considering as distinctive the two parameters of impersonality and grammatical voice. On the other hand, the subsequent continuum will consider various lexicogrammatical structures (some of which also appear in the Cartesian coordinate system), taking as main distinctive parameters those of impersonality and authorial visibility.

The literature review, the corpus-based study (and the Cartesian coordinate system proposed), and the close textual analysis (and its two-parameter continuum) presented in this thesis are aimed at shedding some light on the complexity associated with the notion of impersonality and its multi-layered representation in scientific English. To do so, I have considered various research questions, which are as follows:

1) Is a medical Research Article (in particular, a Randomised Controlled Trial) impersonal because of its extensive use of the passive voice?
2) Does there exist any other lexicogrammatical resources which contribute to the –(im)personalisation of medical discourse? If so, how do these patterns relate to (im)personalisation, grammatical voice, and authorial (in)visibility?
3) Is there any correlation between the dichotomies *personal* vs. *impersonal* and *active voice* vs. *passive voice*? Is there any correlation between these two concepts and authorial (in)visibility?
4) Is the Cartesian coordinate system the best way to frame and categorise all possible lexicogrammatical resources to delve into the notions of (im)personality, grammatical voice, and authorial (in)visibility in medical writing, or is a continuum better?

II) Outline of the Thesis Structure

Once having stated the main research questions that will lead to the development of this dissertation, I now turn to summarising its main chapters. Chapter 1 delves into the importance of scientific discourse in constructing scientific knowledge and understanding the social and natural worlds. It essentially focuses on how scientific writing enables the description of human experience and the construction of a very determined and defined reality by means of very characteristic patterns, such as technical lexis and nominalised technical grammar.

Chapter 1 also highlights the importance of English as a global *lingua franca* and how this relates to the construction of epistemological communities, or discourse communities, which share a very specialised, enclosed, and even restricted type of linguistic genre to express their communal practices. The fact that this linguistic genre is portrayed by some as obscure, distant, and too impersonal has originated the development of new approaches that advocate for a more personal, accessible, and reader-focused writing style. The Plain Language Movement is the main responsible for this incipient change of perspective supporting active voice and first-person pronouns

over the reviled passive voice, short sentences over excessively long strings of words, and simplified lexis over jargonised language, for example.

In the case of the medical scientific community, a large community of practice and a very prominent and prolific discourse community indeed, writing according to the customary patterns consolidated as inherent is crucial to actively participate in the development of any medical field. As seen in Chapter 2, the medical scientific community linguistically frames their research activity within distinct epistemological typologies. The main one is the Research Article, a piece of text which embodies the scientific research process itself, serving thereby as the canonical means to disseminate and transfer linguistically new advances in each medical field.

It is through these linguistic strategies that Chapter 3 attempts to delve into the notion of voice by revealing the traces of identity overtly expressed and purposefully hidden in medical writing. This chapter focuses on the notion of voice and connects authorial voice to the dimensions of interpersonality and impersonality, two concepts inextricably woven in the medical register. Chapter 3 also presents the main textual and linguistic representations of (im)personality and explores the theoretical assumptions expressed by the various strategies of depersonalisation authors use to both respect the rhetorical conventions found in medical literature and portray themselves as valid transmitters of valuable specific and specialised information.

Chapter 4 describes the methodology used in this dissertation. First, it delves into the rationale for using a corpus-based approach to descriptively analyse the notion of grammatical voice and other forms of impersonality, making use of Corpus Linguistics as a methodological tool to explore language and discourse. Then, a description of the corpus of randomised controlled trials in Medicine, the RCTs corpus, is provided, together with a brief description of the CQP interface used for the search of patterns, and the Simple Query Syntax and the CQP Syntax needed to explore them in the corpus.

Once the corpus-based study is dealt with, Chapter 4 describes the underlying complexity and problematisation when it comes to explore the notion of grammatical voice through specific patterns by means of a corpus-based study and describes the methods used for the design of a Cartesian coordinate system, aimed at scientifically representing the notions of impersonality and grammatical voice in medical writing. A close study of these textual patterns is regarded as a complementation to the corpus-based analysis, and the main reasons for using a close textual approach as a comprehensive extension of a corpus-based study are provided. The way these patterns are represented in Chapter 6 is through a continuum ranging from more personal to more impersonal-like structures. The rationale for the use of the continuum in this particular chapter is also provided, as well as a description of its design.

Chapters 5 and 6 account for the linguistic studies carried out in this dissertation. Chapter 5 consists of a corpus-based study which analyses the frequencies and the uses of the various active and passive patterns in the RCT corpus, paying special attention to the sections in which they tend to appear, and relating them to the typology of verbs most associated with them. This chapter also accounts for the problematisation implicit in the study of grammatical voice, particularly when using Corpus Linguistics as a methodology, and suggests the need to design a close textual analysis on the use of very specific patterns linked to the notions of impersonality and authorial visibility, in eight samples of randomised controlled trials. Before the textual analysis is carried out, a proposal of graphic representation of the notions of (im)personality and grammatical voice is given, through the use of a Cartesian coordinate system including these two notions as distinctive parameters (axes x and y, respectively).

Chapter 6 deals with the aforementioned close textual analysis. Indeed, these samples of randomised controlled trials serve to take an in-depth look at the frequent linguistic choices which carry (im)personalisation and assist in demoting the role of the agents in the text from a more textual and thus contextualised perspective. After the exploration and analysis of each linguistic pattern in its natural context, a continuum is designed so as to visually depict the relationship between these linguistic choices and the various degrees of impersonalisation and authorial intervention being used by medical authors.

As the patterns are introduced, the continuum is expanded and is seen to evolve into a relatively complex interplay of resources available to writers, resources which seem necessary to adhere to the conventions expected in medical discourse. The continuum in this chapter is thought to represent the most traditional visual approach to impersonality, ranging from more personal structures (leftwards) to more impersonal ones (rightwards).

As the description of the patterns will go from left to right, their subsequent addition onto the continuum would be in the same direction, so the continuum will be gradually completed from left (more personal) to right (more impersonal). The continuum then seems to be well-justified since it allows the inclusion of patterns perceived through the close textual analysis that were not identified in the corpus-based analysis.

Once both studies are presented and developed, important concluding remarks are made in connection to the overall analysis of the most prominent set of lexicogrammatical choices authors make use of in their medical writings. Special attention will be devoted to the two graphic ways that have served to classify pictorially the notions of (im)personalisation, grammatical voice, and authorial (in)visibility; that is, the Cartesian coordinate system (in Chapter 5) and the continuum (in Chapter 6). Once the study is concluded, I will suggest future research lines in connection to grammatical voice and the analysis of authorial (in)visibility and impersonalisation strategies in specialised settings.

1

Scientific Writing and its Connection to the English Language:

The Importance of Discourse Competence to Access the Scientific Community

---

Science is nothing
but perception.
Plato

When it comes to deal with the notion of *science*, it is important to assert that there is a clear connection between *science* and *knowledge*, since the core meaning of the former entails the core meaning of the latter. Science, 'knowledge' in Latin[1], is considered a system of acquiring knowledge. According to The Science Council (2018), science is "the pursuit and application of knowledge and understanding of the natural and social world following a systematic methodology based on [observable physical] evidence". With the practice of science, scientists are willing to observe, experiment, and thus understand (better) how natural phenomena, and by extension, how the physical and natural worlds operate. This is fully achieved through what is known as the scientific method, which essentially demands accurate observation and experimentation by "stimulat[ing] natural processes under controlled conditions" (Railsback, 2018)[2] in order to test ideas whose validity will serve to produce useful models of reality.

1.1. Science and scientific writing

The multiple facets of science are what really makes it a difficult and complex matter to define or describe. There are, however, some straightforward characteristics that are inherent to any kind of scientific inquiry, which are the following. Firstly, science is exclusively focused on facts happening in the natural world, without ever focusing its attention on supernatural reasons. Secondly, science must be dealt with as a way or 'path to understanding' rather than as a mere collection of facts; science is a way of understanding and learning about the real world, so its scope goes beyond a simple narrow study of raw data.

Thirdly, as a principle when doing science, scientists must formulate hypotheses to generate ideas, which will be accurately analysed and checked if they hold true. Once these ideas are rigorously tested and validated by the scientific community, they become truth claims, which are valid but not absolute at all, and thus expected to be revised and modified in the future by other members of the community. Fourthly and linked to the

[1] The Latin word for English 'science' is *scientia*, which means 'knowledge'.

[2] [accessed in July 2021] https://railsback.org/1122science2.html

third characteristic, scientific studies take part of a community endeavour. Scientists make science keep moving forward not only because of the values and capabilities shown in their work, but also because of their sense of belonging to a given scientific community, which reinforces the idea of science as a collectively constructed area, as it is explored further in this chapter.

Notwithstanding that science[3] is a huge umbrella term that covers a great number of disciplines, it is relevant to highlight that the fields included in this area follow a well-established methodological pattern, known as the *scientific method*. Scientific methodology does not need to include a universal set of steps shared by all disciplines. However, the basic steps it must follow are impartial observation and experimentation to test hypotheses, reasoned induction "to establish general rules or conclusions drawn from facts or examples" (The Science Council, 2018), critical analysis and repetition of experiments to affirm (and/or re-affirm) if the conclusions obtained are of true value, and lastly, verification and testing of the results and conclusions through "critical exposure to scrutiny, peer review and assessment" (The Science Council, 2018).

The core aspect of the so-called scientific method "is the reduction of perceived phenomena to fundamental, testable principles" (Wilson, 2004, p.11). These testable principles need to be written for science to become knowledge; so, science, in fact, becomes knowledge essentially by publication of research results. Writing about scientific phenomena is a vital fact that entails the development of science. Without the

[3] When dealing with the notion of *science*, another important issue that cannot be ignored is the way in which it is sub-divided. Science can be basically classified according to two major types: *natural sciences* and *social sciences*. The former are the ones whose objective is to study in detail the natural and physical world, whereas the latter are those sciences that mainly focus on and analyse human behaviour and society and do so in a systematised way. Another possible – and very used – classification is the distinction between *pure sciences*, those that are restricted to theoretical aspects, such as Mathematics and General Linguistics; and *applied sciences*, which as their name indicates are practically applied and serve to explore and solve experiential problems. Examples of applied sciences would be Sociology or Applied Linguistics. With this classification, one can state that science "seeks knowledge either for the sake of knowledge itself [when pure] or for its application to practical ends [when applied]" (Miettinen, 2015, p.7).

existence of detailed and accurate written scientific documents “based upon observable and replicable data” (Hanauer, 2008, p.51), it would have been extremely difficult – or even impossible – for all scientific disciplines to build up and evolve as they have actually done so far. Indeed, the job of a good scientist does not only consist of designing and conducting research in a laboratory to discover, for instance, a new vaccine to eradicate measles, or to define the side effects of a new antibiotic that is going to be launched by the most powerful company in the pharmaceutical industry. Yet, to be a successful and recognised scientist also means, among – of course – many other things, to transform data into meaningful scientific knowledge through a process of reasoning conducted through linguistic tools (Giere, 1991; Hanauer, 2008).

In fact, writing must serve “as an interactive, constructive process to build knowledge claims, develop compelling arguments, transform data into generalisations, report research results, and persuade colleagues and other people” (Yore et al., 2006, p.116). In other words, “academic publishing primarily aims to advance scientific and medical knowledge” (Joubert & Rogers, 2015, p.14), which means that success in the profession of science is inherently related to the fact of successfully publishing the results of experimental processes. The role of writing is thus essential, since the language used generally serves to “communicate the inquiries, procedures, and science understandings to [the scientific community] so that its members can access the validity of the knowledge claims, make critical decisions about [them], and take informed action on related problems” (Yore et al., 2006, p.114).

It is by making the other members of the scientific community aware of new claims in their field, once they have validated them, that the progression of science becomes a fact. Scientific knowledge is able to expand because it is transmitted linguistically (Joubert & Rogers, 2015) through distinct types of manuscripts, which are meant to “create permanent records to establish [the writers’] priority for discoveries and [these serve] as documented resources for reflection, analysis, and evaluation” (Yore et al., 2006: p.113; Chaopricha, 1997). Science is an area that “depends upon a complex network of inquiries, evidence, knowledge, and people to construct and judge new knowledge claims” (Yore et al., 2006, p.112); it is, defining it broadly, “a complex

interplay of phenomena, data, theories, beliefs, values, motivation, and social context both constituted by, and reflected in, its discourse" (Osborne, 2002, p.206).

Indeed, it is thus worth mentioning that "[t]he development of scientific knowledge is [essentially] dependent on [and directly connected to] the development of a scientific epistemology of knowledge" (Hanauer, 2008, p.51). The scientific epistemology of knowledge consists of

> a set of temporary descriptions and explanations that best fits the existing evidence and current understanding of the real world within the limitations of people's sensory and intellectual abilities. Science knowledge develops epistemologically as a hypothesis, and collected data is utilized to support or refute the hypothesis (Yore et al., 2006, p.113).

The contribution of new science knowledge is certainly the main purpose of scientific research, and it is possible through the discovery of results and their written-based transmission. These results, obtained from detailed and accurate investigation, add to the previous state of knowledge (Malmfors et al., 2004) and will, in turn, serve for other researchers as a basis for their future investigations to continue developing it. In this way, members of the scientific community can take part of a process that allows them to consume, conduct, and produce to again consume, conduct, and produce scientific knowledge. It is through this circular process, mainly plausible because of linguistic interaction, that they can work collectively "to transfer and disseminate knowledge" (Joubert & Rogers, 2015, p.13) for the construction and further development of their research discipline.

In this line, Hanauer (2008, p.21) emphasises the importance of the role of discourse within science asserting that

> [w]hile the collection of empirical data may be central to the work of empirical researchers, this data is not self-explanatory. Rather, (…), it is the **discoursal construct** [emphasis added] within which the data is conceived that makes it meaningful to the development of scientific concepts. It is scientific thinking and, by extension, **scientific discourse** which are crucial to the activity of scientific

> discovery. Essentially, this involves the recognition that science is a field in which theories are socially constructed within a community and that scientific knowledge is **heavily dependent on language** to provide a tool for the expression of scientific theories, hypotheses and conceptualisations (Kuhn, 1970; Taylor, 1996).

Scientific concepts, theories, and hypotheses can be fully developed because there is a truly problem-solving discourse construct that allows them to exist, known as *scientific discourse*, a type of discourse that embodies the scientific method. Scientific discourse is a variety or register of the language "in which certain words, and more significantly certain grammatical constructions, stand out as more highly favoured, while others correspondingly recede and become less favoured than in other varieties of the language" (Halliday & Martin, 1993, p.4). Even though the scientific discourse is a functional variety of the language and, consequently, it is restricted to some structural and linguistic patterns that "mark it off from other varieties of the language" (Halliday & Martin, 1993, p.10), there is in fact certain degree of internal variation, which triggers an absence of fixed and immovable boundaries in such a register.

The writing of science is, however, essentially recognised by distinct patterns of co-occurrence, the "patterns of argumentation and form-function [that] explore relationships among variables and causality among natural elements and events" (Yore et al., 2006, p.110). Such patterns suggest that science is necessarily grammatical because "[truth claims about the natural and social worlds] are chained together into [a narrative syntax, a syntax based on] sequences of agents and causes, relations and consequences" (Halliday & Martin, 1993; in Luke, 1993, p. xiv), which create a very specific type of discourse performance.

Through the use of grammar, science names, constructs, positions and approaches the social and natural worlds, and it does so by building "social relationships of power and knowledge between writers and readers" (Halliday & Martin, 1993; in Luke, 1993, p.xiv). The grammar of scientific discourse enables science to describe linguistically human experience through the connection of both natural and social worlds.

In English, as well as in other languages, writers of scientific language adapted natural language into new forms of wording and did so to construct a "fixed and determinate [reality, that of experimental sciences,] in which objects predominate and processes serve merely to define and classify them" (Halliday & Martin, 1993, p.20). The lexical and grammatical resources that characterise scientific discourse are *technical lexis*, on the one hand, and *nominalised technical grammar*, on the other.

According to Crystal (2003, p.372), scientific discourse is marked off from other types of discourse particularly because of its lexicon. Specialised terminology in scientific discourse was formed "by means of agglutinated classical morphemes" (Pérez-Iglesias, 2004, p.134) from Greek and Latin, because of their "richness (…) in word-forming elements" (Pérez-Iglesias, 2004, p.140). In the case of English, it had to fill the gaps in the lexicon, and did so by borrowing specialised terminology of classical origin and by adding it onto the lexicon of specialised discourse (Millward & Hayes, 2011). This was a process that happened along the 18th and 19th centuries, a period where specialisation began to be a request in any scientific enquiry.

As for the main grammatical device, that of nominalised technical grammar, it results from the continuous process of re-grammaticalisation that verbs (processes) and adjectives (qualities) undergo when transformed into technical nouns, whose prototypical meaning is that of an object. As such transcategorisational process has been highly frequent in the language of science, such variety of the language has become plentiful of technical terms. The interdependency between technical vocabulary and nominalised grammar stems from the fact that "[the creation of] a technical term is itself a grammatical process" (Halliday & Martin, 1993, p.8). When nominalising a verb or an adjective, the resulting nominal group becomes an element in another clause, changing thereby the syntax of the text and, consequently, changing its semantics, through the transformation of a process into an object.

This recurrent process of nominalisation, as will be further explored in this dissertation, is what makes the language of science seem so rigid, precise, objective, and reliable. Nominalisations are essential features in the writing of science because they serve to

construe a reality of something permanent and true, something that is so proven that will persist in time. The register of scientific discourse differs from general language as far as the different devices used to construe reality are concerned. According to Halliday & Martin, (1993, p.15),

> [w]here the everyday 'mother tongue' of common-sense knowledge construes reality as a balanced tension between things and processes, the elaborated register of scientific knowledge reconstrues it as an edifice of things. It holds reality still, (…) and in so doing, interprets it not as changing with time (as the grammar of clauses interprets it) but as persisting – or rather, persistence- through time, which is the mode of being of a noun.

Therefore, while in the general variety of the language there is still predominance for a clausal type of grammar, the grammar used in scientific register tends to be nominalised; it is nominalised because it serves to build up a noun-based reality, a pure reality of facts. Despite the grammatical differences between registers, it is worth noting the high influence the language of science has exerted over the general variety of the language as far as nominalisations are concerned. Nominalisations are so relevant and powerful devices that they have partially taken over clauses in clausal-based discourse; "(…) the language of science has become the language of literacy [and] having come into being as a particular kind of written language, it has taken over as model and as norm" (Halliday & Martin, 1993, p.11). This suggests that we, as humans, have developed a sense of fixed and determinate experiential reality close to that used in the scientific discourse, and hence the highly frequent use of its nominalised grammar in most present-day genres (Halliday & Martin, 1993, p.15).

The presence of a nominalised type of grammar has then become an inherent characteristic of scientific language, and as mentioned, it has become such a powerful device that it has even influenced pervasively general discourse grammar. Despite such an extensive influence over general discourse, the language of science is still an inscrutable area of domain[4] whose target audience is reduced to its consumers (readers)

[4] In line with what Halliday & Martin (1993) suggest, there is a trend which proposes the simplification of the language of science to make it accessible to a wider range of audience. According to them, "people

and its producers (writers). Such a restricted circle is what constitutes the scientific community. Novice members of the scientific community must be highly exposed to scientific discourse in order to acquire the epistemology of science to actively participate in its development and then achieve full membership.

Indeed, Malmfors et al. (2004, p.1) suggest that "research results, (…), do not contribute to knowledge and development unless they are communicated effectively". If such transmission of knowledge were not as effective as it is supposed to be, there would be no progression in the discipline; so, this is an imperative quality for a scientific text to fully achieve its natural purpose. The writers' ability to communicate in an efficient way is, in fact, directly related to the degree of success they may have in the influence of their text, and extensively in their field: the better and the more concise the exposition and explanation of the process and its results, the more readers of the entire paper it might have, the more influence it might exert over others' work, and then the more success and recognition the global procedural and academic performance of the writers will have.

If a paper intended to be published is in an orderly fashion and written conforming to the parameters of the field, the sooner its relevant information will be available to others and, from a general perspective, it would more easily allow the natural advancement of the field. However, even though a vast majority of proficient scientists are aware of the importance writing has in the success of their professional careers, some often find it difficult to accept that such success essentially depends on their skill to communicate their research results. Writing in science must be considered as a kind of "craft that build[s essentially] on clear communication of scientifically research acts" (Rogers, 2014, p.1) and, as such, its pedagogical implications should not be disregarded.

---

are looking for (…) a grammar which, instead of reconstructing experience so that it becomes accessible only to a few, takes seriously its own beginnings in everyday language and construes a world that is recognisable to those who live in it" (Halliday & Martin, 1993, p.21). This change of perspective is seen, for instance, in the foundation of the Plain Language Movement (PLM), which tries to reverse such a complex paradigm through encouraging the simplicity, plainness, and effectiveness in the language of science. The PLM is explored further in Chapter 2.

For a scientific text to be successfully published it may acquire two essential premises: a clear and well-structured organisation of the text, on the one hand, and a concise and skilful use of the register, on the other. As for the former, manuals and journal editors already tend to establish a clear layout of the text and a list of format instructions, and they do so through what is commonly known as their *house style*. For this reason, the structure of the writing composition does not usually pose serious problems to scientists when facing the actual writing process. However, probably the most challenging issue with which novice and even experienced scientists might find some crucial problems is the skilful ability to use language appropriately and concisely as inherently required by the scientific register.

To have their manuscript accepted for publication, writers must conform to the detailed instructions set by the *house style* of the journal they wish to publish in. They also need to master the language they write in and follow the several parameters given so that they can respect the conventions already validated by their discourse communities. For instance, according to Rogers (2014), one of the most difficult issues scientists encounter when writing science in English consists of packing highly complex information into clear and well-structured texts, and such linguistic challenge can only be accomplished with a proficient command of the language, a coherent structure of the text, and the necessary help of the guiding instructions developed by journal editors to acquire the "careful adherence to stylistic conventions used in [their] particular field of science" (Rogers, 2014, p.49).

In the case of Medicine, for instance, there was neither systematisation nor harmony in what the distinct *house styles* required, and this situation made the task of writers extremely challenging, confusing and complex, Nevertheless, this anomalous situation started to change because the *house styles* of important biomedical journals tried to harmonise, with the creation of the *International Committee of Medical Journal Editors* (ICMJE)[5], the dissimilar requirements asked to writers "to improve the quality and

[5] The current members of the ICMJE are *Annals of Internal Medicine, British Medical Journal, Canadian Medical Association Journal, Chinese Medical Journal, Ethiopian Journal of Health Sciences, Journal of*

clarity of manuscripts submitted to [their] journal[s], as well as the ease of editing" (Rogers, 2014, p.55). This proposal to unify criteria was initiated with the document entitled *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication* (http://www.icmje.org), which was basically created to serve as a guideline to make the process of writing more accessible to writers themselves, helping them

> [r]eview best practice and ethical standards in the conduct and reporting of research and other material published in medical journals, (…) [as well as] help[ing] authors, editors, and others involved in peer review and biomedical publishing create and distribute accurate, clear, unbiased medical journal articles. (International Committee of Medical Journal Editors (ICMJE), 2016, p.1)

Processes of harmonisation such as the one developed by the ICMJE bring to the fore the importance of setting a clearly understandable framework for scientists to perform the essential job of writing in much better conditions. To achieve the aforementioned valuable adherence to the stylistic and format conventions of the discipline, the writing of science must be fully "promoted as a central aim of the pedagogical scientific endeavour" (Hanauer, 2008, p.189). It is by making (novice) writers learn the basic "epistemological and procedural components of professional scientific activity" (Hanauer, 2008, p.189) that they can become fully "enculturated into the scientific discourse community related to their research discipline" (Yore et al., 2006, p.114) and show full responsibility for their claims. This would enable a continuous and non-stoppable flow of science construction, which is, as already pointed out, science's main purpose.

---

the American Association (JAMA), Nederlands Tijdschrift voor Geneeskunde (The Dutch Medical Journal), New England Journal of Medicine, New Zealand Medical Journal, Revista Médica de Chile, Public Library of Science (PLOS), Tidsskrift for Den Norske Lægeforening (The Journal of the Norwegian Medical Association), The Lancet, Ugeskrift for Laeger (Journal of the Danish Medical Association), the U.S. National Library of Medicine, and *the World Association of Medical Editors*.

To conclude, it is worth highlighting that science is a fundamental paradigm for the modern configuration of humankind. The advancements in science have caused a major improvement in the way humans perceive and know about the world and thus, in the way we live in it. Humankind has taken advantage from scientific advancements because of linguistic interaction through written-based documents that have enabled the transmission of scientific knowledge from generation to generation. This important body of scientific knowledge has expanded, and it still does, because of the construction of a specialised type of discourse: scientific discourse. The language of science has become extremely complex and specialised and has evolved in parallel with the increase of specialisation of all scientific disciplines. Indeed, it has exerted a very powerful influence over general discourse, applying its distinctive features to most present-day varieties of the language and a clear example of this is the high rate of nominalisations used in most present-day genres.

1.2. The Preponderance of English in Scientific Writing: English as a *Lingua Franca* (ELF)

As a vehicle of communication, efficiency, and progress, the English language has established itself as a global and universal language. Its importance, as well as its leading status, originated along with the expansion of the British Empire around the globe in the early 17th century. At that same time, the English language began to spread all over the British colonies and, consequently, took different forms depending essentially on the geographical area and on the existence, or absence, of other languages in the regions. These factors, among others, shaped different types of English variants, which have been classified into two main categories: 'core' or 'inner circle' varieties of the language; and 'non-core' or 'outer circle' varieties (Kachru, 1992)[6].

[6] On the one hand, core or inner circle varieties of English developed individually and reached high degree of prestige, mostly because they became the language of reference in those areas. In this sense, they acquired almost full status as a language and became the first language spoken in those territories. Examples of core varieties of English would be, for instance, North American and Australian English. On the other hand, non-core or outer circle varieties, also known as 'New Englishes' or 'World Englishes'

With its powerful character and strong presence worldwide, English has constituted the main means of communication in a variety of domains and has established itself as the global *lingua franca*. Its preponderance over other languages essentially derives from the fact that it has been the first language of the most powerful nations in the world, such as the United States and the United Kingdom, which have exerted strong influence on many fields worldwide over the last few decades (Mauranen, 2015). English has been taken as a model, as a medium of communication through which "members of two or more different linguacultures" (House, 1999, p.74) are able to interact.

English has then served, and still serves, as a tool to simplify and facilitate linguistic communication among members who belong to distinct linguistic communities and "for none of whom English is the mother tongue" (House, 1999, p.74). Such a role has led onto the establishment of English as "one of the symbols of our time, together with globalisation, networking, economic integration, and the Internet" (Mauranen, 2009, p.1), as a phenomenon "linked in with the overall trend towards the globalisation of information exchange, communication, and education" (Flowerdew, 2014, p.1). The importance of English as the common international language of communication has led to the creation of the phenomenon called 'English as a *lingua franca*' (henceforth ELF).

ELF[7], as said, refers to the fact that English is elected as "a 'contact language' between persons who share neither a common native tongue nor a common (national) culture"

(McArthur, 1998; Brutt-Griffler, 2002; Jenkins, 2003; Mesthrie & Bhatt, 2008; Melchers & Shaw, 2011; Saraceni, 2015, *inter alia*), did not evolve the same way as core varieties did. The reason for this to happen had to do with the pre-existence of other languages in those territories. Inhabitants, who were native speakers of those languages, began to acquire English as a second language. English was then greatly influenced by those native languages, resulting thereby in quite different variants from the language originally introduced, hence their somehow 'lower' degree of prestige (despite being recognised as official) and their inclusion within the outer circle category. A clear example of a non-core variant is the English spoken in India.

[7] For some time, authors have been using indistinguishably, along with the label 'English as a *lingua franca*', the terms 'English as a global language' (Gnutzmann, 1999; Crystal, 2003); 'World English' (McArthur, 1998; Brutt-Griffler, 2002; Jenkins, 2003; Mesthrie & Bhatt, 2008; Melchers & Shaw, 2011; Saraceni, 2015, *inter alia*), and 'English as a world language' (Mair, 2003). Other labels, which in fact are not that frequent, are 'English as a medium of intercultural communication' (Meierkord, 1996), and

(Firth, 1996, p.240). From a linguistic perspective, features of this English variant are essentially analysed as opposed to the general or standard variant of the language. The latter refers to the social context or framework in which English is used as a vehicle language. From a sociolinguistic perspective, the focus is on the context where this English variant is framed. Indeed, ELF has expanded gradually – and rapidly – because empirical work on the use of English as a common international language has gained importance within the Linguistics framework (Jenkins, 2006)[8]. Examples of research on pragmatics (Meierkord, 1996), phonology (Jenkins, 2000), and lexis and grammar (Seidlhofer, 2004) as well as other relevant contributions to the field (i.e. Widdowson, 1994; Jenkins, 2000, 2007; Seidlhofer, 2001, 2005; Mauranen, 2009; Ferguson et al., 2011, *inter alia*) have brought to the fore the importance of understanding the nature of ELF. Despite the fact that ELF is now recognised as an expanding linguistic area, it is still "the least recognised function of English in the world" (Widdowson, 1994; Jenkins, 2000, 2007; Seidlhofer, 2001; Mauranen, 2009), so much more exploration needs to be made in order for ELF to achieve its full potential[9].

'English as an international language' (Jenkins, 2000). Nowadays, however, "when English is chosen as the means of communication among people from different first language backgrounds, across linguacultural boundaries, the preferred term is 'English as a *lingua franca*'" (House, 1999; Seidlhofer, 2005, p.339).

[8] There has been a rapid growing interest towards ELF from a theoretical, descriptive, and applicational perspective. The theoretical research interest around ELF stems from its inherent nature as a language of contact. Theoretical studies on ELF mainly focus on how linguistic features are manifested within ELF discourse, and specialists try to shed light on the way universals of communication are principled and categorised. From a descriptive point of view, research on ELF tries to capture these features or universals, which are frequent and characteristic of ELF discourse, as well as the main differences and commonalities between such features and the ones typical of Standard English. This description is aimed at understanding the use of English as a second language in real contexts to perceive the changes of discourse in form, function, and meaning. Lastly, from an applicational perspective, the descriptive analysis of features and their inclusion within a theoretical framework are applied to the teaching of real use of the English language (Mauranen, 2008).

[9] The interest in ELF originated with several projects based on corpora, which have contributed to explore the empirical basis for understanding its nature and its pedagogical applications. Beginning with the *London-Lund Corpus of Spoken English* (LLC), the creation of large databases of text to study ELF grew rapidly and derived into the creation of other more recent projects such as the *Brown Corpus* and the *Lancaster-Oslo-Bergen Corpus* (LOB), the *English as a Lingua Franca in Academic Settings* project, and

A crucial aspect about ELF is that most of its users are non-native speakers of English, and in fact they have even outnumbered native speakers, even though both native (in any of its social, regional, and national varieties) and non-native groups of speakers are considered as potential ELF users. This situation positions English in a very paradoxical area because the influence of non-native speakers on native language change cannot be ignored. For this reason, descriptions and analyses of actual instances of English are useful insofar as they serve as pedagogical resources, as tools to improve the use of English among those non-native speakers who use ELF in international settings (Flowerdew, 2014, p.1). In this line, Bruce (2008, p.1) points out that the phenomenon of ELF and the "accompanying demand to learn English as an additional language (EAL), have given rise to exponential growth in the activities of language teaching". Indeed, ELF teaching must be made specific to cover those aspects that are proper of particular communities of practice that use English as their means of communication, in fields such as business or medicine, for instance.

---

the *Vienna-Oxford International Corpus of English* (VOICE). Indeed, the LLC was the first project that was highly determined to begin with the expansion of ELF as a field of linguistic interest. It is a corpus of spoken texts in British English compiled by Jan Svartvik between 1959 and 1990, which in fact derives from two different projects: the *Survey of English Usage* (SEU), at the University College London; and the *Survey of Spoken English* (SSE), at Lund University. The corpus has three different versions: the LLC:O, which is the original corpus, composed of 87 texts; the LLC:s, which is the supplement to the original corpus, formed of 13 texts; and the LLC:c, which is the complete and most updated corpus, composed of 100 texts (it has approximately 1 million words) (Svartvik, 1990). On the other hand, the BC and the LOB were compiled for linguistic research on modern English. The former (Crown University) was compiled in the 1960s by Francis & Kucera and is the first computer-readable general corpus of texts in American English (composed of more than one million words). The latter, also consisting of over a million words, was compiled at the University of Lancaster, the University of Oslo, and the Norwegian Computing Centre for the Humanities (Bergen) in the 1970s and was built up to provide a British English counterpart for the BC (Francis & Kucera, 1964; Leech et al., 1976). Thirdly, the ELFA project (Mauranen, 2008, 2009; Mauranen et al., 2010) is set up in Helsinki. This project is divided into three subprojects, which are aimed at investigating English as an academic *lingua franca*. The first subproject is the ELFA corpus, which also consists of a one-million-word corpus of spoken academic English as a *lingua franca*; the second one consists of the *Written ELF in Academic Settings* (WrELFA) corpus, which is consists of written texts from different genres (institutional, professional, and digital media); and the third subproject is the *Studying in English as a Lingua Franca* (SELF), which takes a more detailed view of ELF, as it is actually used in academic context (Mauranen, 2008).

Indeed, a particular community of practice whose members require a thorough command of English is the medical scientific community. In fact, according to Hersh & Zhang (1999), the world's most widely cited medical journals are published mainly in English, which shows the preponderance of such a language in the medical field. Scientists and clinicians, basically through evidence-based medicine, must make use of English when reading others' research, and the same occurs when they must transmit their own research through the publishing of papers, particularly if they wish to do so at an international level.

As success in Medicine is highly related to international publishing in prestigious journals, and to do so, scientists must be updated with the latest body of literature on the field they wish to publish in, being proficient in English is more than a requirement, particularly at the writing and reading levels. When writing a paper, writers must be aware of the fact that the message needs to be delivered according to the parameters established by the discipline they are writing about, so that they can ensure the transmission of the information in optimal conditions. Failure to comply with the adherence to the discipline may lead to the rejection of the paper by the publishing company and may also trigger a delay in the transmission of relevant information that would otherwise add immediately to the field.

1.3. Exploring the Notion of *Community of Practice* and *Discourse Community*: The Importance of *Discourse Competence* to Access the Scientific Community

The fact that the progress of science depends essentially on interaction among the members of the scientific community is a *fait accompli*. Among many other activities, those members, being either individuals or organisations, generate scientific ideas, test their validity through experimentation, publish the results obtained in scientific journals, organise scientific conferences, etc., and all these activities are aimed at accumulating and sharing knowledge that will serve to make science build on itself. In other words, interaction, and collaboration among members of the scientific community is what

keeps science, in this case, moving forward.

Bearing in mind the crucial role the notion of *community* plays in the construction of knowledge – in this case, scientific knowledge – it is worth considering two relevant concepts. On the one hand, the scientific community needs for a *routine of practice* to exist and achieve its main goals, so it may be considered, in this sense, a Community of Practice; on the other hand, it also requires an *epistemological rhetoric* through a specific type of discourse to spread the message, so it may also be categorised as a Discourse Community.

A Community of Practice (henceforth, CoP) is defined as a group of people "who share interest, a problem area or enthusiasm for a certain subject, look for problem solutions together and intensify their knowledge and expertise in this field via continuous interaction on an ongoing basis" (Wenger et al., 2002, p.4), and who "seem to develop an economical way of referring to things, persons, events and situations in their shared environment" (Cutting, 2000, p.1). For such a community to create and store knowledge, some important aspects are required.

First, members who belong to that community must share relevant information on their field of expertise, insights, and experiences so that they can diagnose problems and add new knowledge to already existing knowledge; second, the form of interaction between these members must be based on cooperation, respect, and willingness to contribute to the expansion of such a field; and third, all these must lead to the development of a communal understanding of their practice to objectively deal with the problems that may arise and "generate tangible, measurable, value-added benefits to the business" (Nickols, 2003, p.4). Taking these general features into account, one may assert that the medical scientific community accomplishes these several requirements and may then be considered a CoP.

Indeed, these general requisites may also apply to define the aims of other types of communities, which are also focused on creating knowledge within a shared type of framework, but which may differ from the notion of CoP as far as their general

perspective is concerned. This is the case of, for instance, Discourse Communities[10] (henceforth, DC), which, according to Swales (1991), are groups whose members have common goals, particular mechanisms of intercommunication, a restricted set of genres and specialised terminology. In addition, one of the main important distinctions between a CoP and a DC is that the former focuses "on the aspect of common practice developed in order to be effective in the domain" (Pogner, 2005, p.9), whereas the latter is based "on the aspect of common discourse in order to be effective in the domain" (Pogner, 2005, p.9). Even though DC and CoP are two different types of communities with a clearly distinguished scope, they are not mutually exclusive concepts, as they both describe the social aspects of knowledge, even if their main aim is different.

Central aspects to consider when dealing with the notion of DC are the use of language within discourse, as well as common thinking, research, and learning. A DC uses various specific patterns of language, which are framed within a set of conventions that have been developed by tradition. Thanks to established mechanisms and media, a DC is not only able to share and exchange information, but also to use common specific terminology and domain-specific discursive genres, which must be well-known by those taking part (or wishing to take part) in the community. Indeed, such linguistic and formal conventions are used as the basis to create stretches of written text on well-based documented proof, and it is through writing that individuals feel connected and experiment the sense of belonging to a particular community and thus demonstrate their identity as members.

Both CoP and DC are tightly linked to the idea of *learning*, since individuals aspiring to be members of the community must learn by gradually taking up its practices and rhetorical discourse, respectively. This gradual implication allows novice members to move from a more peripheral type of membership to a more central one, connecting the

[10] The concept of *Discourse Community* has been widely discussed and debated by Bizzell (1982, 1992); Porter (1986); Freed & Broadhead (1987); Rafoth (1990, 1988); Cooper (1989); Swales (1990, 1993); Killingsworth, (1992); Devitt (2004); Martín-Martín (2004); Hughes et al. (2007); and Hoadley (2012), *inter alia*.

idea of learning to that of empowerment. It is thus important that there is enough space for legitimate peripheral participation so that aspiring members can learn the specific practice and discourse to empower themselves and successfully access the community gradually. This learning process may culminate when the individual achieves an expert-like centrality.

In the case of a DC, it is essential that novice writers develop what is understood as *discourse competence*[11] so that they can "integrate a wide range of different types of knowledge in order to create extended written discourse that is both linguistically accurate and socially appropriate" (Bruce, 2008a, p.1). The learning of discourse competence does not only correspond to the ability to use discrete linguistic units optimally but also to integrate them with pragmatic knowledge and conventionalised forms of communication to create extended coherent and cohesive discourse. In order for a novice writer to get experience and become a discursively competent member, it is necessary that s/he is "situated in authentic practice contexts or practice fields" (Hoadley, 2012, p.290); in other words, that s/he starts writing to generate, apply, and reproduce knowledge so that s/he learns the mastery of the knowledge systems that are necessary to acquire a full member identity within a community.

[11] According to Canale & Swain (1981), *discourse competence* is one of the four components that integrate the model of communicative competence. These components are as follows: the *grammatical competence*, which refers to the mastery of the linguistic code to understand and express the literal meaning of utterances; the *sociolinguistic competence*, which makes reference to the knowledge of rules and conventions of the socio-cultural context in which the linguistic communication is set; the *strategic competence*, which is aimed at enhancing the effectiveness of communication through verbal and non-verbal strategies; and the *discourse competence*, which refers to the mastery of rules to create a meaningful unit of text which is cohesive in form and coherent in meaning.

1.4. The Plain Language Movement: Towards a Simplification of Scientific Writing

When a novice writer faces the task of writing a scientific paper, they may get overwhelmed by the complexity of the scientific register. One of the crucial problems that general and even specialised readers may encounter when reading a scientific paper is the complexity of the style of writing. Scientific writing is usually dense, complex, and difficult to understand because "many scientists believe that they have to write only for colleagues or experts in their own field", without taking into account that readers may not be experts in the topic and may also be "overwhelmed, overloaded, and too busy to wade through dense writing" (Rogers, 2014, p.3). As science is a complex matter, it is assumed that scientific writing and its epistemological repertoire of resources must be complex too. There are then several misconceptions such a type of writing has held through time.

Rogers (2014, p.9) gives account of the frequent shortcomings of scientific texts that result, as said, from myths and misconceptions, which certainly unease the optimal transmission and reception of the text. According to him, in the past, it was an accepted trend that scientific texts had to make use of extremely long and complicated sentences as well as weighty nouns derived from verb forms. Other important misconceptions frequently found in scientific writing refer to the overuse of passive structures, avoiding thereby reference to authorial performance and hence avoidance of first-person pronouns; as well as the inconsistent use of technical terms and units; the mixture of creative and scientific writing; the mixture of results and their discussion; and the poor structuring of text.

Indeed, all these shortcomings have created the impression – or set tradition – that scientific writing is in a way obscure and difficult to understand, even by specialists on the field. This obscure and alienating writer-focused style has started to transform into a more reader-oriented style, where the reader and its understanding of the text have become pivotal for the successful transmission of the message.

This change of perspective stems from an initiative, promoted and reviewed by J. Locke[12], which was labelled as the *Plain Language Movement*[13] (henceforth, the PLM). The PLM attempts to demonstrate "the benefits of writing clearly and concisely in a reader-focused style" (Rogers, 2014, p.5), avoiding dense and obscure writing. In general terms, this movement essentially advocates for the ease of scientific writing by emphasising the logical organisation of the text. It also promotes the correct combination of everyday words and technical terms, avoiding "unnecessary jargon [which only] obscures meaning and wastes the reader's time and patience" (Stuart, 2007, p.73) and usually provokes an interrupting effect. It also favours the use of active voice and first-person pronouns instead of impersonal and passive structures, as well as the use of clear and short sentences instead of long ones with embedded clauses.

According to these authors (Rogers, 2014; Stuart, 2007), the creation of the PLM cannot be mistaken as an attempt to oversimplify scientific writing, but it must be treated as a simplifying approach to help readers of scientific papers easily understand the message and "absorb the ideas effectively" (Stuart, 2007, p.73). It could be considered that the degree of complexity and abstraction of any scientific field is inversely proportional to the degree of complexity and abstraction of the writing style used in such a scientific field. For instance, the writing of Mathematics, a very abstract area, favours the use of active voice and personal pronouns whereas the writing of Medicine makes more use of passive and impersonal constructions (Salazar et al., 2013). According to Rogers (2014, p.1), "[t]he beauty of (…) scientific writing is its ability to express the most complicated concepts in clear words and to point out the beauty of science without distracting decoration".

Indeed, taking Rogers' words into account, for a scientific text to be 'beautiful', it must show some essential features or qualities: it must be "understandable, transparent, clear, credible, efficient, and simple" (Rogers, 2014, pp.1-2). Scientific writing must be

[12] Joanne Locke is the Senior Policy Advisor and Plain Language Coordinator at the US Food and Drug Administration (FDA).

[13] American Medical Writers Association [AMWA] Journal, Vol. 18, [a], 2003; http://www.amwa.org/default/members.only/vol18.1/amwa18.1p5.pdf)

understandable and simple, in the sense that its readers should be able to go through the entire paper easily, without being distracted with complex and ornamental expositions, bearing in mind that most would not be native speakers of the language (Stuart, 2007, p.73); transparent, as published scientific texts are usually the only source available to other researchers to access the research done; clear and credible, since it must inform precisely, accurately, and with an adequate style to specialists (and also non-specialists) of the writers' area of research (Stuart, 2007, p.73); and efficient, since both content and style must conform to the parameters established by journal editors or *house styles*. Scientific texts must not be 'alienating' and need to be made more accessible, according to him.

The language of science must become accessible to anyone interested in the field (Stuart, 2007, p.63) and writers then must try to transform the complex style of this register into an uncomplicated and simplified one. To do so, scientific language must be re-approached from a writer- to a reader-focus, as mentioned earlier. The role of journal editors seems crucial for this change of perspective and prestigious journals offer a section called *Instructions for Authors*, which set all the questions related to format and style to be consulted by writers who wish to submit a scientific paper. Such instructions function as selective parameters that are used to discriminate among the total amount of texts sent for revision and publication. The most prestigious journals tend to go in line with what the *Plain Language Movement* proposes: the natural use of plain and conversational language (Stuart, 2007, p.73).

In medical register, for instance, clear examples of such tendency are the *British Medical Journal*, which proposes writers to write in clear, direct, and active style, or the *Journal of American Medical Association*, which also supports the idea that writing must be clear. However, both messages are too general and somehow vague, and, despite pointing out a trend towards the simplification of scientific language, do not fully achieve its main purpose. As will be seen further in this dissertation, the set tradition of scientific writing as complex and impersonal is still perceived, so one can draw the conclusion that the set tradition of scientific writing as dense and complex is factually solid.

Indeed, it is the journal editor, sometimes supported by the set traditions and the accepted trends in a given field, or sometimes helped by new approaches to the writing of science (like the PLM), who is finally responsible for censoring or accepting the quality of a scientific text. “The distinctive quality of scientific language lies in the lexicogrammar (the ‘wording’) as a whole” (Halliday & Martin, 1993, p.4), which means that if writers aim at producing qualified writing, they must write, apart from other facts, respecting the optimal proportions of technical and non-technical terms “to create a [proper] discourse of organised knowledge”, a text with proper *style*. Day (1998) defines the notion of style as the personality shown by a scientific writing through its consistent wording plus the accomplishment of the editorial requirements set by the particular house style. If the paper accomplishes both requirements, if it passes on pertinent information in a clear and unambiguous style (Rogers, 2014, p.39), it will “facilitate the digestion of the scientific message [by also] shorten[ing] the time-consuming process of editing[; and then it will] help to make the tasks [of both writing and reading] enjoyable” (Rogers, 2014, p.67; Stuart, 2007, p.36).

The PLM considers the content of the scientific paper as a much more important aspect than the use of stylistic devices. For a paper to be understood, it is fully recommended to organise the content clearly and logically, but using a simple writing style will reinforce the understanding of the informational content. It is possible to classify the various aspects recommended by the PLM as far as the writing style of the paper is concerned. These several aspects can be classified into four major categories: technical terminology and jargonised writing, syntax, verbal tenses, and personal and impersonal structures. These four main areas are described as follows.

The first aspect is that of technical terminology and jargonised writing. Technical terminology, as mentioned earlier, is an essential feature of scientific language (Halliday & Martin, 1993, p.4) and it is extremely difficult to be unnoticed. Even when one is not familiar with a specialised type of register, it is easy to label it as such because of the highly frequent use of specialised terminology. According to Rogers (2014, p.6) and Matthews & Matthews (2014, p.17), the use of technical terms cannot be totally omitted to make language of science clearer. Their role is to add information

that is valuable for the understanding of the paper, and they conform a feature that is inherent in such register, so the absence of technical terms in a scientific paper would go against its prototypical and functional nature.

The problem, however, is when technical terminology is overused and misused so that it makes the writing jargonised. Jargonised writing "consists of highly speciali[s]ed technical slang arising from the overuse and misuse of obscure, pretentious, or technical words or phrases" (Matthews & Matthews, 2014, p.126). In such a case, it is difficult for general and even specialised readers to understand the content, and this "can make them feel excluded from something they should be involved in" (Stuart, 2007, p.41). Therefore, the use of technical terminology must be carefully dealt with when writing a scientific article.

To avoid jargonised writing, the PLM advocates for the avoidance of unnecessary technical terms. Scientific terminology must be explained at the same time it is being introduced in the text; however, if this is not possible and writers need to include complex terminology, they must include a glossary explaining in detail the meaning of each of the technical terms mentioned, so that readers do not get lost when coming across one of them. It also recommends the substitution of specialised terms for everyday words (Stuart, 2007, p.42) to make the text accessible to anyone interested in reading it, whether being an expert on the field or not.

An issue directly connected to technical terminology is that of nominal grammar. A nominalisation, as exposed before, is a grammatical device highly used in scientific register. If such grammatical device is overused, that is, if a scientific paper contains too many nominalisations, its reading becomes "dull and heavy going" (Stuart, 2007, p.43), and it comes into contradiction with the plain use of language. Instead of using nominalised forms or "naming words", which are weighty, the PLM defends the use of the corresponding verbs, which Stuart (2007, p.42) calls "doing words". These types of words create a discourse much lighter to cope with and generally make the sentences shorter to achieve a maximal degree of readability.

Furthermore, it is also recommended to alternate the use of short sentences with longer ones (Stuart, 2007, p.36) in order to build up a balanced and non-extremely simplified type of discourse. It is important to bear in mind that "[v]ariation in sentence length and complexity helps sustain reader interest [since] a publication crammed with overly long sentences is difficult to follow, [as well as] a sustained string of extremely short sentences can be choppy and annoying" (Matthews & Matthews, 2014, p.111).

However, the presence of short sentences (*circa* 15-20 words) in a scientific text is necessary, because it makes the text more appealing and, in fact, makes the processing of information easier (Stuart, 2007, p.36). Paragraphs must also be considered as far as length is concerned. Writers must reduce the complexity and length of sentences and paragraphs if they have already used complex and long words, and this trick will serve as a compensation effect (Matthews & Matthews, 2014, pp.109-110).

Another aspect that contributes to jargonised writing is the overuse of hedges and verbiage. On the one hand, hedges serve to protect scientists' "arguments or statements with qualifications that allow for unknown contingencies or withdrawal from commitment" (Matthews & Matthews, 2014, p.113). When hedging, scientists hide themselves from the claims they are making and the text ends up showing certain degree of impersonality, which, to some, may be considered as a lack of self-confidence. Verbiage, on the other hand, consists of the overuse of unnecessary words, which often "leads to a sort of ritualistic, pompous writing style" (Matthews & Matthews, 2014, p.114); that is, instead of using the expression *in light of the fact that*, for instance, writers must try to find a less wordy equivalent expression, such as *because*. As the PLM advocates for the avoidance of jargonised writing, writers are strongly recommended not to use hedges and verbiage because they also go against the plain use of the language.

The second aspect, also highlighted by the PLM, refers to the proper use of syntax in the language of science. Matthews & Matthews (2014, p.125) assert its importance when saying that "more than one interpretation of a sentence or phrase is unacceptable, so careful attention must be paid to both word choice and word arrangement". The

prototypical order of words in a clause, that of the SVO structure, is the one logical and, in fact, expected since it "makes ideas easy to follow because the words appear in the sequence which things happen" (Matthews & Matthews, 2014, p.112) and it does not interfere to effective communication. According to them, readers will understand the content of the scientific text easily if words are arranged logically and if writers tend to avoid pomposity, verbosity, and of course, misunderstandings of any kind (Rogers, 2014, p. 55).

Modifiers, as well as verbal phrase danglers, also play an important role as far as the structure of the sentence is concerned. Modifiers, on the one hand, must be as close as possible to the words, phrases, or clauses they modify (Rogers, 2014, p.59), that is, their position in the sentence must be non-marked; however, there are some exceptions to this rule. In the case of adverbs, even though a traditional rule of grammar says that they must "follow, rather than split, the infinitive of the verb in sentences where the adverb modifies the verb" (Rogers, 2014, p.56), in scientific writing it is common to split the infinitive if the adverb strongly qualifies the verb, because it does so to add emphasis to the nature of the action (Rogers, 2014, p.57).

On the other hand, verbal phrase danglers, such as participle danglers, are also abundant resources that must be avoided at all costs since they contribute to a confusing type of writing. A participle dangles when it implies a subject that does not correspond to the subject of the main clause of the sentence; therefore, "a dangling participle implies an actor but does not specify who or what it is, thereby leaving proper participle-subject matching up to the reader" (Rogers, 2014, p.59).

The third aspect refers to the verbal tenses used in the language of science. "Proper use of tense in scientific documents derives from scientific ethics" (Rogers, 2014, p.41), because it is a basic feature that serves to determine the status of the scientific work that is reported (Matthews & Matthews, 2014, p.48). If the information reported refers to published results, established knowledge, or description of tables and figures the verb tense that must be used is the present simple; however, if the information refers to unpublished data, the methods, materials, and results, and attribution of the writers' own

results, the tense that must be selected is the past simple (Matthews & Matthews, 2014; Rogers, 2014, p.43). Present perfect is essentially used for repeated events and for the description of a situation that persists (Rogers, 2014, p.41).

Indeed, writers are also given information on how tenses are used according to the sections of the scientific paper where they appear. The abstract must be written in past simple tense, because it refers to the authors' own unpublished results; the introduction must be written in present tense since it is a way to emphasise previously established knowledge; the materials and methods, as well as the results section, are written in past tense, because it serves to describe what the authors did and discovered through research; finally, the discussion section uses both present and past simple because it links the authors' research results to established knowledge claimed by previous researchers (Rogers, 2014, p.43).

An aspect that has been a controversial issue for many years and poses several problems to writers of scientific papers is the type of perspective they use when informing readers about their research. As writers, they can choose whether to use personal devices (such as self-mentioning and active voice sentences) to make their texts seem more approachable and personal, or to use impersonal constructions (such as passive voice sentences or inanimate subjects in active sentences), which clearly show a certain degree of detachment. As scientific discourse is essentially built on accountability of empirical facts, according to the PLM, writers tend to overuse an impersonal writing style in their professional publications because they believe their writing will sound more objective, empirical, formal, and rigorous.

By writing with complex and impersonal structures, scientists emulate the writing style considered as prototypical in the set tradition of scientific writing. For a lot of scientific communicators, writing using personal devices, such as personal pronouns, is "inappropriate, even impolite" (Rogers, 2014, p.68) and "unscientific" (Matthews & Matthews, 2014, p.142). They are often taught that impersonality is a necessary feature for a scientific text "to preserve a proper scientific atmosphere of objectivity around an account" (Stuart, 2007, p.18). However, objective, empirical, formal, and rigorous

writing may also be achieved, as advocated by the PLM, through the use of personal devices, which certainly make the text more appealing and easier for the reader to go through. It is true that scientists must be as much objective as possible, "but thinking, believing, considering, and hypothesising are all subjective activities" (Stuart, 2007, p.20), and this reinforces the idea that writing as personally as possible is the most natural way of expressing one's arguments, even in the writing of science.

Nevertheless, although there are a lot of devices that can make a text be either personal or impersonal, the most important device that accounts for the degree of personality of texts is the notion of *grammatical voice*. Grammatical voice refers to "the form of transitive verbs that shows whether the subject acts or is acted upon" (Matthews & Matthews, 2014, p.142). Both active and passive sentences contain either an implicit or an explicit subject. Its position in the sentence will depend exclusively on the type of grammatical voice the writer chooses. In active sentences, the subject is the element that acts, that is, the subject is the agent of the action mentioned by the verb. On the one hand, active sentences are transitive sentences whose structure is composed of an agent, a verb, and an object, and this is considered the prototypical and logical order of a sentence.

On the other hand, passive sentences contain a grammatical subject that does not coincide with the agent that performs the action expressed by the verb. In a passive sentence, there is always a grammatical subject (the subject which is acted upon), a passive verb (the copular verb plus a passive participle form), and optionally, there is the agent (or the cause) introduced by a *by*-phrase. As the set tradition of scientific writing tends to omit agency because it is commonly conceived as unimportant, most passive sentences end up being agentless. The marked order of passive sentences is then used to emphasise other elements rather than the agent of the action, but its overuse makes the writing style become, according to the PLM, certainly plodding and not easily readable.

Taking the set tradition in the language of science into account, the house styles and editors of different prestigious journals have reached an agreement with which complex

writing is certainly threatened, in favour of a more active, direct, and plain style, as exposed in the following quotation.

> Nowadays, most modern grammarians, linguists, and editors agree that the exclusive use of the passive voice is redundant. In this time and age where brevity and conciseness of manuscripts are critical factors, the active voice helps to keep messages lean and clear. Moreover, our time constraints with both the writing and reading of scientific information call for an unambiguous language involving active verbs and personal pronouns whenever possible (Rogers, 2014, p.68).

As seen, scientific writers are strongly recommended to use the active voice, instead of the passive voice, because it seems to make the language clearer, unambiguous, and more dynamic. Apart from this, a passive sentence often uses more words (Rogers, 2014, p.68) and thus creates a much more complex, stuffy, wordy, pompous, obscure, and impersonal message (Stuart, 2007, p.38) when compared to its active counterpart. The overuse of passive constructions certainly "creates an atmosphere of passivity" (Stuart, 2007, p.22), which makes the language of science be more complex and blurred than how it should really be. However, writers must not exchange all passive sentences into active ones because they do not mean the same (Stuart, 2007, p.23). Therefore, writers must be able to know properly when to choose an active or a passive construction to transmit exactly the intended message.

General language considers active voice sentences the norm since "it is the natural voice in which most people speak and write" (Matthews & Matthews, 2014, p.142). As defended by the PLM, the fact of writing actively becomes a good way to approach the language of science to general language, making it sound more natural, friendly, and less alienating. However, although the use of the passive has somehow been restricted, there are some sections of the scientific paper where these structures are necessary. For instance, "descriptions of materials and methods, and statements of results[, survey work and statistical calculations] are meant to be straightforward presentations of facts" (Stuart, 2007, p.20), so these sections require a higher proportion of passive constructions if compared to other sections, because what is important to be emphasised is "who or what was in particular state or had something done to it" (Stuart, 2007, p.22).

In other sections, such as in the introduction or discussion, it is appropriate to mention agents, since it is important to distinguish between what others have previously said and done and what the writers of the paper actually say and do. It is also advisable to account for the percentage of passive structures appearing in a text. To some, to create a well-balanced text as far as active and passive constructions are concerned, it is necessary to mix both types of structures, "bearing in mind that the proportion of passive verbs should not exceed 30%" (Rogers, 2014, p.69).

According to Matthews & Matthews (2014, p.139), scientists also tend to use a very restricted set of verbs, which repeat over and over again. The use of verbs such as *demonstrate*, *exhibit*, *present*, *observe*, *occur*, *report*, *show* makes the writing certainly monotonous and poor, and in most cases, these verbs are coupled with non-human subjects (Matthews & Matthews, 2014, p.142). The sentences formed by a non-human subject plus an active verb are also frequently found in scientific register. Even though in this case the voice of the sentence is active, human agency is hidden behind the meaning of the non-human subject, so it could be treated as an impersonal type of construction.

To sum up, even though technical terminology, jargonised writing (long sentences, hedges, and verbiage), complex word order, and impersonal constructions still predominate, or at least are highly used in scientific writing, editors and houses of style are making a big effort to transition from an impersonal type of style to a clearer, plainer and a more personal-like style in the writing of science. Scientists must be skilful enough to talk about complex concepts in plain and direct style, but the truth is that to do so, they must be pedagogically trained to achieve such level of command.

1.5. Summing up

Scientific writing is the primary form of conveyance of the newly discovered scientific facts and ideas that exist as dominant material that goes hand in hand with the transformation of the environment. The academic credibility, status, and power of scientists and scientific institutions depend on how efficient researchers can communicate their findings. Indeed, the publication of research results through written-based transmission is crucial for the development of any scientific enquiry. Scientists contribute to the construction of new scientific knowledge through writing, and if such a process is appropriately done, it may become a guarantor of success in their discipline of research.

Being proficient in English, the *lingua franca* of science, and knowing the dominant specificities in the register are fundamental issues insofar as they allow members of the scientific community to actively participate in the expansion of their field of expertise. In this regard, writers must be discursively competent if they wish to publish in international journals and thus target a wider audience. English is then seen as a critical instrument influencing the availability of scientific information and serves as the vehicular language when it comes to create and design new theoretical tools and methodologies in science.

The scientific community is a semi-closed circle of various consumers and producers of scientific knowledge. Emphasis has been added on the fact that the scientific community, in general, can be referred to as a community of practice, on the one hand, and as a discourse community, on the other. It is a community of practice because members who make up the community share a routine of practice which, through cooperation and interaction, seek creating and storing scientific knowledge for both a restricted group of members (community), and humankind in its broadest sense.

Besides, it can also be regarded as a discourse community because members share a set of rhetorical and linguistic resources (lexical, grammatical, textual, semantic, and pragmatic) that have become established through use and tradition and are now

common and expected in the register; for instance, a frequent use of stacked nominal noun phrases, a balanced use of active and passive structures, and a proper use of technical terminology, among many others.

Against this monolithic and established tradition, there is a trend promoted by the Plain Language Movement, which advocates for a move towards a more reader-oriented, and thus a more reader-friendly, approach to the writing of science. However, and as will be further explored in this dissertation, the writing tradition is well established and a great number of impersonalisation devices are crucial and distinctive features of scientific writing, in general, and of medical discourse in particular.

2

Medical Writing and the Specificity of the Language in the Field of Medicine:

The Research Article as Embodiment of the Scientific Process

Wherever the art of Medicine is loved,
there is also a love of Humanity
Hippocrates

The design, conduction, and analysis of scientific experiments are steps of paramount importance in the development of any kind of empiric field. The communication of new research results, however, is a fact whose relevance should not be dwarfed. As Malmfors et al. (2004, p.9) point out, "communicating science is as important to the scientific process as designing, conducting and analysing the experiment itself".

Having dealt with the notions of Community of Practice and Discourse Community, having pointed out their particularities and the importance of epistemology in the progression of any scientific field, I will now focus on the medical scientific community, a group of practice that shares a characteristic type of discourse and is recognised as highly prolific as far as its writing is concerned.

2.1. Writing in the Field of Medicine: The Cases of the Research Article and Randomised Controlled Trials

The medical scientific community is a community of practice insofar as it aims at sharing information and experiences so that the field it works on, the field of Medicine, expands through the creation of new knowledge. Thanks to the cooperation of its members, the medical community has developed a wide range of practices, which deal with problems and their solutions to help the community itself learn about new insights that promote the progression of the field. At the same time, the medical scientific community can also be regarded as a discourse community, since it makes use of an epistemological rhetoric that serves as a framework for the communication of knowledge. Such a community has acquired, through tradition, a repertoire of linguistic strategies and mechanisms of intercommunication that must be learned by peripheral members of the community who wish to achieve a centrality in the field, and such an achievement is possible largely due to the acquisition of a particular type of discourse competence.

A medical writer is considered discursively competent when s/he shows in his/her discourse those writing skills that must be inherent within a medical piece of writing.

Those skills are usually specified, described, and exemplified in manuals so that writers intending to have their work published approach the expected linguistic criteria found in real samples of medical texts. Through their writing, writers must first master the domain they are writing about and be aware of the typology of audience to which the discourse is addressed.

Secondly, writers must keep themselves abreast of new crucial information and their search strategy must be highly planned. Third, in general terms, the objectives of the study must also be clearly delimited and organised, and the type of writing as well as the information provided must be succinct, logical, and accurate, and the methods of communication must be transparent and consistent. The use of interpersonal skills such as negotiation, problem-solving and decision-making strategies are fundamental in such a specialised framework. The critical interpretation of results and the limitations found in the study are also expected facts.

The next section will delve into the specific characteristics found in the medical discourse community as well as into the most relevant textual typologies found in the field. Special emphasis will be put on the Research Article, a piece of text which can be considered as the embodiment of the scientific research process.

2.2. Language in Medical Discourse

In Chapter 1, I briefly explored the basic features of scientific language, which were essentially a nominalised type of grammar, on the one hand, and an abundance of technical lexis, on the other. To specifically delve into medical language, I will highlight the linguistic features that are particularly recurrent in such a register, using Maglie's (2009) distinction between lexical (morphosyntactic and semantic) features, and syntactic features, some of which will serve as guiding descriptive parameters for the analysis developed in chapter 3, where I will focus on the particular strategies that are used to impersonalise and make authorial presence somehow invisible.

The lexical features that typify medical English can be grouped into two sub-types: morphosyntactic features, on the one hand, and semantic ones, on the other. As for the *morphosyntactic features*, medical language is expected to contain two types of specialised terminology: anatomical terms, usually words of classical (Latin and Greek), French, and Germanic (Anglo-Saxon) origins, which are internationally standardised; and clinical terms, which need to develop quickly in all medical branches and in fact contribute to a certain terminological chaos (Dzuganová, 2013, p.56). All these terms denote key concepts relevant to the medical field and while the former are used to label and to provide the writing itself with transparency and precision, since most of these borrowings are formed by prefixes and suffixes that are attached to a root (e.g. *endo-* (prefix) *-card-* (root) *-itis* (suffix)); the latter contribute to the idea of medicine as a field constantly changing and updating. Because of that, homogeneity and clearness are two required factors medical writers must endorse.

Derivational processes are the most productive in the formation of words in medical language and allow people working in the medical field manage medical terminology in a very effective way, because "[i]nstead of memorising lists of terms they can easily predict the meaning of others" (Dzuganová, 2013, p.56) by splitting the terms into their basic components. There is, however, another word-formation process that is also highly used in the field, that of compounding.

Indeed, compound nouns are fixed expressions which can be of three different types: two-/three-word compounds, such as *heart attack*, *central nervous system*, *standardised interviewer-administered questionnaires*; hyphenated compounds, like *birth-control*, *life-span*; or one-word compounds, like *haemophilia*, *pseudopolycytemia*. Furthermore, the use of artificial word-forming pseudo-affixes[14] which come from Greek and Latin words and are attached to existing roots contributes to the creation of new words to designate new concepts, devices, symptoms, diseases, and treatments (e.g. *cardi(o)-* (root) *-logy* (suffixoid) and *onco-* (prefixoid) *-logy* (sufixoid)).

[14] Even though these pseudo-affixes (prefixoids – or pseudo-prefixes and sufixoids – or pseudo-suffixes) have certain lexical meaning, they are not used as individual words.

In addition, another feature that typifies the morpho-syntax used in medical English is the densely packed amount of information, since writers need to express a lot of information in the shortest lexical and syntactic forms possible. This is in a way achieved with abbreviations and acronyms, such as *Advance Care Planning* – ACP; *British Thoracic Society* – BTS; *Intensive Care Unit* – ICU; and *High Dependency Unit* – HDU; and stacked noun phrases, which create a highly dense text type. Stacked noun phrases, however, will be dealt with when focusing on the typical syntactic features that characterise medical register.

Regarding semantic features, semantic *univocity* and *denotation* are key concepts highly linked to the aforementioned features of transparency, precision, and shortness. The fact that a word is used to refer to a specific concept with a specific meaning is understood as univocity; denotation, on the other hand, refers to the fact that one word possesses one precise literal meaning, which cannot be misunderstood. In English for Medical Purposes, however, "there is (also) a wide use of words or phrases that depart from straightforward literal language" (Maglie, 2009, p.25). Examples of this phenomenon could be the use of figures of speech such as the metaphor, the catachresis, ellipsis, eponyms, and toponyms[15]. As medicine constantly develops and does so very fast, it needs to make use of borrowings from other languages, registers or other scientific branches so as to designate new scientific concepts, hence that metaphorical items, for instance, could be considered "a special form of borrowings from different areas which, once they become part of the specialised language of the discourse community, are eventually accepted as standard terms" (Maglie, 2009, p.25).

[15] *Catachresis* refers to the expansion of a word beyond the limits of its own meaning; it occurs when a word is used to indicate something completely different from its literal meaning. It can thus be considered a misuse for certain rhetorical effect. Black (1955) defines it as "the use of a word in some new sense in order to remedy a gap in the vocabulary". An example of catachresis in the medical register would be *neck of uterus*. *Ellipsis* occurs when an element in the construction is omitted but it is recoverable from the context. An *eponym* consists of a name (as of a drug or a disease) based on or derived from the name of a person. Examples of eponyms in Medicine would be *Down syndrome* (John Langdon Down), *Chron's dise*ase (Burrill Bernard Crohn), *Dravet syndrome* (Charlotte Dravet). A *toponym* in Medicine would consist of the name of a place so as to designate, for instance, an illness. An example would be *Kinkiang fever*; Kinkiang is a city in China.

When it comes to deal with the syntactic features recurrent in medical register, one cannot refute the high presence of very long sentences that are usually composed of stacked or nominal phrases or nominalised structures. This high number of nouns in the text, as mentioned earlier, creates a lexical density one needs to get familiar with. Moreover, sentence structure is mostly affected by the omission of phrasal elements and the pre-modification of certain elements. Finally, verb tenses and their modality, and the use of passive voice and structures of depersonalisation are also key factors that determine the syntax of medical discourse.

Even though in medical English some instances of clausal structures are longer than those in common language, writers can make use of other strategies such as stacked nominal phrases and nominalisations to pack a vast amount of information so as to economise space (Bartolic, 1978; Salager-Meyer, 1983, 1985; Horsella & Pérez, 1991). In either case, writers avoid leaving any gap that may create certain ambiguity. With the frequent use of stacked nominal phrases as a means to package conveyed information, this information "(…) is presented in a shorter, more direct and condensed fashion, [and] thus ha(s) a greater impact upon the reader" (Carrió-Pastor, 2008, p.29) and by means of this discursive strategy, discourse cohesion and coherence as well as concise referencing are also achieved (Huckin & Olsen, 1983). According to Du Bois (1981), the writer's use of complex noun phrases can be considered as dynamic, in the sense that they are essentially used to build up new information, which is backed up by other pieces of information previously referred to in the text.

In such complex patterns, the head noun within the nominal phrase is modified by internal constituents such as nominals, adverbs, participles, etc., which give relevant grammatical and semantic data. As stated by Maglie (2009, p.33), this type of specification tends to inform about the material of which an element is made, its use, its function or its way of functioning. Some examples found in randomised trials are as follows: *a significant reduction in incidence of incisional hernia*, *decreased vascularity of adipose tissue*, and *a fasting plasma glucose level 6.1 to 6.9 mmol/L [110 to 125 mg/dL]*.

Linked to stacked nominal phrases, nominalisations serve as a strategy to define concepts that are connected to practices or actions. Like in scientific language, in medical writing "(…) objects predominate and processes serve merely to define and classify them" (Halliday & Martin, 1993, p.20). Nominalised structures are thus highly frequent, and the reasons are varied. First, nominalisations certainly create a sense of objectivity because they take previous arguments and show them as an objectified element, which is permanent, established, and true, so it has to be taken for granted (Halliday & Matthiessen, 2014).

Second, as nominalisations tend to rephrase clausal structures previously mentioned in the text, they tend to be placed in thematic position, allowing thereby the natural interpretation from new to already-known information. Third, the presence of a nominalised structure reinforces the idea of an objectified element that is result of a cognitive or experimental process, which is the same that occurs when deriving a noun from a verb form. An example of a nominalisation in a real research article is as follows:

> Participation of surgeons from different specialities might have led to a learning curve in our trial, but this possibility is also a strong advantage of the PRIMA trial; and (…) this impairment probably has an important role in the pathogenesis of distension of the aorta and in formation of incisional hernia in patients after median laparotomy (Jairam et al., 2017, p.574).

Another example which includes a high number of nominalisations while packaging information in a very long string of words is as follows:

> The dramatic effect of lifestyle on both the prevention of incident metabolic syndrome and reduction of its overall prevalence appears to be most strongly related to a reduction in waist circumference and in blood pressure and not, as might have been anticipated, through correction of the lipid abnormalities of triglycerides and HDL cholesterol (Orchard et al., 2005, p.618).

Furthermore, the syntax of medical English tends to omit phrasal elements and relative clauses using a wide range of devices. Nevertheless, specialists in the field can recover

such information from their experience. Complex nominals with adjectival or nominal pre-modification, such as *oxygen delivery devices*, for instance, are favoured over nouns post-modified by relative clauses, like in *devices which deliver oxygen* (paraphrasing the previous example). It is also common to find instances of passive clauses where the relative pronoun and the passive auxiliary are omitted, the so-called bare passives, such as in *(t)he primary empirical approach used was a regression discontinuity design*, instead of *(t)he primary empirical approach* ***that/which was*** *used was a regression discontinuity design*.

Salager-Meyer (1993) points out that the two basic factors determining the choice of verb tense are both the textual typology and the text section. In case reports and research articles, for example, past tenses are favoured over present tenses, which are most used in review articles. As for the text section, purpose, methods, results and case presentations make use of the past tense (either in active or passive forms), whereas in the conclusions, recommendations and data synthesis, present tenses are highly favoured, which is the form to present established knowledge and generalise about specific findings. Vihla (1999) focuses on how modality is employed in medical texts, and asserts research articles, for instance, make greater use of epistemic modality to show conviction, whereas manuals contribute to mark a certain degree of authority through deontic modal expressions.

Two elements which determine the type of syntax found in medical discourse are the high presence of depersonalisation strategies, and above all, the extensive use of passive voice structures. These strategies will be further developed in Chapter 3. While the passive emphasises the results, effects, or processes of an action by fronting them in thematic position, active structures mainly highlight the performer of the action. Even though the use of passive constructions alters the unmarked form of verb phrases in sentences and consequently the ordering of its arguments, and it is thus a syntactically-related phenomenon, Maglie (2009, p.40) argues that “the alternation of active and passive forms cannot [only] be explained in syntactic terms but it depends on complex factors which are pragmatic and textual”.

One of the fuzziest areas as far as the writing of medical texts is concerned is whether authors should write personally or impersonally, actively or passively. It is obvious that scientists must be as detached, objective and empiric as possible in their work; nevertheless, as Stuart (2007, p.20) points out, "thinking, believing, considering, and hypothesising are all subjective activities". In other words, the writers' role is that of putting into words facts, data, and events, which are objective by default, through an act of reasoning (that of writing), which is inherently subjective.

The uncertainty of using a more personal or impersonal structure, an active or a passive form, comes then along with the ideas of objectivity, universality, and modesty, to name a few. A passive structure may transmit a sense of detachment and objectivity, notions which are inherent in scientific events and facts. The use of more personal forms and personal active structures, however, may be considered as a more modest approach to both the text itself and the audience to which it is addressed, if compared to the corresponding passive or nominalisation, which can give the impression that the statements claimed have universal truth value.

Indeed, depending on the section of the text where the verb forms appear, a particular choice of voice can be made.

> Personal style is not appropriate for all sections of a paper. Descriptions of materials and methods, and statements of results are meant to be straightforward presentations of facts. Readers want emphasis on procedures, actions, and outcomes, not on agents. Accordingly, it is usually irrelevant, and sometimes obtrusive, to mention who did what or who observed what in those sections. However, it is appropriate to mention agents in an introduction or a discussion, where it is frequently essential to distinguish clearly between what others have thought or done and what you thought or decided to do (Stuart, 2007, pp.19–20).

Through exposure and learning, writers acquire a certain degree of proficiency of all the above-mentioned writing strategies (lexical, morphological, syntactic, pragmatic, and textual). They must do so to actively take part in their community of practice and thus in their discourse community. However, they do not only need to acquire such writing strategies to conform to the writing conventions of the discipline but must also be aware

of the several typologies of scientific compositions that exist, as well as what type must be used for their studies. The next section explores the different types of medical texts that can be found in the medical literature.

2.2.1. Types of medical texts

Novice members of the medical scientific community must know the most common forms of medical scientific communication to begin to take part in the writing arena of the community to which they wish to belong. It is important to point out the existence of a great number of scientific compositions plus their several sub-types. I would like to highlight and briefly comment on the most relevant ones, which are as follows: the original research article, the review article, the clinical case study, the clinical trial, the perspective article, the opinion article, the commentary, and the book review (Majumder, 2015).

The original research article reports original research through detailed studies, which include a hypothesis, a background study, methodology, interpretation of results, and a discussion of their possible implications. It is classified as primary literature and its length ranges from a 3,000 up to a 12,000-word limit. The review article, on the other hand, is classified as secondary literature and is highly recommended for early career researchers. It serves to illustrate and analyse existing literature on the field about a very specific topic. Its length ranges from 3,000 up to 5,000 words. The review article can be divided into different sub-types: the systematic literature review, which attempts "to identify, appraise and synthesise all the empirical evidence that meets pre-specified eligibility criteria to answer a given research question" (Piper, 2013, p.2); and the meta-analyses, which are "statistical assessments of the data provided from multiple studies or sources that attempt to ask/answer the same question" (Piper, 2013, p.2).

The clinical case study, also recommended for early career researchers, presents actual patient cases that contribute significantly to the field. The text deals with patient cases from medical and/or clinical practice and discusses the signs, symptoms, diagnoses, and

treatment of a disease. The clinical case study is considered as primary literature and its lengths ranges from 3,000 up to 12,000 words. With a similar range of words and a similar layout, experienced researchers tend to make use of the clinical trial, which is aimed at describing the methodology, implementations, and results of controlled studies with large groups of patients. Within this category, it is possible to find three sub-types: randomised controlled trials (henceforth, RCTs), explanatory trials, and pragmatic trials.

RCTs are studies aimed at testing the effectiveness of a medical intervention, and the participants of the study are randomly allocated to two groups: the experimental group (which receives intervention being tested), and the comparison – or control – group (which receives an alternative treatment, a placebo, or no treatment at all). It is also relevant to highlight that the studies "may include an element of blinding (…) [, which means] the group to which the participant has been allocated is either not known by the researcher or not known by the patient (single-blind study), or not known by either (double-blind study)" (Stuart, 2007, p.82). To confirm whether an intervention is effective or not, both groups are followed up and compared.

Explanatory trials differ from RCTs insofar as they are conducted under idealised circumstances, so their results are not as easy to be applied to the real world as those obtained from RCTs. In fact, they are aimed at determining the exact extent to which a particular medical or clinical intervention is effective or not. On the other hand, pragmatic trials are also randomised controlled trials that also aim at determining the effectiveness of an intervention, but in this case, in day-to-day clinical practice; they tend to have less restrictive inclusion and exclusion criteria when compared to explanatory trials.

Perspective, opinion, and commentary articles are scholarly pieces of text that express a more personal opinion or new perspectives about existing research on a particular issue. Through a perspective article, the author expresses a new point of view on problems and issues regarding a specific topic, backed up by original data and personal opinion. In the case of opinion articles, the author, also backed up by evidence, presents their viewpoint on the positive and negative aspects of a hypothesis or a scientific theory, which has

already been published. By doing so, the author aims at challenging the current state of knowledge in the field by promoting a discussion on such issues. Commentaries, the shortest types of articles (around 1,000 and 1,500 words), are written to constructively criticise issues that have already been published in articles, books, or reports, and might be of interest to the readers.

Finally, medical book reviews are academic writing compositions that critically assess the content of a book, its style and its significance to the field in which it is framed through a summarised analysis, so that its readers gain insight into its strengths and weaknesses. It tends to begin with a short introduction to the book followed by an outline of its main contents; once the contents are exposed, the writer highlights issues of interest and evaluates them in detail.

Having explored the main typologies of texts in Medicine, it is of paramount importance to deal with the notion of Research Article (henceforth RA), as it can be considered as the embodiment of the scientific research process itself. The next section explores its main characteristics and analyse the different sections of which an RA is composed.

2.2.2. The Research Article (RA): A Piece of Text as Embodiment of the Scientific Research Process

The prototypical method for communication of new research results in science is a paper published in a scientific journal, which is organised in a standard way. This canon is a conventional framework known as IMRaD[16] (also found as IMRD), which is an acronym that stands for the central sections of which the scientific paper is composed. These four sections are the following: Introduction, Materials and Methods, Results, and Discussion (Malmfors et al., 2004; Matthews & Matthews, 2008; Day & Gastel,

[16] Silvyn-Roberts (2012), in reference to written reports, includes an A for Abstract before the acronym IMRAD, and it becomes aIMRaD.

2011; Davis et al., 2012; Rogers, 2014; *inter alia*), which usually include subheadings that serve to organise the informational content[17].

Indeed, this type of structure is an example of "what is conventional or what the reader or listener expects" (Davis et al., 2012, p.6) and "represents a consensus among scientists about the most effective way to package information" (Heard, 2016, p.74). It is thus the most common format to write up scientific research since "it is a direct reflection of the scientific discovery process (and not simply a convention used by publishers)" (Stuart, 2007, p. 64). In line with this, the rigid structure of the research article makes scientific writing apparently easy and more straightforward (Matthews & Matthews, 2008) because it is a piece of text that embodies the scientific process itself.

2.2.3. The Medical Research Article (RA) and its Internal Structure

The IMRaD structure is presented as the standard and fixed framework commonly used for publication of research articles. However, the acronym itself ignores other sections that commonly tend to appear in research papers and which "connect [the writers'] work to the larger literature and the larges scientific community" (Heard, 2016, p.79); these other sections are the Title, the Abstract, the Keywords, the Conclusion(s), the Acknowledgements, the Disclosures, the References, the Appendices, and the Tables

[17] Toulmin (1958) suggests that any scientific presentation must include five specific components. These are as follows: the grounds or already established information on the topic; the warrant, which consists of qualitative and quantitative facts supporting the hypothesis; the modality, which is the value indicating the degree of certainty regarding the event reported; the claim, the core of the discussion and/or the final conclusion; and the rebuttal, which consists of elements of doubt concerning the claim made. According to Heard (2016, p.77) the IMRaD structure could be illustrated through an hour-shape glass because "a well-written paper follows a predictable change in focus: broad attention to the work's context in a major field at the beginning of the Introduction, narrower definition of the central research question at the Introduction's end, narrowest focus on specific techniques and results at the hourglass's middle in the Methods and Results, and broad context again at the end of the Discussion".

and Figures. The standard logical order for them to be found in a research article is shown in Table 1.

Table 1. Sections of a prototypical Research Article

Order	Heading of the section
1	Title
2	Abstract
3	Keywords
4	**Introduction**
5	**Materials and Methods**
6	**Results**
7	**Discussion**
8	Conclusion(s)
9	Acknowledgements
10	Disclosures
11	References
12	Appendices
13	Tables and Figures

There is, however, some kind of flexibility as far as the layout of the RA is concerned. Although this is the canonical structure of an RA, some of its sections may be blended into other sections or may even be omitted. That is the case, for instance, of the Conclusion, which tends to be included as a final appendix of the Discussion section and not be presented as a section on its own (Malmfors et al., 2004, p.9). Other parts, such as the Acknowledgements, Disclosures, Appendices, and Tables and Figures, are often optional and their inclusion essentially depends on the convenience of the writers. The Title, Abstract, and References, although not being included within the IMRAD acronym, are always present. Indeed, the writers will adapt the different sections to the requirements of the journal they wish to publish in (Joubert & Rogers, 2015).

In the next sections, I will explore the most salient and characteristic structural and linguistic features of each of the core parts of the RA, which are the Introduction, the Materials and Methods, the Results, and the Discussion. I will also include information

on the Title and the Abstract. To do so, I will use as models two samples of each of the sections, extracted from RAs published in prestigious scientific medical journals.

2.2.3.1. Title

The title of an RA is probably the most relevant part of the paper as can be considered the first filter, along with the abstract and a by-line (or list of authors) with their academic affiliations and their addresses (Joubert & Rogers, 2015, p.129; Heard, 2016). These three pieces of information "will make electronic retrieval of the article sensitive and specific" (International Committee of Medical Journal Editors (ICMJE), 2016, p.13), so "its function is advertisement" (Heard, 2016, p.79). A title must thus be informative and descriptive, since it must specifically inform about the study design and briefly summarise the subject of the research carried out.

Accuracy and concision must also be present as inherent features of a title; in fact, a good and balanced title must provide enough informational content of the study to attract, as much as possible, the attention of potential readers, without obviously including the final results of the study. For this reason, writers tend to use titles with densely packed information that are divided into parts by a colon. The first part would consist of a more general opening phrase or clause, and the second half would include a phrase or clause more focused (Heard, 2016, p.79).

Relevant linguistic aspects to consider are, for instance, the great use of nouns and nominalised structures, which are preferred over clause-type constructions (see Figures 1 and 2) and a wide range of specialised terminology and acronyms (see Figure 2). The two samples below are titles that, in one way or another, subscribe the organisational and linguistic aforementioned features.

Figure 1. Example of a Title (Title 1)

TITLE 1
Effects of Vaccination on Invasive Pneumococcal Disease in South Africa. (von Gottberg et al., 2014, p. 1889)

Figure 2. Example of a Title (Title 2)

TITLE 2
Rituximab versus Azathioprine for Maintenance in ANCA-Associated Vasculitis (L. Guillevin and Others) TNEJM. (Biddle et al., 2010, p. 320)

2.2.3.2. Abstract

An abstract is an abbreviated or summarised version of the entire RA. It is the only portion of the article that is "indexed in many electronic databases, and the only portion many readers read" (International Committee of Medical Journal Editors (ICMJE), 2016, p.14); so authors must be skilful in order to reflect in a brief piece of text the content of the entire article in a concise way, including information of each of the sub-sections. An abstract is then a short paragraph aimed at reflecting the theoretical and procedural steps taken to carry out the experiment, as well as at informing about the main results obtained (Matthews & Matthews, 2008). Table 2 shows how the information is distributed along the different sub-sections of the research article.

Table 2. Informational content of a prototypical abstract
and their correspondence with the four main sections of the RA

Order	Content found in a prototypical ABSTRACT	Sections of RA
1	Context or background for the study Hypothesis or question investigated Main purpose of the paper	Introduction
2	Materials used Brief explanation of the design of the experiment and its basic procedures (selection of study participants, settings, measurements, analytical methods)	Materials and Methods
3	Major findings (including key quantitative results with statistical and clinical significance)	Results
4	Discussion and interpretation of the results obtained, highlighting new and important aspects of the study Final or principal conclusion, noting the limitations	Discussion (and Conclusion)

Writers tend to begin their abstracts briefly contextualising the study carried out, narrowing the area of research, and stating its importance, identifying a gap to be filled, and posing a specific research question, which will be answered along the RA. Nevertheless, such a question is not always a direct question, but it is sometimes found in the form of a sentence that states the main outcome(s) of the paper, in a form of a hypothesis. This first part of the abstract corresponds to the most elementary information in the Introduction section.

Then, writers usually include information on the main materials used, and briefly explain the basic experimental design carried out, also including a non-detailed description of the methods and techniques used in an attempt to summarise the approach taken to answer the hypothesis. This second part of the abstract corresponds to the Materials and Methods section of the research paper. The third part includes the major findings of the study presented, which are taken from the Results section. These main results are usually the quantificational answer to the hypothesis or question posed at the

beginning. The final part of the abstract exposes the main points of the Discussion section of the paper and highlights the most relevant conclusion(s) drawn.

In general, there are two fixed typologies of abstracts, as far as their layout is concerned, which can be found in RAs: these are informative abstracts, on the one hand, and indicative abstracts, on the other. The former are commonly written along with papers describing original research and usually include the main objective, the same questions posed in the body of the paper, and some relevant data to support the main conclusions stated in the RA. As opposed, an indicative abstract, also known as a descriptive or topical, also includes the main objectives of the research, but the results are generally suggested, rather than specified. They are usually written to turn the table of contents into sentences for the review of articles or books (Malmfors et al., 2004; Matthews & Matthews, 2008).

As for the informational content, Figure 3 (Abstract 1) consists of a one-paragraph text, which begins with a general and brief explanation of previous research (Introduction) and avoids mentioning the materials and methods that have been used to carry out the experiment. The general result is briefly exposed (Results) and its interpretation is summarised in one sentence, which exposes the general conclusion drawn from the study (Discussion).

Figure 3. Example of an Abstract (Abstract 1)

ABSTRACT 1

Our understanding of the potential cardioprotective properties of nutrition is relatively recent, with most relevant studies completed in the last several decades. During that time, there has been an evolution in the focus of nutritional intervention. **Early trials emphasized** reduction of dietary fat with the goal of preventing heart disease by reducing serum cholesterol. **Results from trials focused** exclusively on dietary fat reduction were disappointing, prompting subsequent studies incorporating a whole diet approach with a nuanced recommendation for fat intake. **The Mediterranean-style diet**, with a focus on vegetables, fruit, fish, whole grains, and olive oil, **has proven to reduce** cardiovascular events to a degree greater than low-fat diets and equal to or greater than the benefit observed in statin trials.

Keywords:

CHD prevention, Low-fat diets, Mediterranean diet, Reducing cholesterol
(Dalen and Devries, 2014, p.364)

On the other hand, Figure 4 (Abstract 2) exposes all the parts of which the research article is composed; that is, Background (which corresponds to the Introduction and states the main outcome), Materials and Methods, Results, and Discussion. In this case, each sub-section is specified with its own heading, so the structure is, in fact, more visual, much clearer, and more organised than in Abstract 1. When comparing the amount of information provided by both abstracts, Abstract 2 offers a higher degree of detailed information and such information is comparatively more structured.

Figure 4. Example of an Abstract (Abstract 2)

ABSTRACT 2

Background

We aimed to describe the determinants of discharge heart rate in acute coronary syndrome patients **and assess** the impact of discharge heart rate on 5-year mortality in hospital survivors.

Methods

French Registry of Acute ST-Elevation or non-ST-elevation Myocardial Infarction (FAST-MI) 2005 is a nationwide French registry that included all consecutive patients with acute myocardial infarction over 1 month in 223 institutions in 2005. **Discharge heart rate was recorded** in 3079 patients discharged alive; all had 5-year follow-up. **Logistic regression was used** to detect predictors of high heart rate at discharge. **Cox's proportional hazards model was used** to assess the hazard ratio for mortality at 5 years. **Heart rate was categorized** into 4 groups by quartiles (<60, 61-67, 68-75, >75 beats per minute). **High heart rate was defined** as ≥75 beats per minute. Landmark analysis was performed at 1 year.

Results

Independent predictors of heart rate ≥75 beats per minute at discharge were female sex, ST-segment elevation myocardial infarction, diabetes, chronic obstructive pulmonary disease, bleeding/transfusion during hospitalization, left ventricular dysfunction, renal dysfunction, and prescription (type, but not dose category) of beta-blockers at discharge. Discharge heart rate was significantly related to mortality at 1 year (hazard ratio 1.13; 95% confidence

interval, 1.03-1.24 per 10 beats per minute, P = .02); this was confirmed by landmark analysis, with a 39% increase (hazard ratio 1.39; 95% confidence interval 1.05-1.84) in the risk of 1-year death for discharge heart rate ≥75 beats per minute vs <75 beats per minute. This relationship was no longer significant between 2 and 5 years.

Conclusions

After acute myocardial infarction, patients discharged with high heart rate (≥75 beats per minute) are at higher risk of death during the first year, but not later, irrespective of beta-blocker use.

Keywords:

Heart rate, Mortality, Myocardial infarction, Registry

(Seronde et al., 2014, p. 954)

Indeed, not only is the layout an indispensable factor to take into consideration when writing the abstract of a paper and submitting it for publication, but also the linguistic features it is composed of. In abstracts, the language used tends to be clear and concise, and the sentences are aimed at getting to the point across quickly, avoiding thereby lengthy background information that would be misleading and would otherwise be found within the research article. Abstract nouns and complex nominalised structures tend to be highly frequent (Abstract 1: *Our understanding of the potential cardioprotective properties of nutrition*; *an evolution in the focus of nutritional intervention*; Abstract 2: *the determinants of discharge heart rate in acute coronary syndrome patients*; *Independent predictors of heart rate ≥75 beats per minute at discharge*)

As for the verb structures commonly found, both active and passive constructions are frequent, although manuals, as explored earlier through the Plain Language Movement, often encourage the use of active voice sentences since they are more straightforward and less wordy. However, only one instance in the two abstracts uses the pattern *we* + active verb or a first personal personal pronoun: *we aimed to describe (…) and assess (…)* (in Abstract 2) and *Our understanding of (…)* (in Abstract 1), respectively. In fact, the subject of most of the active constructions tends to be inanimate (an abstract rhetor), reinforcing thereby the idea of impersonality an objective piece of text is meant to

transmit, as can be seen, for instance, in Abstract 1: *early trials emphasised (...)*, *results from trials focused (...)*, *the Mediterranean-style diet (...) has proven to reduce (...)*.

In Abstract 2, however, passive voice structures are highly favoured over active voice ones, particularly in the Materials and Methods section. The examples are as follows: *discharge heart rate was recorded (...)*, *logistic regression was used (...)*, *Cox's proportional hazards model was used (...)*, *heart rate was categorized (...)*, *high heart rate was defined as (...)*.

As for the tense most frequently used, verb phrases that correspond to the summarised versions of the Materials and Methods section and of the Results tend to be in the past tense, whereas present tense would be basically used in the Introduction section and in the motivation and justification of the topic, the interpretation of Results obtained, and the Conclusion(s) highlighted.

2.2.3.3. Introduction

The introduction is considered to be the first main section of the body of the RA, since it is used to "set[ting] the scene" (Stuart, 2007, p.67) for the writer's work. According to Davis et al., (2012) and Matthews & Matthews (2008), an introduction establishes the topic (or hypothesis) which the writer is to discuss, and also explains to what extent the position of the writer is going to modify pre-existing knowledge in the field of interest. Such pre-existing knowledge may be either challenged or expanded by giving reasoned and empirical arguments. To do so, it provides background information on the subject and highlights results and findings of other similar studies (referencing the most influential ones) to justify the importance of the new knowledge the writer is presenting. An introduction also explains the rationale for the study by listing the objectives and hypotheses to provide the audience with information on what the writer(s) plans to accomplish in the full article.

Glasman-Deal (2009, p.24) divides this section into four basic components: the first one establishes the importance of the field, provides background facts, defines the terminology in the keywords and in the title, presents the problem area and focuses on current research; the second one deals with current research and relevant contributions to the field; the third one locates the gap in the research, describes the problem being addressed in the paper and presents a prediction to be tested; finally, the fourth component describes the paper.

After having briefly analysed the layout of the introduction section, which departs from a description of the general situation and goes on to information that is specific to the topic with which the paper will deal (Stuart, 2007, pp.66–67), it is interesting to focus on the linguistic features. In this respect, Stuart (2007, p.23) points out that "a mixture of personal and impersonal, active and passive constructions" is found. When analysing the two samples provided below (see Figures 5 and 6), one can note the passive voice, and particularly the agentless passive (see passive verb phrases in red below)[18], is highly favoured against active voice structures or more personal expressions. It is only towards the end of the text, where a personal active expression (*we expanded* in Introduction 1; *Our objective* in Introduction 2) introduces the outcome or step carried out.

As for the tenses commonly used, Malmfors et al. (2004, p.12) point out that the motivation and justification is often expressed through present tense, like in the whole Introduction 1; the review of literature is often expressed in past tense, except when it presents "established [and common] knowledge and generally known facts [and] previous results" (Joubert & Rogers, 2015, p.42), in which case the use of present simple or present perfect tense is more justified. By using the present tense, writers mean that "[they] believe [their] findings and deductions are strong enough to be considered as facts or truths" (Glasman-Deal, 2009, p.5).

[18] Note that passive structures which are in red and in italics are reduced relative clauses, which are non-finite and are introduced by present or past participles, which post-modify the noun they go with. The structure is then as follows: Noun + *present/passive participle form of the lexical verb*.

Figure 5. Example of an Introduction (Introduction 1)

INTRODUCTION 1

In vitro fertilization (IVF) and non-IVF fertility treatments (i.e., ovulation induction and ovarian stimulation) constitute major risk factors for the genesis of multiple births (twin, triplet, and higher-order births).[1] IVF procedures, which are defined as procedures in which eggs and sperm are manipulated with the purpose of establishing a pregnancy, represent the overwhelming majority of procedures for assisted reproductive technology. Ovulation induction and ovarian stimulation, which are often included as part of the IVF process, are also coupled with timed intercourse or intrauterine insemination to establish a pregnancy independently of IVF. In ovulation induction, drugs are administered to induce ovulation in women who are otherwise anovulatory.[2] In contrast, ovarian stimulation is a process in which drugs are used to promote multifollicular ovarian development and ovulation in women who are subfertile.[2] Multiple births resulting from fertility treatments are associated with increased health risks for women and their offspring,[3-6] as well as substantial financial costs to families and society.[7,8]

Limiting the number of embryos transferred during IVF cycles is important for reducing multiple births.[9,10] Although the practice of transferring three or more embryos in a single cycle has declined steadily, several factors may constrain efforts at further reduction. First, competition among fertility clinics may result in an intense focus on per-cycle pregnancy rates.[11] Second, financial exigencies encountered by patients may drive up the number of embryos transferred in the hope of maximizing per-cycle pregnancy rates.[12] Finally, patient acceptance of multiple births and the risks thereof may contribute to the transfer of more embryos.[13]

Whereas the use of IVF is documented in a congressionally mandated national registry in the United States, the use of non-IVF fertility treatments is not and thus is not directly ascertainable. Nevertheless, non-IVF fertility treatments are associated with a substantial proportion of multiple births[14-16] and have been shown to account for more multiple births than IVF.[17,18] In this study, we expanded on these earlier observations by conducting a longitudinal analysis to determine the trends in and magnitude of the contribution of fertility treatments to multiple births.

(Kulkarni et al., 2017, p. 2218)

Figure 6. Example of an Introduction (Introduction 2)

INTRODUCTION 2

Radial-head subluxation is a common arm injury among young children and often results in a visit to the emergency department.[1] This type of injury occurs when forceful longitudinal traction is applied to an extended and pronated forearm.[2] Radial-head subluxation is easily recognized by its clinical presentation and can be treated by a simple reduction technique involving hyperpronation or supination

and flexion of the injured arm.[3–7]

Despite the ease of diagnosis and treatment, children with radial-head subluxation often wait hours in the emergency department for a reduction that takes minutes to perform.[8] These visits have direct health care costs and involve time and stress for the child and their family. Early treatment and shorter wait times correlate with patient satisfaction.[9,10] Patient satisfaction is comparable when minor injuries are cared for by a nurse instead of by a physician.[11–13] Nurse-initiated treatments are increasingly a focus of health care.[14–17]

Treatment of radial-head subluxation is an appropriate area to consider nurse-initiated care. Our objective was to determine whether triage nurses, trained in the recognition and treatment of radial-head subluxation, could successfully reduce radial-head subluxation at a rate similar to that of physicians.

(Dixon et al., 2014, p. 317)

2.2.3.4. Materials and Methods

The Materials and Methods section is an important part of the RA because it is aimed at listing, often chronologically or in operational order (Booth, 1993, p.5), the several steps and the materials that were required to do the research experiment presented in the paper. However, Heard (2016, p.89) refuses the use of a chronological order by encouraging a division of the information into three parts: background, experiments or observations, and analysis. Using this structure the writers facilitate the readers' understanding of the section, which is in fact the main purpose of any written text of this type.

The Materials and Methods section also describes, as said, the possible statistical procedures employed in the attainment of the research process, hence that it may sometimes be called *Procedure*, *Experiments*, *Experimental*, *Simulation*, or *Model*, instead (Glasman-Deal, 2009, p.44). In one of the examples provided (see Materials and Methods 2), it is possible to find headings that inform of the steps taken to complete the study. These are as follows: selection of the study participants/subjects, inclusion criteria, exclusion criteria, procedure of randomization, post-intervention examination, statistical analysis, measurement techniques, genotyping, data analysis, etc. (Rutkove, 2016, p.102).

All these steps are of paramount importance insofar as they assert the validity and acceptance of the methodology utilised to carry out the research (Glasman-Deal, 2009, p.44). According to Matthews & Matthews (2008, p.43), "for publication of a new discovery to be 'valid' or scientifically accepted, it must appear in a form so that a trained investigator with considerable experience could repeat the experiment". Davis et al. (2012, p.83) describe it as a detailed "recipe that reveals how [the writer] acquired and used [their] data"; thus, it must serve to replicate the experiment(s) and to assert whether it is credible or not. They consider that both objectives are possible only if the writer provides detailed information on "ingredients, actions, conditions, experimental design, replications, repetitions, and statistical analysis" (Davis et al., 2012, p.83); in other words, all the information that was available at the time the plan or protocol for the study was being written must be included (International Committee of Medical Journal Editors (ICMJE), 2016).

There is also a clear connection between both the Introduction and the Materials and Methods sections. The order of the objectives presented in the Introduction is usually kept as the order of the processes described in the Materials and Methods section, so the writing of this section does not generally pose severe difficulties as far as its organisational structure is concerned. When going through the Materials and Methods section, readers will assess whether the methods used to carry out the experiment are adequate and appropriate to answer the research question(s) posed in the Introduction section (Stuart, 2007).

According to Glasman-Deal (2009, p.67), this section can be divided into four basic components: the first one must provide a general introduction of the materials and methods, restate the purpose of the work, give the source of materials and equipment used, and supply essential background information; the second one must provide specific and precise details about materials and methods, justify the choices made and indicate that appropriate care was taken; the third one must relate both the materials and methods to previous studies, and the fourth one must indicate where the problems occurred. These components may be modified or elided depending on the study typology and the writers' criteria.

The language in this section is usually specific and informative. The first sample provided (see Figure 7, Methods 1) combines instances of both personal active verb forms (*we* + V_{ed}) and agentless passive voice structures ({be} + $V_{pass.\ part}$). Here, the pattern *we* + V_{ed} serves mainly to describe the procedures taken by the research team and emphasis is put on the agent of the action. The agentless passive structures are also frequent, and they tend to appear when the agent is not necessary; when it is, the active counterpart is preferred over the passive. In the Materials and Methods 2 (see Figure 8), however, all procedural verbs are in the passive voice ({be} + $V_{pass.\ part}$) in past tense. Few instances of these passive voice constructions are followed by the agent (underlined in read) (i.e. *The oral examination among these participants was carried out by one trained and calibrated investigator* (Materials and Methods 2)) and all the rest are agentless passives in the preterite.

The use of a past simple tense reinforces the idea that the actions have been taken by the research team (Glasman-Deal, 2009, p.47; Joubert & Rogers, 2015, p.42). Impersonal structures[19] and nominalised grammar are also inherent linguistic features typical of this section. It is worth noting that writers tend to seek for some kind of harmonisation; that is, either the use of active and passive forms is combined throughout the section, and such combination is well-balanced (Methods 1), or the agentless passive voice significantly predominates (Materials and Methods 2).

Figure 7. Example of Materials and Methods (Materials and Methods 1)

METHODS 1

To identify published and unpublished reports of relevant randomised controlled trials we carried out highly sensitive electronic searches of Medline (1966-September, week 3, 2006), Embase (1980-2006, week 38), Medline In-Process (27 September 2006), Biosis (1985-22 September 2006), Science Citation Index (1981-23 September 2006), ISI Proceedings (1990-18 March 2006), Cochrane Controlled Trials Register and Cochrane Database of Systematic Reviews (Cochrane Library, Issue 1, 2006), Database of Abstracts of Reviews of Effectiveness (March 2006), Health Technology

[19] Note the presence of an active voice sentence with an inanimate subject, underlined in green (Materials and Methods 2 – Figure 8).

Assessments database (March 2006), National Research Register (Issue 1, 2006), Clinical Trials (March 2006), and Current Controlled Trials (March 2006). We also searched the proceedings of recent conferences of the European Association of Urology, American Urological Association, and British Association of Urological Surgeons. We defined ablative endoscopic treatments as those that resulted in immediate removal of tissue, usually by resection or vaporisation, and did not include procedures that caused delayed tissue necrosis, such as microwave and radiofrequency therapy. Searches were not restricted by publication year or language and included conference proceedings and abstracts. We scanned the reference lists of all included studies to identify additional potentially relevant studies. (Full details of the search strategies are available from the authors.) Two reviewers (TL, Angela Coutts, or Susan Wong) independently screened the titles and abstracts of identified papers and obtained full text copies of all potentially relevant studies.

Study selection and study characteristics

We included randomised controlled trials if they assessed endoscopic ablative interventions and included transurethral resection as one of the treatment arms. We excluded trials that reported on men without a clinical diagnosis of benign prostatic enlargement; comparisons with conservative management; and trials of interventions not involving tissue removal, such as transurethral incision. Table 1 lists the interventions considered (also see supplementary table 1 on bmj.com). The primary outcome measure was change in symptom score at 12 months after surgery, measured by the international prostate symptom score or the American Urological Association symptom index— these were considered equivalent and therefore we combined the results from trials using these instruments. The international prostate symptom score or American Urological Association symptom questionnaire asks men to rate four voiding symptoms (poor stream, intermittent flow, straining, incomplete emptying) and three storage symptoms (frequency, nocturia, urgency) on a scale from 0 (not present) to 5 (severe) to give a total score ranging from 0 to 35, with severity of symptoms defined as mild (0-7), moderate (8-19), or severe (20-35).

Secondary outcomes were blood transfusion, urinary incontinence, urinary retention, urinary tract infection, loss of ejaculation, erectile dysfunction, quality of life, peak urine flow rate, duration of operation, length of hospital stay, and reoperation. We considered all reports of prespecified complications regardless of their timing. As separate reporting of bladder neck stenosis and urethral stricture was inconsistent we combined these complications.

Validity assessment and data abstraction

the methodological quality of the full text studies using an assessment tool drawing on the schema suggested by the National Health Service Centre for Reviews and Dissemination,[3] Verhagen et al,[4] Downs and Black,[5] and the generic appraisal tool for epidemiology. Judgments on quality were based on the qualitative assessment of the number and type of criteria met by individual studies.

The two reviewers recorded details of methodology, interventions, participants' characteristics, and outcomes on a data extraction form. Any differences that could not be resolved through discussion were decided by an arbiter. For trials with multiple publications we included only the most complete

report for each outcome.

Quantitative data synthesis

For meta-analysis we combined data on dichotomous outcomes using the Mantel-Haenszel relative risk method. For continuous outcomes we used the inverse variance weighted mean difference method and 95% confidence intervals. We intended to report results using a fixed effects model throughout, but we had to use a random effects model for symptom score and peak urine flow rate because of statistical heterogeneity, explored by $\chi 2$ tests and I2 statistics. We used the standard Cochrane software RevMan 4.2.8 for meta-analyses.

(Lourenco et al., 2008, pp. 36–39)

Figure 8. Example of Materials and Methods (Materials and Methods 2)

MATERIALS AND METHODS 2

Selection of the study participants

This was a double-blind randomized parallel study conducted over a period of 2 months from September to October 2012 among 13-20 years old adolescents selected from a residential school in Bhopal city. The permission to conduct the study was obtained from the head master of the concerned school and the ethical clearance was obtained from the Institutional Ethics Committee. A tentative list of 93 adolescents in the age range of 13-20 years was initially prepared. The relevant information facilitating the inclusion and exclusion of study participants was collected from all these participants using the checklist and a data collection sheet. The checklist had 16 questions that specifically elicited the information on oral hygiene practices (method of cleaning, frequency, direction, duration of brushing, frequency of discarding tooth brush, use of mouth rinse, and inter-dental aids) medical history (history of systemic diseases, use of antimicrobials in the last 1 month, history of oral prophylaxis in last 3 months) and tobacco related habits. The checklist was filled by one investigator by means of face-to-face interview with the study participants. The oral examination among these participants was carried out by one trained and calibrated investigator. Plaque score was recorded using full mouth Turesky modification of Quigley Hein index[15] and gingivitis was scored using full mouth Loe and Silness gingival index.[16] The training and calibration of the investigator in the application of plaque and gingival indices was carried out on a group of 20 subjects. The assessment of plaque (Cronbach's $\alpha = 0.985$) and gingiva (Cronbach's $\alpha = 0.989$) showed good intraexaminer reliability. Autoclaved set of instruments were used for oral examination of the study participants.

The selection of participants into the study was based on the following inclusion and exclusion criteria.

Inclusion criteria

Those who give the written informed consent to participate in the study

Aged between 13 and 20 years

Having a minimum of 20 natural teeth which are not indicated for extraction

Whose baseline mean plaque score was more than 1.5 (Turesky modification of Quigley Hein index)

Whose baseline mean gingival index was more than 1.0 (Loe and Silness gingival index)

Who will be staying in the residential home for duration of at least 2 months from the date of screening.

Exclusion criteria

Presence of advanced periodontal disease (periodontal pockets, mobility, gingival recession, furcation involvement, etc.)

Presence of dental appliances (removable or fixed)

History of antibiotic use in previous 3 months

Those using mouth rinses containing chemical agents or inter-dental cleaning aids on a regular basis

Presence of any systemic diseases

Presence of malocclusion traits

Deleterious habits like smoking or use of other tobacco related habits

Individuals failing to offer written informed consent.

Eighteen participants were excluded from the initial list based on the inclusion and exclusion criteria. The reasons for exclusion are summarized in Table 1.

Procedure of randomization

A total of 75 participants fulfilled the inclusion and exclusion criteria. These eligible participants were then randomly assigned to either Group A or Group B using a random allocation sequence by one of the faculty members. The staff member who carried out the process of randomization was not involved in the data collection and statistical analysis. The information on the group allocation and the sequence of random allocation was concealed from the main investigator and the statistician to ensure double blinding in the study. Each participant was offered a coded oral hygiene kit containing the assigned dentifrice and a toothbrush. The soft bristled toothbrush with uniform make was distributed to all participants to ensure uniformity with regards to mechanical cleaning aid. Oral hygiene kit for Group A contained a high cost dentifrice whose unit cost was 1.4 international normalized ratio (INR) (Indian rupee)/g (100 g for 140 INR). The kit assigned to Group B contained a low cost dentifrice whose unit cost was 0.21 INR/g (100 g for 21 INR). The details of the commercial name or the ingredients in the dentifrice were not revealed to the participants and the investigator. The participants were instructed to refrain from all other unassigned forms of oral hygiene, including nonstudy toothbrushes or toothpastes, dental floss, chewing gum or oral rinses for 1 month following the intervention. The study participants were instructed to continue with their existing oral hygiene habits (frequency and method of brushing) with no specific instructions to modify their routine habits owing to the difficulties anticipated in adopting to new oral hygiene instructions overnight. The oral hygiene aids used by the study participants prior to intervention were collected in an effort to maximize the compliance for the study. Adverse effects during the intervention period, if any, were recorded. The comparison of the distribution of the study participants based on frequency of brushing, duration of brushing (in each brushing) at pre- and post-intervention examinations between the participants in two groups was done

to assess the confounding effect of oral hygiene practices.

Postintervention examination

At the end of 4 weeks following the intervention, all the participants were assembled in a waiting room. Then, each participant was examined by the same investigator who carried out the baseline examination for plaque and gingival scores employing the same indices which were used at baseline. The unique ID of the participant was entered after completion of oral examination. The information on the oral hygiene practices, use of antimicrobials, etc., during the intervention was collected using the same checklist that was used at baseline. The final analysis included only 65 participants. The dropout rate was 13.3%.

Statistical analysis

Data were entered onto a personal computer and statistical analysis was performed using SPSS version 19 (IBM, Chicago, IL, USA). Paired t-test was used for comparing the mean plaque and gingival index scores between baseline and postintervention in each group. The independent sample t-test was used for comparing the difference in the mean plaque and gingival scores between the groups. The statistical significance was fixed at 0.05.

(Ganavadiya et al., 2014, pp.382–384)

2.2.3.5. Results

The Results section is also an important part of the RA because it establishes the writers' credibility (Davis et al., 2012). Indeed, it includes the raw data – that is, the figures obtained – from the analysis or study carried out, which will confirm or deny the hypothesis or question exposed in the Introduction section and examined throughout the paper. It tends to be a section in which writers present essential data in tables, graphs, images and/or equations (see text in grey) to make the reader easily understand the resolution of the study in question.

Not only will graphic information appear in the Results section, but it may alternatively include narrative texts, which will objectively refer to the information included in the pictorial elements that are usually spread across the entire RA (particularly, along the Results and the Discussion sections with reference to the methodological procedures used). Data are displayed with a logical order so that relevant results are highlighted, but results do not speak for themselves, and they must be described in the text.

Glasman-Deal (2009, p.123) establishes four basic components of the Results section, which are as follows: first, writers must revisit the research aim and existing research on the topic and the methodology used, and give a general overview of the results; second, they must invite the readers to view the results and provide key results in detail, compare their results with those in similar research and compare them with model predictions; third, writers must include if they have encountered any problems with results; and fourth, they may succinctly comment on possible implications of results. In this section, the use of past tense is strictly necessary (Joubert & Rogers, 2015, p.42). In addition, writers do not include their own interpretation of the results, unless this section is combined with the Discussion section.

In the sample Results 1 (see Figure 9), it is possible to find a highly frequent use of procedural verbs in personal active voice constructions (*we* + V_{ed}), except for when it refers to the Material and Methods section, in which there is a change of perspective and passive voice structures are highly favoured. In the case of the sample Results 2 (see Figure 10), it is possible to find passive structures ({be} + $V_{pass.part}$) in the footnotes referring to the figures and tables, but not in the rest of the text, where simple clausal structures are found together with general and simple verbs, like the verb *to be*. The reason for this to happen may be because of the complexity of the text, in which there is a high number of symbols and figures. According to both samples, one can point out that the simpler the content of the text, the more complex the linguistic structures.

Figure 9. Example of Results (Results 1)

RESULTS 1

As glucans are well-known stimulators of nonspecific immunity, we decided to first evaluate the possible effects of the glucan-Se combo on phagocytosis. Using 2-hydroxymethacrylate particles known for low nonspecific adherence to the cell membrane, we found that Se alone had no effects. Glucan alone showed dose- dependent stimulation of phagocytosis. The addition of Se caused a small increase in phagocytosis, which was significant in the case of 200 μg of glucan with Se#1 [Figure 1].

Next, we focused on the stimulation of cytokine production. Using IL-2 release as a model, we found strong synergistic effects of glucan combined with Se#1 at all glucan-Se ratios [Figure 2].

For subsequent experiments, we used only 100 µg of glucan, either alone or with Se. The amount of Se (10 µg) was the same as before. As mentioned in the Material and Methods section, blood samples were collected at the beginning and the end of supplementation in all the tested groups. No significant changes were observed in the levels of total white blood cell (WBC), lymphocytes, or eosinophils. All tested materials, either alone or together, caused a significant increase in the number of neutrophils (the highest effects were found in case of glucan/Se#1) and a decrease in the number of basophils. The number of monocytes increased only with application of glucan/Se#1 combination [Table 1].

In the next step, we evaluated the role of tested combinations in cancer development. First, we used mice inoculated with the Ptas64 mammary carcinoma cells. Our results found the Se alone had no effects on…

FIGURE 1 - GRAPH

Figure 1: Effects of Se, glucan, or glucan/Se combination on phagocytosis of synthetic microspheres by peripheral blood granulocytes. Each value represents the mean ± SD. *Represents significant differences between glucan alone and glucan/Se combination at $P \leq 0.05$ level

…the growth of mammary cancer but when combined with glucan, Se significantly suppressed cancer growth [Figure 3]. When we used a model of Lewis lung carcinoma cells, glucan alone caused a significant decrease of lung metastasis and the addition of Se caused an additional 34% suppression [Figure 4].
(Vetvicka and Vetvickova, 2016)

Figure 10. Example of Results (Results 2)

RESULTS 2

Study Population

Of all patients enrolled in the original study (n = 902), hsTnT and copeptin values at admission were available in 882 patients (97.8%). The majority of patients had detectable hsTnT values (65.4%, n = 577; Figure 1), while 34.6% (n = 305) had hsTnT values below the LoD.

FIGURE 1 – GRAPH

Patient flow. In the Biomarkers in Cardiology (BIC)-8 main trial, 902 patients were enrolled. Of these, 20 patients did not have high-sensitivity troponin T (hsTnT) or copeptin values available at admission. Thus, the study population of this BIC-8 substudy consists of 882 patients with

complete biomarker results at admission. Patients were divided by their biomarker test. Relative and absolute frequencies are shown in the respective boxes.

Patient Characteristics

The median age was 54 years (IQR 43-67) and the majority of patients were male (63.5%, n = 560). The median GRACE score was 78 points (IQR 59-98), and 26.6% (n = 230) had known coronary artery disease, 15.0% (n = 131) experienced a prior myocardial infarction, 22.9% (n = 199) a prior PCI, and 5.1% (n = 45) CABG (Table 1).

[table has been excluded]

Clinical Characteristics

The majority of patients presented with chest pain (96.0%; n = 844) of moderate intensity (44.5%, n = 328) and a normal ECG (95.9%, n = 843). ST depressions were evident in 2.6% (n = 23; Supplementary Table, Appendix, available online). The median hsTnT value of all patients was 6 ng/L (IQR 4-9 ng/L), and 4.8 pmol/L (IQR 3.0-8.2 pmol/L; Table 1) for copeptin.

Outcomes

After 90 days, endpoint information was available in 98.3% of patients (n = 867). The primary endpoint occurred in 2.2% (19/867; Table 2). In patients with hsTnT values below the LoD, no myocardial infarction or death occurred (Supplementary Figure 1, Appendix, available online). Conversely, in patients with detectable hsTnT values at admission, death or myocardial infarction occurred in 3.4% (19/563). In the subgroup of patients with hsTnT between LoD and 99th percentile, 1.5% of copeptin-negative patients experienced the primary endpoint (6/410), as opposed to 6.3% in the copeptin-positive subgroup (6/96; Figure 2). In patients with an initially elevated troponin value at or above the 99th percentile, the primary endpoint was reached in 9.7% in copeptin-negative patients (3/31) and 15.4% in copeptin-positive patients (4/26).

[table has been excluded]

[graph has been excluded]

ROC Analysis

In ROC analysis, the AUROC for the prediction of death or myocardial infarction was 0.829 (95% CI, 0.760-0.899) for hsTnT and 0.741 (95% CI, 0.628-0.853) for copeptin. Both markers combined achieved a higher AUROC (0.849; 95% CI, 0.790-0.908; Supplementary Figure 2, Appendix, available online). However, the AUROC of both biomarkers combined was not statistically different from the AUROC of hsTnT alone (P = .349). In patients with hsTnT between LoD and 99th percentile, the AUROC for copeptin was 0.745 (95% CI, 0.612-0.877); in patients with hsTnT above the 99th percentile, the AUROC was 0.583 (95% CI, 0.311-0.588).

Survival Analysis

For the biomarker-combination of hsTnT and copeptin, the HR for the primary endpoint was 2.56 (95% CI, 1.85-3.54; P < .0001; Table 2, Figure 3). In patients with hsTnT between LoD and 99th percentile, copeptin was a significant predictor of death or myocardial infarction (HR 4.39; 95% CI, 1.42-13.61; P = .01). In the subgroup of patients with an initially elevated hsTnT, copeptin failed to

be significant (HR 1.61; 95% CI, 0.36-7.17; P = .536).

[graph has been excluded]

(Vafaie et al., 2016, pp. 276–279)

2.2.3.6. Discussion

The Discussion section "provides meaning or an interpretation of the results and shows relationships with other research" (Davis et al., 2012, p.29). It is essentially aimed at "turn[ing] data into knowledge" (Heard, 2016, p.120) and does so presenting the overall significance of the writers' work by usually showing what the results obtained imply, what their impact in already existing knowledge is and how these fit into similar studies found in the literature. In this line, and according to Malmfors et al. (2004, p.4), in the discussion section "results are interpreted in relation to previous knowledge, the formulated problem and [the writers'] hypothesis", by emphasising what is new in the work presented and why it is relevant to the field. Possible weaknesses and prospects for future progress must also be part of the Discussion (see Figures 11 and 12).

Indeed, Stuart (2007, p.71) emphasises this is the most important section within the research article, since it is where the writers must assert whether their findings support or dismiss the hypothesis formulated in the Introduction, so there is a clear connection between these two parts of the RA. As a central section, the discussion "considers and extends [the writers'] results to claim the strongest interpretation and the broadest importance that [the writers] can legitimately argue" (Heard, 2016, p.120).

In addition, it is common to find a final conclusive statement at the end of the Discussion section, particularly if there is no Conclusion section per se. These final lines must include the main objectives and "a general extent on to which [the writers'] have accomplished them" (Davis et al., 2012, p.29).

Figure 11. Example of a Discussion (Discussion 1)

DISCUSSION 1

The present study found that the combined use of TIVA (propofol–remifentanil) with ramosetron decreased the incidence of PONV (during the first postoperative 24 h) to 19.4% from 61.1% with sevoflurane inhalational anaesthesia. The aetiology of PONV after thyroidectomy is unknown, but *factors including* sex, age and intense vagal stimulation (through surgical handling of the neck structure) are known to be involved.[4] Other risk factors include obesity, history of motion sickness, previous PONV, smoking history and anaesthetic technique.[1,3] There were no between-group differences in patient demographics and surgical details in the present study, therefore the difference in PONV incidence can be attributed to anaesthetic technique and antiemetic (ramosetron) administration. Inhalational anaesthesia is a principal cause of PONV, especially during the early postoperative period.[19,20] The incidence of PONV following sevoflurane anaesthesia was 61.1% in the present study, a finding compatible with the results of others.[4–7]

The use of TIVA with propofol and remifentanil is known to reduce the incidence of PONV.[7,8,21] Propofol has antiemetic properties, acting on area postrema neurons via the γ-aminobutyric acid $(GABA)_A$ receptor to reduce activity and serotonin levels in the area postrema and cerebrospinal fluid.[22,23] Remifentanil is a nonaccumulative ultra-short-acting opioid with known antiemetic effects. Interestingly, the combination of remifentanil and sevoflurane does not increase the incidence of PONV compared with sevoflurane alone.[24]

TABLE 3

Incidence of postoperative nausea and vomiting (PONV) and adverse events in female patients following thyroidectomy with total intravenous anaesthesia with propofol and remifentanil (TIVAR) and prophylactic intravenous ramosetron, or sevoflurane inhalational anaesthesia without use of prophylactic antiemetics.

Although propofol based TIVA is associated with a lower incidence of PONV, its effect is limited to the early postoperative period. The choice of anaesthetic technique (volatile vs. TIVA) may be the main risk factor for early (0–2h) but not late PONV (2–24h).[19] In several types of surgery, the antiemetic effect of propofol-based TIVA was limited to within 6h postoperatively.[7–9] We have previously demonstrated that TIVA with propofol/ remifentanil significantly reduced the incidence of PONV compared with sevo-flurane anaesthesia at 0–6 h postoperatively (30.5% vs 55.9%) but not 6–24 h postoperatively (16.9% vs 22.0%).[7] Others have reported that 30–40% of patients experience PONV after propofol-based TIVA,[7,8,14] therefore the use of other antiemetics as well as TIVA is required to prevent PONV. It is generally recognised that a multimodal approach is more effective in preventing PONV compared with other strategies.[25]

Ramosetron is a 5-HT receptor antagonist that is used as an effective antiemetic in various types of surgery.[13] It has higher $5H_3$ receptor affinity and longer duration of action than other $5HT_3$ receptor antagonists, with longer elimination half-life (9.3 h) than ondansetron (3.5 h), granisetron (4.9 h) and alosetron (3.0 h).[16,17] In the present study, the use of TIVA with ramosetron (TIVAR) resulted in a lower incidence of nausea, vomiting, PONV, severe emesis and use of rescue antiemetics in the early post-operative period (0–6h) compared with the sevoflurane group. The use of TIVA without prophylactic antiemetics decreased the incidence of PONV by 45% (without reducing the incidences of nausea and rescue antiemetic use) compared with sevoflurane anaesthesia.[7] Taken together, these data suggest that the addition of ramosetron to TIVA may decrease the incidence and severity of PONV compared with inhalational anaesthesia.

During the late postoperative period (6–24 h) in the present study, only the incidence of severe emesis was significantly decreased in the TIVAR group compared with the sevoflurane group. Despite the long half life of ramosetron, it did not produce a sufficient antiemetic effect in the late post-operative period. Although the combined effects of changing anaesthetic technique (inhalational anaesthesia to TIVA) and prophylactic ramosetron administration are mainly limited to the early postoperative period, the effect on severe emesis was extended to later in the postoperative period. This result is consistent with a study that showed a decrease in early PONV and late postoperative vomiting, but not late postoperative nausea, with the use of ramosetron.[19]

A major limitation of our study was the lack of a TIVA control group in which patients did not receive prophylactic ramosetron. This is partly offset by the findings of our previous study that evaluated PONV following TIVA or sevoflurane anaesthesia.[7]

In conclusion, a combination of TIVA and prophylactic ramosetron administration appears to decrease early PONV and late severe emesis, compared with sevoflurane anaesthesia.

(Joe et al., 2016, pp.85–87)

Figure 12. Example of a Discussion (Discussion 2)

DISCUSSION 2

Cytisine was superior to nicotine-replacement therapy for smoking cessation among dependent smokers motivated to quit. Self-reported adverse events over 6 months were almost twice as common in the cytisine group than in the nicotine-replacement-therapy group. Types of adverse events in the cytisine group were similar to those seen in previous placebo-controlled trials of cytisine.6-9 Our trial was neither large enough nor long enough to assess the occurrence of uncommon adverse events or those with a long time to onset. Compliance with allocated treatment was modest. Time to relapse was delayed in the cytisine group. During treatment, participants in the cytisine group reported fewer symptoms of tobacco withdrawal, found smoking less rewarding, and reduced the number of cigarettes

smoked per day. The higher quit rate observed in women taking cytisine has not been previously reported in studies of nAChR partial agonists. This finding could be the result of chance (since data were not adjusted for multiplicity) but warrants further investigation, since several reviews of nicotine-replacement therapy have reported lower quit rates in women than in men,[23,24] possibly as a result of biologic and psychosocial differences.

We chose a noninferiority design on the basis of data available at the time, before publication of the placebo-controlled trial by West et al.[25] Our trial population was similar to that of New Zealand Quitline callers overall26 (although Asian, Pacific Islander, and male smokers are slightly underrepresented among Quitline users),[26,27] New Zealand smokers overall,[28,29] and three previous trial populations recruited through Quitline.[30-32] The trial was pragmatic, with broad entry criteria. The criteria for exclusion on the basis of medical conditions reflected the manufacturer's precautions for the use of cytisine and Quitline's policy regarding contraindications for nicotine-replacement therapy. As a precaution, we excluded people who reported that they had schizophrenia, given cytisine's similarity to varenicline (which has a boxed warning related to severe mental health problems). Users of noncigarette tobacco products were not excluded but were probably few in number given that such products are rarely used in New Zealand and that the sale of snus (a moist tobacco powder that is placed under the upper lip) is illegal. We followed the manufacturer's recommended dosing regimen for cytisine, although we are not aware of any published studies that support the regimen.

Our study had several limitations. First, since researchers were aware of treatment allocation, there may have been a reporting bias in favor of cytisine. Second, although adverse events were medically reviewed, they were self-reported. The higher proportion of adverse events in the cytisine group may be due to reporting bias, since the known side effects of nicotine-replacement therapy could have been regarded as "normal" by participants in the nicotine-replacement therapy group who had previously received such therapy and could therefore have gone unreported. We did not collect long-term safety data; in the placebo-controlled trial by West et al.,[25] adverse events reported during a 12-month follow-up period were predominantly related to gastrointestinal effects. Third, because participants were unlikely to have prior knowledge of cytisine, some of the treatment effect might be explained by its novelty. Fourth, verification of self-reported abstinence was not undertaken owing to both the broad geographic dispersal of the study population and budget constraints. In addition, at 1 month participants in the nicotine-replacement-therapy group would have received positive test results for cotinine, a metabolite of nicotine, even if they were not smoking. Abstinence rates may therefore be overreported or underreported (see the Supplementary Appendix for an estimate of validated abstinence), but there should not be a difference in the nature of reporting between groups. Fifth, smokers who call Quitline may be more motivated to quit than other populations of smokers.[33] Sixth, the treatment periods for the two interventions were different; the selection of a 1-month primary outcome should help to ensure comparability. There was also a between-group difference in participants' access to treatments (i.e., cytisine was delivered free, whereas nicotine-replacement therapy was obtained from a pharmacy at a small cost). In a previous trial with [1410] participants in

which the same recruitment method was used, participants were randomly assigned either to receive free delivery of nicotine-replacement therapy by courier or to use of the voucher system.[31] No between-group differences in continuous abstinence at 3 weeks, 3 months, or 6 months were observed. The trial shows that cytisine is an effective smoking-cessation aid for use as a first-line treatment for tobacco dependence. The most common adverse events were nausea and vomiting and sleep disorders. The effect sizes in this trial were similar to those observed in a trial of varenicline versus nicotine-replacement therapy.[34] Given the large difference in the market prices of varenicline and cytisine,[11] a head-to-head, noninferiority trial that includes cost-effectiveness analyses is justified.
(Walker et al., 2015, pp.2359–2361)

As for the main linguistic features found in the two sample Discussions, it is possible to find a mixture of impersonal active constructions (inanimate subject + V), personal active constructions (*we* + V_{ed}), passive constructions ({be} + $V_{pass.part}$) and *there existentials* and other impersonal structures. In the case of the Discussion section, the three types of constructions are usually found, although the passive voice seems to be quite predominant if compared to the other two types.

5. Summing up

Chapter 2 has focused on a detailed description of the main linguistic features found in medical writing and has delved into some of the most important types of manuscripts medical writers can make use of to publish their research results, highlighting, however, the preponderance of the RA over all the rest.

The RA embodies the scientific research process and serves as the canonical means to disseminate and transfer scientific knowledge. The Introduction, the Materials and Methods, the Results, and the Discussion are the core sections of this piece of text, together with other relevant front and back matters, such as for instance the Title and the Abstract, among others. Using two samples of each of the sections extracted from prestigious journals of a wide diversity of medical sub-disciplines, an overall description of the informational content has been provided and an exploratory analysis of the use of active and passive voice structures has been carried out.

As seen, the use of active and passive voice structures is an important aspect when it comes to determining the degree of (im)personality a medical text portrays. Chapter 3 will explore the notion of voice from a theoretical and an applied approach and will also relate medical texts with two notions worth considering: interpersonality and impersonality. Taking impersonality as an inherent feature in medical discourse, I will highlight the relevance of the use of voice as a grammatical representation of impersonality and I will frame authorial presence (or absence) through the description of the major linguistic devices used by writers aimed at diminishing or promoting their presence in the text.

3

The Notion of ‘Voice’.

Traces of Identity in Scientific Medical Writing:

A Linguistic Approach to Interpersonality, Impersonality, and Authorial Presence in Medical Discourse. The Use of Voice as a Grammatical Representation of Impersonality

---

The most difficult thing in life
is to know yourself.
Thales

It is true that the notion of voice is a term frequently used in current writing research. However, as seen in this chapter, this construct may be understood and used in many ways and in several contexts. Voice refers to a very broad concept and it is somehow considered as "diffuse and problematic" (Stock & Eik-Nes, 2016, p.89), since it is open to constant discussion among academics. This has led to a disparity of views on what (academic) voice means and how it is represented in academic texts. Indeed, a wide set of connotations have been attached to such a notion, and this has driven to a constant redefinition of the concept, both from more theoretical to more applied-like approaches.

Chapter 3 is essentially aimed at exploring how a wide array of authors have considered and identified features of voice through the analyses of written texts and what linguistic features are to be (and have been) considered when analysing the authors' identity and visibility in the text. Purely theoretical perspectives that have contributed to different concepts of voice (Prior, 2001; Sperling & Appleman, 2011; Tardy, 2012b, 2012a) are essentially referred to so as to give an overall account of this notion in the academic sphere.

Nevertheless, the purpose of this thesis is not to deal with the notion of voice from a theoretical perspective and add more shades of meaning to this concept. In this dissertation, the term voice will be considered as an umbrella term covering aspects related, essentially, to the presence (or absence) of writers in their texts. Even though the term 'voice' will be the one used in the first part of the chapter, the literature review, I will then turn to use other labels such as 'authorial presence', 'authorial visibility' and 'authorial identity', which may be seen as less controversial, more clearly identifiable, and somehow, more related to the analysis carried out in this research project. By doing so, I aim at escaping from interpretations of voice that may lead to some confusion.

After presenting core ideas to the study of the notion of voice from different perspectives, I relate voice to central concepts in scientific medical writing, such as interpersonality and impersonality. It is here where the concept of 'voice' is narrowed down into a more specific and restricted sense of the term, that of the grammatical voice. Here, I explore how the choice of grammatical voice, either the active or passive

voice, and the use of some other frequent strategies of (de)personalisation contribute to the degree of (im)personality portrayed in a text.

This chapter finishes with the idea that a cline, as also shown by other academics in the field (Herrando-Rodrigo, 2014, 2019, 2022; John, 2005, 2007; Martínez, 2001, 2005a; Tang & John, 1999a; Bordet, 2013; Lehman, 2018; Lehman et al., 2022; Lehman & Anderson, 2021a; Lorés-Sanz, 2008) is the best way to account for the degree of involvement or detachment authors show in their texts. The focus put on active and passive structures is swiftly shifted towards a more refined analysis in which grammatical voice is somehow blurred, in favour of a more fine-grain description of more personal and more impersonal-like structures. Hence, the need of a continuum, ranging from more personal to more impersonal structures, seems a positive departure point to the study of (in)visibility and authorial presence (or absence) writers portray in their written material.

3.1. A Conceptualisation of the Notion of 'Voice' and its Connection to Authorial Identity and Visibility.

The notion of voice has become a topic of theoretical discussions in academic writing and has attracted significant attention from scholars. Tardy (2012b, 2012a), for instance, argues that voice offers a very powerful way of understanding writing, and claims that the concept of voice is central to discourse and practice of writing (Tardy, 2012a). Elbow (1994) also discusses voice as complex and emphasises the fact that such a concept seems to mean different things and to function in different ways at different times, hence the divergent approaches to voice in Applied Linguistics. It is true that there has been little agreement on what is understood by the concept of 'voice'; however, what is undeniable is that the notion of voice portrays emotional overtones and attitudes, as well as modes of self and person.

According to Elbow (1994), writers construct identities for themselves when they take various subject positions and create voices, while at the same time he acknowledges that

voice is difficult to define and always deeply subjective. Other researchers, such as Charles (1999), Cherry (1988), Tang (2004) and Tang & John (1999a), Herrando-Rodrigo (2014, 2019, 2022), *inter alia*, have conducted empirical studies dealing with the notion of self-representation in texts, essentially basing those analyses from an expressionist approach to voice. In sum, they have tried to demonstrate how voices are myriad and how they do not necessarily relate to one single, genuine, self-identity.

As mentioned earlier, the fact that the notion of voice is complex and somehow seen as problematic in academic writing stems from its several different interpretations, understandings, and uses. The notion of voice is controversial insofar as it depends on the perspective from which different authors see it. The construct of academic voice "varies according to the perspective adopted" (Castelló et al., 2012, p.98). For example, Stapleton & Helms-Park (2008) deal with the notion of 'individualised voice', whereas Matsuda & Tardy (2007, 2008) focus on how voice is shaped in accordance with the role of the reader, highlighting thereby the social component of the concept and how readers contribute to its construction, an aspect that will be accounted for later in the chapter.

It is not until the late 1990s that a growth of interest into the ways that writing involves identity is really seen. This is possible, particularly, thanks to Ivanič's (1998) influential work *Writing and Identity*. This interest in voice-related issues extends also into a re-thinking of identity. Before that, Foucault (1972) had explored, for example, how the subject that we are becomes fragmentary in the academic or institutionally oriented contexts where we work. In the early 1990s, Fairclough (1993, 1995) had considered the writer's identity through the lens of several available identities, and by doing so, he explored the relationships of power and the extent of their impact on discourse construction and self-identity.

As seen, the foundation for recognition of diverse identities in academic writing can be traced further back to earlier studies, such as Ivanič & Roach (1990), Ivanič & Simpson (1992), Ivanič (1994, 1995) and Fairclough (1995), *inter alia*. In their work, they recognise various identities in academic writing. In Ivanič & Simpson's (1992) work,

for example, they argue that the researcher should use multiple methods to find out which self is to be hidden by which other selves, while for others, such as Butler (2006), the self is nothing other than a performative in academic writing. For Ivanič (1998), the writer's linguistic and textual features, influenced by convention, constitute the writerly self. In fact, these insights led to the identification of four main functions of such writerly self: the *autobiographical self*, the *discoursal self*, the *self as author*, and the *possibilities of selfhood in the sociocultural and institutional contexts*.

First, the *autobiographical self* refers to the fact that each individual is unique. Clark & Ivanič (1997) propose that personal life history plays a significant role in the act of writing, noting that, at any one time, writers are the sum of their experiences up to that point. As Ivanič (1998, p.24) comments, this autobiographical dimension is socially constructed, and it is in constant evolution as an individual moves through life. The sense of the autobiographical self as relatively central is also stressed by Elbow (1995), who suggests that the crucial struggle faced by writers of most academic texts is the tension between being oneself as a writer and being an authoritative figure required by texts in the academy.

Several studies (cf. Hyland, 2010b, 2011, 2012) also underline the complex kinds of selfhood performed by writers of particular academic genres in disciplines throughout the academy and show the extent to which identity is socially constructed through the linguistic choices made by writers as unique individuals. The right choice of such linguistic strategies is tightly linked to the inclusion of writers as rightful members of the community. Second, the *discoursal self* refers to the writer's identity that is crafted in response to the textual setting. Writers' texts articulate specific stances and personas in contexts shaped by a range of discourse conventions (Cherry, 1988).

Third, the *self as author* is a writerly identity that is also invested with a sense of authorship, where the writer's stance is authoritative, where opinions and attitudes in discourse are strongly articulated in a way to which Bartholomae (1995) refers as the 'power of authorship'. Fourth and last, the *possibilities for selfhood within sociocultural and institutional contexts* refer to the aspects where selfhood intersects with the

demands of the situation. Within the institutional setting of scholarly publications, novice academics, for instance, may be seen as less authoritative in their first published academic articles, as they may feel uncertain about their scholar identity. This is then a type of self that must be developed through exposure, time, and practical experience.

Having dealt with Ivanic's proposal of the four aspects linked to the notion of 'self' as far as identity is concerned, it is essential to highlight the major contributions to this field in the literature. Indeed, research into identity in discourse is largely focused on the investigation of self-mentions, where, according to (Hyland, 2005), the deliberate inclusion or absence of explicit authorial reference entails a stance that shapes an authorial identity. When researchers wish to claim their identity and make their texts their own, they resort to self-mention, referring to the extent of explicit authorial presence in the text measured through the frequency[20] of first-person pronouns and possessive adjectives (Brno, 2005; Dontcheva-Navrátilová, 2013; Harwood, 2005a, 2005c; Hewings & Coffin, 2007; Hyland, 2001b; Ivanič, 1998; Martínez, 2005a; Tang & John, 1999b; Vergaro, 2011; Fløttum, 2003; Harwood & Hadley, 2004a; Hyland, 2001a, 2002a; John, 2009; Kuo, 1999; Lorés-Sanz, 2008; Lorés-Sanz & Murillo, 2007; Carciu, 2009; Stotesbury, 2006; Vassileva, 1998, *inter alia*)[21]. However, research in the linguistic sphere into identity markers beyond first-person pronouns, whether singular

[20] Even though a frequency analysis is usually carried out when covering the use of personal pronouns (and possessive adjectives) in written discourse, some authors disagree with this methodological procedure. Zhao (2013) points out that qualitative analyses are preferred over frequency-based rubrics when it comes to measure voice strength. According to her, analysing certain specific features may be more useful than counting their frequency. In this same line, Gross & Chesley (2012) also consider that categorising and measuring the frequency of predefined categories to evaluate voice strength is somehow questionable.

[21] Some authors who are concerned with the use of first-person pronouns and consider such use as the main category in the analysis of 'voice' are: Dontcheva-Navrátilová (2013), Hewings & Coffin (2007), Nunn (2008), Pérez-Llantada (2009), Lorés-Sanz (2011a; 2011b), Stotesbury (2006). Some other authors who focus on the use of first-person pronouns as one of the categories are: Breivega et al. (2002a), Hyland (2008), Ivanič & Camps (2001), Rahimivand & Kuhi (2014), Zhao (2013), Fløttum (2009), Helms-Park & Stapleton (2003), Kuhi & Behnam (2011).

or plural, remains scarce (Breivega et al., 2002a; Charles, 1999, 2004; Hewings & Coffin, 2007; Ivanič & Camps, 2001; John, 2005, 2007, 2009; Callies, 2013)[22].

Needless to say, the examination of linguistic and textual features of written texts offers a clear view of the distinct ways in which identity is constructed and authorial positioning developed within discourse. One of the earliest key works in this respect is Gosden's (1993) investigation of the clause in scientific research articles, which focused on how distinguishable linguistic features were used in academic discourse to express the intentions of writers and establish authority among the specialist scholarly community. This was further built upon by Ivanič's (1998) research, in which she explores aspects such as the clausal structure, lexical density, verbs (focusing essentially on 'tense' and 'mood'), nouns and nominalised structures, modality, and lexis. She deals with the different ways in which these linguistic instances contribute to the creation of authorial identity and to the expression of a 'scholarly voice'.

In this same line, Ivanič & Camps (2001) investigate writers' types of voices and how writers position themselves in the text. This study draws on Halliday's ideational, interpersonal, and textual positioning of authorial identity[23] and it aims at showing how writers relate themselves within the particular scholarly community they are addressing to. Other major contributions to the field are Hyland's (1999a, 2004, 2005) and Hyland & Tse's (2005) analyses of hedges, boosters, emphatic markers, attitude markers, relational markers, personal markers, and self-mentions, linking all these notions with that of 'stance'. Indeed, when referring to hedges, Pho (2008) defends how these

[22] Hewings & Coffin (2007) criticise the connection established between the use of first-person pronouns and the notion of voice. Also, according to Stock & Eik-Nes (2016, p.95), "the focus on first-person pronouns as a feature of voice sems to contribute to the confusion about the concept of voice and to reducing voice to debates about the use of 'I'". Stapleton (2002) also criticises the 'overemphasis on voice' presupposing that the use of such a personal pronoun is considered as a signal for a powerful voice in existing writing research.

[23] Ideational positioning is related to how writers position themselves to the topic they are writing about (similar to the notion of 'stance' in some studies); interpersonal positioning refers to the way writers interact with their readers; textual positioning is related to "the mode of communication itself (Ivanič & Camps, 2001; Stock & Eik-Nes, 2016).

linguistic devices related to modal verbs might have different functions in different moves in abstracts. Hedges can also be considered a way to demonstrate the researchers' commitment to "social norms for scientific conduct" (Gross & Chesley, 2012, p.97) when presenting their findings cautiously. By doing so, authors show "the desire to embody in their prose the organised scepticism the scientific community so values" (Gross & Chesley, 2012, p.97) .

Moreover, Zhao (2013) considers that the identification of hedges may be somehow problematic and that these may cause some trouble when it comes to analyse voice. Hedges may be seen as problematic insofar as they can express a signal for both a weak and a strong voice, insisting thereby that a lack of these devices may result in a strong and authoritative voice. In Zhao's (2013) view, when measuring voice strength, qualitative evaluation is much more useful than frequency-based rubrics, an approach I will take in the study of this dissertation.

Charles (1999, 2004) focuses her analyses on verbs, adverbs, and adjectives and examines the extent to which such linguistic units shed light on the presence of authorial voices in the text. She highlights that these linguistic mechanisms are used to assert authorial presence and establish authority within the discourse. Similarly, Breivega et al.'s (2002a) exploration of first-person pronouns, metatextual comments, explicit and implicit references, and lexical items adds depth to our understanding of how authors in the scientific community position themselves and make affiliations with readers.

In the same line, Harwood (2005a, 2005c) and Harwood & Hadley (2004) delve into how first personal pronouns are used to develop a self-promotional tenor in their writing. In my view, his most relevant contribution is the identification of the several functions personal pronouns can perform, particularly as far as the relationship between writers and readers is concerned. According to him, first-person pronouns do not only show authorial presence, but they operate at a pragmatic and discursive level by, for instance, creating a research space or by organising the discourse, among many other

functions[24]. Furthermore, Harwood & Hadley (2004b, p.1027) highlights the idea that even the so-called 'author-evacuated' articles in the hard sciences can be seen "to carry a self-promotional flavour with the help of personal pronouns".

Another relevant contribution to the study of personal pronouns and their pragmatic implications is that of Vladimirou (2007). She contributes to the study of personal pronouns in the field of Linguistics and focuses on the pragmatic functions they perform to unveil how writers express their positioning within the text and the ways they negotiate their relationship with the informational content presented, as well as with their target audience. Indeed, Lorés-Sanz (2011a, 2011b) also deals with the connection between personal pronouns and voice, when asserting that that these pronouns play a crucial role insofar as they can be used to modulate the voice of the writer in academic texts.

Wise's (2005) analysis of the identity claims made by students in weblogs also complemented this discussion, as did Dickey's (2004) study of the effects of innovative teaching tools on the language and identity of students. Again, Tang (2004) focuses on linguistic reflexes of negotiability, authority, and solidarity brought in valuable insights into identity construction across discourse communities.

In Yamamoto's (2006) examination of the respective grammatical machinery of Japanese and English when projecting agency and impersonality, she peels another layer in uncovering the complex linguistic and cultural underpinnings of identity construction across a range of sociolinguistic contexts. Following a more social perspective, Hyland (2011) argues that individuals define themselves within the context of their community, taking up the linguistic resources of specific discourses and practices. In the academic environment, it is true that individuals negotiate a self which

[24] According to Harwood (2005c, 2005a) and Harwood & Hadley (2004b), first personal pronouns, such as 'I' or 'we', can publicise the writer and their work. The functions these pronouns can perform are as follows: create a research space, organize the discourse, outline procedure and/or methodology, explain the researcher's previous work, report or summarise findings, dispute other research findings, and indicate potential future directions for research.

is coherent and meaningful to both themselves as authors and the group they are part of, their community of practice.

John's (2005, 2007, 2009) wide-ranging study of organising functions, nominal *versus* clausal citation sequences, construal of participant roles of moves, material *versus* mental process types, passive *versus* active voice structures, and agentive *versus* non-agentive voice in academic discourse also offers a window to the subtility of the construction of authorial voice in academic discourses that can arguably only be appreciated as the complex interplay between a variety of linguistic and textual features in identity negotiation.

According to John (2005), the writers' identity is closely connected to such other terms as *ethos, individualism, self, ownership, persona, stance, voice,* and *authority*. The first three terms (*ethos, individualism,* and *self*) relate to writers' individuality, whereas the other terms (*ownership, persona, stance, voice,* and *authority*) refer to the actions writers undertake in their texts. This latter group can be further divided into two sub-groups: those which project attitude and positioning towards the text (*persona, stance, authority*), and those which simply indicate the presence of the writers without additional communicative purpose (*voice*).

As pointed out by Matsuda (2001, p.40), the notion of voice in applied linguistics presents a significant challenge due to its holistic nature and its "amalgamative effect" in the construction of authorial identity. It is due to this holistic nature that several authors, such as Pho (2008), have been using interchangeably the concept of voice with that of stance. To him, 'authorial voice' and 'authorial stance' can be used as synonymic expressions. However, other authors, such as Hyland (2012) and Thompson (2012) consider voice as a broader notion which includes stance. According to Hyland (2012, p.136) voice is broader than stance "as it concerns the control of features which readers recognise as legitimate and authoritative".

In this respect, Hyland's (2008, p.7) interactional model to analyse features related to voice emphasises two main domains: 'stance' and 'engagement'. As for 'stance', he

understands the textual voice of the writers, "the ways writers present themselves and convey their judgements, opinions, and commitments", as something achieved by using hedges, boosters, attitude markers, and self-mentions. As opposed to stance, 'engagement' involves the resources writers use to include the reader in the dialogue, reinforcing the social nature of academic texts. Engagement strategies refer, in this case, to reader-mentions, directives, questions to the reader and knowledge references. Thompson (2012, p.132) agrees in considering stance as an aspect of voice, because of its clear contribution "to the impression of the writer[s] in the text", whereas Stotesbury (2006) considers voice as 'personal stance' and highlights its connection with the use of first personal pronouns[25].

Indeed, an important number of studies have explored the array of distinct linguistic realisations writers use to make themselves (and their intervention) present or absent in their texts. These studies range from a more systemic functional approach (Martínez, 2001; Marín-Arrese, 2002, 2008; Rodríguez-Vergara, 2017) to a more discursive and rhetorical approach (Salager-Meyers, 1994, 1998; Connor, 2003; Lafuente-Millán, 2008, 2010).

Even though it is evident there is little agreement on what the notion of voice embodies in writing, it seems clear that such a concept comprises both an individual and a social dimension. It has an individual dimension insofar as writers' manifest themselves (or not) in the text by choosing particular linguistic realisations, such as particular lexico-grammatical choices, which contribute to the (in)visibility of their authorial voice. In addition, it also has a social dimension, as the discursive features and the summative effect they have on readers are ascribed and framed within a specific social context (Ivanič & Camps, 2001).

[25] According to Hyland (1998, p.20), the notion of 'stance' is tightly connected to the way writers use "community-sensitive linguistic resources to represent themselves, their position, and their readers". Other terms that have been used to refer to this same concept of 'stance' are as follows: footing (Goffman, 1981), hedging (Hyland, 1998; Martin & White, 2005), and evaluation (Hunston & Thompson, 2000).

In most of the aforementioned studies, the notion of voice is tightly linked to the way writers represent and position themselves towards the informational content presented in the paper and their target audience. In the case of scientific medical writing, writers are often believed to make use of a powerful set of depersonalising linguistic resources which give off an apparent sense of rigour, objectivity and detachment somehow presupposed, required, and inherent in this specific genre. I will now turn to delve into the notions of interpersonality and (im)personality as two main features of scientific writing.

3.2. An Approach to Interpersonality and Impersonality. Framing both Concepts within Scientific Medical Writing.

The concepts of interpersonality and impersonality are particularly relevant in academic and scientific medical writing. In the case of medical discourse, where the transmission of knowledge needs to adhere to certain standards, writers tend to make use of interpersonal and impersonal linguistic and rhetoric strategies to enhance their linguistic interaction among peers and with readers. The sum of all these linguistic choices and the right balance between interpersonal and impersonal linguistic features is key to effective communication and contributes to the writers' validation as competent transmitters of knowledge in the field.

Interpersonality in scientific discourse refers to the inclusion of the writers' presence and their personal engagement with the subject they are dealing with. It could then be considered as a rhetorical feature which shows the ways in which writers project themselves and their audience in the discourse (Lorés-Sanz et al., 2010). Indeed, this personal engagement can be achieved with the use first person pronouns (Kuo, 1999; Hyland, 2002b; Harwood, 2005c, 2005a; Tang & John, 1999b) or second-person pronouns (Ädel, 2014), for example, or through other types of expressions which may show the authors' viewpoint, their thoughts and/or emotions. Interpersonal linguistic choices are frequent in the introduction, discussion, and conclusion sections of a research article, as these three sections deal with the writers' motivations, their

contributions to the field, and the relevance of their findings. The degree of interpersonal language in a research article must be certainly balanced, because an overload of interpersonal resources may compromise the objectivity of the transmission of the research carried out.

Interpersonality has been an extensively developed topic in English for Academic Purposes research. Aspects such as the role and structure of grammatical subjects and thematic structures (Gosden, 1992; Rodríguez-Vergara, 2017), the use of nominalisations and stacked nominal phrases (Guillén, 1998a; Liardet, 2016), functions of long subject noun phrases (Vande Kopple, 1994), hedging (Crompton, 1997; Hyland, 1994; Salager-Meyer, 1994), writer presence (Breivega et al., 2002b; Harwood, 2005b, 2005d; Herrando-Rodrigo, 2014, 2019, 2022; Lorés-Sanz & Herrando-Rodrigo, 2020; Martínez, 2005b; Tang & John, 1999b), and the use of the passive voice (Baratta, 2009; Duboviciene & Skorupa, 2017; Espinoza, 1997; Hundt et al., 2016, 2021; Marín-Arrese, 1993; Rundblad, 2007a; Tarone et al., 1998; Hiltunen, 2016; Hardjanto & Roselani, 2022, *inter alia*) have been brought to the fore as key concepts when analysing a scientific academic writing discourse type.

Indeed, the perception of scientific texts has certainly evolved through time. A more traditional approach would take scientific papers as merely objective reports of scientific facts obtained through empiric procedural methods. Traditionally, scientific texts, particularly those belonging to the hard sciences, were considered as being faceless and 'author-evacuated' (Geertz, 1983), minimising the social component of scientific writing. This approach, however, has been contested by authors who have defended a clear interpersonal component in scientific communication (Bazerman, 1988; Mauranen, 1993; Hyland, 2005; Hyland & Tse, 2005; Swales, 1990). Considering such social component, scientific writing must bear in mind the vital role readers play in validating the claims put forward by academic authors (Lafuente-Millán et al., 2010).

Even though the addresser and the addressee seem apparently absent (the so-called 'author-evacuated' prose), the truth is that the text is written by and targeted to a particular discourse community, to a 'hypothetical' or 'anticipated' set of readers (Ädel,

2014). To facilitate communication, the writer considers the audience at which their text is aimed and anticipates the readers' reactions by adjusting the writing accordingly (Bathia, 1993, p.9). The role of the readers, as said, is then crucial, as they are susceptible to either accept or reject the writers' claims. For this reason, scientific writing may then be seen as an epistemological typology which is open to subjective interpretation (Hyland, 1999b), despite being considered as solely reporting results obtained through experimental and empirical evidence.

For this reason, although objectivity is an important aspect to consider when dealing with academic writing and needs to be kept as a prototypical feature inherent in scientific texts (Molino, 2010), subjectivity also needs to be added to the equation through the use of interpersonal strategies as a means to better influence the readers' perceptions and reactions to the text. In this line, Hyland (1999, p.15) states that "there is more to writing an RA than merely using words to represent an external objective piece of data (…) [because t]here is no secure means of distinguishing between objective observation and subjective inference". Therefore, the boundary between objectivity and subjectivity becomes, at some point, blurred.

Even though the notion of interpersonality represents a core concept to academic writing, in general, impersonality depicts a key feature of medical discourse, which is the type of discourse dealt with in this thesis. Shifting now the focus to the concept of impersonality, one can perceive that writers strategically minimise their presence by avoiding any trace of personal viewpoint. These impersonalisation strategies (Marín Arrese, 2002, 2003, 2008) refer to a wide range of linguistic means which allow for various degrees of backgrounding of the role of agency. Instances of impersonal linguistic resources are the extensive use of nominalisations and stacked nominal phrases (Guillén, 1998a), and the great use of agentless passive voice sentences. With an agentless passive, for example, as actions are not attributed to specific individuals, a sense of detachment and impersonality is somehow created. Other linguistic strategies include unmarked intransitives, existentials, non-finite present and past participial clauses, impersonal pronouns, and the use of abstract rhetors (non-personal active

sentences whose subject implies some kind of agency by means of metonymic expressions), *inter alia*.

Such linguistic strategies play an essential role in the transmission of scientific medical knowledge since they contribute to the sense of neutrality and objectivity expected in this type of discourse. Indeed, impersonal features allow the target audience to focus on the data and empirical evidence rather than on the authors' experiences or opinions. In this line and according to Breeze (2008, pp.163-164), "textbooks on academic writing (…) have classically taken the approach that writers should cultivate an impersonal, objective style, perhaps as a means of aiding writers in their struggle to acquire an appropriate scholarly tone". The adverb *classically* is here especially revealing, as it reinforces the idea of a strong and very well-established type of epistemology (with its own linguistic resources) which is respected and shared by, in this case, the medical scientific discourse community.

In this line, Hyland (2001a, p.164) suggests that "writers have to construct themselves as plausible members of a discipline, showing respect for its shared values and objectives, but at the same time, (…) prov[ing] the importance of their claims". The way to show commitment to such values is tightly linked to the adhesion to the customary routines as far as the methodological procedures and the writing practices are concerned (Hyland & Tse, 2005; Hyland, 2000; MacDonald, 1992; Shaw, 2003; Shaw & Vassileva, 2009; Vassileva, 2000; Yakhontova, 2006).

Becher & Trowler (2001) suggest that texts produced by academics within their 'tribe' are conceived as models of the discoursal and social practices of the group members. In this line, Lafuente-Millán et al. (2010, p.18) argue that

> [a]cademics belong to different discourse communities which shape their discoursal and rhetorical choices when communicating their research. Their membership to particular disciplinary or cultural academic communities influences their writing choices, which should aim to meet their peers' beliefs, values and expectations.

The epistemological practices of a given discipline are transmitted through the exposure to such practices. Indeed, as pointed out by Hyland (2001, p.224), choices made by academic authors concerning features such as degree of authorial presence, for instance, "are closely related to the social and epistemological practices of their disciplines". As shown in Fløttum et al. (2006), where they focus on constructions involving the personal pronouns *I* and *we* in three languages (English, French, and Norwegian) and in three distinct disciplines (Economy, Linguistics, and Medicine), the interaction between medical authors and their audience is not as explicit as the one shown by economists and linguists towards their specific discourse community. Medical discourse, when compared to other specialised types of academic discourse, can be perceived as utilising a wider set of impersonal mechanisms which apparently end up creating a clear sense of objectivity, rigour, and ultimately, a sense of detachment.

Despite being two clearly distinct notions, interpersonality and impersonality are inextricably woven together in academic scientific medical writing. Knowledge is transmitted in a community that rigorously enforces its standards of excellence within the bound of interpersonal behaviour. Authors manipulate linguistic and rhetorical resources strategically to engage emotionally with their subjects, while keeping a public persona and a respectful distance from them. Indeed, adopting an impersonal style may be seen as "a rhetorical choice that has more to do with how academics present a situation than with how it "really" is" (Molino, 2010, p.86), leaving space to state that "the decision to base one's arguments on objectivity is a matter of social alignment for persuasive purposes" (Molino, 2010, p.86). A careful balance between the combination of impersonal elements and other more personal-like resources is key to obtaining a text which includes an interpersonal dimension valid and trustworthy to the other members of the discourse community.

Since the writers' voice is constrained by the context and the genre in which it occurs, in this case, the medical scientific discourse, it is of paramount importance to explore the main textual and linguistic representations which contribute to giving the research article its character as far as neutrality, impersonality and objectivity are concerned (Martínez, 2001), features expected to appear in medical texts. These resources are

usually linked to an impersonal type of writing style, and for this reason, they are susceptible to be analysed as a means to determine the degree of visibility, latency, or invisibility of authors in their texts. The next section deals with the presence of writers in medical discourse and explores the most salient textual and linguistic representations of impersonality.

3.3. Authorial Presence in Medical Discourse. Textual and Linguistic Representations of Impersonality. Strategies of Depersonalisation.

As previously discussed, medical writing is a unique genre characterised by its precision, objectivity, and highly specialised approach to conveying and transmitting complex information. One of its major distinguishing features is the extensive use of depersonalisation strategies, which are essentially aimed at supressing the presence of the author by minimising their voice and serve as linguistic devices to compact densely complex information in the least space possible. The fact that authors suppress their voice to some extent, through these strategies, does not mean authors do not expect a positive reaction to their words and claims from the community to which they are addressing. On the contrary, they try to exert influence and control over their peers by sheltering those ideas within the umbrella field of Medicine, which constructs an established and well-robust framework that leads to perceive authors as authoritative enough personalities with an unbiased voice.

According to Martínez (2001, p.228), these linguistic structures, particularly when it comes to deal with an analysis of transitivity, “reveal the tension between the writers’ need to distance themselves from the text to present findings objectively, and the need to approximate to it, in the appropriate style in order to persuade readers of their validity”. This section explores an array of linguistic and textual strategies of depersonalisation used in medical scientific writing, by scrutinising how information is packaged in utterances and by closely examining textual and grammatical mechanisms which transmit a certain degree of impersonality. Indeed, chapters 1 and 2 already

explored the main textual and linguistic features present in medical writing: morphosyntactic features, on the one hand, and syntactic features, on the other.

As for the morphosyntactic features, I focused on the specialised type of terminology used in the register, specific clinical terms, formed in most cases by means of derivation and compounding, and the extensive use of nominalisations and stacked nominal phrases. As regards the syntactic features, I delved into information packaging and its clear connection to the choice of grammatical voice (active or passive), and into information packaging related to the great use of nominalisations and stacked nominal phrases, which present the ideas of the text as objectified, permanent, established, and thus, apparently trustworthy (Halliday & Matthiessen, 2014).

Furthermore, and as explained in the introduction to this chapter, the umbrella term of 'voice' will be henceforth substituted by labels such as 'authorial presence', 'authorial visibility', 'authorial representation' or 'authorial identity', restricting thereby the sense of the term to its purely grammatical category, that is, to the active and passive voice. In fact, the following description will mainly focus on the abovementioned syntactic and lexical features frequent in medical writing, that is, information packaging and its relation to the choice of voice, on the one hand, and how this relates to the extensive use of nominalisations and stack nominal phrases, on the other. Other resources, such as reduced relative clauses post-modifying nouns within nominal groups, inanimate subjects with active verbs, and existential *there* sentences will also be referred to as linguistic devices with the purpose of agent demotion. An integrated account of referential authors who have explored these linguistic phenomena will also be included.

To begin with, the first depersonalising element I will focus on is the way information is packaged in an utterance. Even though this dissertation does not put its main focus on the study of theme-rheme analysis from a systemic functional approach[26], I reckon it is

[26] Systemic Functional Linguistics (henceforth SFL) is a social semiotic theory, developed essentially by Halliday (1985, 1994). Other important authors working on this theory are Eggins (1994) and Halliday & Martin (1993). This approach emphasizes the meaning-making property of language and delves into the existing relationship between text and context. It is functional because language is described in terms of

important to explore the informational organisation of the sentence as a means to portray the (in)visibility or presence of writers in their texts.

Referential authors who have made a significant contribution to the study of thematic and rhematic structure are Halliday (1973), Halliday & Matthiessen (1994, 2004, 2014), and Matthiessen (2004), *inter alia*. Halliday & Matthiessen (2014) *Introduction to Functional Grammar* offers a detailed overview of this approach to grammar and discusses on thematic structure and its main function in text analysis and in discourse organisation. Indeed, Halliday (1973) conceptualises the notion of theme and defines the concept as the unit of the utterance which introduces the topic of the sentence. It is essentially presented as the first element in the clause and includes given information. In contrast, the rheme appears as subsequent information, usually used to elaborate on the theme, and is classified as carrying new information. Halliday (1994, p.37) literally describes the theme as "the element which serves as point of departure of the message; it is that with which the clause is concerned [and t]he remainder of the message, the part in which the theme is developed, is called in Prague School terminology, the rheme".

The existence of a thematic and rhematic structure (that is, a combination of given and new pieces of information) is key for writers to structure information in a way readers expect and can assumably digest thoroughly. By organising the sentence in a way that fronts the theme, authors may minimise their own subjective presence and thus end up producing a more seemingly objective and authoritative message, a message which is seen as more established and factual, or taking Martínez's (2001) term, 'objectified'. The notions of theme and rheme and how they interplay and succeed each other in the text can in fact serve as coherence and cohesion devices (Eggins, 2004), implying a textual and pragmatic function that goes beyond the scope of the utterance in which they occur. As Rodríguez-Vergara (2017, p.3) puts it, "the choice of theme as well as

how it is used in cultures and for which purposes, and not in terms of how language knowledge is stored in the speaker's mind. It is systemic as it gives priority to the linguistic choices that are available to speakers when they create and transfer meaning (Rodríguez-Vergara, 2017). SFL suggests that "each stage of a genre exhibits clusters of predictable lexicogrammatical patterns" (Martínez, 2001, p.229)

the choice of grammatical voice plays a role in the thematic development of larger units of discourse". The use of these cohesive methods of thematic-rhematic development and the choice of voice to adapt the message to such a progressive structure facilitates the coherence of the text by making it flow in a more reader-friendly way.

Indeed, connected to the notions of theme and rheme, as mentioned above, is that of the passive voice. The extensive use of passive is considered as a staple of medical prose and represents another layer of theme deployment. Such construction removes the author by shifting the focus in the clause from the researcher(s) to the research itself and its main results. Within a passive voice sentence, the doer of the action is usually downplayed, and new and relevant pieces of information are foregrounded (Swales, 1990). Marín-Arrese et al. (2001) explore the various linguistic strategies that are used for variation in, as they call them, mystification resources to disguise the agent of the action and highlight the function of the passive, among many other resources, as a mystification strategy to background the agent (or the causer) of the action[27]. According to them, the voice as a grammatical category denotes the relationship between the action, marked by the verb, and the patient (person or non-person), marked by the subject of the clause. A passive voice structure shows that the action presented is then subject directed.

In this respect, Solstad et al. (2004) also suggest that a passive construction involves agent[28] demotion and highlight the importance of such a demotion of the agent from subject in the active to optional oblique in the passive. A passive voice structure "allows for the subject of a sentence to be deleted, and thereby deemphasized in terms of its semantic role" (Baratta, 2009, p.1406). This construction, considered a 'discourse marker' (Rundblad, 2007), can involve the deletion of the original subject (as in agentless passives, in which the agent is totally omitted) or the deemphasis of the

[27] They explore linguistic resources such as agentless passives, nominalisations, *-ed* participle, resultatives, impersonal pronouns, incohatives, infinitive clauses, existentials, metonymic expressions, among many others.

[28] Agent here does not only subsume role of agent but also the role of cause.

subject within a long (or prototypical) passive, which presents the agent with a *by*-phrase in final position, backgrounding it.

Indeed, in medical scientific writing, agentless passives (and, to a much lesser extent, passives with an agent) abound and their summative effect caused by these instances in a research article oozes a perceivable sense of detachment and impersonality, as agentless actions may "serve to distance the writer (…) from the text" (Reilly et al. 2005, p.191). In fact, writers can have various reasons to account for the suppression of the agent. First, the agent may not be necessary, as it may be easily inferred from the context (what Givón (1990, p.567) considers as being "anaphorically predictable" or "predictable as the author of the text"), or because it is unimportant or unknown. Second, the agent may be an inanimate object. Third, the relationship established between contiguous sentences may demand placing a new piece of information towards the end of the clause.

The markedness of a passive voice structure entails a topicalising function, insofar as the emphasis is put on an important part of the utterance which is located forward in favour of a much more perceivable visibility of the agent of the action[29]. Medical English shows a particular fondness for passive constructions (Rodman, 1994), as they enable the research to be topicalised and thus become the object of study (Banks, 1994; Givón, 1994, 2017; Ding 2002). Indeed, the use of the passive voice may also serve as a textual cohesive device because it can ease the readability of the information being presented and can fulfil a guiding purpose, since "more than achieving correct syntax, the writer must ensure that the reader is guided from one idea to the next, with the end result being textual comprehension" (Baratta, 2009, p.1410).

[29] The passive voice is considered a means of demoting the agent and promoting the patient. According to Aissen (1999, p.687), "the passive is used if the patient is more prominent than the agent; the active is used if either the agent is more prominent than the patient, or if neither has greater prominence". In this line, "agent demotion and patient promotion, from the information structural point of view, are processes which take place because of the demands of the context" (Solstad et al. 2004, p.10)

According to Rundblad (2007), the use of passive voice structures as well as that of metonymic expressions in subject position imply two functions of impersonalisation. The former set of units accounts for the type of function called *impersonalisation for the purpose of generalization*. In medical writing, the passive may imply that through impersonalising the agent in the clause, the agent could have been any person, connoting thereby the idea that any researcher could have carried out the research and would have drawn the same conclusions as the ones presented in the text (Harman, 1975). This approach could be seen as a defending argument for the so criticised use of the passive voice in obscuring the author's identity in medical register[30], as it outweighs empiricism, replication, and falsifiability (Vande Kopple, 1994) over authorial presence.

The latter devices, metonymic expressions (also known as 'abstract rhetors') performing the syntactic function of subject, which will be seen further in this chapter, portray the function *impersonalisation for purpose of socialization*. Being anonymous may also represent a claim towards recognising full membership of a community of practice. This idea, that of revealing the presence of the author through its absence, is also supported by Baratta (2009, p.1420). Even though the focus is put on the notion of 'passive stance', he suggests that instances of passive structures can reveal what is important to the writer, and by asserting this, he confirms the "passive voice does have a role to play in academic writing, which goes beyond the disguising of the author and can [also] help reveal the author instead".

Leaving aside the potential role of the passive construction as a reveal of stance and authorship and taking it as a grammatical construction showing some degree of impersonalisation, one can logically assume that the active voice structure (whose subject tends to be the personal pronoun *we* in medical discourse) is in fact attracting, to some extent, personalisation to the text in question. According to Martínez (2001), it is the transitivity system and the syntactic resources attached to it that allow the visibility of the agent or its invisibility. In fact, authorial intervention and authorial distancing are achieved through diverse syntactic resources at the poles of a continuum, being the use of personal pronouns (essentially, the pronoun *we* followed by an active voice) placed at

[30] As seen in Chapter 1, Section 1.4, on the Plain Language Movement.

one end (authorial intervention), and the use of passive constructions, placed at the other (authorial distancing), respectively. Indeed, I will further explore the representation of these resources into a cline.

As already said, though, the passive construction including the agent through a *by*-phrase does not topicalise the doer of the action and thus diminishes its importance in the utterance presented. The agentless passive, additionally, omits the agent at a surface level, but this tends to be easily inferred from the context.[31] The objectified and thus impersonalised character of medical discourse is then not only achieved through these two lexicogrammatical resources. In fact, and as previously mentioned, the great use of nominalisations as "metaphorical grammaticalisations of reality" (Martínez, 2001, p.232), and inanimate subjects within active structures, among others, characterise this specific discourse as predominantly impersonal.

Nominalisations, which are key to academic discourse, portray static representations of language that, according to Liardét (2016, p.16), "reorganize dynamic (…) interactions into lexically dense, noun-dominated language" (cf. Hunston, 2002; Hyland, 2008, 2009; Ravelli, 1996). Such dynamic interactions correspond to processes and actions (which in a non-specialised register would usually be expressed through clausal constructions) that are transformed and remapped into nominal groups. According to Hao (2020, p.144), this transformation is considered "a transcategorisation in the grammar, whereby adjectives, verbs, or conjunctions are [converted and] recoded into nouns" that form weighty nominal groups, and that are somehow considered as "static, as given, or created outside the text" (Guillén, 1998, p.364). Likewise, the expression of events, qualities and logical relations are frequently remapped as nominal groups, and their use is particularly suitable for the medical register, as they connote "an appearance of solidity, stability, and fixed factuality" (Banks, 2005, p.350).

[31] According to Martínez (2001, p.234), basing her reflection on Halliday & Martin (1993, p.41), from a pragmatic approach, when the agent is explicitly stated in the clause structure, the interpersonal potential for negotiability is maximal, whereas when the argument is encoded as a nominal construction, the negotiability is minimal.

To put it differently, the usefulness of nominalisations, from a semantic perspective, is that they serve as linguistic resources to present a process as acquiring some of the quality of an entity, and thus "it becomes more "think-like"" (Banks, 2005, p.350). This process is labelled as 'reification' (Banks, 2001a; Halliday & Matthiessen, 2014). From a grammatical perspective, however, the fact that a process is mapped into a nominalisation allows for the inclusion of a set of arguments such as modifiers and quantifiers. Nominalisation then allows a dense amount of lexical information, and, as a unit, it can perform an array of functions in the clause (it can work as the subject, as a complement or as a prepositional completive, for instance) (Banks, 2005, p.350).

Indeed, according to Guillén (1998b, p.348), semantically speaking, the "unmarked function of a verb is to express a process, and the unmarked function of an adjective is to express a quality". Through a nominalisation, processes (verbs) and qualities (adjectives) are transformed into entities, which is the unmarked form of a noun. In this line, "[w]henever realisation steps outside this framework[,] grammatical metaphor is produced" (Banks, 2005, p.348; cf. Halliday, 1994; Halliday & Matthiessen, 2014; Ravelli, 1988). Here, the notion of 'grammatical metaphor' gains relevance because, as pointed out by Halliday (1998), grammatical metaphors are aimed at shaping scientific discourse because of their two main functions: they contribute to developing technicality, and additionally, they are aimed at facilitating logical reasoning.

A scientific piece of text, which describes and interprets scientific processes and phenomena, requires the use of grammatical metaphors "to carry forward the momentum of the argument" (Halliday, 1998, p.202; in Hao, 2020, p.150). By doing so, a preceding element, usually a clause, is packaged into an objectified element, which is expressed through a nominalisation. This allows the subsequent argument to be developed further through a clause (Hao, 2020, p.150). The nominalisation achieved through a grammatical metaphor is discursively relevant and functional as it contributes to a logical flow of discourse and strongly characterises the objectification inherent in the development and transmission of scientific discourse.

Indeed, this discursive flow is tightly linked to the aforementioned textual features, the theme (or given information) and the rheme (or new information), and to ideational meanings in the clause. Halliday (1988) suggests that, as a nominalisation can perform any syntactic function within a clausal structure, "where a nominalisation occurs as rheme (new) in the thematic structure of a clause, (…) it can then be backgrounded as theme (and [g]iven) in the following structure" (Banks, 2005, p.350). This is what causes nominalisations to be very useful resources for the development of logical reasoning in medical discourse. As described by Tyrikkö & Hiltunen (2009, pp.297-298),

> [a]s a noun, a process can function as the head of a noun phrase and, consequently, as the subject or object of a sentence. This makes adjectival modification possible as well as the use of an extended range of cohesive devices such as pronominal reference and helps to focus thematic attention on a process by changing its semantic role.

Halliday (1998) further explores the idea that the grammatical metaphor[32], created through the process of transcategorisation of verbs and adjectives (and also conjunctions) into nominal nouns, as pointed out earlier, configures and develops 'technicality'. The extensive use of technicisms and specialised vocabulary through nominalisations in medical register shows the result of the conversion of processes (expressed through verbs) and qualities (expressed through adjectives) into technical nouns, since as Halliday (1998, p.197) puts it, "the nominal group has, in its grammar, the potential for organising a large quantity of lexical material into functional configurations". Through nominalising, effective configurations of impersonality are created, allowing the text to be more solidified and obtaining a sense of abstraction which recalls the magnanimity of scientific medical writing.

[32] According to Banks (2005, p.348), "the relationship between grammatical metaphor and semantic metaphor is such that whereas in semantic metaphor the grammatical form is maintained (…) but the meaning is changed, in grammatical metaphor, the grammatical form is altered, but the ideational meaning is the same."

In addition, another phenomenon worth mentioning, due to its high recurrence in medical literature, is the use of inanimate subjects followed by active verb forms, or also called 'abstract rhetors' (Halloran, 1984). In these structures, inanimate subjects are meant to replace human subjects considered to be the actual agents of the activity expressed by the verb. They operate, at a textual level, as rhetorical devices and have thus become conventional in academic writing, particularly in medical register. These devices have been subject of several analyses in linguistic research (Johns, 2001; Master, 1991, 2001; Wiraszka, 2023; Khedri, 2016, 2020; McGrath, 2016; Sancho Guinda, 2014; Seskauskiene, 2022, *inter alia*).

Indeed, the abstraction that characterises these nouns, as well as the connection these nouns have with the research processes carried out, is what indirectly relates them to human agency. In this respect, Hyland (1998, p.172) points out that the relationship between these nouns and the research processes semantically hidden in them suggests that the research activities described are independent of the researchers who performed them, reinforcing the ideas of generality and replicability, because "research outcomes would be the same irrespective of the individual conducting it" (2005a, p.181). This idea recalls to the function *impersonalisation for the purpose of generalisation* proposed by Rundblad (2007) (cf. Harman, 1975), when referring to one of the functions of passive constructions and metonymic expressions working as inanimate subjects to reinforce empiricism and assert replication and falsifiability (Vande Kopple, 1994).

By using an abstract rhetor, there is a shift of focus in which the object under study becomes central. Indeed, scientific language is considered to favour the 'inanimate subject + active verb'[33] structure as a linker from experimental or observational evidence to conclusions (Johns, 2001, p.56). In addition to this, abstract rhetors are linguistic devices which also operate as 'writer-oriented' hedges (Hyland, 1998;

[33] Johns' (2001) study reveals that the three most frequent verbs in the 'inanimate subject + active verb' pattern are *indicate*, *suggest*, and *show/demonstrate*.

Seskauskiene, 2022[34]), and because of this they function as diminishers of the authors' responsibilities for their claims, as they tend to be used to "protect [authors] from the potential consequences of the claim being refuted" (Hyland, 1998, p.170). In this line, Wiraszka (2023, p.218), concludes that the "use of inanimate subjects with research-related active verbs is aimed at making academic discourse more objective".

According to Low (1999), and highlighted by Seskauskiene (2003), the pattern 'inanimate subject + active verb' shows metonymical and/or metaphorical extension. It is metonymical because "the meaning shift takes place in one and the same domain between the writer, or the producer of the text, and the text itself (as in the example *the paper suggests*)[; it is metaphorical,] "because we deal with a cross-domain mapping between the paper (or a domain of written works), an inanimate product, and a person, or a human being, in general. So the artefact assumes some characteristics of a living being" (Seskauskiene, 2022, p.86).[35]

Having already dealt with the notions of passive voice constructions, nominalisations, and combinations of inanimate subjects followed by active verbs, it is now time to turn into the fourth main linguistic device used to impersonalise medical writing: the use of reduced relative clauses (and extended phrases), which are usually paired with nominalised units.

[34] Seskauskiene (2022, p.91) uses the Conceptual Theory of Metaphor (CTM) framework to interpret the results obtained in her study. Such a study yielded several major metaphorical extensions: RESEARCH AS A PERSON (to express speaking and reasoning abilities), RESEARCH AS A JOURNEY (to move forward and overcome obstacles), RESEARCH AS CONSTRUCTION/BUILDING (to deal with the foundations and support), RESEARCH AS ECONOMIC ACTIVITY (to provide things), RESEARCH AS (UNCOVERING) A SECRET (to express unexpected but very much appreciated revealing of a fact or tendency), RESEARCH AS A CONTAINER and RESEARCH AS A PICTURE.

[35] Hardjanto & Roselani (2022, p.2076), referring to (Banks, 1996, 2008), point out that these are "constructions whose subjects are inanimate but the verbs are active (Master, 1991). (…) [These] constructions require human agents as the verbs are in the active form. The subjects, therefore, must have an agentive role. Yet, the subject is inanimate. Thus, there is a mismatch between the inanimate subjects and the active verbs, (…) which is called 'untypical animacy'. This serves, as said earlier, to "avoid any responsibility for the truth of the proposition (…) [because the writers'] responsibility is transferred to the inanimate subjects, which then become the source of their claim (Hyland, 1996).

A reduced relative clause is essentially formed by deleting the relative pronoun and the finite auxiliary form of the verb 'be' of a full relative clause (Chomsky, 1957; Ross, 1967, 1972; Smith, 1964), so it is assumed that reduced relative clauses (non-finite forms) derive from full relative clauses (finite forms). As expressed by Harwood (2018, p.430) , "the inflection that the verb *be* would select (progressive aspect or passive morphology) would still be present in the reduced relative clause, but without the overt finite auxiliary". This contributes to the idea that "the interpretation of progressive aspect and passive voice is encoded on the inflections themselves" Harwood (2018, p.431) and thus, this shows that the semantic component expressed through the auxiliary 'be' is somehow 'pointless', as the aspectual information can be retrieved without inconvenience[36].

This relates to the possibility of transformation of nouns followed by past participial and present participial reduced relative clauses[37] into clausal constructions, showing a marked pattern of voice in the former, and an active pattern of voice and the progressive aspect in the latter. In other words, in the structure 'noun + past participial reduced relative clause', the whole utterance can be transformed into a passive construction (known as bare passive (Wanner, 2009; p.170)), whereas in the structure 'noun + present participial reduced relative clause', the whole nominal unit can be easily paraphrased as an active sentence in its progressive form. In the former, grammatical passive voice is overtly expressed when using a full relative clause (but somehow hidden when using a reduced relative clause), and in the latter, grammatical active voice and progressive aspect are overtly expressed when using a full relative clause (but hidden when using a reduced relative clause).

[36] Following Boskovic (2014) and Harwood (2013, 2015), *inter alia*, Harwood (2018, p.442) assumes, that What You See Is What You Get (WYSIWYG). In other words, if a particular aspectual form is absent from the sentence, then its associated projections are absent from the underlying derivation; that is, if perfect tense, progressive aspect, and voice layers, respectively, are absent from the underlying derivation, vP, however, remains, even in the absence of Voice P, to introduce agentivity.

[37] Also referred to as relative *-ed* participle clauses (or bare passive) and relative *-ing* participle clauses, respectively.

In specialised registers, non-finite clauses functioning as post-modifiers of the head noun are frequent linguistic resources (Elsness, 2014, p.91)[38]. According to Hare et al. (2007, p.430), one reason for this to happen is because "highly transitive verbs, particularly those with human agents, are more likely to occur in the passive, and therefore in the reduced relative" (McKoon & Ratcliff, 2003; Merlo & Stevenson, 1998).

A reduced relative clause functions as a post-modifier of the head noun it accompanies, restricting thereby its sense. It consists of a non-clausal phrase which is typically embedded in a noun phrase, creating, as described earlier, complex, and dense nominal groups, (Biber & Gray, 2011)[39]. Indeed, the use of present or past participle reduced relative clauses extends the length of the nominal group in which they occur, and thus contributes to the weight nominalisations and stacked nominal phrases acquire in the text. Even though in this dissertation emphasis is put on these non-finite structures, mainly because of their ability to be easily rephrased into active and passive finite clauses and their commitment to mystify authorial visibility, it is worth noting that they are not the only utterances which function as post-modifiers of the noun. In fact, there is an extensive number of post-modifiers, essentially prepositional phrases, which hugely contribute to the elongation of the nominal group and grant its so-called objectified and static nature.

Finally, the last linguistic resource to provide medical text with some degree of impersonalisation is the use of existential *there*. This structure consists of a sentence in

[38] Noun phrase post-modifiers differ in terms of explicitness (Quirk et al., 1985, p.1243). It is possible to create a continuum of the different post-modifiers a head noun can hold, from more explicit to less explicit. Full finite relative clauses would represent a high degree of explicitness, non-finite clauses (with some elements omitted) would represent an intermediate degree, and lastly, prepositional phrases and adjective phrases would represent the lowest degree on the scale (in Elsness, 2014, p.96).

[39] According to them, the following five grammatical devices have been especially relevant in the development of the densely packed discourse style in medical texts: nominalisations, attributive adjectives, nouns working as nominal pre-modifiers, prepositional phrases as nominal post-modifiers, and appositive noun phrases. As far as I am concerned, they do not refer to non-finite present and past participle reduced relative clauses.

which the canonical subject position is occupied by the 'empty', but grammatical, subject *there*, followed by an intransitive verb "to predicate the existence of occurrence of something" (Sun & Cheng 2015, p.1869). In general, *there* has no meaning, despite performing the syntactic function of grammatical subject (Huddleston & Pullum, 2002). However, the expletive *there* may be considered a 'dummy' element which is used as a guidance of new information, causing the logical subject to occupy the position adjacent to the verb, labelled as 'extension component' (Sun & Cheng, 2015). It is also used to describe or further interpret the information previously presented[40].

It is worth noting that the form of the verb 'to be' that follows the expletive needs to agree in number with the noun presented as a copular argument, not with the expletive *there*, which, as a word, shows no number at all. This is analysed by Basilico (1997, p.292), who states that "agreement on the finite verb is controlled by the number of the postcopular NP, but it is *there* that occupies the subject position". In medical English, the nouns that form this extension component are very restricted (nouns such as *difference(s), evidence*, etc.), as will be further developed in the corpus study presented in this dissertation.

3.4. Summing up

To sum up, all the abovementioned linguistic resources are key to medical writing. Indeed, the abundance of these distinctive lexicogrammatical structures, and their summative effect, contribute to the impersonal character of medical discourse. This impersonal character is achieved through the accumulation of structures whose purpose, among many other reasons, is to diminish authorial visibility and present trustworthy information statically, solidly, and objectively.

As pointed out earlier, the dichotomy *active* vs. *passive* does not directly correspond to the dichotomy *personal* vs. *impersonal*, because a passive sentence may be interpreted

[40] Check Jenkins (1975) for a detailed account on the use of English existential within the framework of Generative Grammar.

as being more personal than an active one. A clear example of that is the case of long passives, which include the agent, when compared to active structures with an inanimate subject. In the former, the agent is overtly expressed, despite being backgrounded and thus diminished. In the latter, although the agent may be inferred from the context, it is not mentioned.

All the linguistic devices presented in this chapter are considered phenomena that do not qualify as either being totally personal or totally impersonal and therefore "they may be viewed as situated somewhere in between" (Solstad et al., 2006, p.2). As also shown by other academics (John 2005, 2007; Tang & John, 1999a, 1999b; Martínez, 2001, 2005; Lorés-Sanz, 2008; Bordet, 2013; Lehman, 2018; Herrando-Rodrigo, 2014, 2019, 2022), and taking into consideration the complex ways in which writers' make themselves visible or non-visible in their texts, I have concluded that a cline is an optimum way to account for the degree of the writers' involvement or detachment as far as the informational content of the propositions transmitted is concerned, as will be shown in Chapter 6.

Through a continuum, ranging from more personal to more impersonal-like structures, I aim at blurring the line between active and passive structures in favour of a more refined conceptualisation which will be suitable to account for the real degree of impersonalisation these lexicogrammatical choices carry with them.

4

Methodology

In everything,
there is a share of everything.
Anaxagoras

The methodology described in this next chapter has served to carry out the two different studies that are presented in this dissertation: on the one hand, a corpus-based study on the use of active and passive voice structures in medical RCTs (Chapter 5), and on the other, a close textual analysis on the use of a particular set of linguistic strategies which may serve to portray medical discourse as neutral, objective, and detached because of their impersonalising effect (Chapter 6).

The **corpus-based study** (emphasis added) is aimed at quantifying a set of active and passive structures recurrent in RCTs. They are quantified through corpus searches using two distinctive types of syntaxes: the Simple Query Syntax, and the CQP Syntax. The high frequency of these structures, after scrutinising a set of real RAs, allows their inclusion within the restricted set of lexicogrammatical patterns typically occurring in medical register. Bearing in mind their frequency of occurrence is pivotal to assess the extent to which authors rely on them when writing their research. Once the quantificational data are retrieved, their subsequent qualitative analysis will follow.

This qualitative analysis consists of exploring and checking whether there is a systematic semantic connection between the lexical verbs used in the patterns and the patterns themselves. Furthermore, the analysis attempts to account for the relationship between the typology of highly frequent lexical verbs and the choice of tense. Another step taken is that of relating the extent to which active and passive structures relate to (im)personality and authorial (in)visibility. Within the quantitative and qualitative analyses, information on the limitations encountered when relating active and passive structures to the notions of (im)personality and authorial (in)visibility is also provided. The corpus-based study will conclude with a proposal to scientifically categorise the structures analysed in the corpus-based study through a Cartesian coordinate system, in which mainly two distinct parameters or axes will be used, namely *personality* (x axis) and *voice* (y axis).

Through the corpus analysis, the retrieval of information was successfully achieved despite the limitations encountered. However, I considered the need to complement this study with a **close textual analysis** (emphasis added) in which the overall picture of the

RCT paper and the linguistic strategies used in it could be better perceived within a more framed context. The main objectives of this contextualised study are conceived as an attempt to disentangle the complex and multi-layered dimension of impersonality by exploring the extent to which the strategies of (im)personalisation described interact to textually characterise medical discourse as being considerably, or even deeply, impersonal.

Through a manual classification of the lexicogrammatical patterns analysed, essentially those dealt with in the corpus-based study, I attempt to assess the degree of (im)personality they carry with them by analysing the extent to which these instances relate to authorial presence (or absence) in the text, an aspect which is partially tackled in Chapter 5. Raw frequencies and normalised frequencies per thousand words are also provided to account for their distribution in the RCTs and to acknowledge frequency as an important analytical parameter.

This contextualised study is also aimed at assessing the pragmatic functions these patterns perform within the text as far as authorial (in)visibility and (im)personality are concerned. This is achieved through an exploration of the interplay between these lexicogrammatical patterns within the text, an analysis which will serve to evaluate how such interaction affects the transmission of knowledge and how (im)personal this transmission of knowledge is. In this case, I opted to show the gradation from more personal to more impersonal-like patterns through a continuum. This continuum is provided at the end of the analysis of each specific pattern, and it is built up as each pattern is being introduced and assessed. The end of the chapter provides the final cline with all the various structures analysed so that the overall picture of both (im)personality and authorial (in)visibility (also including information on grammatical voice) in RCTs (and by extension, in medical discourse) is visually shown.

Both studies are essentially aimed at deciphering the interrelatedness between the notions of impersonality, authorial (in)visibility, and the use of grammatical voice (essentially the extensive use of the passive voice) in medical discourse using a mixed-method approach, in which quantitative data are subsequently qualitatively analysed. It

is true, however, that the effect other lexicogrammatical patterns exert over this impersonal perception also needs to be considered, as they may be used to either reinforce this impersonal reality or counteract it. To explore the correlation between these patterns and three distinctive analytical dichotomised parameters, namely *personality* vs. *impersonality*, *active voice* vs. *passive voice*, and *authorial visibility* vs. *authorial invisibility*, is then crucial to unravel the complex notion of impersonality, and both a Cartesian coordinate system and a cline prove to be optimum ways to do so.

This chapter includes the rationale for using a corpus-based approach highlighting the status of Corpus Linguistics as a methodological tool to explore language and discourse. Then, information on the methods for data collection and analysis will also be provided. First, by describing in detail the RCTs corpus used in Chapter 5. Second, by succinctly describing most of the syntactic expressions used to search for the distinctive set of patterns. These two aspects will lead to the operationalisation of the notions of grammatical voice and impersonality through a corpus-based approach using a Cartesian coordinate system. A summary of the main limitations encountered and how they have been sorted will also be provided.

The second part of this chapter consists of a brief rationale for using a close textual analysis approach as a complementation to the corpus-based study, accounting for the way the notion of voice has been operationalised as a representative of (im)personality and authorial (in)visibility through a continuum including all the lexicogrammatical patterns analysed. This continuum is in line with the several clines already developed in the literature, as it is acknowledged to be one of the best ways to represent the fuzziness and the multi-layered nature of the notion of impersonality, in particular. The continuum presented in this thesis shows a gradual representation of impersonality by means of lexicogrammatical linguistic choices, and it attempts to revamp other clines by adding a bidimensional character with the inclusion of authorial (in)visibility as a defining feature.

Finally, a brief evaluation of research will follow, stating whether the results were conclusive and whether the choice of methodology was effective in practice.

4.1. The rationale for using a corpus-based approach: Corpus Linguistics as a methodological tool to explore language and discourse

The methodological approach to linguistic studies has experienced a tremendous evolution in the last decades. Corpus-based research has gained recognition among scholars (Sinclair, 1991, 2004; Hunston & Francis, 1999; Conrad, 2000, 2004; McEnergy & Wilson, 2001; Tognini-Bonelli, 2001; Meyer, 2002; Hunston, 2006; McEnergy et al., 2006, *inter alia*) as the best method known to provide linguistic studies with accurate empirical contextualised evidence to describe the way language actually operates (Meyer, 2002).

The intuition-based approach, framed within Chomsky's Generative Grammar, in which native speakers' introspection and elicitation, informant testing, and text linguistics were the basic methods to carry out linguistic enquiries, has given way to Corpus Linguistics as a new methodological approach to linguistic analyses. As Leech (1992) points out, Generative Grammar focuses on linguistic competence and is interested in establishing linguistic universals. Its studies frequently use qualitative analyses, encompassing thereby a more rationalist approach to scientific enquiry. In contrast, Corpus Linguistics offers a more empirical approach, as it focuses on linguistic performance and linguistic description, making use of both quantitative and qualitative methods to hypothesise about linguistic use[41] (Meyer, 2002, p.4).

Corpus linguists describe language by making use of large bodies of naturally occurring texts, which are electronically stored. When a word or a linguistic pattern (passive verb phrases, for instance) is searched for in the corpus, concordance lines, displayed as key words in context (KWIC), return a specific number of occurrences in which that word or linguistic unit appears. The immediate context is also provided in each concordance line and, when the node word or linguistic unit is clicked, further context is usually

[41] While a generative grammarian theorises about language so as to expand theoretically the knowledge of universal grammar, a corpus linguist attempts at accurately describing language performance (Meyer, 2002, p.4).

given. Information on frequency of occurrence, distribution across the sections of the text, and frequent collocates are also relevant pieces of information one can retrieve when using corpora. All this empirical information can help linguists "explore the way in which language users make particular selections for particular occasions and particular tasks" (Sinclair, 2004, p.4), and in more general terms, "gain a better understanding of how language is actually used rather than how language is perceived to be used" (Tsui, 2004, p.40).

Even though Corpus Linguistics has established itself as a trendy methodological tool in linguistic studies, a corpus-based approach does not need to disregard the more traditional intuition-based approach; in fact, both can complement each other (McEnery et al., 2006, p.7). Corpora are texts from which naturally occurring instances come, but intuition and introspection can enhance the previous and subsequent linguistic analyses. Before the corpus is used, the linguist must carefully consider the elements to be analysed. While using the corpus, such considerations can be made more specific and precise. The corpus reveals raw data which, filtered and narrated by the linguist, can become relevant linguistic knowledge about authentic language use. A corpus may then be considered as a means to obtain raw linguistic data which will be transformed by the specialist in refined linguistic knowledge. Corpora are then the means, not the end in themselves. This has been the approach taken in this dissertation.

Corpus Linguistics is a "way of doing linguistics" (Meyer, 2002, p.xi), "a set of methods and principles of how to apply corpora in language studies" (McEnery et al., 2006, p.7) by observing relatively large amounts of naturally occurring language, patterns of association between word groups or between a pattern and a text type, and relative frequencies (Hunston, 2006). Some scholars, however, argue that Corpus Linguistics has established itself as an independent linguistic discipline (Tognini-Bonelli, 2001) because it has surpassed the methodological function to become a domain of research in itself. Some others, despite assessing Corpus Linguistics as a methodology, recognise its theoretical status (McEnery et al., 2006; Gries, 2008).

Hunston (2006, p.244), for instance, refers to Corpus Linguistics as 'theory neutral', but recognises that

> (…) the practice of doing corpus linguistics is never neutral, as each practitioner defines what is meant by 'feature' and what frequencies should be observed, in line with a theoretical approach to what matters in language.

Although Corpus Linguistics may apparently hold some theoretical status as it contributes to linguistic theory, this thesis sees Corpus Linguistics as a methodological tool to explore actual instances of language. Firstly, because Corpus Linguistics is not restricted to a specific linguistic domain, nor does it require analyses of a particular nature. Secondly, because its universality proves that linguists of all persuasions and working from different approaches can make use of corpora for their linguistic studies (Meyer, 2002; Hunston, 2006). Both facts shape Corpus Linguistics' methodological nature.

Applying a corpus-based approach to linguistic studies presents numerous advantages and benefits. A methodology based on corpus meets the standards of any scientific enquiry insofar as it produces pieces of evidence that can be used to corroborate or refute hypotheses. In a way, one could state that Corpus Linguistics is to Linguistics what the scientific method is to science. Indeed, the information obtained from corpora can be statistically processed, which means that the linguistic discussion can include both quantitative and qualitative analyses, ensuring thereby the empiricism and objectivity required in scientific research (McEnery & Wilson, 2001). If a corpus includes authentic texts and is representative[42] enough of the language (or language variety) it is aimed to represent, the results obtained can be extrapolated to the entire language (or language variety) so that generalisations can be made and conclusions can be drawn (Biber, 1993).

[42] According to McEnery et al. (2006, p.7), "(…) the essential qualities of a corpus include machine-readability, authenticity and representativeness."

The rapid and easy access to large amounts of texts is also a point in favour to use a corpus-based methodology. On the one hand, published corpora are available online and do not usually present excessive use restriction. Some are publicly available; some others only request an official registration. On the other hand, the linguist also has the possibility to compile his/her own corpus. Collecting texts in electronic format on the Internet or in databases is time-consuming, but it is indeed an affordable task. Knowledge on how to use linguistic software tools and other computerised methods are, however, fundamental to exploiting the resources corpus-based methodology offers.

Corpus Linguistics has been used as a methodological tool to extensively explore and exploit various linguistic areas. For instance, Corpus Linguistics is tightly linked to the advent of Computational Linguistics, human language processing, and machine translation (Grisham, 1986; Barnbrook, 1996; Brants et al., 2002; Hammond, 2003; Hockey, 2004; Gries, 2009b, 2009a; Kilgarriff, 2009; Svensson, 2010; Owens, 2011; Jensen, 2014, *inter alia*), as they are essentially based on corpora and statistical methods. Computational models are applied to give shape to linguistic data and are used to deal with relevant information to perform large-scale statistical analyses. A process which certainly shapes linguistic data is that of morphosyntactic annotation and parsing, for example. A corpus that is annotated for part-of-speech (POS) tags accepts searches which include those tags, as a way to restrict the grammatical category of the words being searched for. Clause-type and longer structures, such as active and passive voice instances, can be searched for when the corpus is parsed. Annotation at any linguistic level is thus aimed at transforming raw plain text into a real repository of linguistic information.

Corpus-based analyses have also been used in the field of Historical Linguistics (Pahta et al., 2007; Kawaguchi et al., 2011; Biber & Reppen, 2015; Smitterberg & Kytö, 2015; Kesäniemi et al., 2018, *inter alia*) to essentially explore the nature and direction of language evolution, change, and variation through corpora that cover different periods. With the creation of parallel and comparable corpora, Corpus Linguistics has also contributed to expand the field of contrastive translation studies (Baker, 1993; Gellerstam, 1996; Lauridsen, 1996; Barlow, 2000; Botley et al., 2000; Xiao, 2003, *inter*

alia), providing linguists with a wide range of sources to establish parallelisms and dissimilarities between languages.

Language acquisition and learning and language teaching have also greatly benefited from the advent of Corpus Linguistics (Celce-Murcia, 1991; Ellis, 1998; Partington, 1998; Tribble, 2000; Tan, 2002; Granger et al., 2002; Nesselhauf, 2003, 2005; Aston et al., 2004; Sinclair, 2004; Römer, 2005; Scott & Tribble, 2006; McEnery et al., 2006; O'Keeffe et al., 2007; McEnery & Xiao, 2010, *inter alia*), essentially in the areas of lexis, grammar, and semantics. The interest in the description of real instances of language has encouraged the creation of corpus-based classroom materials, informing textbooks, other teaching materials and grammar references (Quirk et al., 1985; Biber et al., 1999).

The use of corpus-based techniques has also revolutionised the making of dictionaries. Corpora have provided lexicographers with primary data about the usage of words, their different senses, their changes in meaning, their frequency in specific registers, and their association with other words (Kilgarriff, 1997; Hunston & Francis, 1999; Hunston, 2006; Atkins & Rundell, 2008; Kilgarriff & Kosem, 2012). Such a contextual approach has encouraged, for instance, the treatment of phrases and collocations in a more systematic way, such as in Hunston & Francis' (1999) *Pattern Grammar*, "a specific application of the observation that phraseology and meaning are linked" (Hunston, 2006, p.242), an aspect which will be explored in this dissertation.

Corpus-based lexical research has greatly contributed to the analysis of lexical items and their collocational patterns (Levin, 1991; Kilgarriff, 1993, 1997; Kilgarriff & Rundell, 2002; Nesselhauf, 2005; Evert, 2008; Gries, 2008, 2013; Granger & Paquot, 2012; Nguyen & Webb, 2016; Gablasova et al., 2017), both for pedagogical purposes and linguistic description. In this line, lists of frequent words have also been developed in academic discourse (Coxhead, 1998, 2000; Coxhead & Hirsch, 2007; Gardner & Davies, 2014, 2016), which cover fields such as History, Social Sciences, Science and Technology, Law and Political Science, among other areas. Wordlists have also become common resources in other more specific disciplinary fields, such as Civil Engineering

(Gilmore & Millar, 2018), Medicine (Lei & Liu, 2016), or Education (Mozaffari & Moini, 2014), to name a few.

Accessing corpora via grammatical patterns and not via word forms is now possible due to advances in automatic tagging and parsing. Corpus-based studies that focus on grammatical aspects, that is, on syntactic units rather than on individual words, have provided empirical evidence on the interrelatedness of lexis (and meaning) and syntax, linguistic domains treated, in the past, as completely independent one from the other in both language description and language pedagogy. The detailed exploration of language through corpus-based methods has bolstered the view that language is highly patterned, and that the phrase, rather than the word, is "the normal primary carrier of meaning" (Sinclair, 2008, p.409).

Bearing in mind these two premises, various research strands have extensively explored the intertwining between Lexis (and Semantics) and Grammar: the Idiom Principle (Sinclair, 1991), Pattern Grammar (Hunston & Francis, 1999), Lexical Priming (Hoey, 2004, 2005), Lexical Bundles (Biber et al., 2004; Biber, 2006), Construction Grammar (Goldberg, 1995, 2006) and Collostructional Analysis (Stefanowitsch & Gries, 2003; Gries & Stefanowitsch, 2004)[43]. Despite differences in terminology, the above-mentioned approaches, making use of corpus-based methodology, mainly aim at integrating Grammar and Lexis as one same domain by systematically identifying the most common phraseological occurrences in English and in different English sub-languages.

The Pattern Grammar approach, for instance, describes all the patterns of English and the lexical items that "govern them" (Hunston, 2006, p.244). It is aimed at describing the English grammar by "exploiting the connection between phrase and meaning" (Hunston, 2006, p.243) through a notation that accounts for the colligational behaviour of lexical items, an exploration of how semantically-related lexical items share a pattern type, and a classification of lexical items according to their meaning. Pattern Grammar

[43] See Römer (2009) for a brief account on each of the major approaches to the interface between Lexis and Grammar.

claims that “the different senses of words will tend to be distinguished by different patterns (…) and particular patterns will tend to be associated with lexical items that have particular meanings” (Hunston & Francis, 1999, p.83), clearly showing that “there is a strong tendency for sense and syntax to be associated” (Sinclair, 1991, p.65).

Similar theoretical approaches to Pattern Grammar are, for example, Collostructional Analysis (Stefanowitsch & Gries, 2003; Gries & Stefanowitsch, 2004), which is aimed at statistically measuring the degree of association between words and the larger construction they tend to occur in; and Lexical Priming Theory (Hoey, 2004, 2005), which defends that words typically occur with specific collocations in specific textual environments and “that an individual’s experience of words in context, over many years, ‘primes’ each lexical item for use in a particular collocational or colligational configuration or for playing a particular role in a text” (Hunston, 2006, p.245).

All these approaches, in one way or another, explore “the pervasiveness of co-selection features and collocations” (Römer, 2009, p.148), a fact which asserts the patterned nature of language. Language is made up of a large proportion of formulaic (or phraseological) expressions, which are highly frequent and may present certain degree of fixedness. These formulaic expressions are combinations of words whose association is statistically significant. They tend to co-occur more often than other word combinations and can thus be quantitatively assessed through statistical methods (McEnery et al., 2006). Among many other advances, the use of Corpus Linguistics as a methodological tool has made it possible to describe frequent patterns of co-occurrence and assess the methods employed to identify them through corpora (Sinclair, 1991; Hunston & Francis, 1999; Gries, 2008).

Furthermore, it has been demonstrated that phraseology is register-specific, so each register makes use of its own typical word combinations (i.e. phraseological expressions in Academic English vary considerably from those of spoken English (Biber et al., 1999; Stubbs & Barth, 2003)). This means that the acquisition and understanding of the typical phraseological expressions used in one particular linguistic register are fundamental so as to effectively engage with the discourse community that uses that

register (Ellis et al., 2008; Vincent, 2013). Knowledge on appropriate multiword units is crucial for the production or understanding of a text, because the rhetorical progression as well as textual features such as cohesion and persuasion heavily depend on the proper use of recurrent co-occurrences or collocations (Bordet, 2013).

This is clearly the case of a scientific medical writer, who needs to master the conventions expected in his/her discourse community. The malleability and universality of Corpus Linguistics as a methodological approach to the study of language makes it a perfect tool for this thesis. Through a corpus-based approach, active and passive voice patterns occurring in the sub-language of scientific medical writing will be analysed quantitatively and qualitatively using the RCTs corpus, which is representative of medical discourse. Frequency of occurrence of active and passive voice patterns, their distribution in the different sections of the Research Article (specifically, Randomised Controlled Trials), and their collocational and colligational patterns will be explored so as to have insights on how the notion of voice and the interrelatedness with that of (im)personality is exploited in medical writing.

4.2. Description of the Randomised Controlled Trials (RCTs) Corpus: The Use of the CQPWeb and its Syntactic Languages

The Randomised Controlled Trials Corpus (henceforth, RCTs Corpus) is a collection of 1,485 medical Research Articles composed of 1,241,161 words, which are distributed across the several sections inherent in such a textual typology. These texts are specifically randomised controlled trials.

Randomised controlled trials (henceforth, RCTs) are clinical texts which aim at testing whether a treatment is more effective than another one through the comparison of two (or more) groups of participants: on the one hand, the control group, hence that they are labelled as *controlled*, and on the other hand, the experimental group. The former group of participants receives either the best existing treatment, no treatment at all, or a placebo; and the latter group receives the new treatment that is being tested. The

treatment given to each of the participants is randomly chosen, hence that they are also called *randomised*, because neither the researcher nor the participants have voice in such an election[44].

The articles that make up the RCTs corpus come from the most prestigious medical journals. Their impact factor (2018)[45] ranks them as the top-five journals in the medical scientific discipline (see Table 3).

Table 3. Top-five medical journals that compose the RCTs corpus, their abbreviation or acronym, and their position in the rank of medical journals, according to their impact factor (2018)

Journal	Title / Acronym	Rank	Impact factor
The New England Journal of Medicine[46]	NEW ENG J MED	1	55.873
The Lancet[47]	LANCET	2	45.217
Journal of American Medical Association[48]	JAMA J AM MED ASSOC	3	35.289
Annals of Internal Medicine[49]	ANN INTERN MED	4	17.81
The British Medical Journal[50]	BRIT MED J	5	17.445

The RCTs corpus is one of the six corpora that form the *Biomedicine and Health Sciences* set of corpora, available at www.cqpweb.bham.ac.uk (University of Birmingham)[51] through CQPweb (Corpus Query Processor Web), a web-based corpus

[44] [accessed in July 2019] www.ctu.mrc.ac.uk/about_clinical_trials/what_is_a_randomised_controlled_trial/

[45] [accessed in July 2019] https://impactfactor.weebly.com/medicine.html#

[46] [accessed in August 2019] http://www.nejm.org

[47] [accessed in August 2019] http://www.thelancet.com

[48] [accessed in August 2019] https://jamanetwork.com/journals/jama

[49] [accessed in August 2019] http://annals.org/aim

[50] [accessed in August 2019] https://www.bmj.com

[51] First, I was granted access to the RCTs corpus through http://cqpweb.wetware.ca, thanks to Dr. Neil Millar; however, the corpus got out of service without any notice. After doing some research, I found out that the same corpus had become available at the Centre for Corpus Research (CCR) at the University of

analysis system (Hardie, 2012). The server also gives access to the other five corpora, which are focused on other areas of Medicine (such as *Acupuncture*, *Midwifery and Perinatal Care*, *Chiropractic*, and *Osteopathy*) and on prospective meta-analysis (*Cochrane Meta Analyses*).

The metadata available on the RCTs corpus are shown in Table 4.

Table 4. RCTs corpus metadata

Metadata	RCTs Corpus
CQPweb's short handles for this corpus	rct / RCT
Total number of corpus texts	1,485
Total words in all corpus texts	1,241,461
Word types in the corpus	29,963
Type: token ratio	0.02 types per token

As for the text metadata, the corpus metadata database stores detailed information on each text collected in the corpus. When a given word or structure is searched for, the query returns a concrete number of occurrences. They are presented through key words in context (KWIC), and each occurrence is ranked and labelled by the filename or "text identifier used in the **<text>** tags of the input text" (Hardie, 2012, p.394). When double-clicked, the meta-information is presented in the form of a table, which includes the following:

a) A **filename**, which is a text identification label, consisting of a three-component code: **la356637a** → la[i] + 356637[ii] + a[iii]

   i) ***la*** refers to the abbreviation of the medical journal from which the RCT was extracted. The five medical journals are coded as follows: ***ne*** stands for *The New England Journal of Medicine*; ***la***, for *The Lancet*; ***ja*** for

Birmingham (UoB) (www.cqpweb.bham.ac.uk). Two years later, the RCTs corpus was withdrawn from the CCR's webpage, and no access was possible.

Journal of American Medical Association; ***an*** for *Annals of Internal Medicine*; and ***bm*** for *The British Medical Journal*.

ii) **356637** corresponds to the number given to the RCT where the node word (or node structure) occurs.

iii) ***a*** corresponds to the first letter of the section in which the node word (or node structure) searched for is found. The core sections are coded as follows: ***a*** stands for Abstract; ***i*** for Introduction; ***m*** for Materials and Methods; ***r*** for Results; and ***d*** for Discussion.

Therefore, the example provided – **la356637a** – shows that the node word (or node structure) searched for in the corpus occurs in the abstract section, ***a***, of the article number ***356637***, which has been published in *The Lancet*, ***la***.

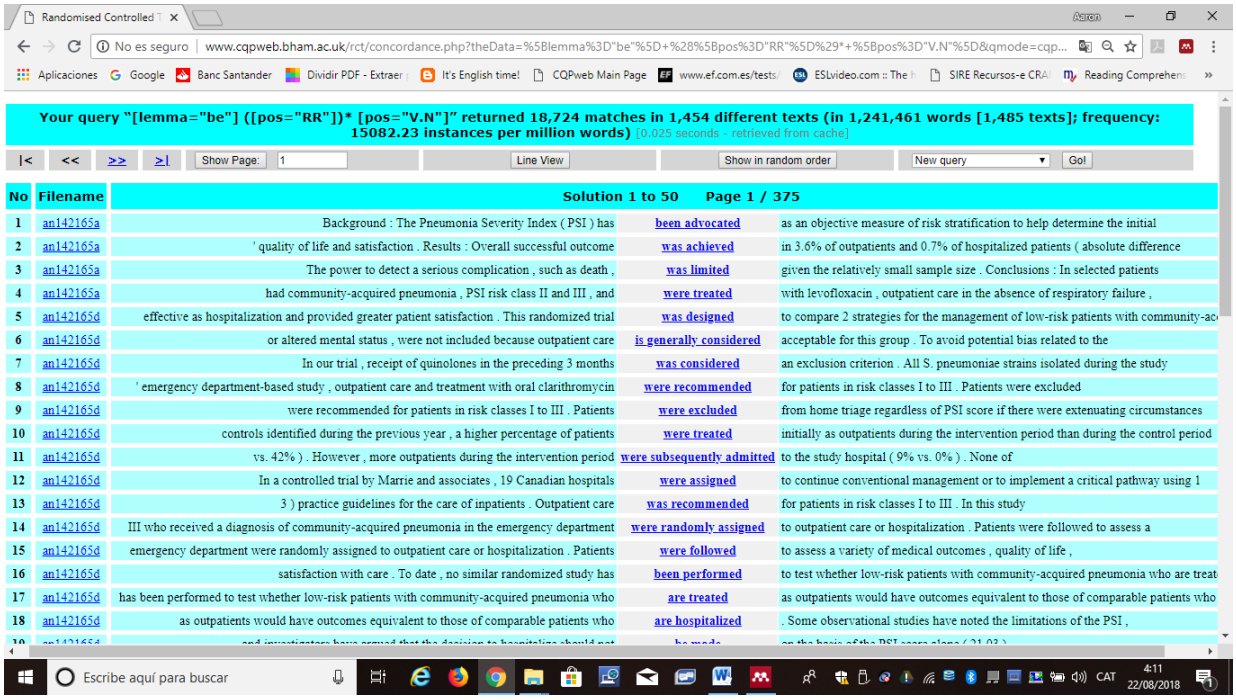
b) The **section** where the node word (or node structure) searched for occurs. Indeed, the primary classification of texts in the corpus is based on the section of the RA. In this case, the category section is represented through numbers from 1 to 5, which represent the order appearance of each section in the paper: **1** refers to the Abstract, **2** to the Introduction, **3** to the Materials and Methods, **4** to the Results, and **5** to the Discussion.

c) The complete name of the **medical journal** where the RA has been published, the **year of publication** of the article, and its **genre type**, which in this case corresponds to RCTs.

d) The list of all the **authors** and the ***title*** of the paper.

e) The digital object identification, known as **doi**; the URL or **link** to the article in .pdf format; the **reference**, including volume, issue, and pages; and the overall **number of words** in the text.

Figure 13. First concordance lines (KWIC) of the search [lemma="be"] ([pos="RR"])* [pos="V.N"] (any form of *be*, plus any number of optional adverbs (zero or more), plus a past participle participle form of any lexical verb (including *do* and *have*)).



Randomised Controlled T

No es seguro | www.cqpweb.bham.ac.uk/rct/concordance.php?theData=%5Blemma%3D"be"%5D+%28%5Bpos%3D"RR"%5D%29*+%5Bpos%3D"V.N"%5D&qmode=cqp...

Aplicaciones · Google · Banc Santander · Dividir PDF - Extraer · It's English time! · CQPweb Main Page · www.ef.com.es/tests/ · ESLvideo.com :: The · SIRE Recursos-e CRA · Reading Comprehen

Your query "[lemma="be"] ([pos="RR"])* [pos="V.N"]" returned 18,724 matches in 1,454 different texts (in 1,241,461 words [1,485 texts]; frequency: 15082.23 instances per million words) [0.025 seconds - retrieved from cache]

|< << >> >| Show Page: 1 Line View Show in random order New query Go!

No	Filename	Solution 1 to 50 Page 1 / 375		
1	an142165a	Background : The Pneumonia Severity Index (PSI) has	been advocated	as an objective measure of risk stratification to help determine the initial
2	an142165a	' quality of life and satisfaction . Results : Overall successful outcome	was achieved	in 3.6% of outpatients and 0.7% of hospitalized patients (absolute difference
3	an142165a	The power to detect a serious complication , such as death ,	was limited	given the relatively small sample size . Conclusions : In selected patients
4	an142165a	had community-acquired pneumonia , PSI risk class II and III , and	were treated	with levofloxacin , outpatient care in the absence of respiratory failure ,
5	an142165d	effective as hospitalization and provided greater patient satisfaction . This randomized trial	was designed	to compare 2 strategies for the management of low-risk patients with community-ac
6	an142165d	or altered mental status , were not included because outpatient care	is generally considered	acceptable for this group . To avoid potential bias related to the
7	an142165d	In our trial , receipt of quinolones in the preceding 3 months	was considered	an exclusion criterion . All S. pneumoniae strains isolated during the study
8	an142165d	' emergency department-based study , outpatient care and treatment with oral clarithromycin	were recommended	for patients in risk classes I to III . Patients were excluded
9	an142165d	were recommended for patients in risk classes I to III . Patients	were excluded	from home triage regardless of PSI score if there were extenuating circumstances
10	an142165d	controls identified during the previous year , a higher percentage of patients	were treated	initially as outpatients during the intervention period than during the control period
11	an142165d	vs. 42%) . However , more outpatients during the intervention period	were subsequently admitted	to the study hospital (9% vs. 0%) . None of
12	an142165d	In a controlled trial by Marrie and associates , 19 Canadian hospitals	were assigned	to continue conventional management or to implement a critical pathway using 1
13	an142165d	3) practice guidelines for the care of inpatients . Outpatient care	was recommended	for patients in risk classes I to III . In this study
14	an142165d	III who received a diagnosis of community-acquired pneumonia in the emergency department	were randomly assigned	to outpatient care or hospitalization . Patients were followed to assess a
15	an142165d	emergency department were randomly assigned to outpatient care or hospitalization . Patients	were followed	to assess a variety of medical outcomes , quality of life ,
16	an142165d	satisfaction with care . To date , no similar randomized study has	been performed	to test whether low-risk patients with community-acquired pneumonia who are treat
17	an142165d	has been performed to test whether low-risk patients with community-acquired pneumonia who	are treated	as outpatients would have outcomes equivalent to those of comparable patients who
18	an142165d	as outpatients would have outcomes equivalent to those of comparable patients who	are hospitalized	. Some observational studies have noted the limitations of the PSI ,

Escribe aquí para buscar

CAT 4:11 22/08/2018

The following Tables (5 and 6) are samples of the detailed information provided for each of the RCTs in the corpus.

Table 5. Sample of detailed information provided for each of the RCT, within the RCTs corpus

Text Identification Label	an142165a
Section	1 (Abstract) (an142165**a**)
Journal	Annals of Internal Medicine (**an**142165a)
Year	2005
Genre	RCT
Type	RCT
Authors	Carratala J, Fernandez-Sab N, Ortega L, Castellsague X, Roson B, Dorca J, Fernandez-Aguera A, Verdaguer R, Martinez J, Manresa F, Gudiol F
Title	Outpatient care compared with hospitalization for community-acquired pneumonia: a randomized trial in low-risk patients.
DOI[52]	Not found
URL[53]	http://www.annals.org/cgi/reprint/142/3/165.pdf
Reference	142(3): 165-172
No. of words	362

[52] DOI: Digital Object Identification.
[53] URL: Uniform Resource Locator.

Table 6. Sample of detailed information provided for each of the RCT, within the RCTs corpus

Text Identification Label	bm3301239m
Section	3 (Methods) (bm3301239**m**)
Journal	British Medical Journal (**bm**3301239m)
Year	2005
Genre	RCT
Type	RCT
Authors	Rivero-Arias O, Campbell H, Gray A, Fairbank J, Frost H, Wilson-Macdonald J
Title	Surgical stabilisation of the spine compared with a programme of intensive rehabilitation for the management of patients with chronic low back pain: cost utility analysis based on a randomised controlled trial
DOI	10.1136/bmj. F
URL	Not found
Reference	330(7502): 1239-1244
No. of words in the text	1136

It is worth noting that words in the RCTs corpus are annotated with lemma, part-of-speech tag, and Semantic Tag (USAS tagset)[54]; and the primary tagging scheme is Part-of-speech tag. The corpus is also parsed. All this information is included within the CQP Interface.

The CQPweb consists of a web-based graphical user interface whose design emulates the BNCweb interface to the British National Corpus (Lehmann et al. 2000; Hoffmann & Evert 2006; Hoffman et al. 2008). This user-friendly interface makes use of a database which allows the addition of any corpus, a fact which enhances its accessibility. Among the many functions and characteristics CQPweb has, it is worth highlighting the most relevant ones[55]:

[54] www.ucrel.lancs.ac.uk/usas/
[55] [accessed in April 2024] These features have been extracted and summarised from https://cwb.sourceforge.io/cqpweb.php

1) CQPweb uses the Common Elementary Query Language (henceforth, CEQL), which allows the simplification of complex syntactic queries. If one searches for the first-person pronoun, *we*, it is possible to make use of the underscore and the word (or tag attached to this category): either **we_** or **_PPIS2**. It also gives access to lemma and POS information.
2) It also uses the CQP Syntax. The search for the lemma *be* would be **[lemma="be"]**, for an any adverb **[pos="RR"]**, and for a particular word, in this case the word *by*, **[word="by"%c]**. %c is a flag which includes the retrieval of upper- and lower-case variants of the query.
3) All the queries are cached, so it facilitates to get back to the queries searched for in the past.
4) Queries can be sorted by adjacent words (or tags) and can also be thinned based on words (or tags) appearing at specific positions relative to the query 'hit'.
5) A quantitative summary of the collocations is given for every search. A range of collocation statistics (Z-score, Mutual Information, and Log-likelihood) appears; in fact, "[c]licking on any collocate creates a query containing just those hits from the original query that have the collocate in question in their context" (Hardie 2012, p.346).
6) The results for a query are distributed across the text categories in the corpus.
7) The frequency breakdown of a query "shows how often particular forms occur as the 'hit' word (and/or as the primary annotation of the 'hit word')" (Hardie 2012, p.346).
8) Several hundred words of context are given so that all queries are contextualised in the original text.
9) There is a database of information on each text (metadata).

Because of all these main characteristics, the RCTs corpus, included within the CQPweb, has proved to be an excellent tool for the retrieval of all the linguistic information presented in Chapter 5. As this dissertation deals with complex syntactic colligational patterns, the development of the syntactic expressions used to retrieve the patterns has become a challenge. Both the Simple Query Syntax and the CQP Syntax

have become crucial for the search of the patterns and their refined subsequent quantitative and qualitative analyses. It is true, however, that these two syntaxes have also posed some issues, which will be exposed in the following section.

4.3. Operationalisation of the Notions of Voice and Impersonality through a Corpus-Based Approach

Making use of the notions of voice and impersonality through a corpus-based study is somehow complex. If the notion of voice is solely restricted to its grammatical sense, one can say that voice refers to the form of a verb that indicates when a grammatical subject performs the action expressed by the verb (active voice) or when the grammatical subject receives that action and is acted upon (passive voice). Grammatical voice is then easy to identify because of the marked alignment of arguments, in which the patient performs the function of subject, the verb is composed of the finite verb form *be* in any tense, followed by a past participle form, and the agent is optionally expressed through a *by*-phrase.

Pragmatically speaking, the active voice may be considered as a more personal form if compared to the passive, because in an active sentence the doer of the action usually appears to be an animate entity and is often the first element presented in the clause. In a passive voice sentence, however, as the grammatical subject becomes the patient, usually inanimate, the resulting conception of the utterance seems to be more impersonal. The agent (doer of the action) becomes backgrounded (in agentive passives) and in most cases even omitted (in agentless passives). Because of this, the sense of impersonality attached to the passive voice structure seems well-justified.

Nevertheless, this is partially true. An active sentence can have an inanimate entity as the subject, performing a human(ised) action. Due to the inanimate nature of this subject, the whole utterance can be perceived as being more impersonal than a prototypical active sentence with an animate subject performing a particular (human) action, despite being both active sentences. This construction, in which the subject is an

inanimate entity and performs a human action by means of metonymy, is labelled, in this thesis, as an impersonal active sentence or, what is the same, as an abstract rhetor. Indeed, a passive voice including the agent (agentive passive) can be considered as being more personal than an impersonal active construction, because in the former the doer of the action is explicitly mentioned in the utterance, whereas in the latter the doer of the action is unnamed, although its inference may be obtained through the context. These two examples prove that there is no straightforward connection between *active voice* and the notion of *personality*, on the one hand, and *passive voice* and the notion of *impersonality*, on the other.

In addition, if a third parameter, such as authorial (in)visibility, is added onto the equation, the complexity of the matter increases. One of the pre-established thoughts on medical writing is that it is highly impersonal because of the extensive use of passive sentences. Medical authors are thought to overuse passive structures to avoid full responsibility for their claims, among many other different functions the passive seems to perform. However, in medical writing, when a passive includes the agent, this rarely refers to the authors of the RA. When authors wish to hide themselves, they make use of other linguistic resources, such as the agentless passive or the impersonal active, rather than the agentive passive.

The notion of grammatical voice, and its materialisation in linguistic instances, is clearly retrievable from the corpus used in this study. However, the notion of impersonality needs to be operationalised. This operationalisation process aims at measuring the fuzzy phenomenon of *impersonality*, not easily measurable, through the inference from other phenomena that can be easily identified, such as the notion of voice (because of the grammatical voice of the patterns) and that of authorial (in)visibility (because of the context and the general and linguistic knowledge one may have). Empirical observation through the analysis of different active and passive structures and their quantification in the RCTs corpus will allow the distinction, measurement, and understanding of the notion of impersonality in medical texts. Relying on grammatical voice as a distinctive criterion (and complementing the analysis

with contextualised instances to relate it to authorial (in)visibility) seems to be a rational and practical way to approach the notion of impersonality in medical discourse.

4.3.1. The Use of Syntactic Languages (Simple Query Syntax and CQP Syntax) and their Limitations to Account for Impersonality

The different active and passive structures that will be searched for and analysed in the corpus are as follows: active structures (namely, the **personal active pattern**, including the personal pronoun *we*; and the **impersonal active pattern** or abstract rhetor); finite passive structures (the **prototypical agentless and agentive passives**); non-finite passive structures (the ***as*-passive**, the **reporting passive**, and the **past participial reduced relative clause**); and the **existential *there*** pattern.

The description of these patterns and the way the notion of impersonality will be operationalised through these active and passive instances will involve the introduction of the two distinct types of syntax used to search for the queries, the Simple Query Syntax (henceforth, SQS), and the CQP Syntax (henceforth, CQPS). I will also account for the limitations imposed by these patterns and how I managed to solve them.

The **personal active pattern** (emphasis added) consists of the pronoun *we*, followed by any number of adverbs (zero or more), followed by any lexical verb (including *do* and *have*). The syntactic expressions to search for the various forms of this pattern differ from each other depending on the tense of the lexical verb and its polarity. For this reason, the personal active pattern requires as many different queries (or syntactic expressions) as tenses it shows. To describe this pattern, I will now only deal with its preterite and present simple sub-patterns, which have been found to be the most frequent. A detailed description of each sub-pattern including the rest of the tenses is fully provided in Chapter 5.

When in the preterite, the affirmative form of the personal active pattern is searched for using the queries **_PPIS2 (_{ADV})* _V?D** (SQS) and **[pos="PPIS2"]**

([pos="RR"])* [pos="V.D"] (CQPS). Both represent the pronoun *we*, followed by any number of adverbs (zero or more), followed by the past tense form of any lexical verb (including *do* and *have*). The negative past form, however, requires the auxiliary *did*, and this causes the whole query to change. To retrieve negative instances in the past, one must search for the pronoun *we*, followed by *did not*, which may be preceded or followed by any number of adverbs (zero or more), followed by the infinitive form of any lexical verb (including *do* and *have*), which results in **_PPIS2 (_{ADV})* _VDD _XX (_{ADV})* _V?I** (SQS) and **[pos="PPIS2"] ([pos="RR"])* [pos="VDD"] [pos="XX"] ([pos="RR"])* [pos="V.I"]** (CQPS).

The same occurs when dealing with the present simple tense. The description of the positive present pattern is the pronoun *we*, followed by any number of adverbs (zero or more), followed by the present tense of any lexical verb (including *do* and *have*). The corresponding syntactic expressions are as follows: **_PPIS2 (_{ADV})* _V?0** (SQS) and **[pos="PPIS2"] ([pos="RR"])* [pos="V.0"]** (CQPS). In the negative form, as the auxiliary *do (not)* is needed, the pattern totally changes into **_PPIS2 (_{ADV})* _VD0 _XX (_{ADV})* _V?I** (SQS) and **[pos="PPIS2"] ([pos="RR"])* [pos="VD0"] [pos="XX"] ([pos="RR"])* [pos="V.I"]** (CQPS). The description of this pattern would then be the pronoun *we*, followed by its negative auxiliary *do not*, which may be preceded or followed by any number of adverbs (zero or more), followed by the infinitive form of any lexical verb (including *do* and *have*).

Indeed, the first limitation appears with the present simple tense. When this pattern is used **_PPIS2 (_{ADV})* _V?0** (SQS) or **[pos="PPIS2"] ([pos="RR"])* [pos="V.0"]** (CQPS), the present perfect tense, like in ***we have** shown*, is also retrieved. In this case, I must restrict the query so that only lexical verbs appear, by substituting the tag **_V?0** for **_VV0** (SQS) and **[pos="V.0"]** for **[pos="VV0"]** (CQPS).

As seen in Table 7, the SQS requires the use of **{braces}** to retrieve the lemma, like in **{ADV}**. The use of **(parentheses)** allows the inclusion of other elements, like the **asterisk ***, showing optionality (hence the expression *zero or more* in the description). The tag **(_{ADV})*** would refer to any number of optional adverbs, including none.

CQPS does not require {braces}; it uses **[square brackets]** and uses elements like part-of-speech **[pos]**, word **[word]**, or lemma **[lemma]**. It also makes use of **"inverted commas"** to introduce each of the tags.

Both syntaxes use the UCREL CLAWS7 Tagset[56], like in the **PPIS2** tag for the pronoun *we*, **RR** for adverbs, **XX** for the particle *not*, etc. If these tags are in SQS, they must be preceded by an **_underscore symbol**. Furthermore, in SQS, the **? interrogation mark** serves as a non-restrictive element. In CQPS, it is the **. full stop**. **_V?I** (SQS) or **[pos="V.I"]** (CQPS) would imply the search for all types of infinitive verbs (including *do* and *have*). If the second element in the tag is **V** instead of **.** , like in **_VVI** or **[pos="VVI"]**, the query would restrict the retrieval to lexical verbs only.

Table 7. Sample of instances to describe the particularities of SQS and CQPS
(Past Simple, negative polarity)

_PPIS2 (_{ADV})* _VDD _XX (_{ADV})* _V?I (SQS)
[pos="PPIS2"] ([pos="RR"])* [pos="VDD"] [pos="XX"] ([pos="RR"])* [pos="V.I"] (CQPS).

In fact, when analysing the tag **_VDD** or **[pos="VDD"]**, it is possible to divide it into three distinctive parts: the first element refers to the category, and as it is a verb, it is represented with the **V**; the second element refers to the type of verb, in this case it is the auxiliary *do* **D** (if it were *have*, it would be **H**, for example), and the third element refers to the tense, in this case, the past simple **D** (if it were in the present tense, for example, it would include a **0**). Therefore, the tags **_VDD** or **[pos="VDD"]** refer to the auxiliary verb *did* in SQS and CQPS, respectively.

Keeping the analysis within the active grammatical voice, the other problem encountered appears when dealing with the **impersonal active voice** (emphasis added). Subjects must be inanimate, and explicitness on the inanimate nature of the noun having the function of subject is mandatory in this case. To my knowledge, even though the

[56] [accessed in July 2023] https://ucrel.lancs.ac.uk/claws7tags.html

corpus is semantically tagged, there is no automatic way to restrict the retrieval of only inanimate nouns. The syntactic expressions used to retrieve this pattern are **_NP? (_{ADV})* _V??** (SQS) or **[pos="N.."] ([pos="RR"])* [pos="V.."]** (CQPS), which consist of any singular or plural noun, followed by any number of optional adverbs, and an active verb in any tense. However, the non-*we* personal active structure, that is, the one including an animate subject as the performer of the action, like in ***other investigators*** *have shown that (…)*, is subsumed within this pattern.

Apart from this issue, another problem arose. Most of the cases happened to be finite forms in the past or present simple tenses. However, some non-finite instances were retrieved when searching for this query, because of the use of the tag **_V??** (SQS) and **[pos="V.."]** (CQPS) (like in *patients* ***receiving***). A solution was to only search for those patterns in the preterite tense, which was shown to be the one extensively used, through the syntactic expressions: **_NP? (_{ADV})* _V?D** (SQS) or **[pos="N.."] ([pos="RR"])* [pos="V.D"]** (CQPS). However, even though the problem with the finiteness of the verb form was sorted, the problem with animacy continued. This is why I opted for restricting the query to two specific inanimate subjects frequently used: *study* (*studies*) and *trial* (*trials*). Obviously, many other instances appearing in this pattern (like *figure X* or *table X*, among many others) could not be retrieved, and thus, could not be analysed either. Nevertheless, the results of the searches made could be extended so that the overall understanding of this pattern is shown.

As for the **prototypical passive voice** structures (emphasis added), as it is a colligational structure subject to variability, different syntactic expressions may be needed. Here, a distinction must be made between agentless and agentive prototypical passives. Because of the complexity of these two sub-patterns, particularly that of the agentive passive, I only made use of CQP syntax.

The **agentless prototypical passive voice** (emphasis added) can be described as any form of the verb *be* (lemma); optionally followed by any number of adverbs (zero or more); optionally followed by the negative particle *not* and by any number of adverbs (zero or more); and followed by the past participle form of any lexical verb (including

do and *have*). The query used is then: **[lemma="be"] ([pos="RR"])* ([word="not"%c])* ([pos="RR"])* [pos="V.N"]**.

For the search of the agentive passive, information on the agent must be added as an extension to the agentless pattern (in blue): **[lemma="be"] ([pos="RR"])* ([word="not"%c])* ([pos="RR"])* [pos="V.N"] [word="by"%c] (([pos="AT"])* ([pos="JJ"])*)? [pos="N.."]**. The part in blue, the one including the agent, consists of the preposition *by*, followed by a noun (either singular or plural), which may be optionally preceded by an article and any number of adjectives (zero or more). This structure introduces the optionality of different linguistic elements which may appear in combination or separately. Hence the inclusion of outer **(parentheses)** to include all the linguistic elements within the expression, and the inclusion of inner **(parentheses)** to refer to each of the POS tags and to allow the optionality, marked through * and **?**, of the tag: **(([pos="AT"])* ([pos="JJ"])*)?**.

One problem encountered is that the agentive prototypical passive query is subsumed within the agentless prototypical passive syntactic expression, so that the query of the former is included within the query of the latter. This can be solved by searching for the most restrictive one: the agentive prototypical passive. By explicitly searching for the agent, agentless prototypical passives seem not to appear and, although it is a laborious task, it seems possible to account for the two different types of prototypical passives.

However, two other considerations must be made in this respect. First, nominalisations and stacked nominal phrases appear within the *by*-phrase, so agency is hidden within nominalised expressions derived from verbal processes (where the agent can easily be inferred by means of metonymy, like in *is suggested **by previous studies***) or within stacked nominal phrases (where agency is somehow denoted, also by means of metonymy, like in *was approved **by the institutional review boards***). The discrimination between explicit human agency, inference of human agency, or absolute absence of human agency is important to delimit the degree of impersonality extended by the pattern itself. Second, a high number of instances including a *by*-phrase are formed of *by* plus present participle verb forms, which means that the *by*-phrase is not

introducing the agent of the action but is giving information on the way the action expressed by the verb is made.

By searching for the query **[lemma="be"] ([pos="RR"])* ([word="not"%c])* ([pos="RR"])* [pos="V.N"] [word="by"%c] ([pos="RR"])* [pos="VVG"]**, adjuncts of manner are then included. However, in these cases, agency is omitted, and the resulting prototypical passive must then be considered as agentless, and not as agentive, despite the inclusion of a *by*-phrase. In some instances, the present participle form following *by* and expressing manner gives a hint on the performer of the action (like in *were performed **by using***). On some other cases, the preposition *by* is followed by a nominalisation (like in *was defined **by hospitalisation***), or a stacked nominal phrase (like in *were analysed **by Wilcoxon rank sum tests***), in which information on manner is also transmitted.

The same limitations observed in the prototypical passive voice, as far as the retrieval of patterns is concerned, are also encountered when retrieving instances of the ***as*-passive voice** (emphasis added), so the same information described as for the prototypical passive can be applied to the non-finite *as*-passive construction.

The other structure analysed is the **reporting passive** (emphasis added). A reporting passive structure consists of the personal pronoun *it*, optionally followed by any number of adverbs (zero or more), followed by the lemma *be*, optionally followed by any number of adverbs (zero or more); followed by the past participial form of any lexical verb. The corresponding query is as follows: **[word="it"%c] ([pos="RR"])* [lemma="be"] ([pos="RR"])* [pos="V.N”]**. To be considered as a reporting (impersonal) passive, the personal pronoun *it* must have no referent within the previous context. In case it had a referent, it would describe a prototypical passive structure (like in ***It is associated** with*). In order to restrict the query, the word form *that* must be added, making the pattern result in **[word="it"%c] ([pos="RR"])* [lemma="be"] ([pos="RR"])* [pos="V.N”] [pos="that"%c]**. By the refinement of the query through the addition of a *that*-clause, the resulting structure is that of the reporting passive.

In addition, the description of the non-finite **past participial reduced relative clause** (emphasis added) consists of a noun (either singular or plural), optionally preceded by an article and any optional number of adjectives (zero or more); optionally followed by any number of adverbs (zero or more); followed by the past participle form of any lexical verb (including non-auxiliaries *be*, *do*, and *have*). It is searched through the following pattern: **((_AT)* (_JJ)*)? _N?? (_RR)* _V?N** (SQS) or **(([pos="AT"])* ([pos="JJ"])*)? [pos="N.."] ([pos="RR"])* [pos="V.N"]** (CQPS).

This query also encounters a problem when being retrieved. Because of some inaccuracies in the POS-tagging, preterite forms may appear instead of past participle forms, particularly when both forms show syncretism. This causes some active sentences to be retrieved when this query is searched for (like in ***researchers recruited*** *participants*), which is not an example of a noun being post-modified by a past participial reduced relative clause, but an example of an active sentence with SVO structure. Another problem related to this pattern has to do with the form of some post-positive adjectives, which were included instead of past participle forms (like in *the percentage of* ***outpatients satisfied*** *with*). As for the inclusion of the agent, the same limitation as the one described for the agentive prototypical passive was encountered.

The **existential *there* pattern** (emphasis added) posed no limitations, as the syntactic expression used is exactly the combination of the expletive subject *there* followed by any form of *be*, preceded by any optional adverb (zero or more). The query used to search for this pattern is **[word="there"%c] ([pos="RR"])* [lemma="be"]**.

Despite the obvious limitations shown in the retrieval of these patterns, I have tried to provide, to my view, reasonable solutions to each of the problems encountered. In fact, these issues have also fostered the design and the execution of a close textual analysis, so that these nuances could be contextually examined. In the corpus-based study presented in Chapter 5, not only has the complexity of the syntaxes been a challenge, but also the variability that most linguistic patterns show, even within a constrained framework such as medical writing.

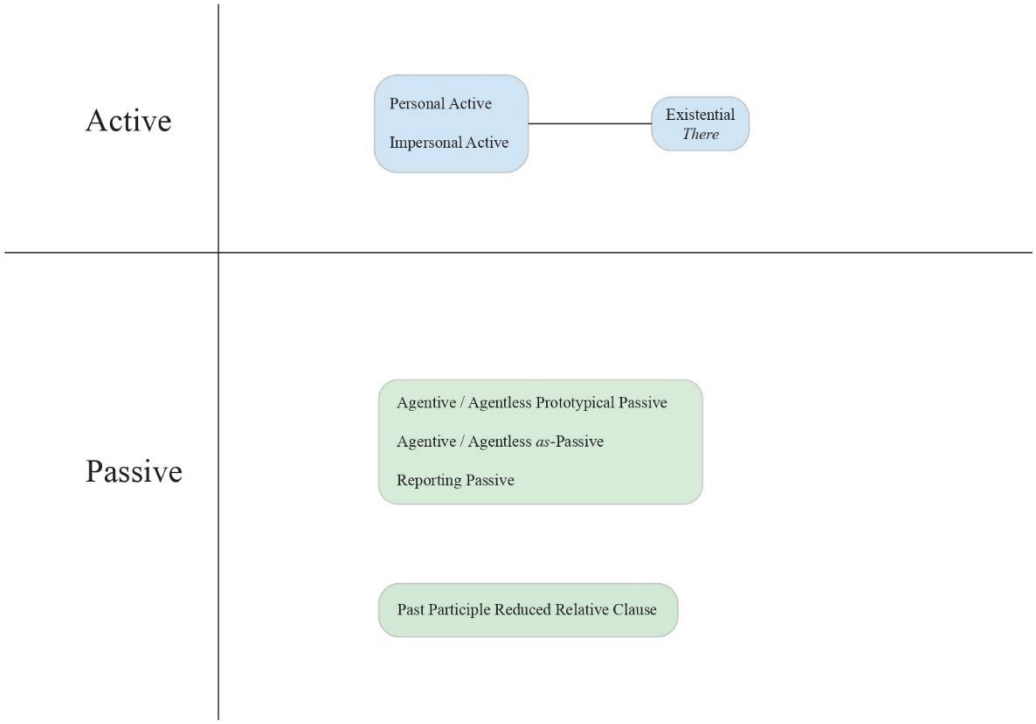
4.3.2. Representations of Grammatical Voice and Impersonality within a Cartesian Coordinate System

Once having described the different patterns used in Chapter 5, and once having considered the limitations imposed by SQS and CQPS, it is worth highlighting the pictorial representation of the information presented above. The first step will consist of only taking into consideration the grammatical voice as the main distinctive criterion to classify all the patterns analysed in the corpus-based study, leaving aside the notions of impersonality and authorial (in)visibility as part of the defining parameters of the structures.

As seen in Figure 14, the grammatical active voice includes the personal active and the impersonal active pattern. I will consider existential *there* structures as active, as the expletive *there* performs the function of grammatical subject. However, this pattern is placed in isolation as it is considered extremely impersonal if compared to the rest of the patterns dealt with, and it deserves a special treatment.

The passive, on the other hand, includes the agentive and agentless prototypical passives, the agentive and agentless *as*-passives, as well as the reporting passive. In addition, the past participial reduced relative clause is placed within the passive category, as it can be easily paraphrased as a prototypical passive structure. It is placed separately from the rest as it shows certain degree of mobility, since it can be embedded within other types of structures.

Figure 14. Types of structures analysed in the corpus-based study



If the notion of impersonality is added onto Figure 14, a bidimensional figure is optimal to show the positioning of all the structures. Through a Cartesian coordinate system, divided into four distinct Quadrants, I will attempt to show how the operationalisation of the notion of impersonality is possible through a corpus-based study, which will be afterwards complemented and refined by a close textual analysis.

The Cartesian coordinate system, shown in Figure 15, is defined by an ordered pair of perpendicular axes: x (the abscissa, in green) and y (the ordinate, in purple). In order to label the axes, I will make use of the terms *Personality* and *Voice*. The notion of Personality will be represented by the x-axis (the abscissa, in green), horizontally and oriented to the right, whereas the notion of (grammatical) Voice will be represented by the y-axis (the ordinate, in purple), vertically and oriented upwards.

The x-axis is considered to be a gradual cline from the notion *impersonality* (-x) to the notion *personality* (x), represented by the green line (the abscissa). In the figure, this

gradation is shown twofold: first, through the discontinuous black arrow going from one notion to the other; second, through the purple discontinuous line. It is worth highlighting, however, that the purple discontinuous line does not represent the x-axis Personality, but the y-axis Voice (the ordinate), hence that it is coloured in purple. The purple line is marked as discontinuous only to visually show the fuzziness of the two concepts forming the x-axis Personality.

On the other hand, the y-axis Voice corresponds to the notion of grammatical voice, either *active voice* (y) or *passive voice* (-y). As there is no possibility to show a gradation of the notion of voice (as a verb form is either active or passive[57] in English) the green continuous line visually serves this purpose. However, again, one must bear in mind that the green continuous line does not represent the y-axis Voice, but the x-axis Personality (the abscissa), hence that it is coloured in green. The green line is marked as continuous only to visualise the definiteness of the two concepts forming the y-axis Voice.

Once the Cartesian coordinate system is designed, it is worth noting that I will attempt to locate in the plane all the patterns dealt with in Chapter 5.

Indeed, it is pivotal to describe the four different Quadrants that form the Cartesian coordinate system. The I Quadrant (x, y) would correspond to personal active structures, the II Quadrant (-x, y) would correspond to impersonal active structures, the III Quadrant (-x, -y) would include impersonal passive structures, and finally, the IV Quadrant (x, -y) would include personal passive structures.

[57] It may be passive if the verb is transitive.

Figure 15. Cartesian coordinate system used to represent active and passive forms according to their connection to impersonality

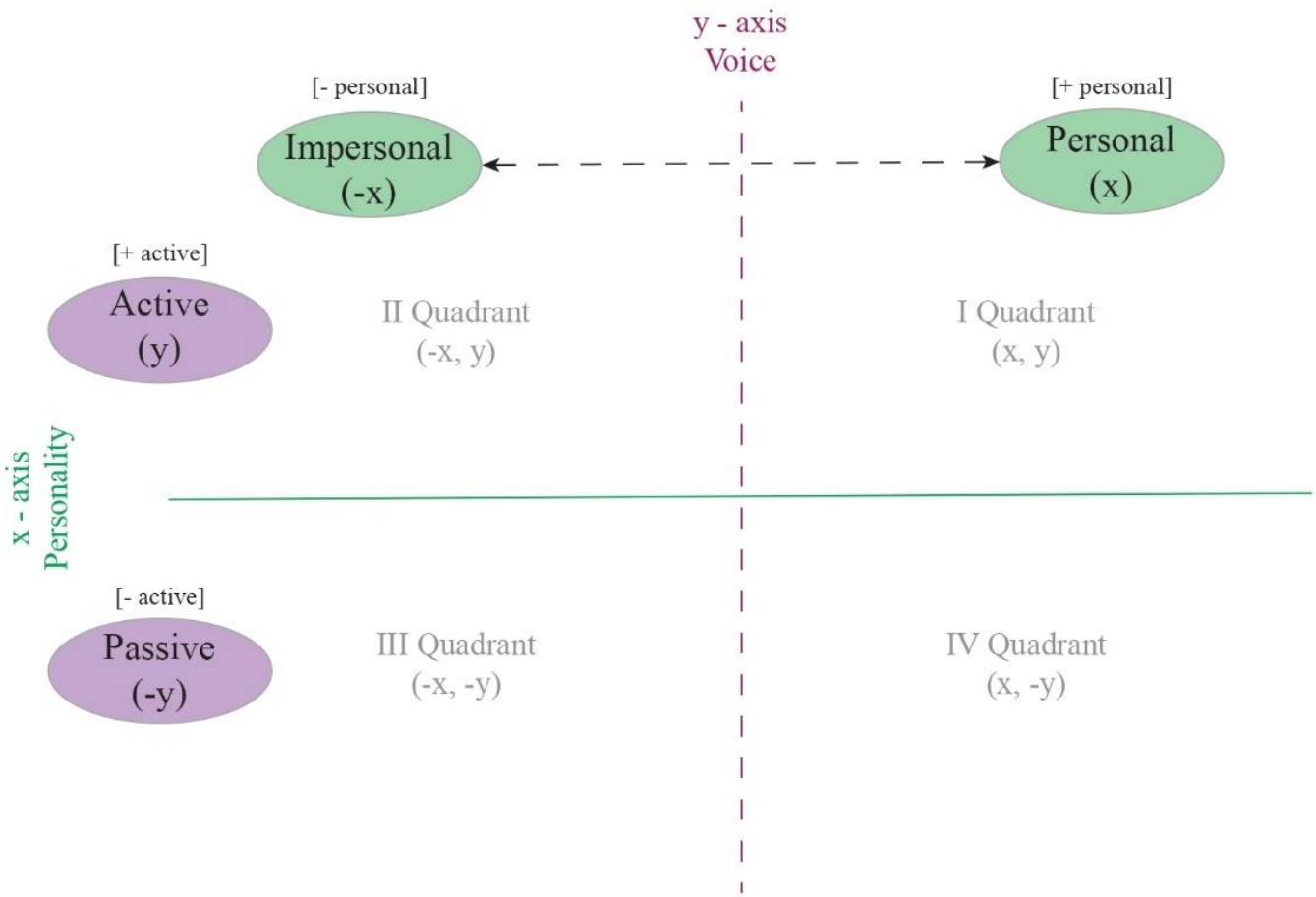


Table 8 shows the different Quadrants with their corresponding literal parameters (those referring only to (+/-) *personality* and (+/-) *active*)[58]. These literal parameters are translated into the actual dichotomised parameters that I will be using throughout this dissertation. These are *personal* vs. *impersonal* and *active voice* vs. *passive voice*.

Table 8. Quadrants of the Cartesian coordinate system, and their corresponding axes described through literal and used parameters – [(x) corresponds to Personality; (y) corresponds to Voice]

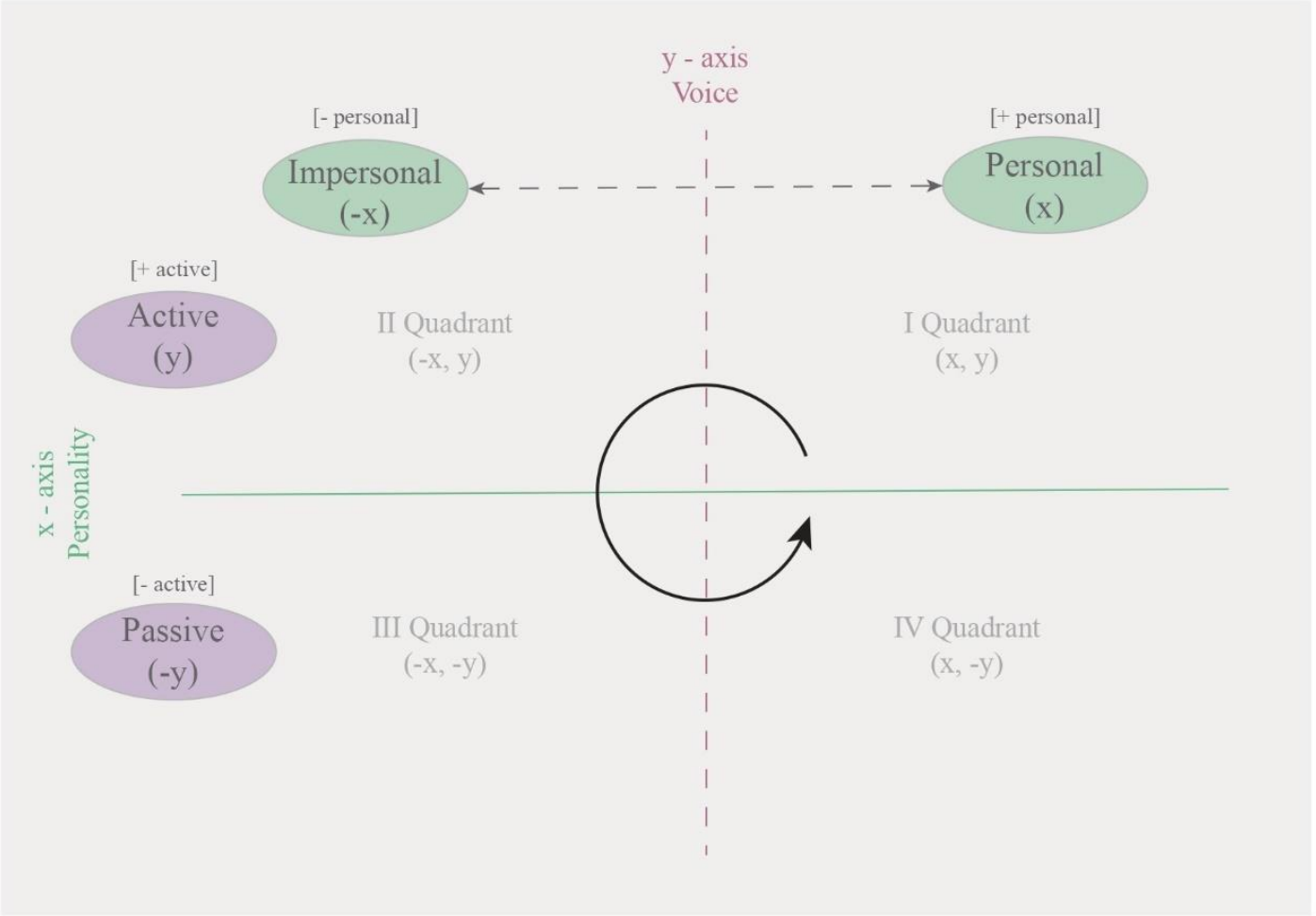
Quadrants	Axes	Literal Parameters	Used Parameters
I Quadrant	(x, y)	(personal, active)	(personal, active)
II Quadrant	(-x, y)	(-personal, active)	(impersonal, active)
III Quadrant	(-x, -y)	(-personal, -active)	(impersonal, passive)
IV Quadrant	(x, -y)	(personal, -active)	(personal, passive)

As seen in Figure 16, through the use of the Cartesian coordinate system it is possible to see a circular path (instead of a linear one, like in a continuum), first from personal active structures (shown in I Quadrant) to more impersonal active structures (shown in II Quadrant), to more impersonal passive structures (shown in III Quadrant), to more personal passive structures (shown in IV Quadrant), back to more personal active structures (shown, again, in I Quadrant).

Impersonal structures are then shown in II and III Quadrants, personal constructions are shown in I and IV Quadrants, active constructions are represented within I and II Quadrants, and lastly, passive constructions are included within III and IV Quadrants.

[58] As for the literal parameters, also detailed in Figure 15, it is worth noting that I make use of the unmarked labels: *personality* (instead of impersonality) and *active* (instead of passive). To my view, it is much easier to understand the concepts [+personality] or [-personality], rather than the concepts [+impersonality] or [-impersonality], the latter meaning [+personality] because of the double negation.

Figure 16. Cartesian coordinate system used, showing its circular path from I Quadrant, to II Quadrant, to III Quadrant, to IV Quadrant, and back to I Quadrant



The Cartesian coordinate system serves, in this case, to represent very well the circularity attached to the notion of (im)personality, bearing in mind this parameter together with that of grammatical voice. One can always keep one of the parameters and change the other one, and this would allow the circularity to gradually flow from one Quadrant to the next one (counterclockwise). To put it differently, a personal active pattern (I Quadrant), for example, can be made impersonal by changing the animacy of the subject keeping the same active pattern (by changing the feature *personality* and keeping the feature *voice*). Following this argument, an impersonal passive (agentless) structure can be made more personal by the addition of a *by*-phrase including the agent. In this case, the feature *voice* would be kept, as both are in the passive voice, but the feature *personality* would be modified.

This circularity, however, does not literally mean there is a continuum showing structures from I Quadrant, to II Quadrant, to III Quadrant and to IV Quadrant to represent the notion of impersonality. The x-axis Personality is what really shows the continuum of (im)personality, as structures may be located in the plane depending on their degree of (im)personality. In contrast, the y-axis only serves to categorise the structures appearing in the Quadrants as either belonging to the active or to the passive voice.

Bearing this in mind, one can assess that the impersonal distinctive feature shown in II and III Quadrants, on the one hand, and the personal distinctive feature shown in I and IV Quadrants, on the other, will allow the gradual classification of the patterns analysed in the study. Specific information on the representation of the patterns under discussion will be given in the concluding remarks of Chapter 5.

4.4. The Rationale for Using a Close Textual Analysis Approach as a Complementation to the Corpus-Based Study

Even though the corpus-based study described in Chapter 5 focuses on the analysis of active and passive voice structures, which are quantifiable elements retrievable from the different RCTs conforming the RCTs corpus, if one attempts to analyse the extent to which these linguistic choices are connected to the notions of impersonality and authorial (in)visibility, a corpus-based methodology may not be enough to delve into such connection. A corpus-based methodology surely uncovers significant and interesting results, because it provides real instances of linguistic choices, in this case, made by medical writers when transmitting their research. However, and in line with what Hunston (2002, p.23) considers, the "absence of visual and social context for the interpretation of concordance lines" may be put forward as an argument for the implementation of a different typology of study to supplement a corpus-based study.

In this same sense, Widdowson (2002) asserts that corpus data are somehow dissociated from the communicative context in which language happens to occur, particularly with large-scale corpora (as is the case of the RCTs corpus), so framing the retrievable and quantifiable information extracted from the corpus within a specific context, in which other linguistic and extra-linguistic features co-occur, may give the analyst a more precise and accurate perception of the reality of the specific items aimed to be analysed. According to Flowerdew (2004, p.324), "this lack of contextual features is particularly problematic for the corpus analyst when dealing with pragmatic features of text, which may only be recoverable from the socio-cultural context", pragmatic features such as (im)personality and authorial (in)visibility. This dissertation sees this as the main argument to complement the corpus-based study with a close textual analysis.

While combining quantitative and qualitative approaches, a more comprehensive understanding of the use of certain linguistic strategies can be achieved through a close textual analysis, as it is aimed at scrutinising individual instances so that the analyst can understand specific contextualised nuances and infer and describe the logic of these occurrences in the text, essentially basing the analysis on the contextual, pragmatic, and

semantic reasons behind these choices. A close reading also allows for the interpretation of the patterns aimed at uncovering particular meanings and intentions behind their use.

In the case of this thesis, the notions of (im)personality and authorial (in)visibility, whether connected or not to the notion of grammatical voice, need to be evaluated through the interpretation of real examples framed within a specific context, as these two notions cannot be limited to the presence or absence of one concrete linguistic element. The concept of grammatical voice is the starting point from which (im)personality and authorial (in)visibility can be tackled. The analysis of specific lexicogrammatical patterns as far as their degree of impersonalisation and authorial manifestation are concerned will provide a holistic understanding of the language used in medical writing as far as the dimension of impersonality is concerned.

Indeed, one can say that one particular linguistic choice is personal or impersonal because of authorial presence or absence, or because of the use of an active verb form instead of a passive verb form, for instance, but in order to deal with impersonality as a whole, one must bear in mind the specific contextualisation from a social and from an epistemological perspective. Not only does contextualisation play a crucial role, but also the interplay of different strategies which may contribute to the personalised or impersonalised character of a text. The cumulative effect some features seem to show has a clear effect on whether the text is perceived as personal or as impersonal. The text-based study presented in Chapter 6 aims at providing this contextualisation and is intended to show the interplay of features, without devaluing the results obtained through a corpus-based methodology.

4.5. Operationalisation of the Notions of Impersonality and Authorial Visibility through a Close Textual Approach: Representations within a Continuum

As said earlier, the notion of grammatical voice and its representation in language through specific lexicogrammatical instances is clearly identifiable, essentially because of the verb form (either active or passive), and because of the alignment of arguments within the clausal construction. However, the existence of other structures which directly relate to the phrasal level, and which have not been (fully) considered in the corpus-based study, are seen to also have an impersonalising effect within the whole textual configuration, as it is the case of the impersonal active structure or abstract rhetor, which consists of an inanimate subject followed by an active verb by means of a metonymic expression.

In this case, apart from operationalising the notion of impersonality, I will include in this operationalisation process the notion of authorial (in)visibility. Through the close scrutiny of the text, and also accounting for their raw and normalised frequency, different linguistic choices will be selected and contextually analysed. The aims of this analysis are twofold: on the one hand, the identification and classification of structures within a continuum, bearing in mind two dichotomised parameters, *personality* vs. *impersonality*, and *authorial visibility* vs. *authorial invisibility*, which will be presented at the poles of the continuum; on the other hand, the interrelatedness between the different structures by means of partially overlapping boxes. Each box will include a specific pattern, and the overlapping of boxes will show how these structures interconnect one with the other in a wide range of examples.

The structures that will be described are as follows: the personal active (*we* + active verb), **the non-*we* personal active**, the agentive and the agentless passives, **stacked nominal phrases**, **abstract rhetors** or **impersonal active constructions**, **nominalisations**, existential *there* patterns, and **Present** and Past **Participial Reduced Relative Clauses**. The ones marked in boldface were not included in the corpus-based study but are included in the close textual analysis, as they are relatively frequent in the

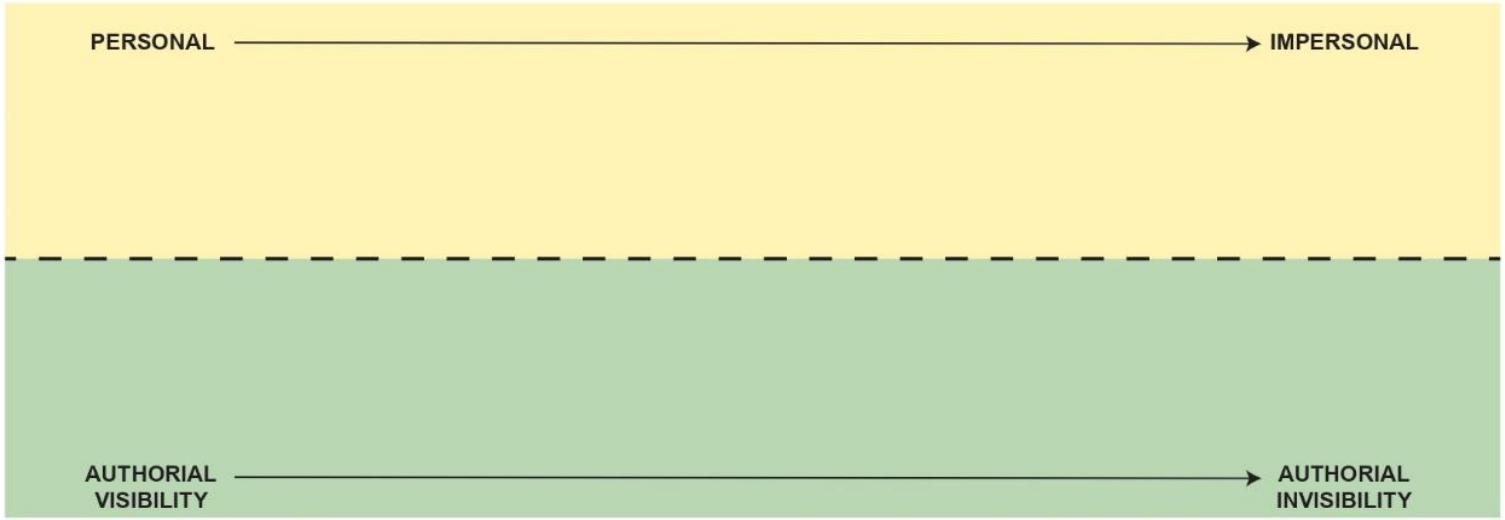
samples of RCTs analysed. Some linguistic choices, despite not being very frequent, are included to show contrast because of the high frequency of their counterparts, like in the case of non-*we* active sentences, which are included to show contrast with their *we*-active counterpart.

As seen in Figure 17, the different linguistic realisations will be placed within a continuum, which is formed of two equally shaped rectangular dimensions, one representing the dichotomy *personality* vs. *impersonality*, and the other representing the dichotomy *authorial visibility* vs. *authorial invisibility*. The former is coloured in yellow, whereas the latter is coloured in green. Both dimensions are divided through a black discontinuous line, showing the fuzziness of the concepts.

The left pole, represented by the features *personality* and *authorial visibility* will include structures in which the agent is promoted and in which the negotiability presented is maximal. The right end of the continuum, represented by the features of *impersonality* and *authorial invisibility*, will include lexicogrammatical constructions which denote total demotion of the agent, and which express no negotiability at all. Those constructions placed in the middle areas of the continuum will express partial or almost total demotion of the Agent, and partial levels of negotiability.

As said, each lexicogrammatical pattern will be included within a box, and the box will be placed within either the yellow dimension (showing the degree of (im)personality) or the green dimension (showing the degree of authorial (in)visibility). Structures that may show both pragmatic notions will be represented on the discontinuous line, showing they are instances representing both (im)personality and authorial (in)visibility. Apart from this, as already mentioned, when structures tend to co-occur, they are shown to partially overlap. As the notion of grammatical voice cannot be included as a third systematic parameter within the continuum, I will make use of colours to identify whether the structure belongs to either the active or passive voice. Not coloured boxes will contain structures that are verbless, and no information on grammatical voice can be given.

Figure 17. Continuum from more personal to more impersonal linguistic choices, and from more authorial visibility to more authorial visibility



Promotion of Agent	Partial demotion of Agent through backgrounding	Almost total demotion of Agent	Total demotion of Agent
Maximal negotiability	Minor degree of negotiability	Minimal negotiability	No negotiability

4.6. Summing Up

This chapter has described the methodological approach taken in the two studies presented in this dissertation: the corpus-based study (Chapter 5), on the one hand; and the close textual analysis (Chapter 6), on the other.

As described, Chapter 5 uses a corpus-based methodology to retrieve and analyse real instances extracted from the RCTs corpus. The main distinctive criterion to account for the set of structures analysed has been the notion of grammatical voice, so impersonality has been essentially operationalised through the distinction between active and passive voice structures. The virtual development of real linguistic strategies has been achieved through two different syntaxes (SQS and CQPS) and this has enabled the retrieval of a large number of instances that have been quantified and qualitatively analysed as regards their degree of impersonalisation. Therefore, bearing in mind the notions of grammatical voice and impersonality, a Cartesian coordinate system has been thought as a good empirical method to classify the different structures found in RCTs.

The notion of authorial (in)visibility, despite being referred to within the qualitative analysis of the structures, has been left aside as a primary classifying parameter because of the difficulty encountered when adding it onto the Cartesian coordinate system, which is an aspect that will be dealt with in the final conclusions to this dissertation. For this reason, and also because of the inherent complexity found in the notion of impersonality, Chapter 6 aims at identifying the linguistic elements seen in the corpus-based study, including others which would not actually fit within the *active* vs. *passive* system. A close textual analysis is developed to quantify and qualitatively analyse the various structures in their natural context, structures through which the notions of (im)personality and authorial (in)visibility have been operationalised.

Once the distinct patterns have been analysed, a continuum including two distinct but related dimensions has been designed. As these dimensions represent (im)personality and authorial (in)visibility, grammatical voice acquires a secondary role in the

exploration of the structures, blurring the line between active and passive constructions in favour of a more pragmatic dichotomy: that of personal and impersonal structures.

In my judgement, despite the many difficulties encountered, particularly when dealing with the retrieval of patterns in the corpus-based study and the impossibility to access the corpus to reconfirm the data previously obtained, results in Chapters 5 and 6 are conclusive and add new insights to the analyses of the impersonal dimensions of medical writing. The different methodologies used, as well as their operationalisations through visual representations, such as the Cartesian coordinate system and the two-parameter continuum, are aspects of novelty worth highlighting.

5

A Corpus-Based Study on the Use of Active and Passive Voice Structures in Randomised Controlled Trials:

Quantitative and Qualitative Analyses

The only thing we have power over
is our own thoughts.
René Descartes

5.1. Introduction to the Corpus-Based Study and its Main Outcomes

The corpus used to carry out the corpus-based study presented in this chapter is the *Randomised Controlled Trials Corpus* (RCTs). As already described, it is a collection of 1,485 randomised controlled trials extracted from the top five medical journals: *The New England Journal of Medicine*, *The Lancet*, *Journal of American Medical Association*, *Annals of Internal Medicine*, and *The British Medical Journal*.

The study will begin by stating the main outcomes. Then, I will provide the identification of each of the patterns explored and will include their description as well as the syntactic expressions used to search for them in the corpus. Once the active and passive patterns have been described, I will account for the quantitative information found in the RCTs and will qualitatively analyse them, illustrating them with contextualised examples. The first set of patterns will be the active ones. I will start the analysis with the personal active structure (*we* + verb in the active voice), and I will follow with the impersonal active structure or abstract rhetor.

Once the active structures have been analysed, I will explore the passive constructions, beginning with the most relevant one, the prototypical passive. The analysis will follow with the non-finite *as*-passive, the reporting passive, and the past participial reduced relative clause. After dealing with finite and non-finite passives, important information on the existential *there* structure will also be provided.

To begin with, it is essential to state the main objectives of this corpus-based study. They are presented as follows:

1) To quantify the active and passive structures in the RCTs corpus through Simple Query Syntax and CQP Syntax, according to their distribution in Randomised Controlled Trials (Research Articles).
2) To explore and qualitatively analyse the various active and passive structures.
3) To semantically classify the typology of verbs associated with active and passive structures, whenever possible.

4) To explore if there is a link between the semantics of the verbs and the choice of tense with the grammatical voice, whenever possible.
5) To relate to what extent active and passive structures are linked to authorial (in)visibility and (im)personality.
6) To account for the limitations encountered when relating active and passive structures to the notions of (im)personality and authorial (in)visibility through a corpus.
7) To propose a way to graphically represent the interrelatedness between the notion of voice and those of (im)personality and authorial (in)visibility.

An important aspect to bear in mind is that I will present the results on a par with their discussion to optimally explore and describe the patterns analysed and their contextualisation within the RCTs. Illustrating this set of patterns with examples will also allow an in-depth qualitative analysis. Once all the patterns have been dealt with, I will summarise the main findings and address to what extent these structures are linked to (im)personality and authorial (in)visibility. I will also highlight the main limitations encountered when dealing with a corpus-based approach to account for these two notions.

5.2. Identification, Description, and Quantitative and Qualitative Analyses of a Set of Active and Passive Voice Structures.

This corpus-based study is essentially aimed at quantifying the different active and passive structures found in the RCTs corpus, and at exploring and qualitatively analysing them in relation to their connection to (im)personality and authorial (in)visibility. As for the quantitative analysis, both raw and relative frequencies will be considered. Exploring the overall number of occurrences in which each pattern appears in the RCTs corpus as well as in each of the sections conforming the RAs is crucial to determine the preponderance of one specific pattern over the rest. It is then important to state the overall number of words in the RCTs corpus, and the distribution of these words within the different sections of the RAs. As shown in Table 9, the corpus is

formed of 1,241,461 words, which, according to the section of the RA, are distributed as follows: Abstract (104,092 words), Introduction (99,009 words), Methods (413,717 words), Results (316,837 words), and Discussion (307,806 words).

Table 9. Total number of words according to each section of the RA, and overall number of words in the RCTs corpus

Section of the RA	Total number of words
Abstract	104,092
Introduction	99,009
Methods	413,717
Results	316,837
Discussion	307,806
TOTAL	1,241,461

I will also account for the semantic typology of verbs appearing in the various patterns, and check if there is a connection between the lemmas used, the choice of tense, and the different active and passive patterns dealt with in this study.

Indeed, the notion of grammatical voice will serve as the main distinctive criterion to categorise all the structures as belonging to either active or passive. As already mentioned, as for the active voice structures, I will explore the personal active pattern, essentially the one formed by the personal pronoun *we* followed by an active verb; and the impersonal active pattern or abstract rhetor, in which a non-human entity performs the function of subject and is followed by an active verb.

As for the passive structures, I will account for finite passive structures, such as the prototypical passive (both agentless and agentive), and the reporting *it*-passive structure; and the non-finite passive structures, namely *as*-passives and past participial reduced relative clauses. Another pattern which will also be analysed is the existential *there*, which is impersonal. Despite being grammatically active, as the expletive *there* functions as the grammatical subject of the clause and is followed by a verb (usually *be*) in its active form, I will simply consider it as a distinctive impersonal structure and will not include it within the active category.

5.2.1. Active Voice Structures

Despite the absolute omnipresence of passive voice sentences in medical writing, the existence as well as the importance of active instances in medical discourse cannot be dwarfed. As it is well documented in many studies (Rodman, 1994; Banks, 1994; Givón, 2017; Ding, 2002; Rundblad, 2007; Vande Kopple, 1994; Reilly et al. 2005; *inter alia*), scientific medical writing favours the use of passive sentences, which heavily outnumber active ones, and this somehow creates a sense of neutrality and objectivity proper of such a register. Nevertheless, the overall frequency of active constructions reveals both personal and impersonal hints which counteract and reinforce, respectively, the impersonal dimension of medical writing.

By using an active sentence, the subject is occupied by either human beings (usually the researchers of the study under discussion), or by non-human entities (usually entities whose designers are the researchers themselves – or other researchers – and which somehow acquire humanised characteristics). The former are labelled as personal active structures because the subject is human and thus *personal*, whereas the latter will be identified as *impersonal*, because authorial agency is masked through the metonymic use of simple noun phrases, nominalisations, and stacked nominal phrases. The heads of these nominal units acquire the human features expressed through the active verb that follows them. I will first explore personal active structures, and the analysis of impersonal active structures will follow.

5.2.1.1. Personal Active Voice Structure: *we* + active verb

As for the personal active constructions, it is possible to identify two main types: *we* personal active constructions, on the one hand, and non-*we* personal active constructions, on the other. The *we* personal active structure is the most frequent within all the personal structures occurring in the corpus, and it is the one where authors' presence is more overtly manifested. It serves, as previously stated, to counteract impersonality in medical discourse, as it is the only pattern tightly related to the notion

of personality and explicit authors' visibility. On the other hand, non-*we* personal actives use the active voice to topicalise other agents different from the authors of the text, although these are somehow related to the study being presented. These other agents also perform the function of subject in the active clause, but their presence is not linked to authorial visibility in any case. Because of this and because of the complexity to retrieve the various nouns performing the function of subject in this pattern, I will disregard them from this analysis. It is true, however, that their particular use indirectly contributes to emitting a sense of personality worth considering.

The personal active voice structure appears an overall of 5,282 times in the RCTs corpus, which corresponds to 4,254.66 per million words[59]. As seen in Table 10, this pattern appears 4,029 times (3,245.37 instances per M words) in the past simple tense, out of which 3,739 times it occurs in its affirmative form (3,011.77 instances per M words) and 290 times in its negative form (233.60 instances per M words). To a much lesser extent, it also appears an overall of 490 times in the present simple tense, which corresponds to 394.70 instances per M words (458 times in its affirmative form, 368.92 instances per M words; and 32 times in its negative form, 25.78 instances per M words) (see Table 11).

Table 10. Total number of occurrences of the Personal Active structure, according to tense. Raw frequencies and frequencies per million words.

Tense	Occurrences	Fq x M words
Past Simple	4,029	3,245
Present Simple	490	394.70
Present Perfect Simple	334	269.04
Future Simple	137	110.35
Total (of these 4 tenses)	4,990	4,109.50

[59] Henceforth, per M words.

The personal active pattern also appears in the present perfect simple tense (334 occurrences, 269.04 instances per M words), and in the future simple tense (137 occurrences, 110.35 instances per M words). These four tenses (past simple, present simple, present perfect simple, and future simple) are essentially the main ones used in medical writing. The past perfect simple, modal verbs and continuous forms have been disregarded because of their overall low frequency (292 occurrences, 235.21 instances per M words). This is why the overall result (5,282 occurrences) does not correspond to the one presented in Table 10 (4,990 occurrences). Figure 18 shows the quantificational distribution of this pattern according to the aforementioned tenses.

Figure 18. Distribution of tenses in the Personal Active pattern (*we* + active verb).

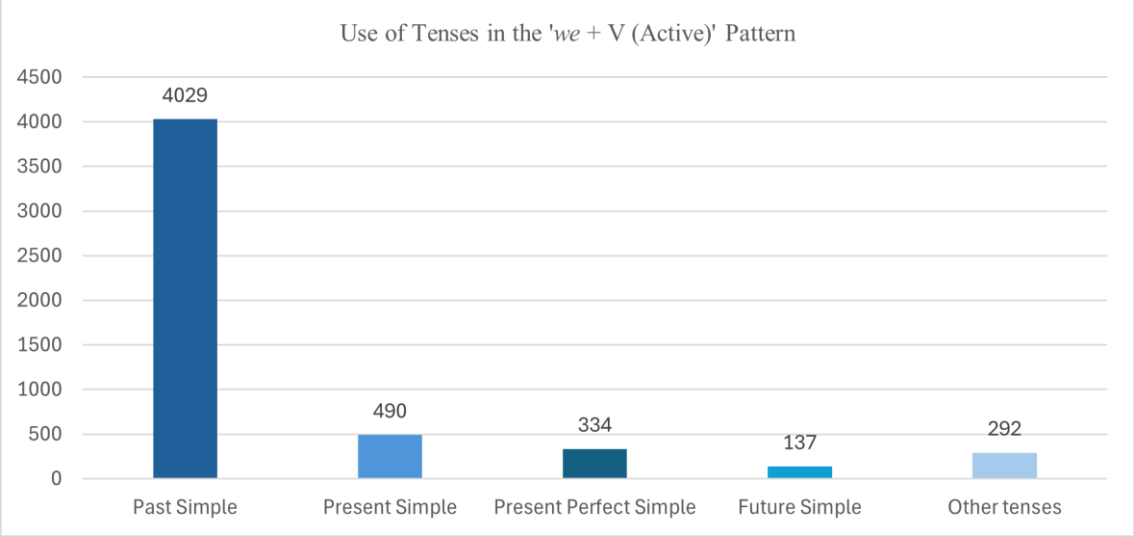


Table 11. Total number of occurrences of the Personal Active structure, according to tense and polarity. Raw frequencies and frequencies per million words.

Tense	**Past Simple**		**Present Simple**		**Present Perfect S**		**Future Simple**	
Polarity	Occ	Fq x Mw	Occ	Fq x Mw	Occ	Freq x Mw	Occ	Freq x Mw
Positive	3,739	3,011.77	458	368.92	334	269.04	137	110.35
Negative	290	233.60	32	25.78				
Total	4,029	3,245.37	490	394.70	334	269.04	137	110.35

As for the polarity of the verb structures in the past and in the present simple tenses, distinct searches have been used to quantify the affirmative and the negative forms. The syntax used to search for these specific patterns is highly different, as present and past simple forms require the auxiliary *do* (or *does*) and *did*, respectively, to mark the negative polarity of the verb phrase. However, in the present perfect simple and in the future simple tenses, the negative polarity has been included as an optional tag to the positive one by adding the particle *not*, expressed through **(_XX)?** (if the Simple Query Syntax is used) or **[pos="XX"]*** (if CQP Syntax is used). In this case, the particle *not* is merely added because the auxiliaries *have* and *will* are already included within the affirmative pattern. The non-distinction between positive and negative polarity in these two tenses essentially stems from their low number of occurrences in the RCTs corpus.

Indeed, the syntactic language used to search for the patterns corresponds to both the Simple Query Syntax (henceforth, SQS) and the CQP Syntax (henceforth, CQPS). The reason why I am providing both types of syntax for each search stems from the fact that some of the patterns cannot be properly searched for if only one of the syntaxes is used, and this causes the need to make use of the other type of syntax. To account for the particularities of each of the patterns, specific information on some problematic issues when searching for some of them will also be included in the analysis, as were described within Chapter 4 (Methodology).

As for the different variants in the RCTs corpus accounting for the personal active structure (*we* + active verb), and bearing in mind the four main tenses dealt with, it is possible to identify the following sub-patterns:

Table 12. Patterns accounting for the positive and negative polarity of the Past Simple tense, using Simple Query Syntax and CQP Syntax, including the description of the patterns

Past Simple	**Simple Query Syntax / CQP Syntax**
Positive	**_PPIS2 (_{ADV})* _V?D** or **[pos="PPIS2"] ([pos="RR"])* [pos="V.D"]**
	Pronoun *we*, followed by any number of adverbs (zero or more), followed by the past tense form of any lexical verb (including *do* and *have*).
Negative	**_PPIS2 (_{ADV})* _VDD _XX (_{ADV})* _V?I** or **[pos="PPIS2"] ([pos="RR"])* [pos="VDD"] [pos="XX"] ([pos="RR"])* [pos="V.I"]**
	Pronoun *we*, followed by *did not*, which may be preceded or followed by any number of adverbs (zero or more), followed by the infinitive form of any lexical verb (including *do* and *have*).

As for the affirmative form of the pattern, if only lexical verbs (V) are searched for, the syntactic expression to be used is as follows:

SQS	**_PPIS2 (_{ADV})* _VVD**
CQPS	**[pos="PPIS2"] ([pos="RR"])* [pos="VVD"]**

If only the verb *do* (D) is searched for, the syntactic expression will vary:

SQS	**_PPIS2 (_{ADV})* _VDD**
CQPS	**[pos="PPIS2"] ([pos="RR"])* [pos="VDD"]**

If only the verb *have* (H) is searched for, the syntactic expression will be:

SQS	**_PPIS2 (_{ADV})* _VHD**
CQPS	**[pos="PPIS2"] ([pos="RR"])* [pos="VHD"]**

For instance, to account for the search of the personal pronoun *we*, followed by the verb form *were*, followed by an optional negative particle *not*, followed by any number of adjectives (in their plain or comparative form), each of which may optionally be preceded by any adverb, the resulting pattern will be:

SQS	**_PPIS2 _VBDR (_XX)? (_{ADV})? _JJ***
CQPS	**[pos="PPIS2"] [pos="BVDR"] [pos="XX"]* ([pos="RR"])* [pos="JJ"]**

Table 13. Patterns accounting for the positive and negative polarity of the Present Simple tense, using Simple Query Syntax and CQP Syntax, including the description of the patterns

Present Simple	Simple Query Syntax / CQP Syntax
Positive	**_PPIS2 (_{ADV})* _V?0** or **[pos="PPIS2"] ([pos="RR"])* [pos="V.0"]**
	Pronoun *we*, followed by any number of adverbs (zero or more), followed by the present tense of any lexical verb (including *do* and *have*).
Negative	**_PPIS2 (_{ADV})* _VD0 _XX (_{ADV})* _V?I** or **[pos="PPIS2"] ([pos="RR"])* [pos="VD."] [pos="XX"] ([pos="RR"])* [pos="V.I"]**
	Pronoun *we*, followed by *do not*, which may be preceded or followed by any number of adverbs (zero or more), followed by the infinitive form of any lexical verb (including *do* and *have*).

If the present simple pattern is searched for using this formula, structures in the Present Perfect also appear. To my knowledge, there is no way to restrict the appearance of the verb *have* after the subject, neither using the SQS nor the CQPS. The only possibility here is to restrict the query to lexical verbs by using the tag **(_VV0)**. By doing so, the verbs *do* and *have* are excluded. The pattern searched for would then be as follows:

SQS	**_PPIS2 (_{ADV})* _VV0**
CQPS	**[pos="PPIS2"] ([pos="RR"])* [pos="VV0"]**

This pattern now includes all lexical verbs except for *have* and *do*. If this alternative pattern is searched for instead, the query is then restricted and allows for the omission of the auxiliary verb *have* of the present perfect simple tense. However, at the same time, the query also restricts those cases in which the lexical verb *do* appears as a lexical verb. Yet, as there are not many instances with *do* as a lexical verb, I will not take it into consideration.

Table 14. Pattern accounting for the positive and negative polarity of the Present Perfect Simple tense, using Simple Query Syntax and CQP Syntax, including the description of the pattern

Present Pf Simple	**Simple Query Syntax / CQP Syntax**
Positive and Negative	**_PPIS2 (_{ADV})* _VH0 (_XX)? (_{ADV})* _V?N** or **[pos="PPIS2"] ([pos="RR"])* [pos="VH0"] [pos="XX"]* [pos="RR"]* [pos="V.N"]**
	Pronoun *we*, followed by the auxiliary *have*, which may be preceded or followed by any number of adverbs (zero or more), and which may be followed by the negative particle *not*, followed by the past participle of any lexical verb (including *do* and *have*).

Table 15. Pattern accounting for the positive and negative polarity of the Future Simple tense, using Simple Query Syntax and CQP Syntax, including the description of the pattern

Future Simple	**Simple Query Syntax / CQP Syntax**
Positive and Negative	**_PPIS2 (_{ADV})* will (_XX)? (_{ADV})* _V?I** or **[pos="PPIS2"] ([pos="RR"])* [pos="VM"] [pos="XX"]* [pos="RR"]* [pos="V.I"]**
	Pronoun *we*, followed by the auxiliary *will*, which may be preceded or followed by any number of adverbs (zero or more), and which may be followed by the negative particle *not*, followed by the infinitive form of any lexical verb (including *do* and *have*).

With the '*we* + active verb (future simple)' pattern, if SQS is used, it is possible to restrict the query to the instances that contain the auxiliary form *will*. This restriction, however, is not possible if CQPS is used. The only way to identify the appearance of *will* is by including the tag **_VM**, which would include any modal verb. This makes, however, other modal verbs (like *can*, *may* or *might*) be included within the results obtained.

Once these four sub-patterns have been assigned with their corresponding formulae and descriptions, I will now turn into analysing each of them, essentially accounting for their raw and relative frequencies (normalised frequencies per M words) in the RCTs corpus. I will introduce them from the most to the least frequently used.

The **past simple tense** (emphasis added), as seen in Table 14, is the most frequent tense in the RCTs corpus, with a total amount of 4,029 occurrences. As for the raw frequencies, it is extremely used in the Methods section (2,287 occurrences), distantly followed by the Results section (730 occurrences), and the Abstract section (570 occurrences). The Introduction section is the one in which this pattern scarcely appears if compared to the rest, with only 78 occurrences. As for the polarity, the unmarked positive form is extremely much more frequent if compared to the negative one, and this can be found in all the sections of the RA (3,739 occurrences in the affirmative form, 3,011.77 instances per M words; against 290 occurrences in the negative form, 233.60 instances per M words). It is also interesting to highlight that this pattern tends to appear after punctuation, either paragraph-initially or sentence-initially.

When relative frequencies are considered, the Methods section is the one whose frequency per million words is higher (5,527.93 instances per M words), followed by the Abstract section (5,475.93 instances per M words). The great use of the personal active voice pattern in the Methods section is well justified because authors recount the steps taken in their experiment, and in this case, do so by overtly manifesting their previous intervention in the set of experimental procedures carried out (see examples 1 and 2). In addition, this same fact is also found in the Abstract section. The high relative frequency in Abstracts also indicates that authors make use of this personal structure

when succinctly presenting the same experimental procedures when providing a summarised version of the RA.

Table 16. Quantificational distribution of the '*we* + Active verb' in the Past Simple tense, according to positive and negative polarities, and overall. Raw and relative frequencies according to sections and overall.

Sections of RA	Words	+ polarity Occ		- polarity Occ		TOTAL	Fq x Mw
Abstract	104,092	554	5,322.21	16	153.71	570	5,475.93
Introduction	99,009	76	767,61	2	20.20	78	787.81
Methods	413,717	2,153	5,204.04	134	323.89	2,287	5,527.93
Results	316,837	643	2,029.43	87	274.59	730	2,304.02
Discussion	307,806	313	1,016.87	51	165.69	364	1,182.56
TOTAL	1,241,461	3,739	3,011.77	290	233.60	4,029	3,245.37

The Introduction is the section in which the '*we* + active verb (past simple)' pattern shows a much lower relative frequency (787.81 instances per M words). This section is mostly related to setting the scene, and the past tense is barely used for this purpose in favour of the present simple tense, which tends to be relatively more frequent. The presentation of background information related to the topic under study is frequently expressed through the present simple tense, whereas at the end of the Introduction, the past simple tense is essentially used to summarise the aim(s) of the study and state the experimental approach taken (see example 3). Examples 4 and 5 show the use of this pattern in the Results and in the Discussion sections, respectively.

(1) **We recorded** baseline serum creatinine, cystatin C, spot urine albumin-to-creatinine ratio, and 24-h ambulatory blood pressure before randomisation. **[Personal active pattern in the past simple; Methods; la370697m]**

(2) First, **we assessed** the age- and sex-specific incidence rates of all types of first TNAs. Subsequently, **we assessed** the association between baseline characteristics and risk of TNA with adjusted Cox proportional hazards models. Then **we assessed** the association between occurrence of TNA and risk of adverse events with time-dependent Cox proportional hazards models, comparing the risk of an adverse event developing after TNA to the risk of an adverse event developing in TNA-free

participants. **We included** occurrence of TNA as a time dependent covariate in the model, which means that the TNA status changed from absent to present at the exact moment during follow-up that it had occurred. **We adjusted** for baseline covariates age and sex (model1) and additionally for a propensity score 25 that was based on C-reactive protein, systolic blood pressure, carotid intima-media thickness, (…). **[Personal active patterns in the past simple; Methods; ja617141m]**

(3) **We determined** the effectiveness of six or 24 lessons in the Alexander technique, massage therapy, and advice from a doctor to take exercise (using an exercise prescription) with nurse delivered behavioural counselling for patients with chronic or recurrent back pain. **[Personal active pattern in the past simple; Introduction; bm289347i]**

(4) **We conducted** our study from January 2005 to February 2008. **[Personal active pattern in the past simple; Results; an467035r]**

(5) Despite our best intentions, **we inadvertently included** nine children (7% of randomly assigned children) in whom persistent or atopic asthma symptoms developed during the study period. Second, **we were unable** to identify characteristics of patients that could modulate the risk-benefit ratio for preemptive treatment. Third, **we tested** parent's initiation of preemptive treatment with fluticasone as it would happen in real-life practice. **[Personal active patterns in the past simple; Discussion; ne284154d]**

Out of all the sections, the **present simple tense** (emphasis added) is more frequently used in the Discussion section (213 occurrences, 691.99 instances per M words) (see Table 17), as it is where authors draw conclusions from the facts revealed from the study and highlight the need to continue their research line (see examples 6-10). Indeed, the personal active pattern serves not only to overtly express their involvement in the research, but also to claim their responsibility in their study by giving valuable conclusions of the research process carried out and to give themselves the credit for what has been investigated.

Table 17. Quantificational distribution of the '*we* + Active verb' in the Present Simple tense, according to positive polarity, negative polarity, and overall. Raw frequencies and relative frequencies according to sections of the RA, an overall.

Sections of RA	Words	+ polarity Occ		- polarity Occ		TOTAL	Fq x Mw
Abstract	104,092	31	297.81	3	28.82	34	326.63
Introduction	99,009	41	414.10	4	40.40	45	454.50
Methods	413,717	85	205.45	0	0.00	85	205.45
Results	316,837	88	277.75	4	12.62	92	290.37
Discussion	307,806	213	691.99	21	68.22	234	760.22
TOTAL	1,241,461	458	368.92	32	25.78	490	394.70

The other two sections in which the '*we* + active verb pattern (present simple tense)' is relatively frequent are the Results (88 occurrences, 277.75 instances per M words), and the Methods sections (85 occurrences, 205.45 instances per M words). In these two sections, the present simple tense is frequently used to give general statements, to describe the characteristics of the materials and procedures used, and to refer to conclusive aspects extracted from the results that will be deeply commented in the Discussion section. Although Abstracts and Introductions do not include a high number of active patterns in the present simple tense, these structures mainly serve to highlight conclusive arguments drawn by authors (see examples 11 and 12).

(6) The significant associations that **we report** between positive screens for disability and access to nutrition and early-learning opportunities call for further investigation. With the growth of interest in child health and development, **we hope** to see continued improvements in global health research and policy to both reduce risk of disability and improve opportunities and outcomes for children with disabilities in countries with low and middle incomes. **[Personal active patterns in the present simple; Discussion; la800346d]**

(7) **We show** that our model exhibits excellent dis-crimination with a c-statistic of 0.85 for SMM or worse and displays superb calibration over the entire range of predicted risk. **[Personal active pattern in the present simple; Discussion; ja478536d]**

(8) In contrast to these studies, **we demonstrate** that the urinary NGAL level identifies acute kidney injury in abroad patient sample with different mechanisms of injury. In

addition, **we demonstrate** that the urinary NGAL level remained highly diagnostic even when the timing of injury was unknown, making NGAL potentially diagnostic of kidney disease for many clinical presentations. **[Personal active patterns in the present simple; Discussion; ja754812d]**

(9) Third, although we recruited patients from 6 a.m. to midnight, potentially producing a cohort biased toward medical conditions associated with acute referrals, **we believe** the effect to be minimal because the mean 12- to 18-hour length of stay in our emergency department allowed us to enroll admitted patients regardless of when they first presented. **[Personal active pattern in the present simple; Discussion; ja754612d]**

(10) **We think** this may be because most other studies recruited patients from hospitals or asked the general practitioners to call a stroke consultant, whereas we had direct access to general practitioners' files and could also question the participants in person. **[Personal active pattern in the present simple; Discussion; ja164698d]**

(11) **We describe** the results of a multicentre phase II trial, testing in parallel 2 novel combination therapies, predefined molecular markers, and tumor rebiopsy at progression. **[Personal active pattern in the present simple; Abstract (Results); la257451a]**

(12) **We describe** cost-effectiveness based on 7-yearfollow-up from the largest of the 4 trials of AAA screening — the Multicentre Aneurysm Screening Study (MASS). **[Personal active pattern in the present simple; Introduction; ja280941i]**

The **present perfect simple tense** (emphasis added) is much less frequent than the past simple form (334 occurrences, 269.04 instances per M words; against 4,029 occurrences, 3,245.37 instances per M words) (see Table 18). Indeed, it is also less frequent than the present simple (490 occurrences, 394.70 instances per M words). The present perfect simple is mainly used in the Results section (118 occurrences, 372.42 instances per M words; see example 13), followed by the Discussion and the Methods sections (86 occurrences, 279.40 instances per M words, see examples 14-16; and 85 occurrences, 205.45 instances per M words, respectively). If the relative frequency is considered, the present perfect simple shows almost the same relative frequency in the

Introduction (373.70 instances per M words) as the one shown in the Results section (372.43 instances per M words). This tense is essentially used to express actions which began in the past that are still strongly relevant to the present time.

Table 18. Quantificational distribution of the '*we* + Active verb' in the Present Perfect Simple tense, including both positive and negative polarity. Raw frequencies and relative frequencies according to sections of the RA, and overall.

Sections of RA	Words	+ / - polarity Occ	Fq x Mw
Abstract	104,092	8	76.86
Introduction	99,009	37	373.70
Methods	413,717	85	205.45
Results	316,837	118	372.43
Discussion	307,806	86	279.40
TOTAL	1,241,461	334	269.04

(13) In our primary statistical analyses **we have only included** calculated crude coverage from multicountry surveys for which the microdata are available and reported estimates from surveys for which the sample size has been reported. For these surveys, multistage sampling design information is not available, so **we have assumed** that they have the 90th percentile design effect noted in the multicountry survey programmes. **[Personal active patterns in the present perfect simple; Results; la368845r]**

(14) In conclusion, **we have shown** that genetic variation has an effect on pharmacologic and clinical responses to clopidogrel. **[Personal active pattern in the present perfect simple; Discussion; ne472632d]**

(15) **We have shown** that mutations in NR5A1 are associated with human ovarian insufficiency, an observation consistent with the hypoplastic ovaries and infertility of mice lacking Nr5a1 in their granulosa cells. **[Personal active pattern in the present perfect simple; Discussion; ne174685d].**

(16) **We have shown** that the MRKAd5 HIV-1 gag/pol/nef vaccine elicited a higher CD8+ T-cell response rate and magnitude than did that reported for any of the candidate immunisation regimens tested over the past 15 years,1 although

immunological assays have changed greatly during this time. Furthermore, pre-existing Ad5 immunity substantially affected responses to the vaccine antigens. **[Personal active pattern in the present perfect simple; Discussion; la348758d]**

(17) It included diagnostic criteria for transient attacks of neurological dysfunction, which in the present article **we will call** transient neurological attacks (TNAs). **[Personal active pattern in the future simple; Introduction; ja649331i]**

(18) Although our findings are limited by only 1 year of data, in future years **we will be able to compare** whether countries are experiencing compression or expansion of morbidity similarly. **[Personal active pattern in the future simple; Discussion; la641359d]**

As seen in Table 19, the **future simple tense** (emphasis added) expresses actions that have yet not been performed (see examples 17 and 18). As expected, it is not very frequent if compared to the rest of the tenses dealt with (138 occurrences, 111.16 instances per M words), and it is quite restricted to the Methods section (115 occurrences, 277.97 instances per M words)[60]. In the rest of the sections, its frequency is much lower.

Table 19. Quantificational distribution of the '*we* + Active verb' in the Future Simple, including both positive and negative polarity. Raw frequencies and relative frequencies according to sections of the RA

Sections of RA	Words	+ / - polarity Occ	Fq x Mw
Abstract	104,092	2	19.21
Introduction	99,009	7	70.70
Methods	413,717	115	277.97
Results	316,837	9	28.41
Discussion	307,806	5	16.24
TOTAL	1,241,461	138	111.16

[60] As access to the RCTs corpus became restricted, the only instances that I had previously collected accounting for this pattern appear in other sections different from the Results. Besides, they do not contain the most frequent lemmas in this pattern.

Once the different sub-patterns corresponding to the personal active voice structure have been quantified and analysed bearing in mind the tense in which they appear, I will now highlight the most common lemmas found in each of them, as shown in Tables 20 and 21. Those lemmas which are frequently used in the different tenses analysed are marked with the same colour to ease their identification. After presenting the lemmas, I will quantify them according to their tense and I will organise them semantically to check whether the choice of verb triggers a particular choice of tense.

Table 20. Top ten list of lemmas, according to their tenses (Past Simple, Present Perfect Simple, and Future Simple), including polarity

	Past Simple		**Present Perfect Simple**	**Future Simple**
	Affirmative	*Negative*	*Affirmative / Negative*	*Affirmative / Negative*
1	USE	FIND	USE	USE
2	DO	APPLY	REPORT	SEEK
3	FIND	INCLUDE	INCLUDE	PRESENT
4	CONDUCT	IDENTIFY	SEEK	EXPLORE
5	CALCULATE	PERFORM	FIND	CONTACT
6	ASSESS	CONSIDER	ASSUME	ASSESS
7	COMPARE	IMPOSE	IDENTIFY	REPORT
8	ESTIMATE	USE	MAKE	CALCULATE
9	EXCLUDE	ENTER	REVIEW	ANALYSE
10	OBSERVE	SUMMATE	PRESENT	CONSIDER

Table 21. Top ten (and top four) list of lemmas, according to the Present Simple tense, including polarity

	Present Simple	
	Affirmative	*Negative*
1	BELIEVE	KNOW
2	SUGGEST	RECOMMEND
3	RECOMMEND	BELIEVE
4	NEED	THINK
5	PRESENT	
6	CONSIDER	
7	KNOW	
8	FEEL	
9	CONCLUDE	
10	PLAN	

The most common lemmas used in the past simple tense are related, essentially, to verbs of experimental procedure (like *use*, *find*, *conduct*, *calculate*, or *include/exclude*, among many others) and are usually presented in the affirmative form. This means that authors use the personal active pattern in the past simple tense to narratively explain the different steps carried out in their research process. They do so by listing, within the text, these different steps, and make use of this pattern to actively promote their involvement in the research process being described (see Table 22). Indeed, very similar lemmas are also very typical of sub-patterns in the present perfect simple and in the future simple tenses (see Tables 20 and 23).

Notwithstanding, considering all the lemmas found in these four sub-variants, the most radical semantic distinction is seen with the lemmas in the present simple tense (see Table 21), as they do not overtly express material procedures but are tightly related to a more cognitive and mental approach to the process being described. The structure '*we* + active verb' in the present simple tense is the one which really serves to position authors towards the claims being made, and it overtly makes them take side with what is being exposed in the text. It may be stated that it is the most volitional sub-pattern existing in the corpus.

As already stated, Tables 22 and 23 show the most frequent instances of the pattern '*we* + active verb' in the past simple tense (Table 22) and in the present perfect simple and future simple tenses (Table 23). The verb *use* is the most frequent one in the three different tenses analysed. The rest of the verbs are essentially related to procedural steps. Among the ones related to procedures, it is worth noting this set of verbs can be divided into two sub-fields: those denoting neutral research acts (like *calculate*, *perform*, or *conduct*, among many others) and those denoting cognitive research acts (like *exclude*, *compare*, and *undertake*, among many others). There are three other semantic groups of verbs accounting for these sub-patterns: intentional verbs related to purposive research acts (like *seek* and *aimed*, among others), evaluative verbs related to assessing research acts (like *assess*, *evaluate*, and *examine*, among many others) and reporting verbs tightly linked to conclusive research acts (like *find*, *consider*, and *observe*, among many others) (see Table 25 in this section).

Table 22. List of top instances of '*we* + Active verb (Past Simple tense)', divided by polarity and including the number of occurrences and their frequency per million words

	Past Simple					
	Positive	**Occ**	**Fq x Mw**	**Negative**	**Occ**	**Fq x Mw**
1	we used	383	308,51	we did not find	50	40,28
2	we did[61]	236	190,10	we did not apply	24	19,33
3	we found	207	166,74	we did not include	22	17,72
4	we conducted	108	86,99	we did not identify	16	12,89
5	we calculated	103	82,97	we did not perform	13	10,47
6	we assessed	94	75,72	we did not consider	12	9,67
7	we compared	86	69,27	we did not impose	11	8,86
8	we estimated	76	61,22	we did not use	9	7,25
9	we excluded	73	58,80	we did not enter	9	7,25
10	we observed	67	53,97	we did not summate	7	5,64
11	we obtained	66	53,16			
12	we included	61	49,14			
13	we randomly assigned	58	46,72			
14	we measured	56	45,11			
15	we undertook	54	43,50			
16	we performed	49	39,47			
17	we considered	49	39,47			

Table 23. List of top instances of '*we* + Active verb (Present Perfect simple tense)' and of '*we* + Active verb (future simple tense)', including the number of occurrences and their frequency per million words

	Present Perfect Simple	**Occ**	**Fq x Mw**	**Future Simple**	**Occ**	**Fq x Mw**
1	we have used	19	15,30	we will use	20	16,11
2	we have reported	18	14,50	we will seek	15	12,08
3	we have included	17	13,69	we will also present	12	9,67
4	we have sought	13	10,47	we will explore	8	6,44
5	we have found	13	10,47	we will contact	8	6,44
6	we have assumed	10	8,06	we will assess	6	4,83
7	we have identified	10	8,06	we will report	3	2,42
8	we have made	9	7,25	we will present	3	2,42
9	we have reviewed	8	6,44	we will calculate	3	2,42
10	we have presented	6	4,83	we will analyse	3	2,42
11	we have considered	6	4,83	we will consider	3	2,42
12	we have excluded	5	4,03	we will perform	3	2,42
13	we have attempted	5	4,03	we will follow	3	2,42
14	we have pooled	5	4,03	we will conduct	3	2,42

[61] Non-modal auxiliary verbs like *did* are included in the results obtained because of the high frequency of the negative polarity of some lexical verbs. Out of 236 occurrences, 131 are negative ones (we did not (*)).

Table 24. List of top instances of '*we* + Active verb (Present Simple tense)', divided by polarity and including the number of occurrences and their frequency per million words

	Present Simple					
	Positive	**Occ**	**Fq x Mw**	**Negative**	**Occ**	**Fq x Mw**
1	we believe	37	29,80	we do not know	15	12,08
2	we suggest	24	19,33	we do not recommend	3	2,42
3	we recommend	18	14,50	we do not believe	3	2,42
4	we need	18	14,50	we do not think	3	2,42
5	we present	16	12,89			
6	we consider	16	12,89			
7	we know	16	12,89			
8	we feel	15	12,08			
9	we conclude	14	11,28			
10	we plan	11	8,86			
11	we hope	10	8,06			
12	we use	9	7,25			
13	we think	9	7,25			
14	we define	9	7,25			
15	we find	8	6,44			

Table 24 shows the list of top instances of the personal active in the present simple tense. As said, the verbs appearing in the present simple tense are essentially restricted to reporting verbs denoting conclusive research acts.

As previously mentioned, the verbs appearing in the personal active pattern in the RCTs corpus can be divided into procedural verbs, denoting neutral research acts and cognitive research acts. The former essentially describe the experimental, material, and physical procedures carried out in the research process. Despite also describing experimental and material procedures, the latter also involve a more cognitive approach to the research being presented. Apart from procedural verbs, it is also possible to find a set of intentional verbs denoting purposive research acts, where authors express their intentions in the research process. Authors also evaluate through the use of evaluative verbs related to assessing research acts, and also draw conclusions through reporting verbs denoting conclusive research acts.

Indeed, it is important to relate these four different semantic classifications of verbs to the nature of the involvement of authors in the research being presented. Procedural verbs related to neutral research acts essentially make use of the past simple tense, although some instances in the future simple and the present perfect simple tenses also occur. The present simple tense does not account for any instance referring to neutral research acts. Furthermore, those verbs related to cognitive research acts are also essentially used in the past simple tense. Here, some instances are also found in the present perfect simple tense, and it can be asserted that there are no instances in the future simple. As for intentional verbs that refer to purposive research acts, the main tenses used are the past simple, as well as the present perfect simple and the future simple tense; therefore, no present simple tense instances are found with this particular meaning.

Verbs referring to evaluation and assessing research acts are used in their past simple and future simple tenses. There are no present simple and present perfect simple instances with verbs related to evaluation. Indeed, when evaluating, authors make use of the past tense to refer to assessments carried out in the research under discussion or point out future particular research evaluations that will be shown later in the text or that will have to be developed in the future to fill some gaps that have emerged from the study that they are presenting.

It is worth highlighting that the more subjective group of verbs, reporting verbs related to conclusive research acts, are the only ones which make extensive use of the present simple tense. However, it is also true that this semantic sub-group of verbs is the only one which makes frequent use of all the four tenses analysed.

As shown in Table 25, a continuum can be created considering the involvement of authors in the type of research acts being undertaken. Even though all the sub-patterns of the active structure '*we* + active verb' are related to the notion of personality and authorial visibility, as authors overtly manifest themselves in the research being described, a certain gradability of their involvement in the research process can also be perceived. Procedural verbs refer to objective neutral actions, and such neutrality and

objectivity gradually decreases as the semantic typology of verbs changes into verbs of intention, followed by verbs of evaluation, and followed by reporting verbs. Each change towards a new semantic category strengthens and reinforces the subjective implication of authors in their research. At the end of the continuum, reporting verbs are seen to promote the role of authors to its full extent, as they are visibly seen as authoritative representatives of the conclusive information they are transmitting.

To sum up, it can then be stated that even within the most personal active structure found in the RCTs corpus, there exists a cline from more neutral and impartial (impersonal) tones, which exclude the use of the present simple tense in favour of the rest of the tenses analysed; to more involving, biased, and partial (personal) shades, in which the present simple tense, particularly, plays a crucial role.

Table 25. Semantic classification of verbs appearing in the pattern '*we* + active verb', according to the tenses analysed (Past Simple, Present Simple, Present Perfect Simple, and Future Simple).

Procedural Verbs							**Intentional Verbs**			**Evaluative Verbs**		**Reporting Verbs**			
Neutral Research Acts			Cognitive Research Acts				Purposive Research Acts			Assessing Research Acts		Conclusive Research Acts			
use	use	use	assume		assume		aim			assess	assess	find		find	
search			exclude		exclude		intend			evaluate		consider	consider	consider	consider
calculate		calculate	include		include		attempt	attempt	attempt	examine		resolve			
perform		perform	compare				seek	seek	seek	analyse	analyse	present	present	present	present
conduct		conduct	undertake							check		report		report	report
record			plan	plan							explore	observe			
make	make		apply								investigate	obtain			
combine			impose										believe		
enter													suggest		
summate													recommend		
		contact											know		
		follow											feel		
		add											conclude		

PtS
PrS
PrPfS
FS

NEUTRALITY - OBJECTIVITY → SUBJECTIVITY based on EVIDENCE

5.2.1.2. Impersonal Active Voice Structures

The impersonal active structure is a pattern which is formed of an inanimate subject and a verb in the active voice and is also referred to as abstract rhetor. It is a linguistic strategy that medical writers frequently use, and in fact, it extends a sense of detachment and impersonality over medical writing, as it introduces any kind of research activity (or results) that is often carried out by the authors themselves or by other researchers in the field, but that is described as occurring independently from them. The nouns (or nominalisations or stacked nominal phrases) which perform the function of subject are inanimate but are humanised by means of metonymic expressions, in which the head acquires human features described by the verb in the active voice.

Although this pattern is quite recurrent in academic writing, its retrieval in the corpus is complex. There are various limitations when it comes to searching for this pattern. The first and the most relevant one is that there is no automatic way to differentiate animate subjects from those which are inanimate. If one searches for any singular or plural noun followed by any number of optional adverbs and an active verb in any tense, through the syntaxes **_NP? (_{ADV})* _V??** (SQS) or **[pos="N.."] ([pos="RR"])* [pos="V.."]** (CQPS), the figures obtained are not accurately representative enough of the impersonal active pattern (see Table 26). Indeed, some instances of personal active patterns are automatically retrieved because it is not possible to restrict the subject to only non-human entities.

Table 26. Distribution of 'Noun (+/- Human) + Active verb in any tense' pattern across the sections of the RA. Raw and relative frequencies per million words

Noun (+/-human) + Active Verb in any tense			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	2,466	1,986.37
Introduction	99,009	3,693	2,974.72
Methods	413,717	8,170	6,580.96
Results	316,837	6,682	5,382.37
Discussion	307,806	10,197	8,213.71
TOTAL	1,241,461	31,208	25,138.12

As regards the number of occurrences of the pattern 'noun (+/- human) + active verb' in any tense, there are 31,208 instances in the RCTs corpus (25,138.12 instances per M words) and it is highly frequent in the Discussion section (10,197 occurrences, 8,213.71 instances per M words). However, as said, these figures cannot be used to determine the frequency of the impersonal active pattern.

An additional limitation encountered when searching for this pattern is the verb forms found. In most of the cases, the verbs are finite forms in the past simple or in the present simple tenses. However, in some instances non-finite forms like the present participle also appear. This is because the tags used to retrieve the verb are **_V??** (SQS) and **[pos="V.."]** (CQPS) and, therefore, non-finite forms, despite being infrequent, are included among the retrieved instances. For all these reasons, I decided to search for this pattern only in the most common tense found, which is the preterite (see Table 27). The syntactic expressions used are then **_NP? (_{ADV})* _V?D** (SQS) or **[pos="N.."] ([pos="RR"])* [pos="V.D"]** (CQPS).

Table 27. Distribution of 'Noun (+/- Human) + Active verb in the Past Simple tense' pattern across the sections of the RA. Raw and relative frequencies per million words

Noun (+/-human) + Active Verb in the Past Simple			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	1,942	1,564.29
Introduction	99,009	2,083	1,677.86
Methods	413,717	7,174	5,778.68
Results	316,837	5,398	4,340.85
Discussion	307,806	6,616	5,329.20
TOTAL	1,241,461	23,204	18,690.88

The figures are similar to the ones including all the tenses, and the limitation is again found when dealing with the presence or absence of animacy in the noun working as the subject. For all these reasons, I decided to restrict the queries to the most recurrent nouns working as the subject of this pattern: *study* (or *studies*) and *trial* (or *trials*).

Table 28. Distribution of '*Study* / *Studies* + Active verb' pattern across the sections of the RA. Raw and relative frequencies per million words

***Study* + Verb**		***Study* (singular)**		***Studies* (plural)**		***Study* / *Studies***	
Distribution	**Words**	**Occ**	**Fq x Mw**	**Occ**	**Fq x Mw**	**Occ**	**Fq x Mw**
Abstract	104,092	88	845.41	302	2,901.28	390	3,746.69
Introduction	99,009	58	585.81	89	898.91	147	1,484.71
Methods	413,717	99	239.29	621	1,501.03	720	1,740.32
Results	316,837	806	2,543.89	2,309	7,287.66	3,115	9,831.55
Discussion	307,806	258	838.19	598	1,942.78	856	2,780.97
TOTAL	1,241,461	1,309	1,054.40	3,919	3,156.76	5,228	4,211.17

As shown in Table 28, the nouns *study* or *studies* appear 5,228 times (4,211.17 instances per M words). They are highly frequent in the Results section (3,115 occurrences, 9,831.55 instances per M words), followed by the Discussion section (856 occurrences, 2,780.97 instances per M words). The most frequent form of all is the plural noun form '*studies* + active verb' in its preterite form, which appears in 2,309 occurrences in the Results section (7,287.66 instances per M words).

Very similar results occur when searching for the noun *trial* in its singular and plural forms and followed by an active verb form in the past simple tense (see Table 29). It appears a total number of 4,285 times in the RCTs corpus (3,451.58 instances per M words). Overall, the plural form of the noun is more frequent than the singular form (3,238 occurrences, 2,608.22 instances per M words; against 1,047 occurrences, 843.36 instances per M words; respectively). As it occurs with the subject *studies*, the most used pattern is '*trials* + active verb' in its preterite form in the Results section (1,557 occurrences, 4,914.20 instances per M words), followed by the same pattern in its singular form (681 occurrences, 2,149.37 instances per M words).

Table 29. Distribution of '*Trial* / *Trials* + Active verb' pattern across the sections of the RA. Raw and relative frequencies per million words

***Trial* + Verb**		***Trial* (singular)**		***Trials* (plural)**		***Trial* / *Trials***	
Distribution	**Words**	**Occ**	**Fq x Mw**	**Occ**	**Fq x Mw**	**Occ**	**Fq x Mw**
Abstract	104,092	74	710.91	452	4,342.31	526	5,053.22
Introduction	99,009	16	161.60	40	404.00	56	565.61
Methods	413,717	99	239.29	661	1,597.71	760	1,837.00
Results	316,837	681	2,149.37	1,557	4,914.20	2,238	7,063.57
Discussion	307,806	177	575.04	528	1,715.37	705	2,290.40
TOTAL	1,241,461	1,047	843.36	3,238	2,608.22	4,285	3,451.58

Bearing in mind the data presented in the following Tables (30, 31, 32, 33), it is possible to state that abstract rhetors, at least those including the nouns *study* (*studies*) and *trial* (*trials*) contain active verbs in the preterite and present simple tense, which can be classified using the same semantic typologies found in the personal active pattern, a fact that reinforces the idea of humanisation of the subject. Procedural verbs accounting for neutral research acts, like *use*, are combined with those relating to cognitive research acts, like *include* or *compare*. Evaluative verbs linked to assessing research acts are also frequently used, like *measure* and *examine*. Intentional verbs connected to purposive research acts are barely used (see example 19), and reporting verbs, linked to conclusive acts, are by far the most varied, like *report*, *show*, *meet*, *indicate*, and *suggest* (see examples 20-23). However, in examples 22 and 23, the subject is different from the nouns analysed: *findings* and *analysis*, respectively, to show there are plenty of nouns fulfilling this pattern.

Table 30. List of top instances of '*study/studies* + active verb (past simple tense)', including the number of occurrences and their frequency per million words

	***Study* + V (preterite)**	1,217	980.30	***Studies* + V (preterite)**	2,954	2,379.45
		Occ	**Fq x Mw**		**Occ**	**Fq x Mw**
1	study reported	194	156.27	studies reported	383	308.51
2	study compared	70	56.39	studies used	195	157.07
3	study found	64	51.55	studies included	182	146.60
4	study used	61	49.14	studies compared	139	111.96
5	study showed	59	47.52	studies showed	75	60.41
6	study included	59	47.52	studies met	66	53.16
7	study met	23	18.53	studies found	65	52.36
8	study demonstrated	16	12.89	studies that compared	51	41.08
9	study which measured	15	12.08	studies measured	50	40.28
10	study that compared	14	11.28	studies examined	40	32.22
11	study examined	13	10.47	studies that reported	36	29.00
12	study looked	13	10.47	studies that met	34	27.39

Table 31. List of top instances of '*study/studies* + active verb (present simple tense)', including the number of occurrences and their frequency per million words

	***Study* + V (present S)**	92	74.11	***Studies* + V (present S)**	965	777.31
		Occ	**Fq x Mw**		**Occ**	**Fq x Mw**
1	study reports	11	8.86	studies suggest	35	28.19
2	study provides	6	4.83	studies show	27	21.75
3	study shows	6	4.83	studies report	22	17.72
4	study estimates	4	3.22	studies increasingly employ	15	12.08
5	study suggests	3	2.42	studies indicate	13	10.47
6	study means	3	2.42	studies provide	12	9.67
7	study indicates	2	1.61	studies need	11	8.86
8	study demonstrates	2	1.61	studies that compare	8	6.44

Table 32. List of top instances of '*trial/trials* + active verb (past simple tense)', including the number of occurrences and their frequency per million words

	***Trial* + V (preterite)**	956	770.06	***Trials* + V (preterite)**	2,404	1,936.43
		Occ	**Fq x Mw**		**Occ**	**Fq x Mw**
1	trial reported	170	136.94	trials reported	338	272.26
2	trial compared	91	73.30	trials compared	182	146.60
3	trial included	54	43.50	trials included	179	144.18
4	trial used	47	37.86	trials used	137	110.35
5	trial found	39	31.41	trials that compared	89	71.69
6	trial showed	36	29.00	trials found	69	55.58
7	trial involved	18	14.50	trials showed	57	45.91
8	trial suggested	12	9.67	trials met	52	41.89
9	trial that compared	11	8.86	trials that included	32	25.78
10	trial stated	11	8.86	trials measured	29	23.36

Table 33. List of top instances of '*trial/trials* + active verb (present simple tense)', including the number of occurrences and their frequency per million words

	***Trial* + V (present S)**	91	73.30	***Trials* + V (present S)**	834	671.79
		Occ	**Fq x Mw**		**Occ**	**Fq x Mw**
1	trial reports	18	14.50	trials suggest	21	16.92
2	trial suggests	7	5.64	trials show	12	9.67
3	trial indicates	3	2.42	trials need	10	8.06
4	trial provides	3	2.42	trials report	10	8.06
5	trial appears	2	1.61	trials provide	9	7.25
6	trial that includes	2	1.61	trials use	9	7.25
				trials that compare	8	6.44
				trials that address	7	5.64
				trials include	7	5.64
				trials become	6	4.83

(19) Subsequently, **several studies tried** to differentiate hESCs along the keratinocyte lineage. [Introduction; la84763i] **[Impersonal active pattern; Introduction; la84763i]**

(20) Thus, **the LARGE study showed** that B16 Arg/Arg and B16 Gly/Gly patients with asthma had similar and substantial improvements in airway function when

salmeterol was added to inhaled corticosteroid therapy. **[Impersonal active pattern; Discussion; la621047d]**

(21) The hypothesis that gluten given in small amounts at 5 to 6 months of age would protect at-risk children from developing celiac disease was addressed in a randomized placebo-controlled intervention trial; however, **that study produced** null results. **[Impersonal active pattern; Discussion; ja87210d]**

(22) **These findings do not support** the use of ventilation with sustained inflations among extremely preterm infants, although early termination of the trial limits definitive conclusions. **[Impersonal active pattern; Results; ja720365r]**

(23) **Post hoc Kaplan-Meier analysis** (eFigure 4 in Supplement 3) **showed** excess mortality in the first week of life (log-rank test: first week: P=.001; entire curve: P=.11). **[Impersonal active pattern; Results; ja87623r]**

As concluded with the personal active pattern, the impersonal active structure can also be presented within a continuum showing the involvement of authors in the type of research acts being used. However, all these instances are related to the notion of impersonality and authorial invisibility, as authors do not overtly manifest themselves in the research being described but hide themselves through metonymic expressions containing an inanimate entity which performs the action described by the verb. Authorial involvement, despite not being overtly manifested, is inferred from the context. Nevertheless, it is important to point out that these authors do not really need to be the ones writing the actual RCT and may make reference to other authors connected to the field under discussion.

Indeed, reporting verbs indicating results and conclusions, essentially within the Results section, tend to be shaped using this inanimate subject plus verb in the active voice pattern, as it is a frequent way to avoid claiming full responsibility for what is being presented while impregnating the discoursal construct with an impersonal tone.

5.2.2. Passive Voice Structures

As already known, one of the main fruitful purposes of the use of the passive voice in medical discourse is to highlight the object which experiences the action rather than giving importance to its doer. The passive then serves as a means to present the ideas under discussion by giving them more prominence and by avoiding the omnipresence of authors in the text. Writers have a repertoire of passive and passive-like strategies that can be used to reinforce the notion of objectivity and neutrality characteristic of medical writing. Among these linguistic strategies, the most basic and frequent one is the prototypical finite passive, which can include or exclude the agent of the action. Other less frequent strategies related to the passive are the non-finite *as*-passive structures (also including or excluding the agent), the finite reporting passives (also including or excluding the agent), and non-finite past participial reduced relative clauses, or bare passives, which serve as post-modifiers of the nouns they usually go with. These are the linguistic strategies that will be dealt with in this sub-section.

5.2.2.1. Prototypical Passive Voice Structures (Agentless and Agentive Passives)

The prototypical passive voice pattern is a finite structure that is composed of the auxiliary *be* in any tense and a past participle form of any lexical verb. This structure is, by far, the most frequent pattern of all in medical writing and essentially serves to present information in an unbiased way. The use of the passive tends to be linked to the notions of impersonality, objectivity, empiricism, and neutrality proper of scientific discourse. In fact, the extensive number of occurrences in the passive retrieved in the RCTs corpus corroborates the set tradition which establishes the passive as prototypical in scientific medical writing.

The CQP syntax used to retrieve the **agentless** prototypical passive in the RCTs corpus is as follows:

[lemma="be"] ([pos="RR"])* ([word="not"%c])* ([pos="RR"])* [pos="V.N"]

This sub-pattern can be described as any form of the verb *be* (lemma); optionally followed by any number of adverbs (zero or more); optionally followed by the negative particle *not* and by any number of adverbs (zero or more); and followed by the past participle form of any lexical verb.

The CQP syntax used to retrieve the **agentive** prototypical passive in the RCTs is as follows:

[lemma="be"] ([pos="RR"])* ([word="not"%c])* ([pos="RR"])* [pos="V.N"] [word="by"%c] (([pos="AT"])* ([pos="JJ"])*)? [pos="N.."]

This sub-pattern can be described as any form of the verb *be* (lemma); optionally followed by any number of adverbs (zero or more); optionally followed by the negative particle *not* and by any number of adverbs (zero or more); followed by the past participle form of any lexical verb; followed by the preposition *by*, and followed by a noun (either singular or plural), which may be optionally preceded by an article and any number of adjectives (zero or more).

After presenting the main figures, I will come back to these two sub-patterns and describe the problems and limitations encountered when searching for them in the corpus.

As seen in Table 34, according to the RCTs corpus, the prototypical passive voice is highly frequent in medical writing, as it appears in 19,511 occurrences (15,716.16 instances per M words). Its extensive use is deeply revealing, particularly when comparing it against the personal active voice counterpart, which accounts for 5,282 times (4,254.66 instances per M words). The passive is pervasively used in the past

simple tense (11,929 occurrence, 9,608.77 instances per M words), followed by the present simple tense (3,192 occurrences, 2,571.16 instances per M words). Less used are the instances in the present perfect simple tense (1,138 occurrences, 917.05 instances per M words), and in the future simple tense (189 occurrences, 152.38 instances per M words).

It is worth noting that some tenses are more frequently used in particular sections of the RCT. For example, the passive voice in the past simple tense is mostly used in the Results section (5,598 occurrences, 17,668.39 instances per M words), followed by the Methods section (3,937 occurrences, 9,516.17 instances per M words) (see Table 35). This is due to the nature of these sections, as they present descriptions of the results obtained and the methods used to obtain them; the past simple tense is primarily used to indicate actions occurring at a specific moment in the past. The section in which the passive in the past simple is less frequent is the Introduction (307 occurrences, 3,100.73 instances per M words).

Table 34. Total number of occurrences of the Passive Voice structure, according to tense. Raw frequencies and frequencies per million words

Passive Voice		
Tense	**Occurrences**	**Fq x Mw**
All tenses	**19,511**	15,716.16
Past Simple sg	4,957	3,993.28
Past Simple pl	6,971	5,615.49
Past Simple	11,929	9,608.77
Present Simple sg	1,672	1,346.69
Present Simple pl	1,947	1,568.55
Present Simple	3,192	2,571.16
Present Perfect sg	493	397.30
Present Perfect pl	645	519.76
Present Perfect	1,138	917.05
Future Simple	189	152.38
Future *be going to*	2	1.72

Table 35. Total number of occurrences of the Passive Voice structure in the past simple tense, according to RAs sections. Raw frequencies and frequencies per million words

Passive Voice in the Past Simple			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	902	8,665.41
Introduction	99,009	307	3,100.73
Methods	413,717	3,937	9,516.17
Results	316,837	5,598	17,668.39
Discussion	307,806	1,185	3,849.83
TOTAL	1,241,461	11,929	9,608.03

When dealing with passive structures in the present simple tense (see Table 36), the Introduction is, however, the section in which the passive is more frequently used (902 occurrences, 9,110.28 instances per M words), followed by the Discussion section (787 occurrences, 2,556.81 instances per M words), and the Results section (606 occurrences, 1,912.66 instances per M words). When using the present perfect simple (see Table 37), the Introduction is also the section in which the passive voice is more frequently used (393 occurrences, 3,969.34 instances per M words). Both tenses, the present simple and the present perfect simple, are more frequent in the Introduction section because this section of the RCT deals with the state of affairs and the justification of the research being carried out. Through the present simple, generalisations can be made, whereas through the present perfect simple, actions occurring in the past but relevant to the present are being described.

Table 36. Total number of occurrences of the Passive Voice structure in the present simple tense, according to RAs sections. Raw frequencies and frequencies per million words

Passive Voice in the Present Simple			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	406	3,900.40
Introduction	99,009	902	9,110.28
Methods	413,717	491	1,186.80
Results	316,837	606	1,912.66
Discussion	307,806	787	2,556.81
TOTAL	1,241,461	3,192	2,571.16

Table 37. Total number of occurrences of the Passive Voice structure in the present perfect simple tense, according to RAs sections. Raw frequencies and frequencies per million words

Passive Voice in the Present Perfect Simple			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	106	1,018.33
Introduction	99,009	393	3,969.34
Methods	413,717	179	432.66
Results	316,837	160	504.99
Discussion	307,806	300	974.64
TOTAL	1,241,461	1,138	915.05

Even though the passive voice is not highly frequent in the future simple tense (see Table 38), it is worth noting that, when it appears, it does so essentially in the Methods section (110 occurrences, 265.88 instances per M words). The future simple serves to indicate an action that is expected to take place at a future time, and its frequency in the Methods section may be linked to prospective research lines that can be considered in subsequent research activities.

Table 38. Total number of occurrences of the Passive Voice structure in the future simple tense, according to RAs sections. Raw frequencies and frequencies per million words

Passive Voice in the Future Simple			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	3	28.82
Introduction	99,009	23	232.30
Methods	413,717	110	265.88
Results	316,837	26	82.06
Discussion	307,806	27	87.72
TOTAL	1,241,461	189	152.24

Table 39. Top list of lemmas, according to their tenses (Past Simple, Present Simple, and Present Perfect simple), including raw frequencies and relative frequencies per million words

	Past Simple			Present Simple			Present Perfect Simple		
	LEMMA	**Occ**	**Fq Mw**	**LEMMA**	**Occ**	**Fq Mw**	**LEMMA**	**Occ**	**Fq Mw**
1	REPORT	694	559.02	NEED	546	439.80	USE	74	59.61
2	INCLUDE	574	462.36	GIVE	351	282.73	SHOW	69	55.58
3	USE	529	426.11	PRESENT	363	292.40	REPORT	50	40.28
4	EXCLUDE	415	334.28	USE	280	225.54	SUGGEST	41	33.03
5	FIND	352	283.54	REPORT	271	218.29	PUBLISH	33	26.58
6	ASSESS	278	223.93	ASSOCIATE	202	162.71	PROPOSE	20	16.11
7	PERFORM	268	215.87	REQUIRE	197	158.68	DEVELOP	20	16.11
8	CONSIDER	250	201.38	INCLUDE	164	132.10			
9	IDENTIFY	247	198.96	BASE	150	120.83			
10	GIVE	181	145.80	DESCRIBE	143	115.19			
11	COMPARE	175	140.96	KNOW	137	110.35			
12	CONDUCT	163	131.30	CONSIDER	129	103.91			
13	SEE	148	119.21	SHOW	120	96.66			
14	CALCULATE	147	118.41	LIST	100	80.55			

The most common lemmas used in the passive cannot be easily classified according to their tense. A procedural verb denoting a neutral research act which is extremely used is the verb *use*, which is frequent in the three tenses analysed. Procedural verbs denoting cognitive research acts, like *include* or *exclude*, are also frequently used in the passive voice. Furthermore, reporting verbs which involve conclusive research acts are also frequent, like *report*, *present*, *find*, *consider*, or *show*. The lemmas that have not been coloured are quite specific to one particular tense, as they do not appear frequently in the rest of the tenses.

As stated earlier, there is an extensive number of agentless passive structures found in the RCTs corpus. Only a few have been selected as representatives of this structure. Examples 24-30 include instances of the passive in the past simple tense, and make use of verbs like *analyse*, *measure*, *find*, *use*, and *include*, whereas examples 31-36 show instances of the passive in the present simple tense (and an instance in the present perfect tense), and the verbs commonly used are *show*, *present*, *need*, *report*, and *perform*, among many others. All the examples below are instances of agentless

passives since no explicit agent is included within a *by*-phrase argument of the verb. With an agentless passive, prominence is given to the affected subject and impersonality is achieved through complete demotion of the agent. It is true, however, that even though authorial intervention is not overtly manifested, the involvement of authors can be easily inferred through the context.

(24) Data for the primary end point **were analysed** on an intention-to-treat and per control basis **[Agentless prototypical passive; Methods; an142165m]**

(25) A fasting blood sample was taken at each annual examination, from which serum glucose concentration and plasma lipid levels **were measured**. **[Agentless prototypical passive; Methods; an142313m]**

(26) Somewhat weaker ORs **were found** for being female (1.0-2.9, significant for 1 outcome) and for being a homemaker (1.2 - 2.9, significant in only 1 outcome). **[Agentless prototypical passive; Results; ja114320r]**

(27) 264 units of blood products **were used** in the first 15 h of the major incident including a total of 130 units of packed red blood cells, 46 units of fresh frozen plasma, 70 units of cryoprecipitate, and 11 pools of platelets. Typical blood product usage runs at around 240 units of all blood products per day. **[Agentless prototypical passive; Results; la476231r]**

(28) We found that meta-analyses of prevalence estimates were similar when generalised linear mixed models or standard inverse variance aggregation methods **were used**. **[Agentless prototypical passive; Results; bm310045r]**

(29) Because 97% of primary articles examined by the excluded studies were at high risk of overlap with studies that **were included** in the analysis, however, we do not believe that their omission would have substantially changed our findings. **[Agentless prototypical passive; Discussion; bm876240d]**

(30) In one recent study, regulatory approvals for first indications **were found** to be more likely than those for supplemental indications to be based on at least two pivotal trials. **[Agentless prototypical passive; Discussion; bm461087d]**

(31) The PSI accurately stratifies patients into 5 risk classes with 30-day mortality rates ranging from 0.1% in class I to 27.0% in class V. This index **is viewed** as an objective measure of risk stratification to help determine the initial site of care for

community-acquired pneumonia. **[Agentless prototypical passive; Introduction; an142165i]**

(32) Immunogenicity data **are reported** for up to 76 months after first vaccination. **[Agentless prototypical passive; Methods; la179654m]**

(33) To date, no similar randomized study **has been performed** to test whether low-risk patients with community-acquired pneumonia who **are treated** as outpatients would have outcomes equivalent to those of comparable patients who **are hospitalized**. **[Agentless prototypical passive; Discussion; an142165d]**

(34) Parameters calculated on a segmental basis **are shown** in TABLE3. A sample case without stenosis **is presented** in FIGURE 2. Cases with matched positive readings **are presented** in FIGURE 3. **[Agentless prototypical passive; Results; ja143652r]**

(35) Efforts **are needed** to identify optimal interventions for primary and secondary prevention of suicidality. **[Agentless prototypical passive; Discussion; ja114320d]**

(36) The major limitation to the clinical interpretation of the study is that blood pressures after the trial **are not known**. **[Agentless prototypical passive, Discussion, an142342d]**

Once the most relevant quantificational data on the prototypical passive have been introduced, and as pointed out earlier, I will now come back to the two distinct syntactic expressions used to search for the two types of passive structures analysed in this section: agentless passives and agentive passives.

Table 40. Syntactic expressions of the Agentless Passive and the Agentive Passive, using CQP syntax

Agentless Passive
[lemma="be"] ([pos="RR"])* ([word="not"%c])* ([pos="RR"])* [pos="V.N"]
Any form of the verb *be* (lemma); optionally followed by any number of adverbs (zero or more); optionally followed by the negative particle *not* and by any number of adverbs (zero or more); followed by the past participle form of any lexical verb.
Agentive passive
[lemma="be"] ([pos="RR"])* ([word="not"%c])* ([pos="RR"])* [pos="V.N"] [word="by"%c] (([pos="AT"])* ([pos="JJ"])*)? [pos="N.."]
Any form of the verb *be* (lemma); optionally followed by any number of adverbs (zero or more); optionally followed by the negative particle *not* and by any number of adverbs (zero or more); followed by the past participle form of any lexical verb; **followed by the preposition *by*, and followed by a noun (either singular or plural), which may be optionally preceded by an article and any number of adjectives (zero or more).**

As can be seen in Table 40, the syntax used to retrieve the agentless passive coincides with the syntax used to retrieve the agentive one (the part which is not in blue). In blue one can identify the addition of the agent by means of a *by*-phrase. The problem here is that there is no way to restrict the agentless pattern to avoid the occurrences showing the agent. It is possible, though, to only search for the agentive pattern to restrict those instances in which the agent is not mentioned (see Table 41), but another problem arises. When doing so, the *by*-phrase can be found to introduce a nominalisation, where authorial intervention is not overtly manifested but is inferred through the context.

Table 41. Total number of occurrences of the Agentive Passive Voice structure, according to RAs sections. Raw frequencies and frequencies per million words

Agentive Passive Voice			
Distribution	**Words**	**Occ.**	**Fq x Mw**
Abstract	104,092	45	432.30
Introduction	99,009	75	757.51
Methods	413,717	623	1,505.86
Results	316,837	114	359.81
Discussion	307,806	260	844.69
TOTAL	1,241,461	1,117	899.75

The passive including the agent is much less frequent than the agentless passive (bearing in mind that agentless passive's figures actually include those accounting for the agentive passive), as there are only 1,117 occurrences in the RCTs corpus, which implies a normalised frequency of 899.75 instances per M words. The agentive passive is much more frequent in the Methods section (623 occurrences; 1,505.86 instances per M words), distantly followed by the Discussion section (260 occurrences; 844.69 instances per M words).

It is worth noting that when the agent is included in the passive, it rarely makes reference to the authors of the paper; if authors wish to overtly express their involvement in the research being presented, they make use of the personal active voice instead. If they wish to focus on what has been done without explicitly mentioning themselves, they can make use of the agentless passive or other linguistic devices where agency is totally or partially demoted. Taking this into account, the agentive passive subtly provides a sense of impersonality insofar as it foregrounds the object of study and backgrounds the agent performer of the action. Yet, this structure does not serve to analyse authorial manifestation and authorial visibility, because it has nothing to do with the authors of the RCTs, but with other agents somehow involved in the research being presented (except for example 43).

This can be seen in the number of examples, (37-46), encountered in the RCTs corpus. In these examples, the human agents (37-43) or the stacked nominal phrases hiding human agency (44-46) are underlined.

(37) Social desirability bias, defined as over-reporting of condom use because that is what is perceived **to be desired by the interviewer**, also contributes to reduced estimates of effectiveness of condoms. **[Agentive prototypical passive; Discussion; an143707d]**

(38) If the problem was deemed non-modifiable and full recovery was unrealistic, the patient stayed on temporary work disability and follow-up **was done by the patient's primary care physicians**. **[Agentive prototypical passive; Methods; an143404m]**

(39) Race and ethnicity **were assessed by the study coordinator** from patients and chart information to determine the degree of diverse representation in the study population. **[Agentive prototypical passive; Methods; ja2941625m]**

(40) Main Outcome Measures. Gallbladder and gallstone disease outcomes **were directly reported by the participants** in the semiannual safety monitoring questionnaire. **[Agentive prototypical passive; Methods; ja293330m]**

(41) However, according to Centers for Disease Control and Prevention criteria, infection can be present without laboratory confirmation and, in our study, the blinded wound evaluator considered any of the following as confirmation of infection: purulent drainage, with or without laboratory confirmation; organisms isolated from an aseptically obtained culture of fluid or tissue; at least 1 of the following signs or symptoms of infection (pain or tenderness, localized swelling, redness, or heat, and the incision **was deliberately opened by surgeon**, unless incision was culture-negative); or independent diagnosis of incisional SSI by the surgeon or attending physician. **[Agentive prototypical passive; Discussion; ja2942035d]**

(42) Wound infections **were diagnosed by blinded investigators** using Centers for Disease Control and Prevention criteria. **[Agentive prototypical passive; Abstract; ja2942035a]**

(43) Data **were collected by the pharmaceutical sponsor** and **were assessed jointly by the authors and the sponsor**. **[Agentive prototypical passive; Methods; la365139m]**

(44) Exenatide injection **is approved by the U.S. Food and Drug Administration** as adjunctive therapy to improve glycemic control in patients with type 2 diabetes who are taking metformin, a sulfonylurea, or a combination of metformin and a sulfonylurea but have not achieved adequate glycemic control. **[Agentive prototypical passive; Introduction; an143559i]**

(45) The study **was supported by the National Institutes of Health (NIH)** and **was approved by the institutional review boards** at each of the participating institutions, with each patient providing informed consent to participate in the substudy. **[Agentive prototypical passive; Methods; an143714m]**

(46) This strategy **is suggested by previous studies** that have shown that physicians are often not aware of their patients AE preferences, and that simple interventions can improve communication. **[Agentive prototypical passive; Methods; ja294211m]**

As stated earlier, when the *by*-phrase is included in the pattern by means of the syntactic expression **[word="by"%c] (([pos="AT"])* ([pos="JJ"])*)? [pos="N.."]**, another problem arises. Instead of human agents, nominalisations and stacked nominal phrases are also retrieved as if belonging to the aforementioned agentive pattern, but in this case, the preposition *by* does not include the agent of the action, but the manner in which something is done. The noun cannot be considered an agent, and the preposition *by* can be substituted for *through*. In this case, then, the passive is not agentive but agentless, as the argument introduced by *by* is an adjunct of manner (see examples 47-51).

(47) Comparison of microfilaraemia between doxycycline and placebo groups at baseline and subsequent follow-up assessment **were analysed by Wilcoxon rank sum tests** (Mann-Whitney U tests) of raw data. **[Agentless prototypical passive; *by* (manner) + stacked nominal phrase; Results; la3652116r]**

(48) Staff awareness of the introduction of the MET system **was maintained by the use of regular reminders** until the first day of the study period, after which awareness and education became the responsibility of the individual hospitals. **[Agentless prototypical passive; *by* (manner) + nominalisation; Methods; la3652091m]**

(49) Unstable angina **was defined by hospitalization** for new chest pain or changes in the patient's usual chest pain in combination with electrocardiographic ST-T wave changes indicative of ischemia or accompanied by elevations in enzyme markers not

diagnostic of myocardial infarction. **[Agentless prototypical passive; *by* (manner) + nominalisation; Methods; ne3521637m]**

(50) Because study participants were randomly assigned to receive either 0 mg of atorvastation or 40 mg of pracastatin daily, we had the additional opportunity to assess the relative effect of these two agents on the reduction in CRP levels and to assess whether the main effects observed in the total cohort according to LDL cholesterol and CRP levels achieved **were modified by the choice of statin therapy**. **[Agentless prototypical passive; *by* (manner) + nominalisation; Results; ne35220r]**

(51) These changes **were reflected by increases** in parathyroid hormone levels in both groups at three months (16.31 pg per milliliter in the zoledronic acid group and 11.19 pg per milliliter in the risedronate group) and six months (7.15 pg per milliliter and 2.12 pg per milliliter, respectively), and the changes were significantly greater at six months in the zoledronic acid group (P=0.002). **[Agentless prototypical passive; *by* (manner) + nominalisation; Results; ne353898r]**

Another limitation when accounting for the search of the agentive pattern is that the part of the syntactic expression introducing the noun (shown in blue) can include nouns with *-ing* suffixes derived from verbs:

[word="by"%c] (([pos="AT"])* ([pos="JJ"])*)? [pos="N.."]

Therefore, this allows the retrieval of instances in which the preposition *by* introduces an adjunct of manner that is expressed through a present participial form. In order to identify the sub-patterns which include a gerund, I made use of the following syntactic expression:

Table 42. CQP Syntactic expression for the retrieval of the Agentless Prototypical Passive, followed by *by* + gerund form. Description of the pattern

Agentless Prototypical Passive, followed by an adjunct of manner (*by* + V*ing*)
[lemma="be"] ([pos="RR"])* ([word="not"%c])* ([pos="RR"])* [pos="V.N"] [word="by"%c] ([pos="RR"])* [pos="VVG"]
Any form of the verb *be* (lemma); optionally followed by any number of adverbs (zero or more); optionally followed by the negative particle *not* and by any number of adverbs (zero or more); followed by the past participle form of any lexical verb; followed by the preposition *by*, followed by any number of adverbs (zero or more), and followed by a past participial form of any lexical verb.

Table 43. Total number of occurrences of the Agentless Passive voice structure, followed by *by* + present participial form, according to RAs sections. Raw frequencies and frequencies per million words

Agentless Passive Voice + *by* V*ing*			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	2	432.30
Introduction	99,009	5	757.51
Methods	413,717	116	1,505.86
Results	316,837	14	359.81
Discussion	307,806	17	844.69
TOTAL	1,241,461	154	899.75

This agentless sub-pattern is not highly frequent. It appears 154 times (899.75 instances per M words). When it appears, it does so essentially within the Methods section (116 occurrences, 1,505.86 instances per M words). Although the raw frequencies in the Discussion and the Introduction are not very high, the normalised frequencies per M words are worth considering (17 occurrences, 844.69 instances per M words; and 5 occurrences, 757.51 instances per M words; respectively) (see Table 43). The following contextualised examples (52-58) illustrate this sub-pattern, which essentially includes as gerunds procedural verbs related to neutral research acts.

(52) To assess differences in the frequency of outcomes in the 2 treatment groups, relevant variables **were compared by using** the Fisher exact test, overall and by hospital. **[Agentless prototypical passive, *by* (manner) + gerund; Methods; an142165m]**

(53) These analyses **were performed by using** SPSS for Macintosh, version 10 (SPSS, Inc., Chicago, Illinois). **[Agentless prototypical passive, *by* (manner) + gerund; Methods; an142165m]**

(54) Finally, the proportional hazards assumption of the Cox regression models **was checked by testing** the interaction of each baseline factor with duration of follow-up. **[Agentless prototypical passive, *by* (manner) + gerund; Methods; an142342m]**

(55) The total number of pain medications reported by participants in these 2 groups **was compared by using** generalized estimating equation models with an unstructured correlation matrix and the robust variance estimator at week 1 and after week 12. **[Agentless prototypical passive, *by* (manner) + gerund; Discussion; an14233d]**

(56) In the current study, hypoglycaemia **was fairly well controlled by reducing** the dose of sulfonylurea, was rarely severe, and did not cause patient discontinuation in either treatment group. **[Agentless prototypical passive, *by* (manner) + gerund; Discussion; an143559d]**

(57) School attendance **was calculated by dividing** the hours that the patient attended lessons in the previous week by the hours that the patient should have attended. **[Agentless prototypical passive, *by* (manner) + gerund; Methods; bm33014m]**

(58) Setup error **was minimized by obtaining** daily portal images throughout the first phase of treatment, imaging the bones and also the metal markers within the Lucite probe that lay against the anterior rectal wall. **[Agentless prototypical passive, *by* (manner) + gerund; Discussion; an142342d]**

Considering all the quantificational information and the qualitative analysis carried out in this section, it is undeniable to assert that the extensive use of the passive voice identifies it as a staple in medical writing. The passive voice serves to represent an important layer of theme deployment as it foregrounds new information and backgrounds, and even omits, information which is already known, such as the agency of the actions being presented. This shift of focus, which highlights the research itself

and its main results, is what really reflects the impersonal character of any passive voice structure. The summative effect of these linguistic resources radiates a notorious sense of detachment and impersonality, which extends onto a sense of objectivity, impartiality, and empiricism.

The tenses in which passives occur are varied; however, the high frequency of passives in the past simple tense reinforces the idea that the passive is used to essentially express research procedures carried out and research results obtained in the research process. When reporting verbs are used, for instance, the passive tends to intensify the sense of detachment, as authors may present the information without taking full responsibility, and this is one main function the passive is thought to perform. As already seen, the lemmas used in the passive are also varied, and they are difficult to be semantically classified. As the passive may serve different functions, there is no clear relationship between the tenses used, the lemmas included, and the semantic field to which they belong. It is then difficult to create a continuum such as the one presented with the personal active voice, because the main function of this personal pattern is clearly to identify, above all, the doer of the action, presenting known information (thematic information) in first position, and new information (rhematic information) afterwards.

Indeed, it is important to point out, however, that when analysing the passive using a corpus-based methodology, some important limitations arise, and these do not allow a purely refined analysis of the passive in medical discourse. As seen in this section, the agentive passive, which is more restricted, is included when one searches for the agentless passive, which is less restricted. A specific syntactic expression must be used to retrieve the agentive passive, which would include the agent. However, when retrieving it, there is no way to restrict the typology of nouns appearing within the *by*-phrase; they could be human-related nouns (agents of the action, usually different from the researchers themselves) or non-human-related nouns, which appear in the form of stacked nominal phrases or nominalisations, which at the same time imply human agency.

Apart from this fact, in some instances, the preposition *by* can be substituted for *through*, and this means that an adjunct of manner is being introduced. These instances including a *by* adjunct expressing manner are always agentless, and they cannot be then counted as agentive. When searching for this pattern, some of the nouns appearing after *by* are nominalisations derived from verbs and show the inflection proper of the present participial form, the *-ing*. The number of occurrences which include the preposition *by* followed by a gerund form of a verb also relate to agentless passives.

It is undeniable that a corpus-based methodology serves to deal with the passive voice in general, because it allows to retrieve a huge number of occurrences as well as to collect a huge amount of quantificational information. It also helps determine the typology of lemmas found, the tenses in which the passive is frequently used, and the sections in which the passive voice is extremely frequent, among many other relevant aspects. However, when it comes to relate the passive voice to the notions of authorial (in)visibility, the explicit mention or omission of agency, and the notion of impersonality they transmit, the limitations posed by a corpus-based study are patent.

5.2.2.2. *As*-Passive Voice Structures (Agentless and Agentive Passives)

An *as*-passive is a non-finite form of a passive structure. This pattern is formed of the conjunction *as*, which expresses the meaning of ‘in the way that’, and the past participle form of any lexical verb. The omission of both the subject, in most cases the personal pronoun *it*, and the finite form of the auxiliary verb *be* prototypical of the passive voice, cause the non-finiteness of the structure. It is possible to state that this pattern is an alternative to the passive voice and hence its inclusion in this corpus-based analysis.

The syntactic expressions used to search for this pattern in the RCTs corpus are as follows:

Table 44. Syntactic structures of the *as*-Passive structures, using Simple Query Syntax and CQP Syntax.

***as*-Passive Voice**
as (_{ADV})* _V?N (_{ADV})* (if SQS is used) **[word="as"%c] ([pos="RR"])* [pos="V.N"] ([pos="RR"])*** (if CQPS is used)
The conjunction *as*, optionally followed by any number of adverbs (zero or more), followed by the past participle form of any lexical verb (including the non-auxiliary verbs be, do, and have), optionally followed by any number of adverbs (zero or more).

Even though this pattern seems to be a good alternative to the prototypical passive voice, it is worth highlighting it is not extremely frequent (see Table 45). Indeed, the total amount of occurences in the RCTs corpus is 1,093, which accounts for a normalised frequency of 889.41 instances per M words. As the prototypical passive, the *as*-passive is essentially found within the Results section (290 occurrences, 1,230.92 instances per M words), followed by the Discussion section (260 occurrences, 844.69 instances per M words), and followed by the Methods section (307 occurrences, 742.05 instances per M words).

Table 45. Total number of occurrences of the *as*-Passive voice structure, according to RAs sections. Raw frequencies and frequencies per million words

***as*-Passive Voice**			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	94	903.05
Introduction	99,009	42	424.20
Methods	413,717	307	742.05
Results	316,837	390	1,230.92
Discussion	307,806	260	844.69
TOTAL	1,241,461	1,093	880.41

As access to the corpus was not possible at one point in time, information on the most frequent lemmas used in this pattern could not be retrieved. However, bearing in mind the instances found in the RCTs corpus (see examples 59-66), one can state that

reporting verbs denoting conclusive research acts are usually found in this structure, like *as found*, *as observed*, and *as described*. Procedural verbs like *prescribe* and *compare* also tend to appear.

Indeed, it is worth noting that most of these structures serve as guiding discourse indicators related to time, for example when referring to an idea that has been already dealt with in the discourse, like *as described previously* (example 60); or as guiding discourse indicators related to place, for example to refer to the physical space where an aspect, a fact, or an idea is dealt with, like *as shown in Figure 3* (example 62) or *as observed in the present study* (example 63).

(59) Older patients can also have considerable problems with adhering to their drug regimens: up to 50% of prescribed drugs are estimated to be not taken **as prescribed**. **[Agentless *as*-passive; Introduction; bm33093i]**

(60) We measured levels of total and high-density lipoprotein cholesterol, triglycerides, glucose, apolipoprotein B, creatinine, and serum aspartate and alanine aminotransferases by using standard laboratory procedures, **as described** previously in detail. **[Agentless *as*-passive; Methods; an143337m]**

(61) In the planned stratified Cox regression analysis, significant factors that were associated with improved survival included chemotherapy **as compared** with observation (hazard ratio for the difference survival, 0.67; 95 percent confidence interval, 0.51 to 0.9; P=0.006) and squamous histologic features **as compared** with adenocarcinomas (P=0.0005). **[Agentless *as*-passive; Results; ne352589r]**

(62) **As shown** in Figure 3, tiotropium fairly uniformly reduced exacerbations for all subsets included in the analyses. **[Agentless *as*-passive; Results; an143317r]**

(63) The main assumptions behind the estimates were conservative: that mortality not due to abdominal aortic aneurysm is unaffected by screening and that screening has a net effect on mortality due to abdominal aortic aneurysm only between 1.5 and five years after screening, with difference in mortality due to such aneurysms per 1000 years of 0.9 (0.40 to 1.37), **as observed** in the present study. **[Agentless *as*-passive; Discussion; bm330750d]**

(64) **As expected**, and consistent with their improved glycemic control, patients in the inhaled insulin plus 2 oral agents group and the inhaled insulin monotherapy group

had more hypoglycemic events than those in the dual oral agent therapy group. **[Agentless *as*-passive; Discussion; an143549d]**

(65) Reductions in heavy drinking, **as observed** in this study with long-acting naltrexone, can be expected to lead to improvements in various areas of health and in the quality of life in alcohol-dependent patients although direct evaluation of these outcomes is needed. **[Agentless *as*-passive; Discussion; ja2931617d]**

(66) It is possible that the effect of increased access on pregnancy rates is truly negligible because EC is not as effective **as found** in the single-use clinical trials, or because women at highest risk do not use EC frequently enough or at all. **[Agentless *as*-passive; Discussion; ja29354d]**

As seen in Table 46, the syntax used to retrieve the agentless *as*-passive coincides with the syntax used to retrieve the agentive one (the part which is not in blue). In blue one can identify the addition of the agent by means of a *by*-phrase, but there is no way to restrict the agentless pattern to omit the occurrences showing the agent. If the agent is explicitly searched for, the number of occurrences radically decreases.

Table 46. Syntactic expressions of the Agentless Passive and the Agentive Passive (CQP Syntax)

Agentless *as*-Passive Voice
[word="as"%c] ([pos="RR"])* [pos="V.N"] ([pos="RR"])*
The conjunction *as*, optionally followed by any number of adverbs (zero or more), followed by the past participle form of any lexical verb (including the non-auxiliary verbs be, do, and have), optionally followed by any number of adverbs (zero or more).
Agentive *as*-Passive Voice
[word="as"%c] ([pos="RR"])* [pos="V.N"] ([pos="RR"])* **[word="by"%c] (([pos="AT"])* ([pos="JJ"])*)? [pos="N.."]**
The conjunction *as*, optionally followed by any number of adverbs (zero or more), followed by the past participle form of any lexical verb (including the non-auxiliary verbs be, do, and have), optionally followed by any number of adverbs (zero or more); **followed by the preposition *by*, and followed by a noun (either singular or plural), which may be optionally preceded by an article and any number of adjectives (zero or more).**

The *as*-passive including the agent appears a total number of 179 occurrences (144.19 instances per M words) (see Table 47). Like the rest of the passive structures analysed, the section in which this pattern is most frequently used is the Methods (79 occurrences, 190.95 instances per M words), followed by the Results (44 occurrences, 138.87 instances per M words), and the Discussion section (41 occurrences, 133.20 instances per M words).

Table 47. Total number of occurrences of the Agentive *as*-Passive Voice structure, according to RAs sections. Raw frequencies and frequencies per million words

Agentive *as*-Passive Voice			
Distribution	**Words**	**Occ.**	**Fq x Mw**
Abstract	104,092	6	57.64
Introduction	99,009	9	90.90
Methods	413,717	79	190.95
Results	316,837	44	138.87
Discussion	307,806	41	133.20
TOTAL	1,241,461	179	144.19

The following examples (67-71) include a *by*-phrase, however, not all of them introduce the agent of the action. In examples 67 and 68, the doers of the action are explicitly mentioned through a *by*-phrase. However, they do not refer to the authors of the text, but to the intervention of other agents in the research process being described. The other three examples (69-71) use a *by*-phrase which include a nominalisation, so agency is demoted and hidden through the nominalised structure. At the same time, these same instances can be considered as being agentless, as the preposition *by* introduces an adverbial adjunct expressing manner.

(67) Patients' compliance was 100% **as assessed by the study coordinators and hospital-based and home-care nurses** who help to administer the intravenous antibiotics, and follow-up was complete for all randomised patients. **[Agentive *as*-passive; Results; la366463r]**

(68) Of three secondary outcome measures, the first was clinical global improvement during on-time, **as measured by the examiner**, at week 1 compared with baseline

(scale: -3 = markedly improved; -2 = moderately improved; -1 = minimally improved; 0 = unchanged; 1 = minimally worse; 3 = markedly worse). **[Agentive *as*-passive; Methods; la365947m]**

(69) These characteristics place them in the most severe kidney disease group, **as defined by large demographic studies**. **[Agentive or Agentless *as*-passive; Discussion; an143777d]**

(70) One study concluded that smoking **as measured by self report** was likely to be inconsistent and that in the future biochemical validation would yield more reliable data. **[Agentive or Agentless *as*-passive; Methods an143777m]**

(71) Volunteers had to be in good health **as confirmed by physical examinations** at screening and routine laboratory tests within normal limits. **[Agentive or Agentless *as*-passive; Methods; la36539m**]

If the sub-pattern *as*-passive in which the agent is excluded and in which the *by* is followed by an adjunct of manner in its present participial form is searched for, a very few instances are retrieved from the RCTs corpus (see Tables 48 and 49, and examples 72-73).

Table 48. CQP Syntactic expression for the retrieval of the Agentless *as*-Passive Passive, followed by *by* + gerund form. Description of the pattern

Agentless *as*-Passive Voice, followed by an adjunct of manner (*by* + V*ing*)
[word="as"%c] ([pos="RR"])* [pos="V.N"] ([pos="RR"])* [word="by"%c] ([pos="RR"])* [pos="VVG"]
The conjunction *as*, optionally followed by any number of adverbs (zero or more), followed by the past participle form of any lexical verb (including the non-auxiliary verbs *be*, *do*, and *have*), optionally followed by any number of adverbs (zero or more); followed by the preposition *by*, followed by any number of adverbs (zero or more), and followed by a past participial form of any lexical verb.

(72) Patient utility **as estimated by using** the EuroQol EQ-5D questionnaire at several time points and used to calculate quality adjusted life years (QALYs). **[Agentless *as*-passive; Abstract; bm3301239a]**

(73) Inclusion criteria for the trial were an age at recruitment of 3-6 years, stuttering **as diagnosed by using** standard procedures17 and at least 2% of syllables stuttered, and proficiency in English for children and parents. **[Agentless *as*-passive; Methods; bm331659m]**

Table 49. Total number of occurrences of the Agentless *as*-Passive structure, followed by *by* + present participial form, according to RAs sections. Raw frequencies and frequencies per million words

Agentless *as*-Passive Voice + *by* V*ing*			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	1	9.61
Introduction	99,009	0	0
Methods	413,717	1	2.42
Results	316,837	1	3.16
Discussion	307,806	0	0
TOTAL	1,241,461	3	2.42

Bearing in mind the information presented in this quantitative and qualitative analyses of the *as*-passive, it is possible to assert that this structure may serve as an alternative to the prototypical passive voice. However, its low frequency shows authors do not make extensive use of it and keep it, essentially, as a guiding indicator showing temporal and locative information to which the text refers.

An interesting characteristic about this particular pattern is that it syntactically operates at a different level if compared to the prototypical passive. If followed or preceded by a comma, no clausal argument is topicalised and full relevance is given to the lexical verb form. In this case, new or rhematic information tends to appear after the pattern, and the pattern merely serves to signposting (see examples 62 and 63). In the rest of the examples, the non-finite *as*-passive construction post-modifies a noun (for instance, see examples 67 and 72).

The same limitations observed in the prototypical passive voice, as far as the retrieval of patterns is concerned, are also encountered when retrieving the *as*-passive voice structure.

To conclude, the effect this pattern has over the text is the same as the passive as far as the transmission of impersonality is concerned. Yet, its much lower frequency diminishes its impersonalising influence over medical writing. What is true, however, is that the summative effect of these patterns together with the extensive use of the passive voice contribute to the impersonal dimension of medical writing.

5.2.2.3. Reporting Passive Structures

The reporting passive structure is an impersonal finite construction, which is introduced by the third personal singular pronoun *it*, followed by the finite verb form of *be*, and followed by a reporting lexical verb in its past participle form (see Table 50). This type of pattern may serve as an alternative to the passive voice when reporting facts; however, its low frequency overall in the RCTs corpus suggests that authors make use of other linguistic strategies when it comes to reporting.

Table 50. CQP Syntactic expression for the retrieval of the Reporting Passive.

Description of the pattern	
Reporting Passive	
[word="it"%c] ([pos="RR"])* [lemma="be"] ([pos="RR"])* [pos="V.N”]	
The personal pronoun *it*, optionally followed by any number of adverbs (zero or more), followed by the lemma *be*, optionally followed by any number of adverbs (zero or more); followed by the past participial form of any lexical verb.	

As for the frequency, 856 occurrences of this pattern appear in the RCTs, which account for a normalised frequency of 689.51 instances per M words. As already seen with the rest of the passives, the section in which this pattern is more highly frequent is the Results section (220 occurrences, 694.36 instances per M words), followed by the Introduction section (211 occurrences, 2,131.12 instances per M words) (see Table 51).

Table 51. Total number of occurrences of the Reporting Passive structure, according to RAs sections. Raw frequencies and frequencies per million words

Reporting Passive			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	73	701.30
Introduction	99,009	211	2,131.12
Methods	413,717	191	461.67
Results	316,837	220	694.36
Discussion	307,806	161	523.06
TOTAL	1,241,461	856	689.51

However, it is worth noting that this pattern also poses some problems when it comes to searching for it. The first issue has to do with the pronoun *it*. The pronoun *it* must not include a referent to which it substitutes. It is an empty grammatical pronoun with no reference in the discourse. The second problem is related to the typology of verbs included in this pattern, because they must be only reporting verbs. To my knowledge, there is no way to automatically retrieve this pattern without including agentless passives whose subject is filled by the pronoun *it*, which would recall to known or thematic information that has been previously presented in the discourse. That is why examples, like 74, may appear as belonging to this category, despite being prototypical agentless passives whose subject is the pronoun *it*.

(74) **It is associated** with several develop-mental comorbid conditions (for example, communication impairment) and the potential to restrict activities of daily living. **[Agentless prototypical passive; Introduction, an236572i]**

To delimit the syntactic expression to only retrieve real reporting patterns, the addition of the word *that* after the pattern would restrict the queries and ensure the retrieval of only reporting passives (like in examples 75 and 76). When including the word *that*, the list of reporting *it*-passive structures seen in Table 52 appears. As clearly perceived, the typology of verbs used within the reporting pattern somehow imply human agency, and authorial visibility is shown as demoted. Authorial intervention is hidden through the use of this pattern.

Table 52. Most common verbs used within the Reporting Passive structure.
Raw frequencies and frequencies per million words

	Reporting Passive	Occ	Fq x Mw
1	it was reported that	41	33,03
2	it was planned that	32	25,78
3	it was assumed that	26	20,94
4	it is not known that	25	20,14
5	it was decided that	23	18,53
6	it was found that	21	16,92
7	it is estimated that	17	13,69
8	it was stated that	17	13,69
9	it is associated that	14	11,28
10	it is thought that	14	11,28
11	it is believed that	13	10,47
12	it was intended that	12	9,67

(75) As mentioned in the methods, **it was assumed that** all hospitalized patients suffered from severe dyspnea. **[Reporting passive; Results; bm135674r]**

(76) To use this table, **it was assumed that** age was evenly distributed within the 15- to 19-year age bracket. Therefore, because 6.8% of U.S. males were aged 15 to19 years, **it was assumed that** 2.72% males were aged 18 to 19 years [6.8% × (2/5)]. Similarly, because 6.3% of U.S. females were aged 15 to 19 years, **it was assumed that** 2.64% females were aged 18 to 19 years [6.6% × (2/5)]. **[Reporting passive; Methods; an431009m]**

5.2.2.4. Past Participial Reduced Relative Clauses

Reduced relative clauses are non-finite structures which depend on other elements in the clause, usually nouns. These are considered *reduced* because they lack the relative pronoun and the auxiliary verb *be*, which if stated would transform these non-finite structures into finite ones. Their prototypical function is that of post-modifiers of the noun they accompany, as they are used to semantically restrict its sense. As they are embedded in the noun phrase, they contribute to package a huge amount of specific and descriptive information. The past participial reduced relative clause, also known as *bare passive*, is the non-finite version of a finite relative clause introduced by a subject (the pronoun) and the verb in its passive form.

Table 53. Syntactic expressions of the Past Participial Reduced Relative clause, using SQS and CQP syntax

Past Participial Reduced Relative Clause
((_AT)* (_JJ)*)? _N?? (_RR)* _V?N (if SQS is used) **(([pos="AT"])* ([pos="JJ"])*)? [pos="N.."] ([pos="RR"])* [pos="V.N"]** (if CQPS is used)
Noun (either singular or plural), optionally preceded by an article and any optional number of adjectives (zero or more); optionally followed by any number of adverbs (zero or more); followed by the past participle form of any lexical verb (including non-auxiliaries *be*, *do*, and *have*).

Undoubtedly, the past participial reduced relative clause is a very frequent linguistic resource used by medical authors, as can be seen in Table 24. The total number of occurrences is 6,779, which accounts for a normalised frequency of 5,460.50 instances per M words. This non-finite form is essentially used within the Methods section (2,5052 occurrences, 4,959.91 instances per M words), followed by the Results section (1,872 occurrences, 5.908.40 instances per M words), and followed by the Discussion section (1,749 occurrences, 5,682.15 instances per M words). However, it is important

to highlight that the highest relative frequency is found within the Abstract section (626 occurrences, 6.013.91 instances per M words).

Table 54. Total number of occurrences of the Past Participial Reduced Relative Clause structure, according to RAs sections. Raw frequencies and frequencies per million words

Past Participial Reduced Relative Clause			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	626	6,013.91
Introduction	99,009	480	4,848.04
Methods	413,717	2,052	4,959.91
Results	316,837	1,872	5,908.40
Discussion	307,806	1,749	5,682.15
TOTAL	1,241,461	6,779	5,460.50

The following examples (77-83) show real instances of the past participial reduced relative clause. In all of them, the lexical past participle form post-modifies a noun, which is shown as underlined. It is interesting to point out that this pattern is certainly ubiquitous, as it can be included almost within any clausal argument: it could be part or the subject (like in 77 and 78, for instance), or part of a direct object (like in example 82).

(77) **Direct costs associated** with the disability were estimated to be around 1.6bn in the United Kingdom in 1999, and the condition is estimated to be responsible for close to 120 million UK work days lost per year. **[Past participial reduced relative clause; Introduction; bm3301239i]**

(78) **The 117 participants recruited** in this pilot are included in the main trial report. **[Past participial reduced relative clause; Methods; bm3301003m]**

(79) In the intention-to-treat analysis, the overall odds ratio for **successful outcome linked to** treatment group (outpatient care vs. hospitalization) was 0.76 (CI, 0.37 to 1.54) for both hospitals combined, 0.7 (CI, 0.33 to 1.6) for IDIBELL-Hospital Universitari de Bellvitge, and 0.75 (CI, 0.1 to 3.12) for SCIAS-Hospital de Barcelona. **[Past participial reduced relative clause; Results; an142165r]**

(80) Since death rates between special intervention and usual care participants with similar smoking habits did not differ, **the differences observed** in the groups as a whole were almost certainly due to differential cessation rates. **[Past participial reduced relative clause; Discussion; an142233d]**

(81) Fourth, **patients assigned to** combination treatment had higher medial ALT levels at baseline than **patients assigned to** monotherapy. **[Past participial reduced relative clause; Discussion; an142240d]**

(82) This study found that **patients treated** in hospital had higher rates of confusion and urinary and bowel complications. **[Past participial reduced relative clause; Discussion; bm330129d]**

(83) Structured programme of warm-up exercises to prevent injuries. **The exercises used** in the programme were developed on the basis of previous intervention studies in team handball and other sports, and had been feasibility tested and modified to be suitable for team handball. **[Past participial reduced relative clause; Discussion; bm330449d]**

Nevertheless, when searching for this pattern in the RCTs corpus, some minor problems may also be encountered. The first problem consists of the retrieval of some active instances which share a similar pattern (see examples 84 and 85). The pattern for the past participial reduced relative clause includes the past participle form of any lexical verb. However, when they are regular verbs, past participle forms and preterite forms show syncretism, that is, they share the same shape. As past participle forms of (some) regular verbs seem not to be properly tagged, the result ends up being an active instance whose subject (underlined) performs the action expressed by the lexical verb.

(84) **Researchers recruited** participants from the community using a wide variety of techniques. **[Active voice pattern; Methods; an142233m]**

(85) **Coronary heart disease accounted** for 77 deaths (10.5%), and cardiovascular disease including **coronary heart disease accounted** for 163 deaths (22%). **[Active voice pattern, Results; an142233r]**

A second minor problem that may arise is the semantic distinction between past participle forms and post-positive adjectives. Instances like example 86, in which an

adjective is found as a post-modifier of a noun, can also be found included within this pattern.

(86) **The percentage of outpatients satisfied** with care was high and was close to the figure found by other researchers. **[Noun + Post-positive adjective; Discussion; an142165d]**

The third problem also comes when accounting for the agentive past participial reduced relative clauses. As shown in Table 55, the syntax used to retrieve the agentless past participial reduced relative clause coincides with the syntax used to retrieve the agentive one (the part which is not in blue). In blue, it is possible to identify the addition of the agent by means of a *by*-phrase. The problem here is the same one encountered with the passive and the *as*-passive. There is no way to restrict the agentless pattern to avoid the occurrences showing the agent. It is possible, though, to only search for the agentive pattern to restrict those instances in which the agent is not mentioned (see Table 56), but another problem arises. When doing so, the *by*-phrase can be found to introduce a nominalisation, where authorial intervention is not overtly manifested but is inferred through the context.

Table 55. Total number of occurrences of the Agentive Past Participial Reduced Relative Clause, according to RAs sections. Raw frequencies and frequencies per million words.

Agentive Past Participial Reduced Relative Clause			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	51	489.95
Introduction	99,009	55	555.51
Methods	413,717	200	483.42
Results	316,837	69	217.78
Discussion	307,806	131	425.59
TOTAL	1,241,461	506	407.58

Table 56. Syntactic expressions of the Agentless and Agentive Past Participial Relative Clauses, using SQS and CQPS

Agentless Past Participial Reduced Relative Clause
((_AT)* (_JJ)*)? _N?? (_RR)* _V?N (if SQS is used) **(([pos="AT"])* ([pos="JJ"])*)? [pos="N.."] ([pos="RR"])* [pos="V.N"]** (if CQPS is used)
Noun (either singular or plural), optionally preceded by an article and an optional number of adjectives (zero or more); optionally followed by any number of adverbs (zero or more); followed by the past participle form of any lexical verb (including non-auxiliaries *be*, *do*, and *have*).
Agentive Past Participial Reduced Relative Clause
((_AT)* (_JJ)*)? _N?? (_RR)* _V?N by ((_AT)* (_JJ)*)? _N?? (if SQS is used) **(([pos="AT"])* ([pos="JJ"])*)? [pos="N.."] ([pos="RR"])* [pos="V.N"] [word="by"%c] (([pos="AT"])* ([pos="JJ"])*)? [pos="N.."]** (if CQPS is used)
Noun (either singular or plural), optionally preceded by an article and an optional number of adjectives (zero or more); optionally followed by any number of adverbs (zero or more); followed by the past participle form of any lexical verb (including non-auxiliaries *be*, *do*, and *have*); **followed by the preposition *by*, and followed by a noun (either singular or plural), which may be optionally preceded by an article and any number of adjectives (zero or more).**

The past participial reduced relative clause including the agent is much less frequent than the agentless past participial reduced relative clause (because agentive clauses are in fact included within agentless ones). The agentive counterpart appears 506 times in the RCTs corpus, which accounts for a normalised frequency of 407.58 instances per M words. This agentive structure is usually found within the Methods section (200 occurrences, 483.42 instances per M words), and followed by the Discussion section (131 occurrences, 425.59 instances per M words). Despite these figures, the highest normalised frequency is found within the Introduction (55 occurrences, 555.51 instances per M words), followed by the Abstract section (51 occurrences, 489.95 instances per M words).

As occurring with the passive voice and the *as*-passive structure, when the agent is included in this construction, it rarely makes reference to the authors of the paper (see examples 87-91). Bearing this in mind, the agentive past participle reduced relative clause subtly provides a sense of impersonality insofar as it is tightly linked to its finite form, the prototypical passive, and operates, in this respect, similarly to it.

(87) The pragmatic trial was designed to compare two treatment strategies (spinal stabilization surgery or intensive rehabilitation) for **patients considered by surgeons** to be candidate for surgical stabilization of the lumbar spine. **[Agentive past participial reduced relative clause; Introduction; bm3301233i]**

(88) The total number of **pain medication reported by participants** in these 2 groups was compared by using generalized estimating equation models with an unstructured correlation matrix and the robust variance estimator at week 1 and after week 12. **[Agentive past participial reduced relative clause; Methods; an14310m]**

(89) Patients in the control group continued with **traditional hospital reviews ordered by the rheumatologist** every three to six months according to normal practice, and, as usual, requests for urgent reviews were made by general practitioners through the secretary and accommodated as quickly as possible. **[Agentive past participial reduced relative clause; Methods; bm330171m]**

(90) The inclusion criterion was the issue of a "common diseases" temporary work disability initiation form, with **an MSD-related cause reported by the primary care physician**, within the inclusion period. **[Agentive past participial reduced relative clause; Methods; an143404m**]

(91) This is important because the improvement in certain indexes of mood and cognitive performance with **triiodothyronine substitution reported by Bunevicius and colleagues** was restricted to athyreotic patients with thyroid cancer who received supraphysiologic thyroid hormone doses to suppress TSH secretion. **[Agentive past participial reduced relative clause; Discussion; an142412d]**

Some other examples show stacked nominal phrase or nominalised structures (underlined), which somehow imply human intervention (like in examples 92, 93, and 94) or not (like in 95).

(92) Protocol Patients were eligible for the trial if they were registered with a general practitioner in **the primary care trust served by the community hospital** and were considered by the responsible geriatrician to be medically stable and in need of post-acute rehabilitation care. **[Agentive past participial reduced relative clause; Methods; bm331317m]**

(93) At the end of the season we also obtained information on **prevention training conducted by the control clubs**, including the types and volume of exercises used. **[Agentive past participial reduced relative clause; Methods; bm330449m]**

(94) The main strength of this study lies in **the pragmatic approach adopted by the randomised controlled trial**. **[Agentive past participial reduced relative clause; Discussion; bm3301239d]**

(95) These results support the practice of **empirical antibiotic use guided by symptoms**. **[Agentive past participial reduced relative clause; Abstract; bm331143a]**

As stated earlier, when the *by*-phrase is included in the pattern by means of the syntactic expression **[word="by"%c] (([pos="AT"])* ([pos="JJ"])*)? [pos="N.."]**, another problem arises. Instead of human agents, nominalisations and stacked nominal phrases are also retrieved as if belonging to the aforementioned agentive pattern, but in this case, the preposition *by* does not include the agent of the action, but the manner in which something is done. The noun cannot be considered an agent in this case. Therefore, the past participle reduced relative clause is not agentive but agentless, as the argument introduced by *by* is an adjunct of manner (see examples 96-98).

(96) Measurements: Insulin sensitivity estimated by the oral glucose tolerance test, **subcutaneous and visceral abdominal far measured by single-slice computed tomography, endothelial function measured by flow-mediated vasodilation**, and fasting plasma measurements. **[Agentless past participial reduced relative clause + *by* (manner); Abstract; an143337a]**

(97) The antenatal steroids for term caesarean section (ASTECS) randomized trial therefore sought to evaluate whether giving the recommended two doses of betamethasone before delivery reduces the incidence of respiratory distress in **babies delivered by elective caesarean section** at term. **[Agentless past participial reduced relative clause + *by* (manner); Introduction; bm331662i]**

(98) The sole intervention was **tiotropium given by inhalation** once daily. **[Agentless past participial reduced relative clause + *by* (manner); Methods; an143317m]**

Taking into consideration the quantificational information and the qualitative analysis carried out in this section, one can assert that past participial reduced relative clauses are fundamental in medical writing, and they serve to cautiously express the same semantic information as the one expressed through prototypical passives, without overusing finite constructions. This pattern does not relate to the writers' involvement in the study under discussion and thus does not serve to account for their visibility in the text. Nonetheless, they do portray a hint of impersonalisation to the medical text.

As shown, when analysing the past participial reduced relative clause using a corpus-based methodology, some important limitations arise, and these do not allow a purely refined analysis of this structure in medical discourse. As seen in this section, the agentive type, which is more restricted, is included when one searches for the agentless type, which is less restricted. A specific syntactic expression must be used to retrieve the agentive past participial reduced relative clause, which would include the agent. However, when retrieving it, there is no way to restrict the typology of nouns appearing within the *by*-phrase; they could be human-related nouns (agents of the action, usually different from the researchers themselves) or non-human-related nouns, which appear in the form of stacked nominal phrases or nominalisations, which at the same time may imply human agency.

Apart from this fact, in some instances, the preposition *by* introduces an adjunct of manner. All these instances including a *by* adjunct expressing manner are always agentless, and they cannot be thus dealt with as agentive. For all these reasons, a corpus-based methodology serves to generally deal with these instances, because it allows to retrieve an extensive number of occurrences, to collect crucial quantificational information, to determine the distribution of these patterns within the RA, etc.

Nevertheless, when one tries to relate the pattern to the notions of authorial (in)visibility, the explicit mention or omission of agency, and the notion of impersonality they transmit, a corpus-based analysis poses, again, some relevant limitations.

5.2.3 Existential *There* Structures

The existential *there* structure is probably the most impersonal pattern of all. The expletive subject *there* is followed by any form of *be*, and this form is expected to agree in number with the argument presented after the verb, the notional subject. The syntactic expression used to search for this pattern in the RCTs corpus is as follows:

[word="there"%c] ([pos="RR"])* [lemma="be"]

When this pattern is searched for in the corpus, a total number of 1,515 instances appear (1,220.34 instances per M words) (see Table 57). This pattern is highly frequent in the Results section (742 occurrences, 2,341 instances per M words), followed by the Discussion section (424 occurrences, 1,220.34 instances per M words). A high normalised frequency is found in the Abstract Section (108 occurrences, 1,037.54 instances per M words).

Table 57. Total number of occurrences of the Existential *there* pattern, according to RAs sections. Raw frequencies and frequencies per million words

Existential *there*			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	108	1,037.54
Introduction	99,009	90	909.01
Methods	413,717	151	364.98
Results	316,837	742	2,341.90
Discussion	307,806	424	1,377.49
TOTAL	1,241,461	1,515	1,220.34

Table 58. Total number of occurrences of the Existential *there* pattern in its negative form, according to RAs sections. Raw frequencies and frequencies per million words

Existential *there* in its negative form			
Distribution	**Words**	**Occ**	**Fq x Mw**
Abstract	104,092	53	509.16
Introduction	99,009	19	191.90
Methods	413,717	46	111.19
Results	316,837	391	1,234.07
Discussion	307,806	154	500.32
TOTAL	1,241,461	663	534.05

The negative form of this pattern is found with the existential *there*, followed by the verb form of *be*, and the negative particle *no*. This negative sub-pattern appears 663 (534.05 instances per M words) (see Table 58). This pattern is mostly used in the Results section (391 occurrences, 1,234.07 instances per M words), and followed by the Discussion section (154 occurrences, 500.32 instances per M words). When in the negative form, this pattern tends to collocate with the noun phrases *significant differences*, *difference*, *significant differences*, or *evidence*, as seen in Table 59.

Table 59. Collocates in the negative form of the Existential *there* pattern, according to RAs sections. Raw frequencies and frequencies per million words

Existential *there*, with collocates, in its negative form			
No	**Collocates**	**Occ**	**Fq x Mw**
1	There were no significant differences	152	122.44
2	There was no difference	58	46.72
3	There was no significant difference	54	43.50
4	There were no differences	47	37.86
5	There was no evidence	35	28.19

The instances presented below, (99-104), include the existential *there* pattern. It is possible to assert that this pattern denotes the existence (or inexistence, when in the negative) of an aspect which further appears in the discourse (the argument following

the verb to *be*) (see examples 99-101). Apart from this, it is possible to identify, hidden within this structure, the medical considerations taken by authors of the RCT (see examples 102-104). It is an impersonal way to show medical opinions expressed by the authors without claiming full responsibility. It seems, then, a very powerful linguistic resource to project what authors state without fully claiming for what they state. This may also give universality to the content presented in the clausal structure, reinforcing the truthfulness of the propositional content being expressed.

(99) **There was no validation** by an external data set. **[Existential *there* pattern; Abstract; an473302a]**

(100) At the baseline, **there were 82095 women** in the NHS in 1980 and **45770 men** in the HPFS in 1986. **[Existential *there* pattern; Results; ja246820r]**

(101) In reality, even within districts **there are** probably **areas** not reached by surveillance and vaccination teams, resulting in heterogeneities in population immunity that are not captured in the analysis. **[Existential *there* pattern; Discussion; la403211d]**

(102) Nevertheless, our present study shows for both novel devices that **there is no loss** of 1 year efficacy and safety. **[Existential *there* pattern; Discussion; la352146d]**

(103) Because surgery is the only remaining treatment option with established efficacy (21), **there was no acceptable alternative** to offering surgery if the intervention failed. **[Existential *there* pattern; Discussion; an304215d]**

(104) After an additional fresh or frozen cycle, **there is no significant difference** in live birth rate between elective single and double embryo transfer and the risk of a multiple birth in the elective single embryo transfer group is still much lower than in the double embryo transfer group. **[Existential *there* pattern; Discussion; bm325701d]**

In conclusion, the existential *there* pattern can be considered the most impersonal linguistic resource of all. In all cases, it contributes to an impersonal dimension. In some cases, particularly when the existential *there* pattern is served to introduce concluding remarks, or objective facts observed from experimentation, the choice authors make to present themselves as invisible is patent.

5.3. Conclusions to the Study and Limitations

This corpus-based study aimed at exploring, quantifying, and categorising the wide range of syntactic structures belonging to either the active or the passive voice found in RCTs. As stated in the introduction to this study, the notion of grammatical voice has been the main distinctive criterion to categorise all the structures. However, within the analysis and the classification of such structures, other distinct twofold criteria have also been included: on the one hand, *personality* vs. *impersonality*; and on the other hand, *authorial visibility* vs. *authorial invisibility*.

As for the former, it is worth highlighting that all the structures analysed in this corpus-based study can be classified according to their degree of (im)personality. Their use and their cumulative effect are actually related to the degree of personal and impersonal nuances they evoke in the text. As for the latter, one can state that not all the structures analysed show authorial visibility (or presence) or authorial invisibility (or absence) in the text. Hence that this third criterion can only be used whenever possible. In fact, it is revealing to assert that combining these three distinctive dichotomised parameters (*active* vs. *passive*; *personality* vs. *impersonality*; and *authorial visibility* vs. *authorial invisibility*), through a corpus-based analysis, has resulted in a very convoluted task.

After highlighting the main conclusions obtained, I will recall the many problems encountered when dealing with these various patterns in the RCTs corpus, problems which have already been described within the analysis of each structure.

The personal active pattern is the only structure overtly personal, if compared to all the rest analysed in the study. It can be stated that, except for the personal active pattern, the rest of structures somehow show different levels of impersonality. This emphasises the personalising contribution to the text carried out by personal active structures, which serve to counteract a deeply high degree of impersonality that is inherently impregnated in the whole text. The use of the personal active structure can then be considered a counterpoint to the vast repertoire of impersonalising strategies prevailing in medical register. It is true, however, that there are some other non-syntactic strategies which also

exert a personalising effect over medical texts, such as the use of possessive adjectives like *our*. Because of their non-connection to verbal forms, they have been disregarded from the analysis presented in this thesis, but their personalising function cannot be denied in any case.

Example (105) serves to illustrate and contextualise the typical interplay of structures found within an ordinary main sentence in medical writing. A past participial reduced relative clause with a nominalisation operating as the agent of the action, implying then certain human intervention, is combined with a verbal-adjectival phrase, pre-modifier of a noun (in blue), and an agentless passive structure located within a prepositional phrase (in green). These three linguistic choices add impersonalising hints to the whole clausal construction and, by extension, to the whole text. Within the clause, they are then counteracted by the presence of a personal active structure including a reporting conclusive research act verb, the most personal or subjective typology of all (in red), and the use of the possessive adjective *our* followed by the noun *findings*. This is one of the many examples that demonstrate the intricacy involved in the use of structures recurrent in medical writing, which certainly denote explicit layers of (im)personality.

(105) Because **97% of primary articles examined by** the **excluded studies** were at high risk of overlap with **studies that were included** in the analysis, however, we do not believe that their omission would have substantially changed **our** findings. **[Agentless prototypical passive; Discussion; bm876240d]**

As for the lexical verbs appearing within the personal active structure, it has been possible to identify many types: procedural verbs denoting neutral or cognitive research acts, intentional verbs denoting purposive research acts, evaluative verbs used to assess research acts, and reporting verbs that serve to point out concluding research acts. This semantic classification is directly or indirectly related to the involvement of authors in the research under discussion. The quantitative analysis has revealed that there is a connection between the semantic classification of verbs and the main tenses in which they appear. Procedural verbs related to neutral and cognitive research acts are mainly expressed through the past simple tense. Intentional verbs referring to purposive

research acts also use the past simple, but instances in the present perfect simple and the future simple tenses are also frequent. Verbs expressing evaluation are also used in the past simple and the future simple tenses, and lastly, reporting verbs, considered the main subjective set of all, are the only ones which make extensive use of the present simple tense.

The semantic classification of the most common lexical verbs appearing in the personal active pattern into four sub-types has led to the creation of a continuum showing from little personal (or authorial) involvement to more substantial personal (or authorial) involvement as regards the type of research acts being described. Even though all the sub-patterns of the active structure '*we* + active verb' have shown to be related to the notion of personality and authorial visibility, as authors overtly manifest themselves in the research process being described, a certain gradability of their involvement in the research has also been accounted for and illustrated within a cline.

Procedural verbs refer to objective neutral actions, and such neutrality and objectivity gradually decreases as the semantic typology of verbs changes into verbs of intention, followed by verbs of evaluation, and followed by reporting verbs through which authors usually draw conclusive remarks. In the cline, each rightward move towards a new semantic category reinforces authorial subjectivity in the research process. At the end of the cline, reporting verbs are seen to promote the role of authors to its full extent. Therefore, even within the most personal active structure, there exists a continuum from more neutral (impersonal) patterns, which essentially make use of the preterite tense, to more subjective (personal) constructions, which include present tense reporting verbs that clearly visualise the authors' viewpoint.

In addition, this continuum can also be used as a framework to categorise those lexical verbs occurring within the impersonal active structure. Indeed, inanimate subjects followed by active verbs can all be related to impersonalising structures. Apart from this, this pattern is also related to the notion of authorial invisibility, because authors hide themselves through metonymic expressions containing an inanimate entity which performs the action described by the verb. It is true, however, that authorial

involvement, despite not being overtly manifested as it is through personal active structures, can be easily inferred from the context. Impersonal active structures tend to make use of reporting verbs to indicate results and draw conclusions, and this reinforces the impersonalising character of this specific linguistic strategy, as it helps avoiding full responsibility for the claims being made. Furthermore, a sense of universality and replicability is transmitted by means of this structure, since the result presented is conceived as empiric evidence, away from subjective human intervention.

As for the passive voice, it is undeniably the most frequent impersonal pattern found in the RCTs corpus. The passive is characteristic of medical writing and its marked alignment of arguments, if compared to the unmarked active structure, makes it have a special pragmatic consideration. By foregrounding relevant rhematic information, presented as new, and by backgrounding or even omitting the agent (thematic information), the passive establishes a particular hierarchy of importance of the different arguments of the clausal construction. The subject acquires a special relevance and becomes the point of departure of the message, which deals with the research itself and gains importance over the performers of the action. All prototypical passive instances, whether agentless or agentive, denote an impression of impersonalisation, which contributes to the senses of detachment, objectivity, and neutrality, searched for, and clearly expected in medical writing.

The tense in which the passive typically occurs is the past simple tense, as verbs of procedure are highly recurrent within this pattern and serve to describe the main procedures carried out in the past to fulfil the authors' research process. Reporting verbs are also used, and their usage contributes to the intensification of the sense of detachment, as authors are seen to use the passive to present the information without taking full responsibility for their claims. Apart from this, as the passive serves different functions, there is no clear relationship between the tenses used, the most frequent lemmas, and the semantic field to which they belong. For this reason, a continuum has not been provided including the different types of lexical verbs in this pattern.

The non-finite *as*-passive may be expected to be one of the best alternatives to the prototypical passive. Yet, its low frequency shows authors do not really count it as an alternative. Its main function, as observed in the corpus, is its guiding nature, showing temporal and locative information referred to in the text. This pattern has been seen to operate differently from the prototypical passive insofar as it usually shows anaphoric connections with previous elements mentioned in the text, particularly when followed or preceded by a comma. This reinforces its discoursal function while working as signposts which operate at a textual level rather than just at a clausal one. In this case, no clausal argument is topicalised, and full relevance is given to the lexical verb form, which is followed by rhematic information. In some other cases, it functions as a post-modifier of a noun, giving specific information about it.

The case of the reporting passive is interesting because one may consider that, as it is impersonal, it will frequently appear in medical writing. The truth is that, as shown in the corpus, this pattern is not as frequent as it may seem. Authors rather make use of other reporting patterns, such as the personal active, the impersonal active, or the passive structure, including reporting verbs. It has been seen that the reporting passive must include a reporting verb and must collocate with a *that*-clause. These restrictions allow the search for this specific type of pattern in the corpus. If these restrictions are not considered, a wide number of prototypical passive voice instances may be retrieved instead. This structure is indeed impersonal; however, as the verbs used must all be reporting verbs and, as said, reporting verbs denote a higher degree of subjectivity if compared to other semantic typologies, authorial manifestation seems to be implied, although agents appear as totally demoted.

Another structure analysed has been the past participial reduced relative clause, or the bare passive. This pattern has shown to be extremely frequent in medical writing and it can be the best alternative to the prototypical passive voice. It essentially serves the purpose of adding a huge amount of descriptive information within a phrasal construction to avoid the concatenation of multiple passive simple sentences which would include the same informational content with a major number of words, so its economising nature is highly considered and truly valued by authors. Apart from being

a very powerful resource to load phrases with detailed information which gives specificity to the information presented, the fact that it post-modifies a noun, as it is included within a phrase, allows it to be used in multiple contexts. This fact contributes to its omnipresence in medical texts. It is considered as an impersonal resource because authors have other alternatives to present information in a more personal way but decide to use bare passives to amalgamate nuances that contribute to the lexical density inherent in medical discourse, and this density also results in a sense of objectivity, neutrality, and impersonalisation. This pattern may or may not relate to authorial visibility, as actions described by the past participle form may be connected to the authors of the RCT themselves, to other researchers in the field that authors are referring to, or to other inanimate entities.

The last structure analysed, the existential *there* pattern, may be considered the most impersonal linguistic device used by medical writers in their papers. It highly contributes to impersonalising medical texts. Two types of existential *there* structures have been found. The first type would correspond to instances which focus on the existence (or absence) of some factual events. In this case, these examples do not relate to authorial visibility, but contribute to the impersonal dimension of the text. However, the second type would correspond to instances which introduce opinions derived from the observation of experimental research, totally hiding authors' involvement, and promoting their absolute invisibility. As said earlier, senses of universality, replicability, and objectivity emanate from these structures, as the total demotion of agents makes the result be presented as universal and totally unbiased.

Once having summarised the main results obtained in this corpus-based study, it is important to bear in mind the overall frequency of all these structures, as shown in Table 60. It is important to highlight the omnipresence of the passive voice in medical discourse (15,716.16 instances per M words), distantly followed by the non-finite past participial reduced relative clause (5,460.50 instances per M words). The personal active, the one which is aimed at showing hints of personality and overtly expressing authorial visibility accounts for the 4,254.66 instances per M words, which may seem as relatively low if considering that it is the only personal syntactic resource that authors

have available to neutralise an overwhelming presence of impersonal linguistic choices. The next less frequent is the existential *there* pattern, which accounts for a normalised frequency of 1,220.34 instances per M words. The *as*-passive is the one that follows, which shows relative frequency of 880.41, followed by the reporting passive, which accounts for the 689.51 instances per M words. The figures representing the impersonal active pattern have not been included in Table 60, because of the difficulties encountered to retrieve instances that actually matched a non-human entity followed by an active verb.

Table 60. Structures analysed in the corpus-based study.
Raw frequencies and frequencies per M words.

Active and Passive Voice Structures	Occurrences	Frequency per M words
Prototypical Passive	19,511	15,716.16
Past Participial Reduced Relative clause	6,779	5,460.50
Personal Active	5,282	4,254.66
Existential *There*	1,515	1,220.34
as-Passive	1,093	880.41
Reporting Passive	856	689.51
Impersonal Active (abstract rhetor)	Not determined	Not determined

The use of corpora as a source to analyse very specific (and at the same, time very broad) linguistic material, such as the different active and passive (or passive-like) structures explored in this corpus-based study, has been extremely helpful, insofar as it has allowed the extraction of a huge number of contextualised examples that clearly show the existence of recurrent patterns that are truly genre-specific. Apart from this, it has provided the analyses with a high volume of figures which indicates the reliance that authors show towards some specific linguistic choices to present their research.

However, the automatic retrieval of structures has not been straightforward, essentially because of the complexity and the variability shown by the structures to analyse. There is an enclosed number of syntactic strategies that authors recurrently use, but these strategies show a high degree of internal variability.

As for the active structures, the one which posed more problems when being searched for in the corpus was the impersonal active pattern. As the RCTs corpus was not semantically tagged, there was no automatic way to retrieve instances in which the subject was non-human. Therefore, despite being able to analyse the structure itself and its main peculiarities, real quantificational information could not be extracted; and the one which could be extracted was not representative enough of the nature of this pattern in medical writing.

Indeed, all passive (or passive-like) structures also posed several problems when being retrieved. The first problem was connected to the presence or absence of the agent. In fact, the syntactic expressions (both SQS and CQPS) used to retrieve agentless passives almost totally coincide with the ones used to retrieve agentive passives, because the structure of the agentless passive is subsumed within the structure of the agentive one. To obtain the agentive passive, an extension of the syntactic expression was used to retrieve the agent as included within the pattern. Therefore, to know the exact number of agentless passives, agentive passives had to be deducted from the number of passives, and this indirectly modified the other relevant information on tenses, on the main lemmas used, etc.

Another problematic issue encountered had to do with the typology of nouns appearing as part of the *by*-phrase when searching for the agentive passive. Three types of categories were found as representing the agent: human beings, never the authors themselves but other relevant agents included within the research; nominalisations, which imply human agency but nominalise a process (a verb) through a transcategorisational transformation; and stacked nominal phrases or simple nouns, which do not imply any kind of human agency at all. Being able to distinguish between these three different degrees of human involvement through the corpus would have been extremely helpful, as most of the shades marking impersonality are rooted in the choice of these small details.

The *by*-phrase also resulted to be problematic in a different sense. Instances of adjuncts of manner introduced through the preposition *by* were added onto the equation, and

despite being able to retrieve these instances in isolation, most of them were agentless passives because the *by*-phrase did not include the agent of the action but the way the action had been done. In this case, it was possible to again find nominalisations, stacked nominal phrases, or simple nouns, whose human involvement showed different degrees, impossible to be retrieved automatically.

Bearing in mind all the information presented in this corpus-based study, it is worth noting that all the structures found are related to the notion of grammatical voice, as they are either active or passive forms, and to the notion of impersonality. Considering these two distinctive parameters, it is possible to create a Cartesian coordinate system in which all these patterns are integrated, as a way to graphically represent the interrelatedness between grammatical voice and (im)personality from a scientific perspective.

As explained in Chapter 4 (Methodology), the Cartesian coordinate is defined by an ordered pair of perpendicular axes: the abscissa, represented by **x** (in green); and the ordinate, represented by **y** (in purple). The x-axis represents the notion of *Personality*, and it is shown horizontally and oriented to the right. The y-axis, on the other hand, represents the notion of *Voice*, and it is shown vertically and oriented upwards.

The x-axis can be considered as the gradual representation of *Personality*, ranging from *impersonal* (-x) to *personal* structures (x). This gradation is represented through the black arrow linking both notions (impersonality and personality) and through the purple discontinuous line, which does not actually represent the x-axis Personality, but the y-axis Voice. The purple line is discontinuous only to visually show the fuzziness of the two concepts forming the x-axis Personality.

The y-axis corresponds to the notion of *Grammatical Voice*, either *active voice* (y) or *passive voice* (-y). This feature shows no gradation since structures are either active or passive. Because of this, the green line is shown as continuous, to symbolise this limitation. However, the continuous green line does not represent the y-axis Voice, but the x-axis Personality, and that is why it is coloured in green. The green line is marked

as continuous only to represent the definiteness of the active and passive voice notions, the ones forming the y-axis Voice.

As seen in Figure 19, the system is formed of four distinctive Quadrants: I Quadrant, which includes personal active voice structures; II Quadrant, which includes impersonal active structures; III Quadrant, which includes impersonal passive structures; and IV Quadrant, which includes personal passive structures.

I Quadrant (x, y) = (personal, active) includes the **personal active** pattern *we* + active verb. It is the only personal verbal structure placed in I Quadrant and, as it is the most personal form of all, it is placed at the right side of the Quadrant. If the distinctive feature personal (x) changes into the impersonal feature (-x), two structures are found to be placed in II Quadrant.

The two structures positioned in **II Quadrant (-x, y) = (impersonal, active)** are the **existential *there*** and the **impersonal active** verb form. Both may be considered active (although the existential *there* may pose some problems in this sense)[62] and they show two different degrees of impersonality. The existential *there* structure is placed at the lower left corner of II Quadrant, to mark it as the most impersonal form of all and one which is placed towards the passive area, because of its impersonal character and its unclear classification as an active structure. On the other hand, the impersonal active (or abstract rhetor) is placed within the impersonal category but more towards the right, showing that its impersonality is shown in a lower degree. This is because the subject acquires humanised qualities expressed by the active verb following the inanimate subject.

III Quadrant (-x, -y) = (impersonal, passive) includes a wide array of structures. The degree of personality shown by the structures is determined by their finiteness, going from non-finite structures (towards the left end of the x-axis Personality) to finite

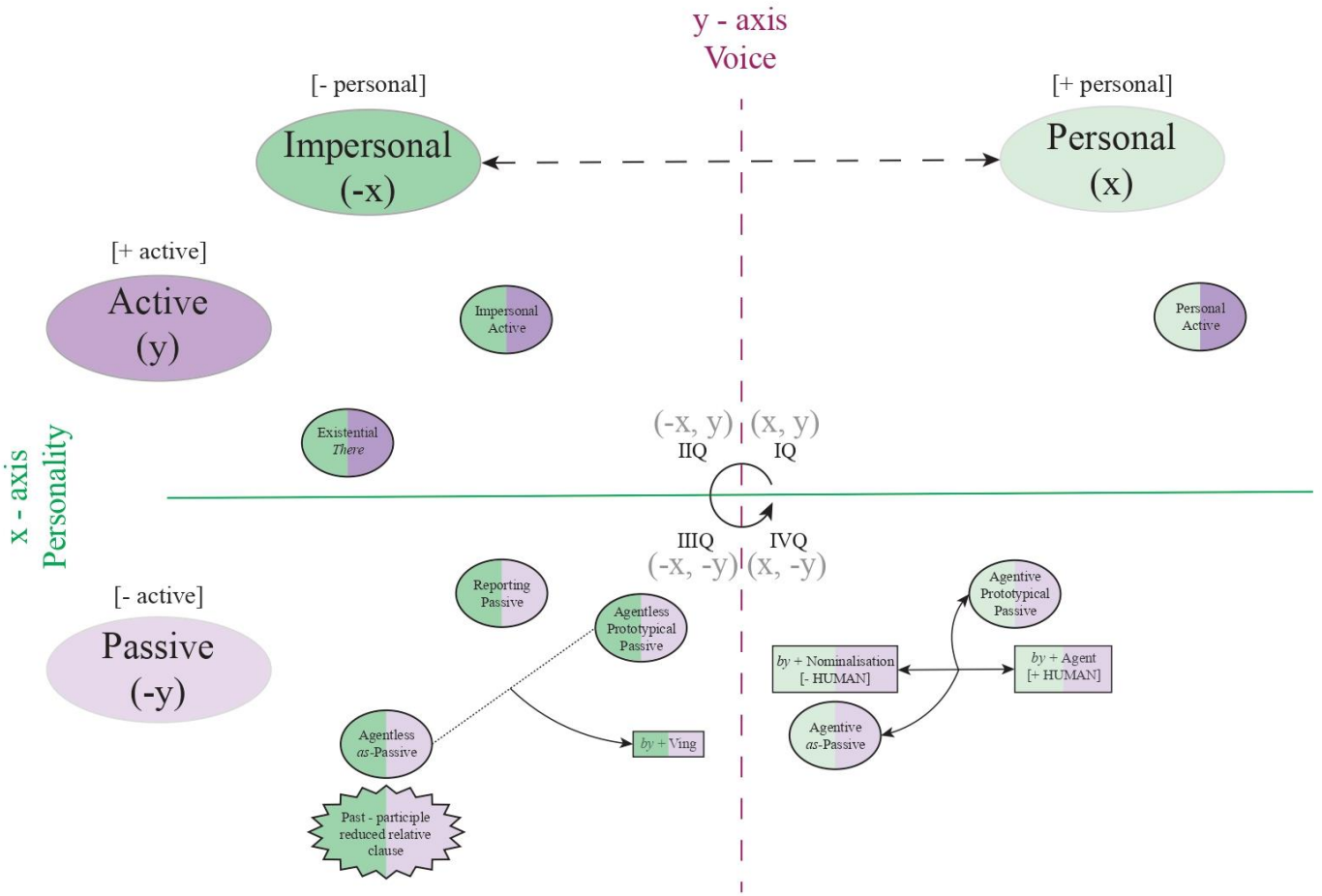
[62] As stated earlier, I include the existential *there* as an active structure, because it follows the structure of a grammatical subject (expletive) followed by an active verb, which agrees in number with the argument that follows.

structures (towards the right end of the x-axis Personality). The fact that the structure is non-finite can be perceived as even more abstract than finite counterparts, and thus as slightly more impersonal.

The non-finite structure **as-passive** is the one placed on the downside of the Quadrant, together with the **past participial reduced relative clause**. The shape of the box corresponding to the past participial reduced relative clause is different from the rest, and that is because it can be embedded within other types of structures, since it is a dependent functioning as a post-modifier and cannot really stand on its own. Placed more on the right side of the x-axis, it is possible to find both the **reporting passive** and the **agentless prototypical passive**. Both are placed more towards the personal area because the agents can be usually inferred through the context. The agentless *as*-passive and the agentless prototypical passive are joined through a discontinuous line, because the former may be considered the non-finite version of the latter. Both can include an adjunct of manner included through a *by*-phrase. This *by*-phrase may be followed by present participle verbs whose agency is perceived as hidden, although its inference is possible through the context.

Lastly, **IV Quadrant (x, -y) = (personal passive)** shows two other structures, the **agentive prototypical passive**, and the **agentive *as*-passive**. Although these structures are passive and thus might be considered as impersonal, the fact that they explicitly mention the agent through a *by*-phrase makes them be considered as more personal. They are placed more to the left if compared to the personal active pattern, to show they are more impersonal. The agentive *as*-passive, as it is non-finite, is represented as slightly more impersonal than the finite agentive prototypical passive. The *by*-phrase also distinguishes the two different types of nouns it introduces. If the noun is an explicit agent [+human], the structure is placed more towards the right, the personal area; however, if the *by*-phrase includes a nominalisation [-human], the structure is placed more towards the left, the impersonal area. It is kept within the IV Quadrant because through a nominalisation, agents can be recovered from context.

Figure 19. Classification of structures within the Cartesian coordinate system



This Cartesian coordinate system does not include any trace of authorial (in)visibility, because it only consists of two distinct axes, and not three. Authorial visibility could be taken as a secondary feature and each pattern could be marked in a specific form to show authorial manifestation or authorial absence. In any case, through this coordinate system, all the structures have been categorised according to the two primary parameters used in this corpus-based study: grammatical voice and (im)personality.

To sum up, the corpus-based study has served to deal with these patterns and to extract insightful information on their use, their context, and their frequency. To my view, however, some of these linguistic choices must be looked at closely so as to deeply explore the multi-layered nature of medical writing as far as the concepts of (im)personality and authorial visibility are concerned. Because of this complexity, a continuum dealing with the notions of *active voice* vs. *passive voice*, *personality* vs. *impersonality*, and *authorial visibility* vs. *authorial invisibility* may shed some light on the intricacy attached to the combination of these three dichotomised parameters.

The relevance of these structures in medical discourse stems from their cumulative effect, the position they occupy within the clause and within the text, the way they present information and how it is linked to other contextual information, and above all, how the different patterns interact with each other. It is then reasonable and sensible to delve into a close textual analysis, which is presented in the following chapter, aimed at refining the findings obtained in this corpus-based study and at creating a continuum showing the different degrees of (im)personality and authorial (in)visibility shown in the text.

6

Contextualised Study on the Linguistic Strategies to Show Impersonality and Disguise Authorial Presence in Randomised Controlled Trials:

Quantitative and Qualitative Analyses

Look closely.

The beautiful may be small.

Immanuel Kant

When outlining and writing a research article (or a randomised controlled trial, in particular), both novice and experienced medical writers may find it arduous to verbalise all the amount of data obtained through experimentation (e.g. raw data, figures, facts, inclusion of tables, interpretation of results, methods applied, evidence-based opinions, etc.). As already mentioned, being successful in putting any kind of medicine-related notion(s) into words certainly entails a great deal of expertise. For any scientist wishing a successful career, the writing of scientific papers and their publication in high impact journals, which essentially use English as their vehicular language, are then determinative.

As explored in Chapter 1, English is the global *lingua franca* of academia and despite not being the first tongue to most of the writers, they need to make extensive use of it. Showing a proficient language command becomes thus more than necessary (Mauranen et al., 2010, p.183). However, to meet the expectations of both the publisher (and its house of style) and the discourse community (and community of practice) to which the text is aimed to be addressed, showing a proficient discourse competence is imperative. To sound convincing to their peers and promote themselves as valid transmitters of valuable scientific knowledge, writers need to acquire a repertoire of discursive and pragmatic strategies that adhere to the conventions of their discourse community.

Vast exposure to medical literature is indispensable to acquire and master the discursive patterns typical of such a register. Those patterns are conventions that have been acquired through time and usage and are, in fact, as important as the informational content added to expand the body of already-existing knowledge in each medical field. The information may be crucial, but if it is not well expressed, it cannot be made available to the community of practice. In the case of medical writing, as highlighted in Chapter 2, the wide spread of certain linguistic mechanisms and strategies contribute to the uniqueness of such a register, which is marked off from other variants of the language. Medical scientists need then to be aware of the repertoire of linguistic choices available to suit their conventionalised type of register.

Indeed, the medical community makes an extensive use of several resources that allow the text to adhere to the standards expected to be found in the medical literature. Most of these linguistic resources serve to transmit impersonality, objectivity and empiricism and are, at the same time, strategically used to disguise or promote the writers' presence in the text. Through the (im)personalising devices described in Chapter 3, like active and passive voice patterns, nominalisations and stacked nominal phrases, abstract rhetors, present and past participle reduced relative clauses, and *there* existentials, authors follow the customary patterns found in medical writing, which help to portray medical writing as detached and somehow neutral.

A balance between promoting and demoting the authors involved in the writing of RCTs and projecting a sense of neutrality and objectivity is then crucial to both respect the epistemological tradition and present themselves as representative of the study they aim to share with the rest of the community. Hence, that both notions, that of impersonality and that of authors' visibility, are chosen as two main parameters to classify the most recurrent lexicogrammatical patterns typical of this specialised register.

After having analysed some of these structures, mainly active and passive voice instances, in the corpus-based study presented in Chapter 5, it is worth having a look at real texts so that these recurrent patterns are analysed in more detail, considering their frequency, their distribution across the RCTs, their interplay with other different linguistic resources, and their connection to both impersonality and authorial visibility. This close textual analysis contributes then to the exploration of these linguistic resources from a more textual-based approach, bearing in mind contextualised instances that will help in portraying the different nuances these choices add to the complex study of impersonality and authorial intervention in the texts.

6.1. Introduction to the Contextualised Study and its Main Outcomes

The study presented in this chapter focuses on a contextualised analysis of the most frequently observed patterns of (im)personalisation found in the corpus-based study presented in Chapter 5. Above all, through this close textual analysis, I aim at both understanding and disentangling the complex dimension of impersonality (and interpersonality) and exploring the extent to which these strategies of depersonalisation interact and textually characterise medical writing as being profoundly impersonal.

The source of the analysis carried out in this study consists of eight different Randomised Controlled Trials (henceforth, RCTs). I have chosen two samples for each of the journals which compose the RCTs corpus – the same one used in Chapter 5 – to explore the distinct patterns showing a certain degree of impersonality.

These RCTs have been electronically obtained from the corresponding journals' websites, which are as follows: *The New England Journal of Medicine* (henceforth, NEJM), *The Lancet* (henceforth, LNC), *The British Medical Journal* (henceforth, BMJ), and *The Journal of American Medical Association* (henceforth, JAMA). However, *Annals of Internal Medicine* (henceforth, AIM), which is also part of the RCT corpus, has not been included in this contextualised study because it has not been possible to freely download articles from its website.

Detailed meta-textual information on each of the RCTs selected is given in the following tables[63]:

Table 61. RCT 1. *The New England Journal of Medicine* (NEJM1)

Journal	**The New England Journal of Medicine**	
Code	**NEJM1**	
Title	*A Randomized Trial of Long-Term Oxygen for COPD with Moderate Desaturation.*	
Authors	The Long-Term Oxygen Treatment Trial Research Group	
Date, volume, issue, and pages	2016 Oct 27 / Vol. 375 No. 17 2016; 375: 1617-27	
DOI	http://dx.doi.org/10.1056/NEJMoa1604344	
Number of words per section	Abstract	**367** words
	Introduction	**391** words
	Methods	**1,282** words
	Results	**819** words
	Discussion	**614** words
	Total no. of words	**3,473** words

[63] The information given for each of the RCTs is as follows: name of the journal, code used to refer to them in the study, title of the RCTs, authors (or group of authors they belong to), date of publication, volume and issue, number of pages, DOI, and total number of words and total number of words per section. As for the number of words per section, it is worth noting I have decided to exclude the words appearing in graphs, tables, visual supports of any kind, and additional comments inserted in coloured boxes on extra information related to the investigation being presented in the RCTs.

Table 62. RCT 2. *The New England Journal of Medicine* (NEJM2)

Journal	**The New England Journal of Medicine**	
Code	**NEJM2**	
Title	*Shiga Toxin-Producing E. coli Infections Associated with Flour.*	
Authors	Crowe, S. J. *et al*.	
Date, volume, issue, and pages	2017 Nov 23 / Vol. 377 No. 21 2017; 377: 2036-43	
DOI	http://dx.doi.org/10.1056/NEJMoa1615910	
Number of words per section	Abstract	**237** words
	Introduction	**221** words
	Methods	**920** words
	Results	**1,006** words
	Discussion	**708** words
	Total no. of words	**3,092** words

Table 63. RCT 1. *The Lancet* (LNC1)

Journal	**The Lancet**	
Code	**LNC1**	
Title	*Prevention of incisional hernia with prophylactic onlay and sublay mesh reinforcement versus primary suture only in midline laparotomies (PRIMA): 2-year follow-up of a multicentre, double-blind, randomised controlled trial.*	
Authors	An P. Jairam, L. Timmermans, *et al.*	
Date, volume, issue, and pages	2017 Aug 5 / Vol. 390 2017; 390: 567-76	
DOI	http://dx.doi.org/19.1016/S0140-6736(17)31332-6	
Number of words per section	Abstract	**362** words
	Introduction	**637** words
	Methods	**1,973** words
	Results	**894** words
	Discussion	**1,744** words
	Total no. of words	**5,610** words

Table 64. RCT 2. *The Lancet* (LNC2)

Journal	**The Lancet**	
Code	**LNC2**	
Title	*Momelotinib versus best available therapy in patients with myelofibrosis previously treated with ruxolitinib (SIMPLIFY 2): a randomised, open-label, phase trial.*	
Authors	Harrison, C. N. *et al.*	
Date, volume, issue, and pages	2018 Feb / Vol. 5 2018; 5: e73-81	
DOI	http://dx.doi.org/10.1016/S2352-3026(17)30237-5	
Number of words per section	Abstract	**544** words
	Introduction	**224** words
	Methods	**1,368** words
	Results	**1,458** words
	Discussion	**1,116** words
	Total no. of words	**4,710** words

Table 65. RCT 1. *The British Medical Journal* (BMJ1)

Journal	**The British Medical Journal**	
Code	**BMJ1**	
Title	*Bivalirudin or unfractionated heparin in patients with acute coronary syndromes managed invasively with and without ST elevation (MATRIX): randomised controlled trial.*	
Authors	S Leonardi, et al.	
Date, volume, issue, and pages	2016 / Vol. 354 2016; 354: i4935	
DOI	http://dx.doi.org/10.1136/bmj.i4935	
Number of words per section	Abstract	**342** words
	Introduction	**303** words
	Methods	**1,556** words
	Results	**998** words
	Discussion	**935** words
	Total no. of words	**4,134** words

Table 66. RCT 2. *The British Medical Journal* (BMJ2)

Journal	**The British Medical Journal**	
Code	**BMJ2**	
Title	*Intramuscular versus intravenous oxytocin to prevent postpartum haemorrhage at vaginal delivery: randomised controlled trial.*	
Authors	Adnan, N.; Conlan-Trant, R.; McCormick, C.; Boland, F.; Murphy, D. J.	
Date, volume, issue, and pages	2018 / Vol 362 2018; 362: k3546	
DOI	http://dx.doi.org/10.1136/bmj.k3546	
Number of words per section	Abstract	**396** words
	Introduction	**377** words
	Methods	**1,908** words
	Results	**616** words
	Discussion	**1,342** words
	Total no. of words	**4,639** words

Table 67. RCT 1. *Journal of American Medical Association* (JAMA1)

Journal	**Journal of American Medical Association**	
Code	**JAMA1**	
Title	*Effect of Deutetrabenazine on Chorea Among Patients with Huntington Disease. A Randomized Clinical Trial.*	
Authors	Huntington Study Group	
Date, volume, issue, and pages	2016 / Vol 316 No. 1 2016; 316(1): 40-50	
DOI	http://dx.doi.org/10.1001/jama.2016.8655	
Number of words per section	Abstract	**433** words
	Introduction	**315** words
	Methods	**1,680** words
	Results	**933** words
	Discussion	**771** words
	Total no. of words	**4,132** words

Table 68. RCT 2. *Journal of American Medical Association* (JAMA1)

Journal	**Journal of American Medical Association**	
Code	**JAMA2**	
Title	*Preladenant as an Adjunctive Therapy With Levodopa in Parkinson Disease: Two Randomized Clinical Trials and Lessons Learned*	
Authors	Robert A. Hauser, *et al.*	
Date, volume, issue, and pages	2015 / Vol. 72 No. 12 2015; 72(12): 1491-1500	
DOI	http://dx.doi.org/10.1001/jamaneurol.2015.2268	
Number of words per section	Abstract	**383** words
	Introduction	**243** words
	Methods	**671** words
	Results	**1,053** words
	Discussion	**1,012** words
	Total no. of words	**3,362** words

Once the RCTs analysed in this study have been introduced, it is worth gathering the information related to the length of each of the journals, and how the number of words is distributed across the different sections of the RCTs. As the size of the different articles (and sections) is different, it is crucial to bear these figures in mind to account for the relative frequency in which they occur, which has been calculated and will be presented throughout the study. Figures are shown in Table 69.

Table 69. Total number of words according to RCT, journal, and section

	NEJM1	NEJM2	LNC1	LNC2	BMJ1	BMJ2	JAMA1	JAMA2	TOTAL
Abstract	367	237	362	544	342	396	433	383	3,064
Introduction	391	221	637	224	303	377	315	243	2,711
Methods	1,282	920	1,973	1,368	1,556	1,908	1,680	671	11,358
Results	819	1,006	894	1,458	998	616	933	1,053	7,777
Discussion	614	708	1,744	1,116	935	1,342	771	1,012	8,242
Total	3,473	3,092	5,610	4,710	4,134	4,639	4,132	3,362	33,152
TOTAL	6,565		10,320		8,773		7,494		33,152

Even though the size of the sample of RCTs randomly chosen is rather limited if compared to the extensive existing amount of medical literature, the fact that they are articles published in the top-five journals in Medicine somehow ensures their representativeness in this register. These articles undoubtedly conform to the writing standards accepted by the journals' houses of style, as they have already been accepted and published by them. Taking this into account, any linguistic aspect showing (im)personality that frequently appears in the samples may serve to give general insights onto how (im)personal linguistic resources are used in the writing of RCTs. To deeply understand the complexity of the dimension of impersonality and how such impersonality is linguistically represented and interpreted in context, I have set the following outcomes for this study:

1) To closely examine the texts to observe and manually identify the main linguistic resources that are used to show a certain degree of (im)personality and analyse to what extent these lexicogrammatical instances relate to authorial presence.
2) To account for their frequency and their distribution on the RCTs under analysis, highlighting the main differences according to the journals and the sections in which they appear.
3) To assess their functions within the text as far as authorial visibility, authorial presence, and impersonality are concerned.
4) To explore the interplay between the several linguistic features analysed through real instances found in the texts and how they account for the realisation of (im)personality.
5) To analyse the transmission of information and the use of these linguistic resources in the structural development of the sections of the RCTs.
6) To develop a continuum, including these linguistic realisations from more personal to more impersonal resources so as to account for their degree of impersonality (and authorial (in)visibility).

A close reading of the RCTs corroborates the recurrent use of a specific set of linguistic devices; namely nominalisations and stacked nominal phrases, agentless prototypical passive structures, prototypical passive structures including the agent, personal active structures whose subject is the personal pronoun *we*, and other active structures whose subject is a human agent different from the author(s), active structures whose subject is an inanimate entity, *there* existentials, and a high number of dependents which pre- and post-modify certain elements in the text. Among these different modifiers, it is worth considering the high frequency of prepositional phrases that load the noun phrase with a dense amount of explicatory information, and present and past participle reduced relative clauses that are used to specify and restrict the semantic extent of the noun they go with.

Out of all these linguistic choices, it is important to highlight, however, that I have excluded nominalisations and stacked nominal phrases from the quantitative analysis. In fact, I have accounted for the different types of nominal phrases in the RCTs, as well as for their main function in the text. The main reason for this exclusion is based, essentially, on the fact that these resources are extremely frequent in almost all the paragraphs in the RCTs. Specifying the exact number of occurrences would not have given highly valuable information because of their absolute omnipresence in the text. In my view, their relevance as regards impersonality stems from their summative effect, the position they occupy in the clause and how such position affects the organisational structure of the informational content presented in the clause (and in adjacent clauses), and above all, the high likelihood they show to interplay with other impersonal resources, which pervasively extends the aforementioned cumulative effect.

I have also excluded from this close-textual analysis the *as*-passive and the reporting passive, because, as seen in the corpus-based study, their frequency is relatively low if compared to the rest of structures. Apart from this, the texts used to carry out this textual-based analysis do not contain many instances of these two patterns. For these reasons, I have not included them in this study.

In addition, as nominalisations and stacked nominal phrases are usually found with other impersonal resources, this interaction may add some complexity when it comes to account for the delimitation of the structure as regards its impersonalising effect. That is, since a nominalisation (or a stacked nominal phrase) is usually densely packed with a huge load of information, which may be also worded using other impersonal resources, the impersonalising textual effect it exerts may be attributed to either the basic part of the nominalisation itself (that is, the main noun objectifying the process) or to the whole extent of the nominalisation including all the modifiers attached to it. This, in fact, adds another layer of impersonality. Furthermore, when the nominalisation is presented as an objectification of a process and is followed by an active verb, what is considered an abstract rhetor, the impersonal effect it exerts over the clause, and by extension over the text, is actually multiplied. An example in which a nominalisation interplays with an abstract rhetor is shown as follows:

(1) **Trace-back investigation [of the two bags of bran A flour (collected from patients in Colorado and Washington)]** revealed that the flour from Colorado was unbleached all-purpose manufactured on November 14, 2015, and the flour from Washington was bleached all-purpose manufactured on November 14, 2015. **[Results, NEJM2]**

In this case, the nominalisation serves to topicalise a huge amount of information, which is presented in the form of a noun followed by an extensive embedded post-modifier (and a pre-modifier nominal, in bold black). The nominalisation is formed of a head noun derived from a verbal process (in blue), followed by a post-modifier prepositional phrase (in grey and square-bracketed) which includes a past participle reduced relative clause (in red and bracketed). This bare passive structure is a post-modifier within a post-modifier, and it is used to demote the agent of the action of 'collecting', which coincides with the same downplayed agent who performed the action of 'investigating'. As said, this nominalisation is related to the structure 'inanimate subject + active verb', insofar as the nominalisation functions as the subject of the active verb underlined. Here, the abstract rhetor and the nominalisation, including all the post-modification dependents, have fused into a unique hallmark of impersonalisation.

Another important aspect about this textual study is that I will not deal with the results and their discussion as two separate entities. I consider showing the results on a par with discussing them is the optimum way to deeply explore the various layers of impersonality textually represented in the text. Referring to quantificational information, relating it to a more qualitative analysis, and illustrating this analysis with examples is a sensible way, to my view, to represent the fuzziness of the concept of impersonality in medical texts.

As it has been already highlighted, all the linguistic resources analysed are, in fact, connected to the notion of impersonality in one way or another. However, not all these resources are tightly linked to the notion of authorial presence in the text. When describing and contextualising each lexicogrammatical pattern, I will refer to their degree of impersonality through the confection of a continuum from more personal to more impersonal structures, and I will also account for the degree of connection between the pattern itself and the notion of authorial visibility.

Not all reduced relative clauses, for instance, have to do with the presence of the author in the text. In fact, some instances of these non-finite relative clauses may be concerned with other agents implied in the research process, for example, or may refer to objects, instruments, or even patients implied in the study being described. In these cases, the correlation ‘personal equalling visible’ and ‘impersonal equalling invisible’ seems somehow diluted and blurred. These cases, although not being tightly related to authorial presence, are in a way connected to authorial stance, insofar as they are linguistic choices made by the writers to transmit a piece of information that could have been worded in a much more personal or impersonal way. The fact these are worded the way they are seems to stem from the specific rhetoric of medical writing and hence its inclusion in the study.

The following section deals with the identification of the linguistic resources recurrent in the RCTs which have been selected to explore impersonality and its possible connection to authorial (in)visibility. First, an account of their frequency as well as their distribution across the sections of the articles will be given. Second, I will explore and

describe their main textual functions and how the interrelatedness found between these several structures enhances the sense of detachment and objectivity characterised in medical writing. Lastly, I will analyse the relationship between these lexicogrammatical patterns and the transmission of informational content, which will be complemented with a continuum grading these structures from more personal to more impersonal-like. Indeed, this cline will build up as the different patterns are being introduced, so that at the end of this chapter, a full continuum accounting for the linguistic resources analysed as regards impersonality will be shown.

6.2. Identification, Description, and Quantitative and Qualitative Analyses of the Set of Structures Portraying (Im)personality Found in RCTs: Interplay between Patterns as a Means of Impersonalisation

As already mentioned at the introduction to this close textual study, the main linguistic resources dealt with are active and passive counterparts, nominalisations and stacked nominal phrases, abstract rhetors or inanimate subjects followed by an active verb, present and past participial reduced relative clauses, and existential *there* patterns.

6.2.1. Active and Passive Voice counterparts

The first typology of structures explored are passive structures that could be rewritten as active (2), and active structures that could be easily transformed into the passive (3), hence that I label them as active-passive counterparts. In this case, the active sentence belongs to the personal type, as the personal pronoun *we* performs the function of subject. The passive sentence, on the other hand, will relate to a more impersonal type of style.

(2) **A total of 56 cases were identified** in 24 states (Figs. 1 and 2); 55 were infections with STEX O121, and 1 was an infection with STEC O26. **[Agentless passive; Results; NEJM 2]**

We identified a total of 56 cases in 24 states (Figs. 1 and 2); 55 were infections with STEX O121, and 1 was an infection with STEC O26. **[Personal active; Rephrased example]**

(3) **We initiated** the PRIMA trial (PRImary Mesh closure of Abdominal midline wounds) in 2009 with the aim to investigate prophylactic mesh reinforcement in high-risk groups (ei. patients with abdominal aortic aneurysm or a BMI ≥27 kg/m^2). **[Personal active; Introduction; LNC 1]**

The PRIMA trial (PRImary Mesh closure of Abdominal midline wounds) was initiated in 2009 with the aim to investigate prophylactic mesh reinforcement in high-risk groups (ei. patients with abdominal aortic aneurysm or a BMI ≥27 kg/m^2). **[Agentless passive; Rephrased example]**

These structures are tightly linked to notions of personality, in the case of the personal active voice, and that of impersonality, in the case of the passive voice. Indeed, it is the writers' choice to use either form or the other, and the cumulation of such choices ends up influencing how the text is transmitted by the writers, and how it is perceived by the readers. If passive voice structures prevail over personal active ones, as it is seen in all the RCTs analysed, a more impersonal character is textually generated, as expected in scientific medical discourse.

It is true, however, that some other instances of active structures whose subject does not correspond to the author(s) of the paper are sometimes used, although their frequency is low when compared to the *we* personal active. The sentence in (4) uses a *non-we* personal active structure whose subject is animate. In this case, the agent 'performing the action' is being generalised by means of the term 'investigator' (meaning 'any investigator who wishes to carry out the research'). In fact, it is a piece of advice given to readers and somehow expected to be followed. It makes use of this subject and uses a deontic type of modality to express strong advisability.

(4) Therefore, **the investigator must select** patients who actually have PD, **confirm** that the patient has true motor fluctuations, **teach** the patient to recognize the PD motor states, and **verify** that the patient understands them. **[Non-*we* personal active; Discussion; JAMA2]**

In scientific medical writing, the passive voice is extremely frequent because it serves to project a sense of objectivity, neutrality, and scepticism, and to spotlight the research over authorial visibility. Even though there are clear indications by most of the houses of style of the journals analysed in favour of a plain type of discourse, somehow postulated by the PLM, the truth is that the passive voice by far exceeds previous expectations one may have when exploring such a pattern in medical writing. After having explored the eight samples and having manually quantified the number of tokens of personal active structures (*we* + active verb), on the one hand, and passive structures, on the other, it is worth mentioning that, as seen in Chapter 5, passive structures prevail to a greater extent over other types of verb structures, and considerably outnumber personal active patterns; a fact which contributes, as said, to the impersonal dimension of medical discourse.

Table 70. Total number of tokens (raw frequency) of '*we* + Active verb' pattern according to journal and section.

***we* + act V**	**NEJM 1**	**NEJM2**	**LNC1**	**LNC2**	**BMJ1**	**BMJ2**	**JAMA1**	**JAMA2**	**TOTAL**
Abstract	3	0	4	1	0	0	0	1	9
Introduction	1	1	2	0	3	1	0	1	9
Methods	3	0	52	1	20	15	0	1	92
Results	2	0	0	0	0	1	0	0	3
Discussion	10	1	10	0	6	8	0	7	42
Total	19	2	68	2	29	25	0	10	155
TOTAL	21		70		54		10		155

Referring to the raw frequency of these three linguistic devices, it is important to point out that a total amount of 647 instances in the passive voice have been found, while only 155 instances appear as personal active forms (see Table 70). Out of the 647 occurrences of passive forms, 95.67% (619 occurrences) correspond to agentless passives, which are passives whose agent is left out because it is obvious, irrelevant, and can be inferred from the context. When using the agentless passive, what gains importance is the process itself and the *thing* acted upon rather than the performer of the action. Figures reveal that passives including the agent are not frequent (28

occurrences), a fact which reinforces the idea that when writers wish to overtly express the doer(s) of the action, the active voice is preferred (see Table 71).

Table 71. Total number of tokens (raw frequency) of Agentive Passive thorugh a *by*-phrase (column '*by*') and Agentless Passives (column 'no Ag'), according to journal and section

Passive	**NEJM 1**		**NEJM2**		**LNC1**		**LNC2**	
(+/- Agent)	**by**	**no Ag**	**by**	**no Ag**	**by**	**no Ag**	**by**	**no Ag**
Abstract	0	6	0	9	0	12	0	10
Introduction	0	8	0	5	0	10	0	1
Methods	1	36	0	34	2	30	6	26
Results	0	13	2	28	0	16	3	16
Discussion	0	4	0	17	0	27	1	15
Total	1	67	2	93	2	95	10	68
TOTAL	68		95		97		78	

Passive	**BMJ1**		**BMJ2**		**JAMA1**		**JAMA2**		**TOTAL**	**TOTAL**
(+/- Agent)	**by**	**no Ag**	**by**	**no Ag**	**by**	**no Ag**	**by**	**no Ag**	**by**	**no Ag**
Abstract	0	4	0	3	0	6	0	3	0	53
Introduction	0	2	1	5	1	2	0	6	2	39
Methods	1	25	3	56	3	58	1	26	17	291
Results	0	6	1	12	0	2	1	12	7	105
Discussion	0	16	1	17	0	19	0	16	2	131
Total	1	53	6	93	4	87	2	63	28	619
TOTAL	54		99		91		65		647	

As shown in Table 72 and as expected, out of the three structures analysed, the agentless passive voice is the most frequent one (18.67 per 1,000 words), distantly followed by the *we* + active verb pattern (accounting for the 4.68 per 1,000 words); whereas the passive voice structure including the agent is barely used (only 0.84 per 1,000 words).

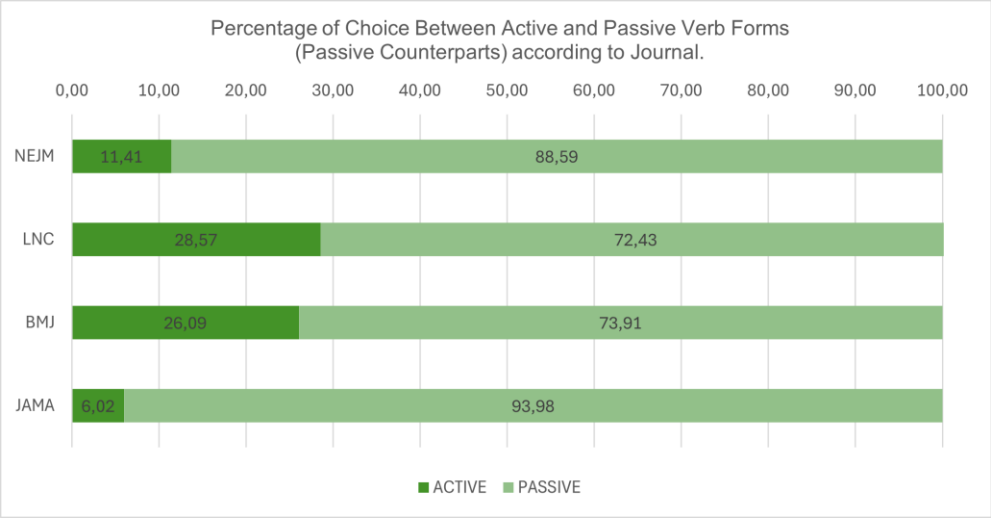
Table 72. Use of '*we* + Active verb' pattern, Agentive Passive and Agentless Passive verb patterns in RCTs according to journal. Normalised results per 1,000 words

Journal	*we* + active V	Agentive Passive	Agentless Passive
NEJM	3.20	0.46	24.37
LNC	6.78	1.16	7.07
BMJ	6.16	0.80	16.64
JAMA	1.33	0.80	20.02
TOTAL	4.68	0.84	18.67

The journals which show a higher degree of disparity between personal active forms and agentless passive structures are the JAMA (1.33 against 20.02 per 1,000 words, respectively) and the NEJM (3.20 against 24.37 per 1,000 words, respectively). Indeed, a relative degree of disparity is shown in the BMJ, which accounts for 6.16 per 1,000 words when dealing with personal forms and 16.64 with agentless passive forms. Almost no disparity is shown in RCTs in the LNC journals, which reveal that the personal active forms are almost equally used as the agentless passives (6.78 against 7.07 per 1,000 words, respectively).

These figures show that although the agentless passives are the most frequent structures, the overt expression of the authors through the personal active pattern cannot be dwarfed, since in two of the four journals analysed, the use of these personal forms is worth considering.

Figure 20. Percentage of choice between Active and Passive verb forms in RCTs, according to journal



When it comes to analyse not the normalised frequency of these forms per 1,000 words but the choice of grammatical voice, it is important to consider the overall number of personal active and passive verb forms found in RCTs (802 verbs[64]). Through the addition of all the instances accounting for the *we* + active verb and the two passive structures, with and without the agent, one can account for the percentage in which one form has been preferred over the other. As said, they are active-passive counterparts, and it is the writers' choice to use either one or the other (see Figure 20).

Out of all these forms, authors in JAMA and in NEJM clearly prefer the use of the passive voice against that of the personal active voice (6.02% for the *we* active structure against 93.98% for any of the passive forms in the BMJ; and 11.41 for the *we* active structure against 88.59% for any of the passive forms in the NEJM). On the other hand, authors in the LNC and the BMJ show a much higher probability of choosing the personal active form (28.57% and 26%, respectively), although its use still differs considerably from the use of passive structures (which accounts for the 72.43% and the 73.91%, respectively).

[64] The number of active and passive counterparts in the RCTS corpus divided by journals is as follows: NEJM, 184 verbs; LNC, 245 verbs; BMJ, 207 verbs; and JAMA, 166 verbs.

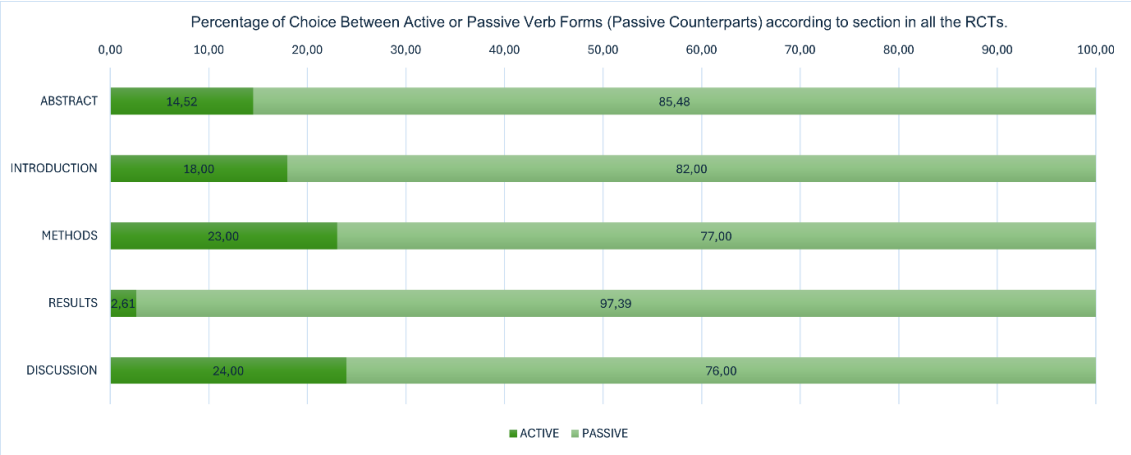
Similar figures result when analysing the use of these structures according to the sections of all the RCTs in which they appear. As shown in Table 73, and as expected, agentless passives outnumber the use of the other two structures (18.67 instances per 1,000 words in the agentless passives, against 0.84 in the passive with an agent, and 4.68 in the '*we* active' pattern). The personal active pattern appears more frequently in the Methods and in the Discussion section (8.10 and 5.10 tokens per 1,000 words, respectively), whereas these sections are the ones in which the passive is also highly frequent (25.62 – which is the most frequent – and 15.89 tokens per 1,000 words, respectively). The nature of these two sections (the Methods section, where all the procedures carried out are being described, and the Discussion section, where all the results obtained are being interpreted and analysed by the authors of the RCTs) explains the relatively high frequency of these two forms.

It is worth noting that the agentless passive is preferred in Abstracts, in Introductions and in Results, where barely any instances of passives with an agent occur. The use of *we* + active forms accounts for 2.94 (in Abstracts), 3.32 (in Introductions), and 0.39 tokens (in Results) per 1,000 words. This thus shows that the '*we* + active' pattern is rarely used in Results (0.39 instances per 1,000 words), where the abstraction found is more perceivable than in other sections, which accounts for a high frequency of agentless passives (13.50 instances per 1.000 words).

Table 73. Use of *we* + active verb pattern, Agentive Passive and Agentless Passive verb patterns in RCTs according to sections (including all journals). Normalised results per 1,000 words

Sections	*we* + active V	Agentive Passive	Agentless Passive
Abstract	2.94	0.00	17.30
Introduction	3.32	0.74	14.39
Methods	8.10	1.50	25.62
Results	0.39	0.90	13.50
Discussion	5.10	0.24	15.89
TOTAL	4.68	0.84	18.67

Figure 21. Percentage of choice between Active and Passive verb forms in RCTs, according to section in all RCTs



When exploring the use of active and passive counterparts[65] taking the sections of the RCTs as the main distinctive parameter (see Figure 21), a high degree of disparity between personal active and passive structures is mainly found in the Results section (2.61% of personal actives over a 97.39% of passives) and in the Abstract section (14.52% against 85.48%, respectively). In the Introduction, Methods, and Discussion sections, the use of the passive prevails over that of the personal active (82.00%, 77.00%, and 76.00%, respectively), but the use of active structures must also be considered, as almost two out of ten instances present this personal pattern over a passive structure (18.00%, 23.00%, 24.00%, respectively).

Once having analysed the frequency of active and passive structures in the RCTs, bearing in mind the journals in which they appear and their distribution along the sections which compose them, I will now focus on the exploration of real instances to assess their function as far as authorial visibility and authorial presence are concerned. Along with this close analysis, I will also check how the interaction between the

[65] The number of active and passive counterparts in the RCTs according to each section (and considering all journals) is as follows: Abstracts, 62 verbs; Introductions, 50 verbs; Methods, 400 verbs; Results, 115 verbs; and Discussions, 175 verbs.

different linguistic resources portrays a more approximating or a more detached sense in the text.

The personal active pattern is the one which more comprehensively represents the authorship of the actions carried out by the researchers and thus the one which extensively projects authorial visibility in the text, as seen in examples 5-9.

(5) Preladenant is an adenosine 2A receptor antagonist that reduced "off" time in a placebo-controlled phase 2b trial in patients with Parkinson disease (PD). **We sought** to confirm its efficacy in phase 3 trials. **[Personal active; Abstract; JAMA2]**

(6) **We originally designed** the trial to test whether long-term treatment with supplemental oxygen would result in a longer time to death than no use of supplemental oxygen among patients who had stable COPD with moderate resting desaturation (oxyhemoglobin saturation as measured by pulse oximetry [Spo2], 89 to 93%). **[Personal active; Abstract (Methods); NEJM1]**

(7) **We therefore designed** the Minimizing Adverse Haemorrhagic Events by Transradial Access Site and Systemic Implementation of Angiox (MATRIX) phase 3 programme in patients with acute coronary syndrome managed invasively via the radial or femoral route. In this large, randomised controlled trial **we assessed** whether bivalirudin is superior to unfractionated heparin with or without concomitant use of glycoprotein IIb/IIIa inhibitors. **[Personal active; Introduction; BMJ1]**

(8) Using a computer-generated random sequence, **we randomised** patients in a 1:1 ratio to receive the bivalirudin strategy or unfractionated heparin strategy, with a random block size stratified by type of acute coronary syndrome— ie, with ST segment elevation versus without ST segment elevation—intended or ongoing use of P2Y inhibitor (clopidogrel versus ticagrelor or prasugrel), and study site. **[Personal active; Methods; BMJ1]**

(9) After obtaining informed consent **we registered** patients via the trial's online process system, in which data were stored securely, and every patient received a unique trial code. **We randomly allocated** participants at the end of the elective midline laparotomy procedure, before closing the abdomen, securing optimum allocation concealment**. We used** a computer-generated randomisation sequence to

allocate patients to one of the three groups (…). **We stratified** randomisation by centre and operation indication. **[Personal active; Methods; LNC1]**

(10) **We describe** the epidemiologic, laboratory, and trace-back aspects of the investigation and **discuss** the public health implications of our findings. **[Personal active; Introduction; NEJM2]**

(11) Given that a specific wheat field was not implicated in this investigation, **we could not evaluate** whether animal intrusion was a source of contamination. **[Personal active; Discussion; NEJM2]**

As shown in these examples, the pronoun *we* explicitly refers to the writers of the RCTs, and is usually followed by material process verbs of procedure (like *randomise, calculate, allocate, use*) or mental process verbs of creation, decision, and ideation (like *design, evaluate, find, exclude*). These verbs are usually in the past tense, which is the most frequent tense within this pattern. This occurs because, on the one hand, authors portray themselves as the agents of the actions carried out in the research, and these actions took place in the past. As seen in examples 7, 9, and 10, the personal active pattern allows for the description of a sequence of procedures carried out by the writers, and each new main step is signalled by the presence of *we* + active verb.

In example 7, for example, authors comment on the fact that 'they designed a programme', and then focus on the fact that 'they assessed whether a specific treatment is better than another one'; each of these main steps concerning the procedures carried out is introduced by the personal active pattern. In example 9, however, although a major number of personal active forms clearly shows the evolutive sequence of steps carried out by the researchers, in between these steps, other types of structures are embedded to inform about other procedures whose importance is somehow mitigated. That is to say, in example 9, authors first 'obtain informed consent', then 'register the patients' (informing that 'the data had been previously stored securely'), and then they 'randomly allocate participants'. The actions of 'registering' and 'allocating' hold more relevance as they introduce the main steps carried out by the authors themselves, and the rest of the steps ('obtaining informed consent', for example) are diluted and downplayed as sub-steps. The active personal pattern is then a textual cohesive device

which serves as a guiding sign for the reader so as to present and describe the research procedures conducted.

Indeed, in the Introduction, the personal active voice pattern is used in 3.32 instances per 1,000 words, and it occurs in 18% of the instances. In general, this pattern tends to appear at the end of the Introduction to introduce the main outcomes of the study and to link them to the Methods section (see examples 12 and 13). After the pattern, the object of study, the procedure or research carried out, appears usually functioning as the object of the clause. By doing so, authors claim for their full authorship and thus maximally promote their visibility in the study under discussion.

(12) **We present** the results of prespecified, randomised comparisons of bivalirudin with heparin in two MATRIX subpopulations with acute coronary syndrome: 4010 patients presenting with persistent ST elevation and 3203 patients presenting without persistent ST segment elevation for whom percutaneous coronary intervention was planned at the time of coronary angiography. **[Personal active; Introduction; BMJ1]**

(13) **We aimed** to establish whether intravenous oxytocin 10 IU is more effective than intramuscular oxytocin 10 IU at preventing PPH at vaginal delivery and whether there is a higher incidence of side effects. **[Personal active; Introduction; BMJ2]**

Three introductions (NEJM1, LNC2, and JAMA1) disguise the authors by using an acronym (NEJM1) followed by a passive voice structure (example 14), and an abstract rhetor, an inanimate noun followed by an active verb (LNC2 and JAMA1), as seen in (example 15).

(14) **The Long-Term Oxygen Treatment Trial (LOTT) was originally designed** to test whether the use of suprasegmental oxygen would result in a longer time to death than no use of supplemental oxygen among patients with COPD and moderate resting desaturation (SPO_2, 89 to 93%). **[Acronym followed by an Agentless Passive voice structure; Introduction; NEJM1]**

(15) **This study tested** the efficacy and safety of deutrabenazine compared with placebo to control chorea while reducing peak concentration adverse effects. **[Abstract rhetor; Introduction; JAMA1]**

It is worth noting that the personal active form at the end of the Introduction, "to indicate actions by writers to fill the research gap that they had previously outlined" (Breeze 2010, p.166; Swales 1981, 1990; Swales & Najjar, 1987), becomes diluted and is transformed into a more impersonal structure at the end of the Discussion (except for in NEJM1 (example 16), and partially LNC1 (example 19)), where authors assess and evaluate if their hypothesis holds true. This could be regarded as a way to disguise authorial visibility aimed at promoting and reinforcing a sense of objectivity and impartiality and at avoiding full responsibility for their claims (examples 17-18, 19-20) (see Table 74). In example 18, for example, it is the data which show the final conclusion, and not the authors themselves, although in this case, there is a weaker mark of visibility through the possessive adjective 'our'.

Table 74. Linguistic resources shown at the end of the Introductions and at the end of the Discussions

Journals	End of the Introduction	End of the Discussion
NEJM1	acronym	we
NEJM2	we	abstract rhetor (our)
LNC1	we	we and abstract rhetors
LNC2	abstract rhetor	passive voice
BMJ1	we	passive voice
BMJ2	we	abstract rhetor
JAMA1	abstract rhetor	passive voice
JAMA2	we	abstract rhetor (our) and passive voice

(16) In conclusion, among patients with stable COPD and resting or exercise induced moderate desaturation, **we found** that long-term supplemental oxygen did not provide any benefit with respect to the time to death or first hospitalization or any sustained benefit with respect to any other measured outcome. **[Personal active; Discussion; NEJM1]**

(17) In 2016, a multistate outbreak investigation in the United States linked infection with STEC serogroups O121 and O26 to contaminated flour from a large domestic producer. **We describe** the epidemiologic, laboratory, and trace-back aspects of the investigation and **discuss** the public health implications of our findings. **[Personal active; Introduction; NEJM2]**

(18) **Our data show** that although it is a low-moisture food, raw flour can be a vehicle for foodborne pathogens. **[Abstract rhetor; Discussion; NEJM2]**

(19) **We initiated** the PRIMA trial (PRImary Mesh closure of Abdominal midline wounds) in 2009 with the aim to investigate prophylactic mesh reinforcement in high-risk groups (ie, patients with abdominal aortic aneurysm or a BMI ≥27 kg/m²). **We also aimed to** assess which mesh position in the abdominal wall should be used to prevent incisional hernia. **The primary aim of the PRIMA trial was to study** the effectiveness of prophylactic mesh reinforcement to prevent incisional hernia. **[Personal active; Introduction; LNC1]**

(20) **The PRIMA trial provides** level one evidence for the prevention of incisional hernia after midline laparotomy in patients at risk for incisional hernia. Closure of laparotomy with onlay mesh reinforcement has the potential to become the standard treatment in high-risk groups, which will reduce the socioeconomic burden of incisional hernia. **The results of the PRIMA trial also offer** future perspectives. **The next step will be a trial** in which onlay mesh reinforcement is combined with the small bites suture technique to lower the incidence of incisional hernia even further, because **the small bite technique has been shown** to be superior in closing midline laparotomy. **[Abstract rhetor, Agentless passive; and other impersonal structures; Discussion; LNC1]**

As opposed to the active personal pattern, where agents show maximal visibility, the passive voice demotes the agent by downplaying it, when it is introduced by a *by*-phrase (examples 21 and 23), or by totally demoting it (example 22), when the agent is omitted.

(21) Clinical, laboratory, and disease assessments were completed at regular study visits. **Abdominal MRI or CT scans were** done every 12 weeks and **rated by a central reader** masked to treatment allocation. All transfusions received during screening and throughout the study were recorded in the patients' diaries. **[Agentive passive; Methods; LNC2]**

(22) **Clinical, laboratory, and disease assessments were completed** at regular study visits. **Abdominal MRI or CT scans were done** every 12 weeks and rated by a central reader masked to treatment allocation. **All transfusions received during screening and throughout the study were recorded** in the patients' diaries. **[Agentless passive; Methods; LNC2]**

(23) We debated the choice of primary outcome and made a pragmatic decision to use PPH of 500 mL or more, although **blood loss in the range 500-1000 mL is well tolerated by most women in a developed world setting**. **[Agentive passive; Discussion; BMJ2]**

It is important to point out, however, that when the passive with an agent is used, the agent is always an external individual and thus does not refer to the authors of the RCTs (examples 21 and 23). Long passives are thus never used to disguise the authors of the text but are aimed to background the external doer of the action. The introduction of the agent can be seen as a recognition of external help in the procedural steps of the study presented (example 21), or as an identification of an agent which is less significant than the item being described (which functions as the subject of the clause) (example 23). Thus, both instances do not relate to authorial visibility.

Indeed, the same occurs with some agentless passives, in which procedural processes are considered to have been carried out by an external individual, which may be unknown or too obvious to explicitly mention it. That is the case of examples 24 and 25. The doers of the action refer to external individuals, whose mention does not add any relevant information to the study and there is no need to inform the reader about them. However, in the case of examples 26 and 27, the procedural processes have been carried out by the authors themselves, and not mentioning them through an agentless passive triggers a topicalization of relevant information, which is presented as new, and downplays the role of the authors involved. These instances do show a clear connection to authorial invisibility because authors opt to promote valuable information over their authorial presence in the text.

(24) A total of 40 women (3.7%) were ineligible after randomisation as **their babies were delivered** by caesarean section. **[Agentless passive; Results; BMJ2]**

(25) **Labour was induced** in over half the women, and one in three women required augmentation with an oxytocin infusion. **Anaemia was diagnosed** in 7% of women during pregnancy, and two thirds reported taking iron supplements, **which are routinely recommended** in Ireland. The spontaneous vaginal delivery rate was almost 70%, with a slightly higher operative vaginal delivery rate in the intravenous group. **[Agentless passive; Results; BMJ2]**

(26) Throughout the study, **the patients, caregivers or study partners, investigators, site personnel, data management staff, and steering committee were blinded** to treatment. **The safety review committee was blinded** to treatment assignment but evaluated treatment groups separately. **Patients were titrated** to an optimal study drug dose level over 8 weeks, followed by 4 weeks of maintenance therapy. **The 12-week treatment period was followed** by a 1-week washout. **Patients were randomized** in a 1:1 ratio to deutetrabenazine or placebo using a computerized randomization algorithm implemented via an interactive web-based randomization system. **[Agentless passive; Methods; JAMA 1]**

(27) **Women who were confirmed to be in labour or ready to have their labour induced were brought** into a room in the delivery suite. **A sample was taken** for a full blood count at the time of siting a venous cannula. **The first and second stages of labour were managed** as usual. **[Agentless passive; Methods; BNJ2]**

As seen in these instances, particularly when it comes to the description of the procedures carried out, the passive voice serves, like the personal active pattern, as a textual cohesive device which serves as a guiding sign for the reader. In addition, these expositions of research phases through agentless passive structures somehow imply the ideas of falsifiability and replicability, as any researcher would have carried out the research and would have got the same results. It is true, however, that all these passive structures, whether mentioning the agent or not, add a sense of impersonality to the text, as the objects under study are highlighted by occupying the subject position of the clause and the presence of agents is consciously mitigated.

A fact which is worth mentioning as far as the personal active and passive structures are concerned is the interplay between these distinct features (and with other features that will be presented in this chapter). This interrelatedness of personal and impersonal features is one of the main causes why medical discourse is seen as complex as far as

the notion of impersonality is concerned. The complexity of the different shades of (im)personality shown in the RCTs add different degrees of layers of interaction between the authors and the readers. The high use of the passive is representative of the discourse, a discursive convention, and portrays a clear sense of impersonality to the text. As for authorial visibility, personal instances come to the surface when authorial presence is desired for a specific purpose (to claim for authorship, for instance), but authorial visibility hides back underneath the surface when the focus is put on the research being described, demoting the presence of the agent.

The natural interplay between the personal active patterns and passive voice structures (and other impersonal structures) enriches scientific discourse and avoids its monotony. As the number of impersonal resources exceeds that of personal ones, when authors use a personal active structure, they display their authorial intervention in the process being described and this connotes a sense of trustworthiness and reliability to the text, as shown in examples 28-30.

(28) Highly symptomatic patients who declined enrolment might have had a different response to oxygen than what **we observed** in the enrolled patients. Second, **the lack of masking may have influenced** some of the patient-reported outcomes; however, it is unlikely to have influenced the primary outcome. Third, **we did not use** uniform devices for oxygen delivery; it is possible that **there was** variability in the amount of oxygen **delivered**. Fourth, **the immediate effects of oxygen on symptoms or exercise performance were not assessed**. **We did not measure** nocturnal oxygen saturation; some patients with COPD and severe nocturnal desaturation might benefit from nocturnal oxygen supplementation. **[Agentless passive, Personal active; Existential *there*, and Past participial reduced relative clause; Discussion; NEJM1]**

(29) Between March, 2009, and December, 2012, **498 patients were enrolled** to the study, of whom **18 were excluded** before randomisation. Therefore, **we included** 480 patients in the primary analysis: **107 were assigned** primary suture only, **188 were allocated** onlay mesh reinforcement, and **185 were assigned** sublay mesh reinforcement. **92 patients were identified** with an incisional hernia, 33 (30%) **who were allocated** primary suture only, 25 (13%) **who were assigned** onlay mesh reinforcement, and 34 (18%) **who were assigned** sublay mesh reinforcement (onlay mesh reinforcement *vs* primary suture, OR 0·37, 95% CI 0·20–0·69; p=0·0016;

sublay mesh reinforcement *vs* primary suture, 0·55, 0·30–1·00; p=0·05). Seromas were more frequent in patients **allocated** onlay mesh reinforcement (34 of 188) than in **those assigned** primary suture (five of 107; p=0·002) or sublay mesh reinforcement (13 of 185; p=0·002). **[Agentless passive, Personal active, Past participial reduced relative clause; Abstract (Results); LNC1]**

(30) In our post hoc analyses of trial 1, **we identified** a large placebo effect in Turkey, India, and Latin America, with numerically greater reductions in off time in these regions in the placebo group than in the preladenant or rasagiline groups. The exact reason for this finding is not known, but **a large placebo response was also observed** in a phase 3 monotherapy trial of preladenant in Latin America, India, Turkey, and Eastern Europe compared with North America and the European Union. **We are also aware** of a phase 2 trial of fipamezole as an antidyskinetic agent in which **benefit was demonstrated** in the United States but not in India. The differences could potentially be owing to clinical trial experience, cultural or language variations, genetic variation, or as yet unidentified reasons. Because **no stratification or block randomization was used**, our subgroup analyses may be subject to bias because they do not represent a fully randomized sample. **A subtype analysis was not performed**, and subtype response variance may also have affected results. **We found** that **there was** a striking difference in results between the first 50% of patients enrolled and the second 50% of patients enrolled in trial 1, the only trial to evaluate this. **[Agentless passive, Personal Active, Existential *there*; Discussion; JAMA2]**

In example 28, for example, there is a sequence of procedures being described. In two out of the four steps mentioned, the personal active serves to guide the reader through the complex procedural method (first and third). In one of them (fourth), the passive voice is used first to characterise the statement as valid and objectively recognised, but then it is somehow reassessed using a personal active structure, which brings to the forefront the role of the authors in the intervention being presented.

In contrast, in example 29, the high presence of agentless passive structures indicates that authors present the information in an objective fashion without making them present in the text. There is, however, one authorial realisation through a personal active structure, which serves to explicitly justify a particular intervention made by the

researchers. Once this intervention is exposed, the rest of the paragraph goes back to the objective and neutral tone with a great sum of agentless passive structures.

The complexity of the notion of impersonality and how this connects to authorial (in)visibility stems from the fact that both notions are represented by very similar or even the same lexicogrammatical resources. These linguistic resources may portray the text as being extremely impersonal and the addition of some hints of personality, mainly through the personal active structure, is used to counteract the abstraction impersonality carries with it.

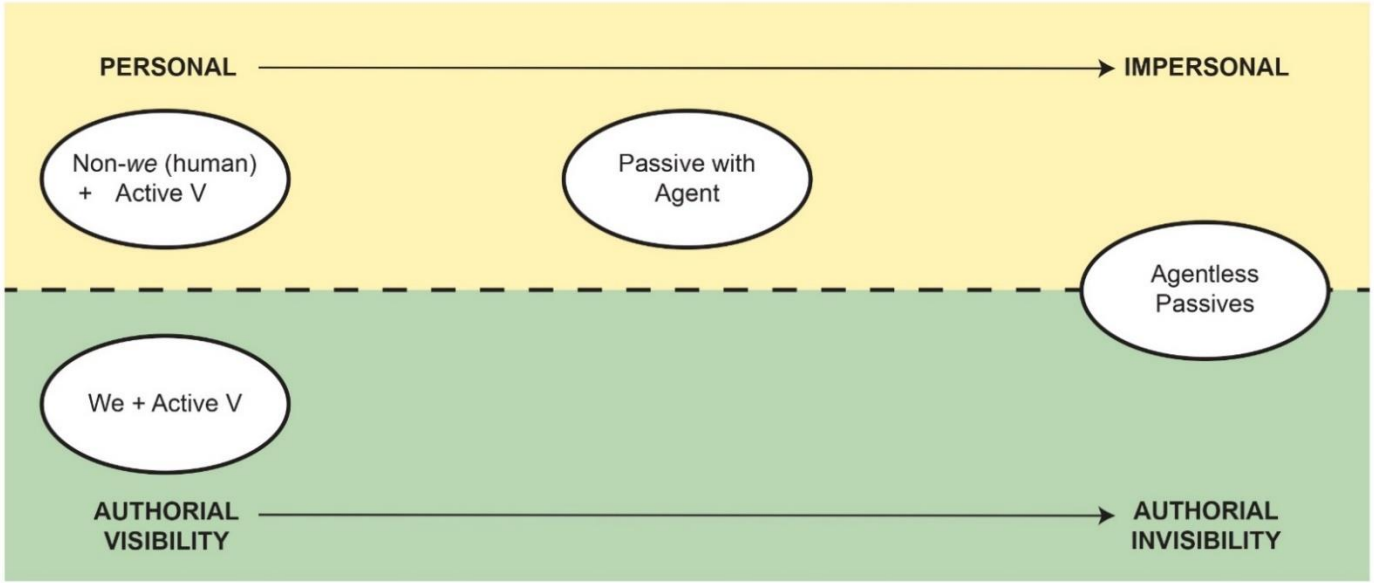
All these patterns can be graded within a continuum to show the layering of the different degrees of authorial visibility and intervention, on the one hand, and of the different shades of impersonalisation expressed in the text, on the other (see Figure 22). For this reason, the continuum in which these three patterns are allocated shows two distinct domains of study. The yellow area represents the notion of (im)personality considering the linguistic resources being dealt with, whereas the green area would represent the extent to which authorial visibility is shown and perceived in the texts.

If the focus is put on the cline from more personal to more impersonal structures, one may consider 'non-*we* (human) + active verb structures' as being the most personal ones, together with the already described '*we* + active verb' pattern. These clearly show the doer(s) of the action by promoting it and use active voice clausal constructions to transmit informational content. An intermediate position between personal and impersonal would be represented by the passive construction including the agent. In this case, importance is given to the foregrounded piece of information, functioning as the subject of the sentence, while the role of the agent is backgrounded and partially demoted. It is overtly expressed but the position it occupies in the sentence makes it become secondary and diminished. The maximal representation, so far, of impersonality is shown by agentless passives, which demote the agent almost totally by not even mentioning it. As it can be easily inferred from the context, the demotion is thus not complete.

On the other hand, if the focus is put on the cline from more authorial visibility to less authorial visibility, the patterns used to express this visibility are restricted to two: *we* + active verb and agentless passives. Pragmatically, the personal active pattern shows maximal negotiability and brings to the fore the role of the authors in the research process being described. The agent is fully promoted, and this highlights the responsibility for what is being claimed and reinforces the active intervention of the authors in the process.

Agentless passives, in contrast, show minimal negotiability and avoid the explicitness of authorial intervention, as it can be inferred from the context. What is interesting here is that agentless passives, indeed, may play a twofold role in the writing of Medicine and this is what makes them be so crucial and so frequently used. On the one hand, all the instances reinforce the impersonal dimension of the medical text and serve as conventionalised rhetorical resources; yet, on the other hand, only some instances reflect authorial intervention in the text and serve as replacement of *we* + active verb patterns when a more impersonal nature wants to be transmitted in the medical text as far as the procedural methods are concerned.

Figure 22. Continuum of lexicogrammatical structures (Personal Active and Passive) from more personal to more impersonal, and from more authorially visible to more authorially invisible



Promotion of Agent	Partial demotion of Agent through backgrounding	Almost total demotion of Agent
Maximal negotiability	Minor degree of negotiability	Minimal negotiability

6.2.2. Nominalisations and Stacked Nominal Phrases

As already justified in the introduction to this study, a quantificational recounting of nominalisations and stacked nominal phrases has not been carried out due to their omnipresent nature in medical discourse and the complexity to delimit its extent considering the various linguistic resources embedded within them. Furthermore, these structures show a high likelihood to be merged with other patterns and this constructs a fixed entity which heavily builds up onto the impersonal nature of medical RCTs.

It is important to point out, however, the distinction between a nominalisation and a stacked nominal phrase. A nominalisation occurs when a process (expressed through a verb) or a quality (expressed through an adjective) is transformed into an objectified entity (expressed through the nominalisation). This transcategorisation (from verbs or adjectives into nouns) results in a more objectified notion which is transmitted and perceived as being more objective, trustworthy, reliable, and permanent. In fact, it is so true that it becomes somehow factual and established (in orange, in examples 31, 32, and 33).

A stacked nominal phrase, on the other hand, is a simple head noun that does not derive from a verb, and it is usually pre-modified, post-modified or complemented by many other dependents (in purple, in examples 31, 34, and 35). The weight of these dependents generates a huge load of information packaged within one same phrase, a noun phrase, and this lexical density loads impersonal traits to the text.

(31) **Calculation of the final required sample** was based on **a time-to-composite event survival model with the use of the log-rank test statistic**. **[Nominalisation and stacked nominal phrase; Methods; NEJM1]**

(32) Therefore, **prevention of incisional hernia** is of paramount importance: it will lead to **reduction of disease** and is, thus, cost-effective. **[Nominalisations; Introduction; LNC1]**

(33) Besides **the physical examinations at 1 year and 2 years,** 283 (59%) of 480 patients also underwent **radiological examinations at 6 months and 2 years**, 60 in

the primary suture group, 115 in the onlay mesh reinforcement group, and 108 in the sublay mesh reinforcement group. Of the 376 patients who completed follow-up, 265 (70%) underwent **radiological examination**, 58 in the primary suture group, 105 in the onlay mesh reinforcement group, and 102 in the sublay mesh reinforcement group. **[Nominalisations; Results; LNC1]**

(34) Calculation of the final required sample was based on **a time–to–composite event survival model with the use of the log-rank test statistic**. **[Stacked nominal phrase; Methods; NEJM1]**

(35) In **patients with acute coronary syndromes managed invasively with coronary angiography and revascularisation**, **the optimal strategy for preventing coronary thrombosis and ischaemia**, while limiting bleeding, is uncertain. **[Stacked nominal phrases; Introduction; BMJ1]**

(36) **Participation of surgeons from different specialties** might have led to a learning curve in our trial, but this possibility is also a strong advantage of the PRIMA trial. **[Nominalisation; Discussion; LNC1]**

The nominalisations presented in these examples use nouns derived from verbs. The noun ‘calculation’ in example 31 derives from ‘calculate’, ‘prevention’ in example 32 comes from ‘prevent’, ‘examination’ in example 33 derives from ‘examine’, and ‘participation’ in example 36 from the verb ‘participate’. All these nouns are followed by post-modifiers or complements which further comment on the head noun they go with. These dependents, if included within a clausal construction, would be arguments of the verb and would therefore perform an explicit syntactic function in the clause. Examples 31, 32, and 36 have been rephrased into clausal constructions to show how these dependents would become different syntactic arguments within the clause (in examples 37 and 38, the argument performs the function of direct object; in example 39, however, it becomes the subject of the dependent conditional clause).

(37) We **calculated <u>the final required sample</u>** through a time-to-composite event survival model with the use of the log-rank test statistic. **[Rephrased from (30), using an SVO structure]**

(38) Even though we physically **<u>examined 480 patients</u>** at 1 year and 2 years, 283 (59%) of whom also underwent… **[Rephrased from (31), using an SVO structure]**

(39) If **surgeons from different specialties had participated (in the study)**, we might have found a learning curve in our trial, but this possibility is also a strong advantage of the PRIMA trial. **[Rephrased from (35), using a conditional clause with an SV structure]**

These nominalisations are connected to practices and actions carried out by the researchers, but are expressed as objectified entities, hiding thereby the authorship of the noun derived from a material verbal process. Nominalisations are largely related to the impersonalisation of medical discourse, insofar as they carry procedural and technical information, in most cases performed or dealt with by the researchers, in a detached fashion. It is true, however, that even though the authors are not explicitly represented, they can be inferred from the context.

Authors could have chosen to present such pieces of information overtly exposing themselves through a *we* + active voice structure (or a non-*we* active structure, in the case of example 38) or hiding themselves through an agentless passive. Instead, when using a nominalisation, they choose to add another impersonal layer to their intervention by objectifying the process and by giving it a full corporeal entity. Because of this, the notion of nominalisation is tightly linked to authorial invisibility, but this only occurs when the noun connects to the procedural intervention of the researchers themselves, which is in fact the case in most of the nominalisations found in RCTs.

In contrast, stacked nominal phrases are not connected to authorial invisibility; in fact, they are not related whatsoever to authors in the text. What distinguishes nominalisations from stacked nominal phrases, apart from the nature of their noun, is that, in most cases, nominalisations imply authorial invisibility, whereas stacked nominal phrases do not. Instead, they usually refer to objects under study, research-related subjects, patients, etc.

(40) Baseline characteristics were similar within **qualifying groups with acute coronary syndrome**, with **burden of risk factors for atherothrombosis and comorbidities (including hypertension, diabetes, hypercholesterolaemia, previous myocardial infarction) higher in patients who qualified for the group without ST segment**

elevation than those who qualified for the group with ST segment elevation (table 1). **[Stacked nominal phrase; Results; BMJ1]**

(41) Safety parameters included **assessment of adverse events, laboratory tests, vital signs, and electrocardiogram** and **assessment based on the following scales**: **UHDRS cognitive, behavioral, and functional scales**, **Epworth Sleepiness Scale**, **Columbia Suicide Severity Rating Scale**, **Swallowing Disturbance Questionnaire**, **UPDRS speech item**, **Barnes Akathisia Rating Scale**, **HADS**, and **Montreal Cognitive Assessment**. These scales were analyzed as changes from day 0 to week 12. **[Nominalisations and stacked nominal phrases; Methods; JAMA1]**

(42) Furthermore, **there is a risk of obstruction and strangulation of the bowel with perforation and possible mortality as a result**. **[Existential *there* and stacked nominal phrase; Introduction; LNC1]**

(43) In **this randomised controlled trial of women who delivered vaginally with active management of the third stage of labour**, **the incidence of postpartum haemorrhage** (PPH, ≥500 mL) was not significantly reduced in those assigned to intravenous oxytocin compared with intramuscular oxytocin but **there was a statistically significant reduction in the incidence of severe PPH (≥1000 mL), need for blood transfusion, and admission to a high dependency unit**. **[Existential *there* and stacked nominal phrase; Discussion; BMJ2]**

These examples, 40-43, clearly reflect the complexity of quantifying nominalisations and stacked nominal phrases. In example 40, it is possible to find stacked nominal phrases which function as post-modifiers or as complements of other stacked nominal phrases, which at the same time, include other elements that cause the phrase to be elongated. The important aspect here then is not the exact number of stacked nominal phrases but the huge load and density of lexical information being introduced in them, and the impersonalising effect they exert over medical writing. As seen, most of the dependents are embedded in the nominal group but, if conceived in isolation, would belong to other types of structures, such as non-*we* active structures, underlined in examples 40 and 43, for instance. In example 41, for example, the two nominalisations used serve to list a wide array of stacked nominal phrases which represent different scales considered for the study under discussion, some of which are expressed through

acronyms. In examples 42 and 43, apart from the stacked nominal phrases, it is possible to find existential *there* structures, which add, even, a more marked sense of impersonality to the whole structure.

(44) A total of 56 cases were identified in 24 states. **Univariable exact conditional logistic regression models of 22 matched sets showed** that infection was significantly associated with the use of one brand of flour (**odds ratio**, 21.04; 95% **confidence interval** [CI], 4.69 to 94.37) and with tasting unbaked homemade dough or batter (**odds ratio**, 36.02; 95% CI, 4.63 to 280.17). **Laboratory testing isolated the outbreak strains** from flour samples, and **whole-genome sequencing revealed** that the isolates from clinical and food samples were closely related to one another genetically. **Trace-back investigation identified** a common flour-production facility. **[Interplay of Nominalisations and Stacked nominal phrases with Abstract rhetors; Abstract (Results); NEJM2]**

(45) **Early studies investigating the use of momelotinib in patients with myelofibrosis showed reductions in spleen volume**, **improvement of disease-associated symptoms**, and **reductions in red blood cell (RBC) transfusion requirements**. **[Interplay of Abstract rhetors, present participial reduced relative clause, Stacked nominal phrases, and Nominalisations; Introduction; LNC2]**

(46) **Patients with more advanced disease were enrolled**, and **cognitive measures suggested** that the **existing baseline cognitive impairment** did not worsen with **deutetrabenazine exposure**. **[Interplay of Stacked nominal phrases and Abstract rhetors; Discussion; JAMA1]**

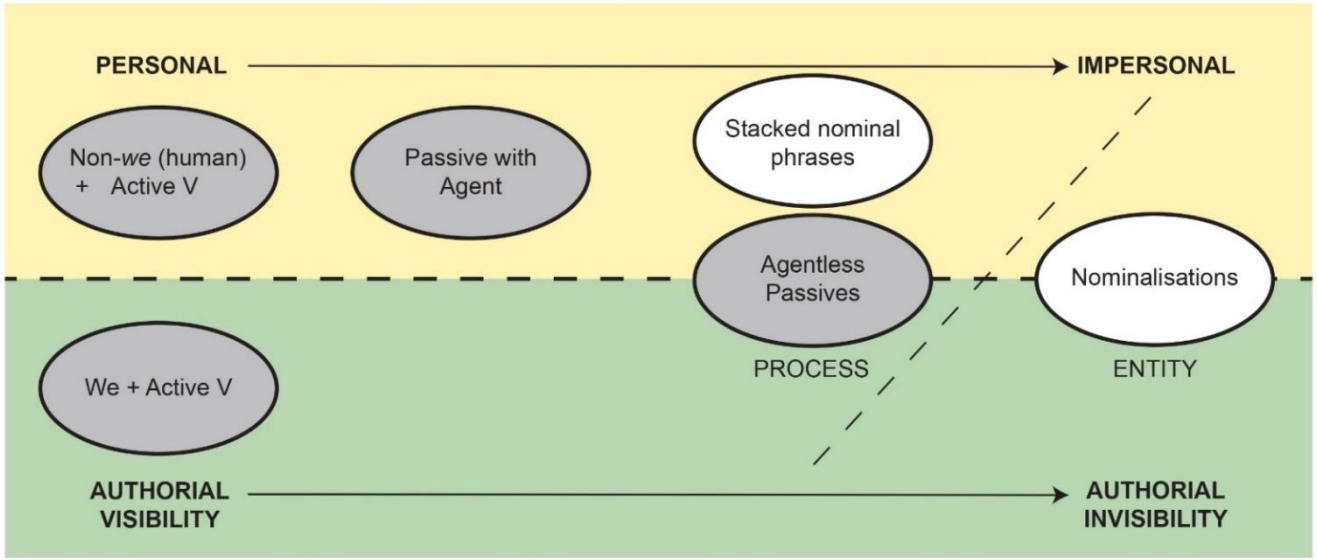
Nominalisations and stacked nominal phrases are strategic linguistic resources that serve to add to sentences a large amount of lexical information in a very condensed way. This, as already shown, triggers the natural interplay between these linguistic resources and other linguistic choices that reinforce the impersonal character of the text. That is the case of examples 44-46, which show an amalgamation of structures, particularly nominalisations and stacked nominal phrases which overlap with abstract rhetors. For the sake of simplicity, as far as the quantificational study is concerned, I have only counted them as abstract rhetors, as both notions cannot be really separated one from the other. The nominalisation or the stacked nominal phrase is presented as an

inanimate subject followed by a verb in the active voice. Because of the extensive number of examples found in the RCTs, despite not having been manually counted, nominalisations and stacked nominal phrases are pervasively the most recurrent lexicogrammatical patterns used in medical register, together with agentless passives. In addition, these forms are really connected to the notion of impersonality, and together with other impersonal structures, derive in a very powerful resource to portray an impersonal textual character typical of medical register.

In the continuum presented (see Figure 23), I have included these two linguistic resources as being part of the strategies of impersonalisation used by medical writers. Nominalisations operate similarly to agentless passives, insofar as they hide the authors' involvement in the research carried out, in this case, by means of a noun, which is perceived as objectified and established. They are represented as entities, whereas agentless passives represent processes, which I consider connote lightly a more personal contribution to the text. That is why I placed nominalisations as being a bit more impersonal than agentless passives, as there is an extra layer of impersonalisation being extended by the transformation of a process into a noun.

In addition, nominalisations are in between the two domains (that of impersonality and that of authorial invisibility). When nominalisations semantically carry information connected to material procedures done by the researchers, they clearly show authorial detachment. However, in some cases, they are not connected to procedures carried out by researchers but to other individuals. In that case, nominalisations are simply impersonal linguistic strategies that reinforce the impersonal nature of the text, a fact which coincides with what happens with stacked nominal phrases.

Figure 23. Continuum of lexicogrammatical structures (Stacked Nominal Phrases and Nominalisations) from more personal to more impersonal



Promotion of Agent	Partial demotion of Agent through backgrounding	Almost total demotion of Agent
Maximal negotiability	Minor degree of negotiability	Minimal negotiability

6.2.3. Abstract Rhetors or Impersonal Active Structures

An abstract rhetor, or what is the same, an inanimate subject followed by an active verb, consists of another linguistic strategy medical writers recurrently use. It extends a sense of detachment and impersonality over research activities on medical writing. In fact, it serves as a rhetorical device which typically introduces any kind of research activity or result, which may have been carried out or obtained by the authors themselves (like in examples 47, 48 and 50) or by authorities or other researchers in the field (like in example 49), but that is described as occurring independently from them.

(47) **This investigation implicated** raw flour as the source of an outbreak of STEC infections. **[Abstract rhetor; Abstract; NEJM2]**

(48) Finally, **subgroup analysis indicates** that a strategy of bivalirudin monotherapy with glycoprotein IIb/IIIa inhibitors restricted to angiographic complications might be useful in patients with ST segment elevation pretreated with heparin and with renal dysfunction. **[Abstract rhetor; Discussion; BMJ1]**

(49) **Current guidelines for treatment of myeloproliferative neoplasm1 recommend** ruxolitinib, a Janus kinase (JAK) inhibitor, for symptomatic myelofibrosis. **[Stacked nominal phrase; Introduction; LNC2]**

(50) **Both the odd ratios and the proportions of responders showed** no significant differences between preladenant or rasagiline vs placebo. **[Stacked nominal phrase; Results; JAMA2]**

(51) **The laboratory component of the investigation also faced** difficulties. **[Stacked nominal phrase; Discussion; NEJM2]**

In the examples provided, it is clear to identify the syntactic function that most abstract rhetors perform within a clausal construction: they usually appear as the subject of the clause, and therefore are placed in initial position. In the case of examples 47, 48, 50, and 51, the head nouns of the nominalisations (examples 47 and 48) or stacked nominal phrases (examples 49-51) are humanised by means of a metonymic expression, in which the head acquires human features described by the verb in the active voice. One the one

hand, examples 47, 48, 50, and 51 show a clear connection between the head noun being humanised and the researchers of the RCT and this connection is hidden by means of the nominalisation (or stacked nominal phrase), connoting thereby authorial invisibility and denoting a sense of detachment and impersonality.

On the other hand, in example 49, although the head noun implies human agency, the hidden agents involved are different from the researchers (usually other researchers, institutions, or other entities). In this case, the abstract rhetor and the noun within it help in the impersonal transmission of informational content but do not suggest any kind of active authorial involvement in the procedural process. Again, like in the agentless passive and in nominalisations, abstract rhetors serve, in all cases, to impersonalise the medical text, but only those instances mentioning the procedures carried out by the researchers are the ones lightly representing their direct authorial participation.

As for the raw frequency concerning abstract rhetors, they appear in 165 occurrences in all the RCTs, so they are relatively frequent (see Table 75). They are frequent particularly in the NEJM2 (55 occurrences). In the rest of the journals, abstract rhetors are almost equally used (39 instances in JAMA, 36 instances in LNC, and 35 instances in BMJ). Overall, the average appearance rate in all the RCTs is 20.63 occurrences per journal. As for their relative frequency, the journal which makes a more extensive use of abstract rhetors is the NEJM (0.84 instances per 1,000 words), followed by the JAMA (0.52 instances per 1,000 words).

Table 75. Total number of tokens (raw frequency) of abstract rhetors according to journal and section

Abstract rhetors	NEJM 1	NEJM2	LNC1	LNC2	BMJ1	BMJ2	JAMA1	JAMA2	Total
Abstract	0	6	0	1	1	0	1	1	10
Introduction	4	1	1	4	1	2	2	3	18
Methods	6	10	5	2	3	4	4	6	40
Results	4	12	3	1	0	2	1	9	32
Discussion	3	9	12	7	8	14	3	9	65
Total	17	38	21	15	13	22	11	28	165
TOTAL	55		36		35		39		**165**

The distribution of abstract rhetors across the RCTs changes depending on the section, but it must be pointed out that they are spread across all the sections of the RCT (see Tables 76 and 77). It is worth noting, however, that the Introduction and the Discussion sections are the ones in which they appear more frequently (0.66 and 0.79 tokens per 1,000 words, respectively). The Introduction deals with setting the scene and framing the medical project presented in the RCT, whereas the Discussion evaluates and reasons the results obtained. Both sections make use of referential works, studies, pieces of research (made by the authors themselves or by other authors) and this may contribute to a higher presence of abstract rhetors in these sections.

Table 76. Use of Abstract Rhetors in RCTs according to journal.
Normalised results per 1,000 words

Journal	Abstract rhetor
NEJM	0.84
LNC	0.35
BMJ	0.40
JAMA	0.52
TOTAL	**0.50**

Table 77. Use of Abstract Rhetors in RCTs according to sections (including all journals).
Normalised results per 1,000 words

Sections	Abstract rhetor
Abstract	0.33
Introduction	0.66
Methods	0.35
Results	0.41
Discussion	0.79
TOTAL	**0.50**

As already mentioned, abstract rhetors emit a sense of impartiality, empiricism, and falsifiability. Their effect is the same as that of nominalisations and stacked nominal phrases, since both linguistic resources somehow overlap, and their addition strengthens an atmosphere of impersonality which serves to present the implications of the studies under analyses in the RCTs in a detached way. Abstract rhetors serve to shift the focus by highlighting the object under study and they do so by linking observational, experimental, and procedural evidence to conclusive and resultative facts, as can be seen in examples 52-56.

(52) **Further analysis of the quality-of-life measures for patients with and without an incisional hernia showed** no differences in scores on the SF-36 or EQ-5D questionnaires (table 4). **[Abstract rhetor; Results; LNC1]**

(53) **Our findings support** the use of intravenous oxytocin rather than intramuscular oxytocin in terms of preventing severe PPH and the need for blood transfusion, with the potential to prevent mortality if implemented globally in developing world settings. **[Abstract rhetor; Discussion, BMJ2]**

(54) **In-vitro studies have shown** potent inhibitory activity against wild-type JAK and the JAK2Val617Phe mutant. **[Abstract rhetor; Introduction, LNC2]**

(55) Notably, **pooled analyses indicated** that patients in the fourth quarter of enrolment were more likely to be placebo responders than early-enrolling patients. **[Abstract rhetor; Discussion; JAMA2]**

(56) **Investigations of future outbreaks will need to account for** the fact that laboratory procedures using lactose fermentation as a screening step for STEC O121 may reduce the likelihood of recovering the pathogen. **[Abstract rhetor; Discussion; NEJM2]**

In these examples, head nouns such as *studies*, *analyses*, or *findings* are seen to typically occur within this construction as the subjects of the clause, a fact which emphasises the transition from procedural processes, hidden behind these nouns by means of a nominalisation or a stacked nominal phrase, into results or conclusive arguments, described by the direct object following the main verb. The verb in the active is chosen to express the degree of achievement (or the strength) of the argument

being proposed as a conclusion. Verbs like *suggest*, *indicate*, and *show*, for instance, represent a more neutral positioning of the authors towards the arguments they (or others) are expressing (like in examples 52, 54, and 55). As opposed to this, verbs like *recommend*, *support*, *account for*, or *investigate* express a stronger authorial stance towards the proposition being made (examples 49, 53, 56, and 57).

It has been made clear that nominalisations and stacked nominal phrases usually overlap, since the subject of the abstract rhetor tends to be fulfilled by a nominalised structure. However, abstract rhetors also interplay with other linguistic resources, such as personal active and impersonal passive structures. In example 57, for example, authors begin by showing themselves after setting the scene. Once the scene is set and their intervention is clearly stated, they make use of abstract rhetors to comment on the rest of the procedures carried out. After that, they surface again to comment on a last procedure.

(57) For the first stage of analysis **we used** descriptive statistics to describe recruited women and to investigate comparability of the trial arms at baseline. **The primary analyses involved** intention to treat comparisons between the two groups for the primary and secondary outcomes **adjusted** for operative vaginal delivery and the stratification variable (parity). **Secondary analyses investigated** the effects of further adjustment for any variables **displaying** imbalance between the arms at baseline. **All analyses used** logistic or linear regression models, with results **presented** as point estimates (odds ratios or difference in means), 95% confidence intervals, and P values. **We calculated** the number **needed** to treat (NNT) with 95% confidence intervals for primary and secondary PPH outcomes. The NNT is the reciprocal of the absolute risk difference, or 1/ adjusted risk ratio. **[Interplay of Abstract rhetors, Personal actives, and Past and Present participial reduced relative clauses; Methods; BMJ2]**

In example 58, however, the interplay between abstract rhetors and other linguistic resources seems much more complex. For example, both abstract rhetors and arguments functioning as direct objects include reduced relative clauses, connoting thereby a depersonalising effect of the propositional content being described. Example 57 is a very good example to see how personal resources of all types interact with a wide array

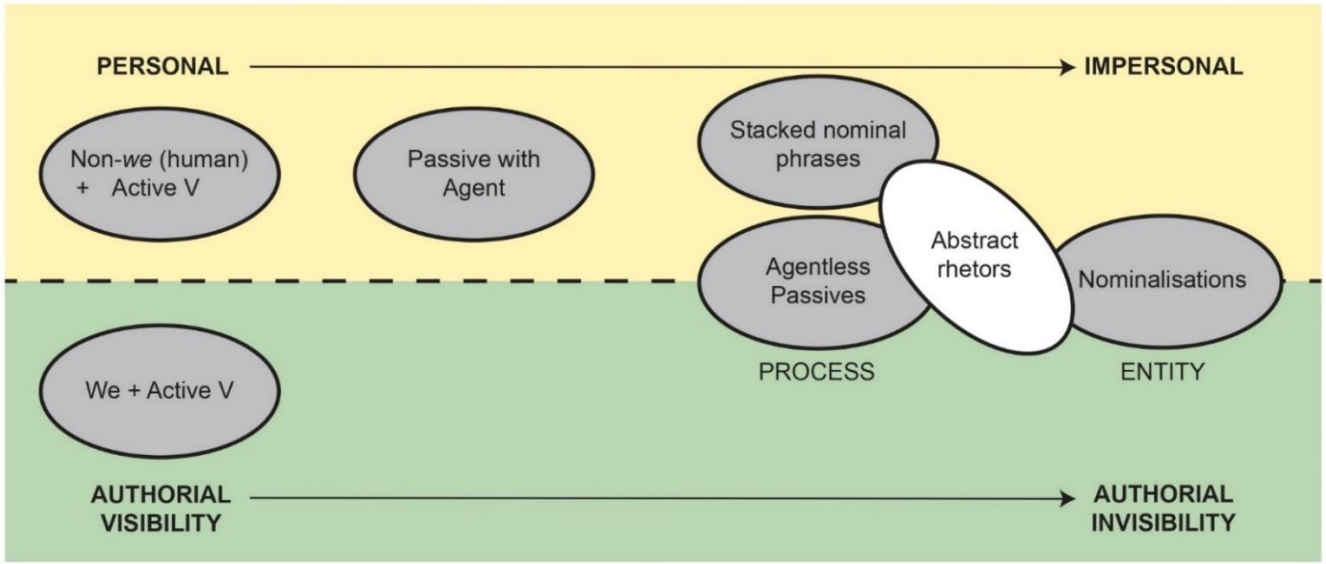
of impersonal linguistic choices, and all this interaction between structures seems to portray to the full extent the complex dimension of impersonality.

(58) **Several Cochrane systematic reviews have addressed** the use of uterotonic agents to prevent PPH at vaginal delivery. **The most recent review, updated in 2015, included** seven studies **comparing** active management of the third stage of labour with physiological management using a range of uterotonic agents, including ergometrine, syntometrine, intramuscular oxytocin, and intravenous oxytocin. **Active management was associated** with a reduced incidence of PPH irrespective of the agent **used**, but **none of the studies were graded** as high quality. **A review in 2013 of 20 studies comparing prophylactic oxytocin with placebo or alternative agents reported** that prophylactic oxytocin at any dose decreased PPH greater than 500 mL. **The authors recommended** the intravenous route, where the evidence was strongest, with the intramuscular route as an alternative, and **concluded** that **further high quality evidence was required**. **A Cochrane review published in 2012 that specifically compared intramuscular oxytocin with intravenous oxytocin for the prevention of PPH at vaginal delivery identified** no randomised controlled trials (published or unpublished). **[Interplay of Abstract rhetors, Agentless Passives, Personal actives, Past and Present participial reduced relative clauses; Discussion; BMJ2]**

To locate abstract rhetors within the last version of the continuum (see Figure 24), it is worth noting that these lexicogrammatical devices must be placed in between the yellow area (representing the (im)personal dimension) and the green area (representing authorial (in)visibility). In any case, all abstract rhetors represent impersonality. However, not all of them can be linked to authorial (in)visibility. When the nouns forming the subject of the abstract rhetor refer to experimental procedures carried out by the researchers, and when these nouns are connected to nominalisation devices, authorial invisibility is shown. If these nouns are simply nouns forming a stacked nominal phrase, abstract rhetors only serve to add a higher degree of impersonalisation to the text. Lastly, if these nouns are connected to nominalisations but they refer to procedures that other members of the medical community have carried out, there is invisibility of these authors. Yet, in this last case, there is no connection with the actual researchers of the RCT.

Not only does the structure used have to do with authors' visibility, but also with the choice of verb. If the verb is a procedural verb, it favours the (in)visibility of the actual writers of the paper. This also happens with agentless passive forms and nominalisations. That is why abstract rhetors, within the continuum, are represented as taking up a space occupied by these other structures.

Figure 24. Continuum of lexicogrammatical structures (Abstract Rhetors)
from more personal to more impersonal



Promotion of Agent	Partial demotion of Agent through backgrounding	Almost total demotion of Agent
Maximal negotiability	Minor degree of negotiability	Minimal negotiability

6.2.4. Reduced Relative Clauses. Present Participial and Past Participial Reduced Relative Clauses

Reduced relative clauses are non-finite subordinate clauses which are dependent on another element in the clause, usually a noun. They are called *reduced* because they lack both the relative pronoun and the verb form of the auxiliary *be*. Their prototypical function is that of post-modifiers of the noun they go with, as they are used to semantically restrict its sense. They are embedded dependents within the noun phrase, and this causes the phrase (and by extension, the whole clause) to be packed with a huge load of semantic information, which, if expressed through simple sentences, would make the text oversimplified, monotonous, and much more extensive. Therefore, these non-finite clauses function as linguistic resources that add specification and descriptive information that enrich the clause and allow for a packaged type of information. These structures represent a sub-layer of information which is posited as giving nuances about the main layers of information, expressed through most of the linguistic resources dealt with in this study.

It is interesting to point out that these two linguistic strategies are somehow connected to both the active and the passive voice, insofar as they can be easily rephrased by using a finite active voice structure, in the case of the *-ing* or present participial reduced relative clause, or a finite passive sentence, in the case of the *-ed* or past participial reduced relative clause, which is also known as a *bare passive*.

Table 78. Total number of tokens (raw frequency) of Present Participial Reduced Relative Clauses according to journal and section

***-ing* RRCl.**	**NEJM 1**	**NEJM2**	**LNC1**	**LNC2**	**BMJ1**	**BMJ2**	**JAMA1**	**JAMA2**	**Total**
Abstract	0	0	1	3	0	0	2	1	7
Introduction	0	0	5	1	2	4	5	2	19
Methods	5	0	4	1	10	7	8	4	39
Results	1	2	1	16	0	0	10	13	43
Discussion	1	2	2	5	0	8	6	2	26
Total	7	4	13	26	12	19	31	22	134
TOTAL	11		39		31		53		**134**

As for the raw frequency of present participle reduced relative clauses, it is worth noting that they are less frequent when compared to their past participle counterparts. According to the different journals analysed in this study, they tend to appear within the Methods, the Results, and the Discussion section. However, it is also true that some sections of some of the RCTs do not account for any instances of *-ing* reduced relative clauses (see Table 78). The journal which shows the highest use of this linguistic device is JAMA, in which there are 53 instances (7.07 tokens per 1,000 words), followed by the LNC (39 occurrences; 3.78 instances per 1,000 words), and the BMJ (31 occurrences; 3.53 instances per 1,000) (see Tables 79 and 80).

Table 79. Use of Present Participial Reduced Relative Clauses in RCTs according to journal. Normalised results per 1,000 words.

Journal	*-ing* Red. Rel. Clause
NEJM	1.68
LNC	3.78
BMJ	3.53
JAMA	7.07

Table 80. Use of Present Participial Reduced Relative Clauses in RCTs according to sections (including all journals). Normalised results per 1,000 words.

Sections	*-ing* Red. Rel. Clause
Abstract	2.28
Introduction	7.01
Methods	3.43
Results	5.53
Discussion	3.15
Total	4.04

When it comes to deal with the relative frequency, the highest number of tokens per 1,000 words appears in the Introduction (7.01 per 1,000 words), followed by the Results section (5.53 per 1,000 words). The main difference between raw and relative

frequencies is obviously the higher number of words that compose the Methods section (11,358 words) when compared to other sections, such as the Introduction (2,711 words) and the Results section (7,777 words). Overall, *-ing* reduced relative clauses represent 4.04 per 1,000 words.

Their presence in medical literature may not be highly frequent; however, it is worth treating them as independent choices writers make to extend the informational content of a clause. Present participle reduced relative clauses appear following nouns, which can also be pre-modified by other elements. All these pre-modifying dependents, the noun, and the *-ing* reduced relative clause form a stacked nominal phrase, a very complex type of noun phrase packed with a dense amount of semantic information.

(59) Onlay mesh reinforcement has the potential to become the standard treatment for high-risk **patients undergoing** midline laparotomy. **[Noun followed by a Present participial reduced relative clause; Abstract; LNC1]**

(60) More **patients receiving** momelotinib (27 [26%] of 103 evaluable patients) had a reduction in total symptom score of at least 50% from baseline based on the modified MPN-SAF compared with **those receiving** BAT (three [6%] of 51 evaluable patients), **indicating** greater symptomatic improvement in **patients receiving** momelotinib (nominal p=0.0006,figure 3). **[Noun followed by a Present participial reduced relative clause; Results; LNC2]**

(61) The number of **patients experiencing** adverse events overall and within psychiatric and nervous system areas were similar in the deutetrabenazine and placebo groups (Table 3). **[Noun followed by a Present participial reduced relative clause; Results; JAMA1]**

(62) The number of **patients reporting** depression or agitated depression as an adverse event were not significantly different in the 2 groups. **[Noun followed by a Present participial reduced relative clause; Results; JAMA1]**

Examples 59-62 show that a typical head noun found in this structure is ‘patients’ (although sometimes this word works as part of the complement of the noun ‘number’, like in ‘a number of patients’).

(63) In total, nearly **250 products containing** flour were recalled. **[Noun followed by a Present participial reduced relative clause; Results; NEJM2]**

(64) In addition, many case patients also reported exposure to chocolate chips, but additional epidemiologic and laboratory **evidence supporting** flour as the source helped to rule out this food. **[Noun followed by a Present participial reduced relative clause; Discussion; NEJM2]**

It is worth noting that all these examples are not related to the writers' involvement in the study being described, so they do not serve to account for their visibility in the text. However, they do reflect a hint of impersonalisation, as writers could have opted for a more personal active construction. Writers may have considered the information presented in the reduced relative clause as descriptively relevant, but not as relevant as to present it within an independent clausal construction. The same occurs in example 65, although in this case "the *-ing* reduced relative clause" is introduced as a pre-modifying adjective, which has been transcategorised from the verb it derives, 'operate'.

(65) The **operating (vascular or gastrointestinal) surgeon, urologist, or gynaecologist** closed the abdomen, not a specialised abdominal wall surgeon. **[Noun pre-modified by an adjective, transcategorised from its verb; Methods, LNC1]**

Present participle reduced relative clauses can also post-modify nouns whose nature connotes procedures carried out by some other researchers, like 'studies' or 'analyses', interplaying here with nominalisations (example 66), and with abstract rhetors (examples 67-69).

(66) The most recent review, updated in 2015, included **seven studies comparing** active management of the third stage of labour with physiological management using a range of uterotonic agents, including ergometrine, syntometrine, intramuscular oxytocin, and intravenous oxytocin. **[Nominalisation followed by a Present participial reduced relative clause; Discussion; BMJ2]**

(67) **Early studies investigating** the use of momelotinib in patients with myelofibrosis **showed** reductions in spleen volume, improvement of diseaseassociated symptoms, and reductions in red blood cell (RBC) transfusion requirements. **[Interplay between an Abstract rhetor, a Nominalisation, followed by a Present participial reduced relative clause; Introduction; LNC2]**

(68) **A review** in 2013 of **20 studies comparing** prophylactic oxytocin with placebo or alternative agents **reported** that prophylactic oxytocin at any dose decreased PPH greater than 500 mL.11 **[Interplay between an Abstract rhetor, a Nominalisation, followed by a Present Participial reduced relative clause; Discussion; BMJ2]**

(69) **Analyses of potential regional differences (Table 3) found that Turkey, India, and Latin America had the largest mean reductions in off time in the placebo group (range, −1.00 hour to −2.15 hours), leading** to numerically greater reductions in off time in the placebo group than in the preladenant or rasagiline groups. **[Nominalisation followed by a Present participial reduced relative clause; Results; JAMA2]**

All the examples presented above suggest that the interplay between all these different linguistic resources adds a great deal of complexity to the dimension of impersonality in medical writing. But this complexity is even more strengthened with the addition of past participle reduced relative clauses. *-ed* clauses are much more frequent than present participial reduced relative clauses, and as said, they are reduced versions of finite passive constructions. As for their raw frequency, *-ed* reduced relative clauses occur 325 times in all the RCTs. The BMJ is the journal in which they appear more frequently (113 tokens; 12.88 instances per 1,000 words), followed by JAMA (80 tokens; 10.68 instances per 1,000 words). It seems that it is a very resourceful device to posit an impersonal atmosphere to the medical text (see Table 81).

Table 81. Total number of tokens (raw frequency) of past participial reduced relative clauses according to journal and section

***-ed* RRCl.**	**NEJM 1**	**NEJM2**	**LNC1**	**LNC2**	**BMJ1**	**BMJ2**	**JAMA1**	**JAMA2**	**Total**
Abstract	1	2	3	7	13	7	9	2	44
Introduction	1	1	2	0	5	2	5	1	17
Methods	3	6	12	14	13	14	21	2	85
Results	4	14	17	21	17	5	3	8	89
Discussion	3	4	10	7	17	20	21	8	90
Total	12	27	44	49	65	48	59	21	325
TOTAL	39		93		113		80		**325**

Table 82. Use of Past Participle Reduced Relative Clauses in RCTs according to journal. Normalised results per 1,000 words

Journals	***-ed* Red. Rel. Clause**
NEJM	5.94
LNC	9.01
BMJ	12.88
JAMA	10.68

Bare passives tend to appear many more times in the Discussion (90 tokens), Results (89 tokens), and Methods (85 tokens) than in the other two sections. However, when focusing on the relative frequency, it is possible to state they are very useful linguistic choices within the Abstract, as they account for 14.36 instances per 1,000 words, the highest, followed by the Results and the Discussion sections (with 11.44 and 10.92 instances per 1,000, respectively). Abstracts tend to make extensive use of these non-finite clauses, as they are reduced versions of finite passive structures and thus are used as packaging devices, which compress the message to its full extent, allowing thereby the transmission of the essential information in a section conceived as restrictively limited (see Tables 82 and 83).

Table 83. Use of Past Participle Reduced Relative Clauses in RCTs according to sections (including all journals). Normalised results per 1,000 words

Sections	*-ed* Red. Rel. Clause
Abstract	14.36
Introduction	6.27
Methods	7.48
Results	11.44
Discussion	10.92
Total	9.80

In the following examples, 70-74, it is possible to assert that this structure usually collocates with the word 'patients', the same one found within structures where the present participial reduced relative clause was used. Indeed, one aspect which is worth considering, despite not being automatically related to this lexicogrammatical pattern, is the repetition of some structures to assure the correct understanding of the data being presented (example 70). In this sense, authors avoid omitting important information, particularly when presenting figures, and it seems they do not mind using the same pattern repeatedly. This suggests authors prime comprehensibility and straightforwardness over stylistic factors.

All these instances also show the clear connection of the past participles with the procedures carried out by the researchers of the study under discussion. Past participles like *assigned*, *defined*, *randomised*, and *assessed* are typically found in these structures. As these are material or mental processes carried out by the authors of the paper, it is important to state that they somehow express their invisibility in the text, the same denoted when using agentless passives. So, these structures do have a connection with the lack, in this case, of authorial presence in the text.

(70) Use of a glycoprotein IIb/IIIa inhibitor in **patients assigned to** heparin was planned at baseline in 30.7% of patients with ST segment elevation, in 10.9% without ST segment elevation, and in no **patients assigned to** bivalirudin. In patients with ST segment elevation, major adverse cardiovascular events occurred in **118 (5.9%) assigned to** bivalirudin and **129 (6.5%) assigned to** heparin (rate ratio 0.90, 95% confidence interval 0.70 to 1.16; P=0.43), whereas net adverse clinical events

occurred in **139 (7.0%) patients assigned to** bivalirudin and **163 (8.2%) assigned to** heparin (0.84, 0.67 to 1.05; P=0.13). **[Noun followed by a Past participial reduced relative clause; Abstract; BMJ1]**

(71) Overall **patients randomized** to preladenant (5 mg or 10 mg twice daily) or rasagiline **enrolled** in the first have demonstrated significant improvement vs placebo in reduction in off time (approximately 1 hour or more) (eTable in Supplement 3). **[Noun followed by a Past participial reduced relative clause; Results; JAMA2]**

(72) The primary **outcome** in the time-to-event analysis, **measured** from randomization, was the composite of death or first hospitalization. **[Noun followed by a Past participial reduced relative clause; Methods; NEJM1]**

(73) This study was designed to detect a 2.7-unit treatment difference in scores. **The secondary end points**, **assessed** hierarchically, were the proportion of patients who achieved success on the Patient Global Impression of Change (PGIC) and on the Clinical Global Impression of Change (CGIC). **[Noun followed by a Past participial reduced relative clause; Abstract; JAMA1]**

(74) No significant difference was noted in the **subgroups defined** according to oxygen prescription, desaturation profile, race, sex, smoking status, nadir SPO_2 during exercise (the **10^{th} lowest SPO_2 observed** during the 6-minute walk), forced expiratory volume in 1 second, BOD index, SF-36 physical component score, body-mass index, or history of anaemia (Table S6 in the Supplementary Appendix). **[Noun followed by a Past participial reduced relative clause; Results; NEJM1]**

Examples 75 and 76 show the interplay between different linguistic strategies. Although they are not directly related, as the *-ed* clause does not post-modify any element belonging to the other linguistic choices, the fact that they appear within the same clausal construction elucidate the intricacy of the dimension of impersonality in scientific discourse. Features representing a more personal-like approach are being turned on and off amidst a wide array of impersonal resources that qualify medical writing as detached and unbiased.

(75) **We found** that **there was** a striking difference in results between the first 50% of **patients enrolled** and the second 50% of **patients enrolled** in trial 1, the only trial

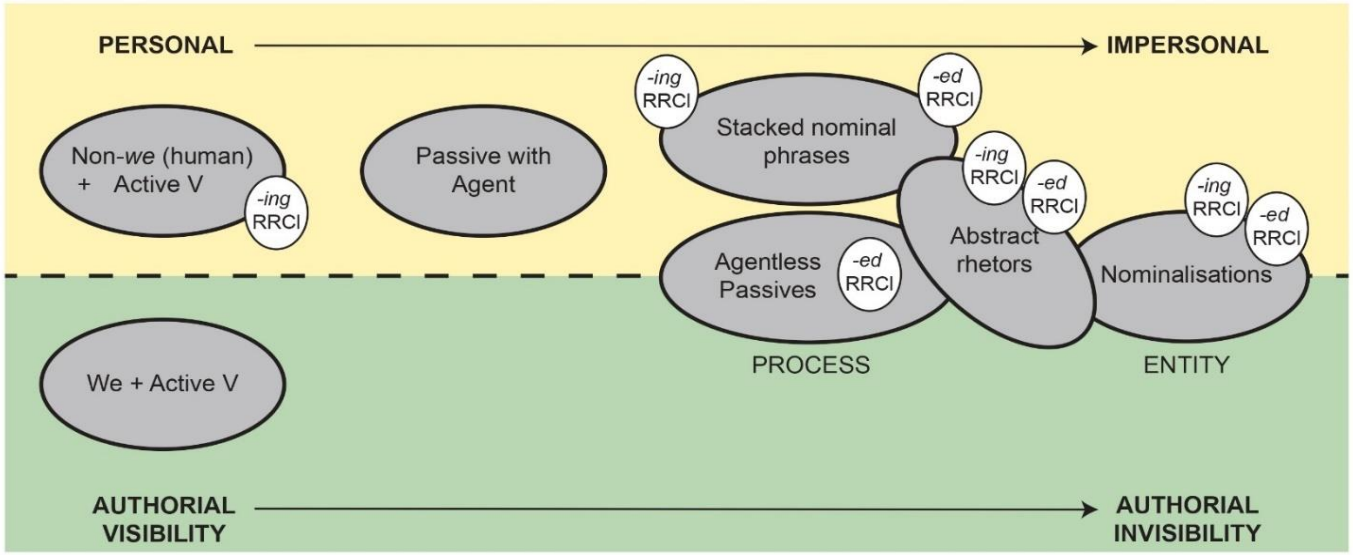
to evaluate this. **[Interplay between Personal active, Existential *there*, and Noun followed by a Past participial reduced relative clause; Discussion; JAMA2]**

(76) **The estimated mean hours per day of use in the supplemental-oxygen group (15.1±6.2 hours per day in the 24-hour group and 11.3±5.0 hours per day in the sleep–exercise group) (Fig. 2)** was similar to the **use observed** in the Nocturnal Oxygen Therapy Trial (17.7 hours per day in the continuous-oxygen group and 12.0 hours per day in the nocturnal-oxygen group). **[Interplay between Stacked nominal phrases and Noun followed by a Past participial reduced relative clause; Discussion; NEJM1]**

When placing these two structures in the continuum, it is not possible to locate them in a fixed and unique place because of their nature (see Figure 25). These structures are always related to other patterns dealt with in this study, either because they often interplay with them (like in the case of abstract rhetors, nominalisations, and stacked nominal phrases), or because they are reduced versions of their finite counterparts (like in the case of non-*we* active structures and, above all, agentless passives). They function as supplementary structures that operate at a lower level than the other aforementioned lexicogrammatical patterns.

In all cases, they add a sense of impersonalisation to the medical text. However, as already stated, not all of them relate to authorial visibility. As with most of the other patterns, only when the verb is connected to verbs of material procedure, like it is the case with abstract rhetors and nominalisations, they portray the invisibility of the researchers, who can easily be inferred through the context. The past participial reduced relative clause (or bare passive) is inserted within the agentless passive category, as the reduced clause may substitute the agentless passive if the information were presented in the form of an independent clause. However, both present and past participial reduced relative clauses are not inserted within the other patterns they are related to, but they are presented as being adjacent to them. This is to represent that non-finite reduced relative clauses do not substitute these patterns, but they perform the function of post-modifiers of their heads, and therefore they are embedded within them.

Figure 25. Continuum of lexicogrammatical structures (Present and Past Participial Reduced Relative Clauses) from more personal to more impersonal



Promotion of Agent	Partial demotion of Agent through backgrounding	Almost total demotion of Agent
Maximal negotiability	Minor degree of negotiability	Minimal negotiability

6.2.5. Existential *There*

An existential *there* pattern may be considered the most impersonal form of all the ones analysed. It consists of a structure with a grammatical expletive subject, *there*, followed by the singular or the plural form of the verb *be*, a form which must agree in number with the argument that follows. As for the raw frequency, it is worth noting that it is not very frequent in RCTs. It only appears 43 times across all the RCTs of all the journals analysed, which accounts for 1.30 instances per 1,000 words (see Table 84).

Table 84. Total number of tokens (raw frequency) of Existential *there* according to journal and section

Existential *there*	**NEJM 1**	**NEJM2**	**LNC1**	**LNC2**	**BMJ1**	**BMJ2**	**JAMA1**	**JAMA2**	**Total**
Abstract	0	0	0	1	0	0	1	0	2
Introduction	0	0	2	1	0	0	1	0	4
Methods	0	0	0	3	0	2	1	0	6
Results	3	1	1	3	1	0	3	3	15
Discussion	2	0	1	2	1	1	6	3	16
Total	5	1	4	10	2	3	12	6	43
TOTAL	6		14		5		18		43

The highest frequency revealed in all the RCTs is to be found in the JAMA journal, which accounts for 2.40 instances per 1,000 words. The one which represents the lowest is the BMJ, with 0.57 instances per 1,000. These structures do not tend to appear neither in the Abstracts (2 tokens, 0.65 instances per 1,000 words), nor in the Methods (6 tokens, 0.53 instances per 1,000 words) sections (see Tables 85 and 86).

Table 85. Use of Existential *there* in RCTs according to journal. Normalised results per 1,000 words

Journals	Existential *there*
NEJM	0.91
LNC	1.36
BMJ	0.57
JAMA	2.40
Total	1.30

Table 86. Use of Existential *there* in RCTs according to sections (including all journals). Normalised results per 1,000 words

Sections	Existential *there*
Abstract	0.65
Introduction	1.48
Methods	0.53
Results	1.93
Discussion	1.94
Total	1.30

In the following examples 77-82, the existential *there* pattern implies the existence of something, and this something is presented as an argument of the copular verb *be*. Here, although representing a deep sense of impersonality, authorial invisibility is somehow hidden. Apart from implying the existence of a particular situation (which is very clear in example 77), underneath the existential *there*, it is possible to find the medical consideration and opinion of the researchers towards a given context (examples 78-81). They position themselves towards what they are describing using this pattern, but they do so by portraying themselves invisible.

(77) Among these patients, **there were** 22 matched sets that each contained one STEC outbreak case and one or more comparison cases. **[Existential *there*; Results; NEJM2]**

(78) Once **there was** no concern of further active bleeding, the blood loss was estimated by weighing all the soiled materials (swabs, pads, disposable sheets) on a scales and subtracting the known weights of these materials. **[Existential *there*; Methods; BMJ2]**

(79) They were then transferred to the postnatal ward if **there were** no apparent complications. **[Existential *there*; Methods; BMJ2]**

(80) Third, we did not use uniform devices for oxygen delivery; it is possible that **there was** variability in the amount of oxygen delivered. **[Existential *there*; Discussion; NEJM1]**

(81) Furthermore, **there is** a risk of obstruction and strangulation of the bowel with perforation and possible mortality as a result. **[Existential *there*; Introduction; LNC1]**

(82) Thus, **there is** an unmet need for patients with myelofibrosis who progress while on treatment with ruxolitinib. **[Existential *there*; Introduction; LNC2]**

Indeed, there is an extension to such invisibility when researchers communicate relevant information connected to the results obtained in the medical study (examples 83-87). In this case, they can be considered as being invisible, because they express these results without claiming their responsibility. The results have been obtained empirically, and the existential *there* is a good resource to erase any trace of the visibility, and thus responsibility authors may hold for such results, as they merely transmit them.

(83) **There was** no heterogeneity of treatment effect on mortality according to qualifying acute coronary syndrome at baseline (P=0.73), with bivalirudin associated with fewer cases of all cause death (0.68, 0.46 to 1.00, P=0.048) and cardiovascular death (0.67, 0.45 to 1.00, P=0.048) compared with heparin in patients with ST segment elevation (…) **[Existential *there*; Results; BMJ1]**

(84) **There were no** significant differences in the primary and secondary efficacy endpoint results across subgroups for the different myelofibrosis subtypes (appendix p 3). **[Existential *there*; Results; LNC2**

(85) **There was** no difference between groups (mean difference of 1.0 unit; 95% CI, -0.3 to 2.3; P= 14, for improvement in the Berg Balance Test, which improved by 2.2

units (95% CI, 1.3-3.1) in the detetrabenazine group and by 1.3 units (95% CI, 0.4-2.2) in the placebo group. **[Existential *there*; Abstract; JAMA 1]**

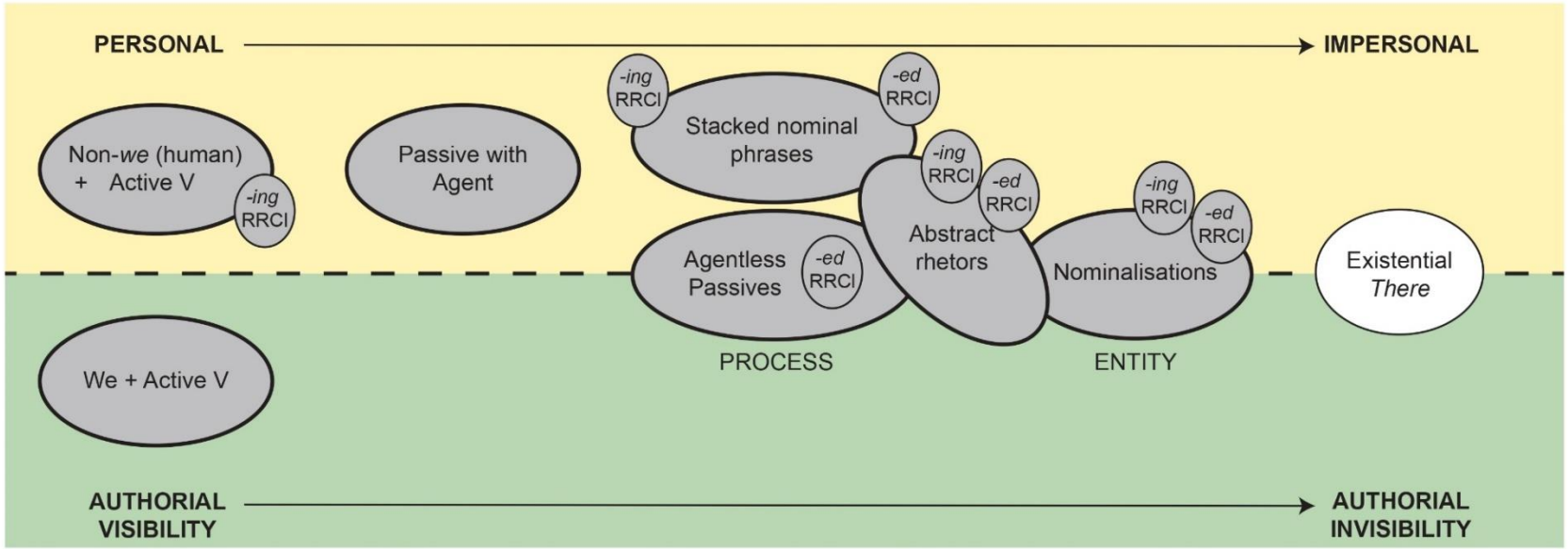
(86) Finally, because hospitalization was recorded from self-report every 4 monts, it is possible that we underestimated the number of hospitalizations; however, **there did not appear to be** systematic bias in follow-up between groups. **[Existential *there*; Discussion; NEJM1]**

(87) Thus, **there is** an unmet need for patients with myelofibrosis who progress while on treatment with ruxolitinib. **[Existential *there*; Introduction; LNC2]**

This is the main reason why this pattern is placed at the end of the continuum, as it may be considered the most depersonalising form out of all the patterns analysed in this study (see Figure 26). Authors are specialists who are linguistically transmitting what exists, on the one hand, and what has resulted from their studies, on the other, and this transmission is performed in a totally detached way. It is so detached and neutral that it is presented as a fact. Indeed, it is so credible and legitimate that readers may interpret that any investigator taking the same procedural steps would obtain the same results, which reinforces the aforementioned ideas of replicability, falsifiability, and universality.

Furthermore, existential *there* patterns are placed, as many other resources, as being in between two different dimensions, that of impersonality and that of authorial visibility. All instances of this pattern portray impersonality, as already said; however, not all the instances portray authorial visibility. When this pattern only accounts for the existence of a particular element, there is no connection to any kind of authorial manifestation. Yet, when this pattern expresses information empirically obtained through a set of experiential procedures, authorial presence seems to be hidden underneath such linguistic realisation.

Figure 26. Continuum of lexicogrammatical structures (Existential *there*)
from more personal to more impersonal



Promotion of Agent	Partial demotion of Agent through backgrounding	Almost total demotion of Agent	Total demotion of Agent
Maximal negotiability	Minor degree of negotiability	Minimal negotiability	No negotiability

6.3. Blurring the Line between Active and Passive Voice Structures: A Continuum Ranging from more Personal to more Impersonal Structures and their Connection to Authorial (In)visibility - Summing Up

Before summarising the results obtained in this close textual study, it is essential to first describe the different parameters used to locate the several linguistic choices recurrently used in RCTs within a continuum (see Figure 27). This cline is aimed at representing the fuzziness accounting for the impersonalisation strategies commonly found in medical articles, and at portraying how these strategies relate to the notion of authorial visibility. Indeed, the interrelatedness between both concepts (that of impersonality and that of authorial visibility) suggests that the dimension of impersonality is extensively and intrinsically deep and complex. In fact, it is much more complex than asserting that the use of the passive voice is what makes a medical text feel impersonal and detached. The passive voice is only one strategic resource out of many to reach this outcome.

All the lexicogrammatical patterns analysed are represented within a horizontal rectangle. Its left area stands for the notions of personality, on the top end, and authorial visibility, on the bottom end. On the opposite side, its right area corresponds to the notions of impersonality, on the top end, and that of authorial invisibility, on the bottom end. Bearing these four descriptive parameters in mind, it is possible to project two distinct axes grading from personal and authorially visible linguistic material to more impersonal and less authorially visible elements. With the addition of these four parameters and the existence of two different axes, two main descriptive areas are created.

Indeed, the yellow area, representing the gradation from more personal to more impersonal structures, is distinguished from the green area, which represents the gradation from more authorially visible to more authorially invisible strategies, through a discontinuous line. This dashed line symbolises the fuzziness between the two dimensions analysed. One can see that there are some structures which are clearly

positioned within one of the areas, meaning these linguistic resources only account for a description within the parameters used in that specific area. In contrast, others appear as belonging to both dimensions, and are located on the line dividing the two-coloured halves of the rectangle. This represents that those patterns may account for any of the two areas dealt with, which would essentially depend on the context.

This complex way to demonstrate the different layers of (im)personality and authorial (in)visibility inherent in the lexicogrammatical patterns analysed is somehow aimed at defying the oversimplified connection between active voice and passive voice structures as being mostly personal and impersonal, respectively. The shades of (im)personality found in these and other linguistic structures go beyond a dichotomised classification between personal and impersonal, and visible and invisible. The different linguistic choices made by authors account for a particular degree of (im)personality and authorial manifestation.

As it can be seen in the continuum, active structures are coloured blue, whereas passive ones are coloured orange. Following the alternation of blue and orange boxes positioned in the continuum, one can clearly see that the dichotomy active vs. passive is totally blurred, and the existence of many linguistic choices spread throughout the continuum makes it impossible to dichotomise the wide array of patterns portrayed as belonging to the two polarised ends of the continuum. The white boxes do not contain any finite verbal structure, and therefore cannot be represented as belonging to the active voice or to the passive voice. An exception to this would be the two types of reduced relative clauses, which are painted in a lighter colour (light blue when indirectly connected to the active voice and light orange when indirectly related to the passive voice).

Focusing on the different linguistic choices, it is worth noting that the most relevant personal structure is the *we* + active verb pattern. This is tightly linked to the notion of personality, and it overtly expresses authorial representation in the text. Then, it is possible to find another personal structure, the non-*we* personal active sentence, whose subject is animate but different from the researchers and writers of the paper. This

structure portrays some hints of personality to the text, but they are not related to authorial manifestations of any kind.

Next to the non-*we* personal active pattern, it is possible to locate the passive including the agent. In this case, the passive voice represents a clearer sense of impersonality if compared to the two personal structures already dealt with. In passives with an agent, importance is given to the foregrounded piece of information, while the role of the agent is backgrounded and partially demoted. In fact, although it is overtly expressed, the fact that it is backgrounded and demoted makes it be less personal than agentless passives. In the case of passives without an agent, the agent is almost totally demoted, but it can be recovered from the context, so such demotion is not complete. Authorial intervention is thus limited if compared to the personal active pattern. Agentless passives operate between the two domains. They are impersonal resources, but they overtly express authorial visibility when the past participle forms are linked to procedural activities carried out by the researchers themselves.

Nominalisations and stacked nominal phrases are represented as equally or even more impersonal than agentless passives. Nominalisations operate very similarly to agentless passives because they hide the authors' involvement in the research carried out, and this is somehow achieved by the objectification of verbal processes into nominal entities, which are perceived as factual, already established, and more solid than their verbal counterparts. Nominalisations are in between the two domains, that of impersonality and authorial visibility, and behave the same as stacked nominal phrases. They always show a major degree of invisibility but connect to authorial manifestation when the head noun relates to procedural activities carried out by the authors. Stacked nominal phrases solely add hints of impersonality to the text but do not relate to authorial contribution in the text.

Abstract rhetors are also represented in between the two dimensions. They represent a great extent of impersonality despite using the active voice in its finite verb form. When the nouns forming the subject of the abstract rhetor refer to experimental procedures carried out by the researchers (that is, when they are nominalisations inserted within the

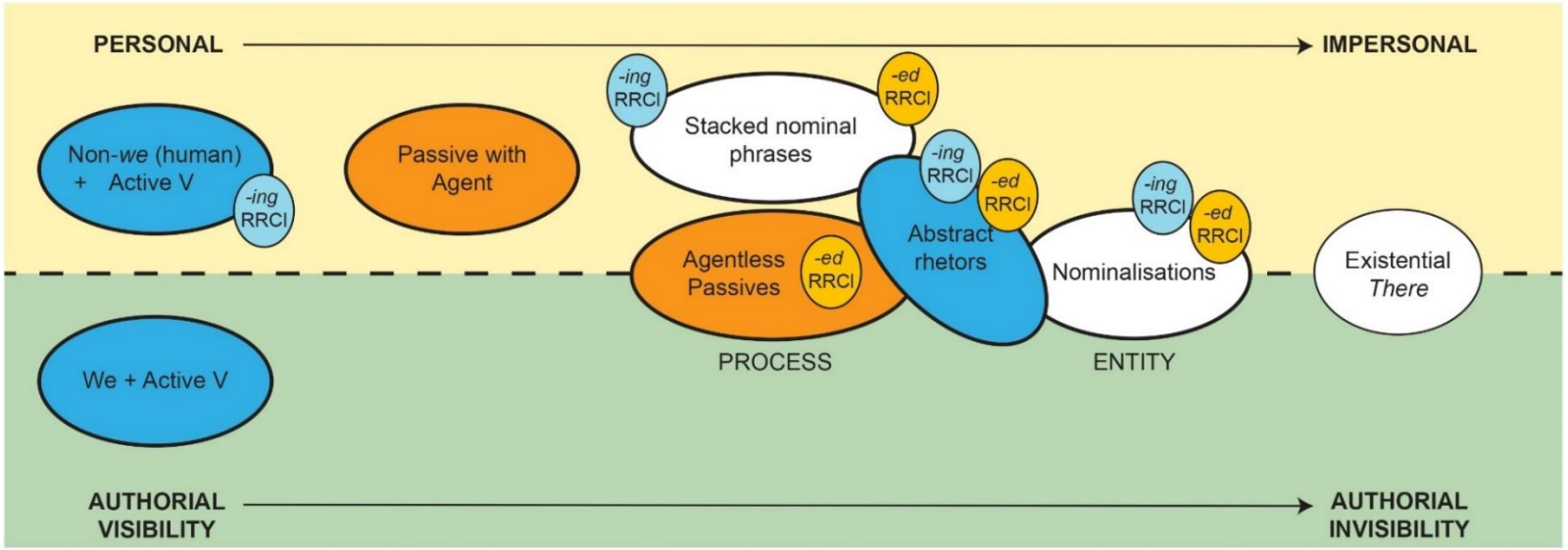
abstract rhetor), a small hint of authorial visibility is portrayed in the text. If these nouns do not refer to procedures (that is, when they are simple nouns or stacked nominal phrases), they only represent impersonality but have nothing to do with authorial manifestation.

The existential *there* pattern is the most impersonal resource if compared to the rest, and that is why it is placed at the most impersonal pole of the continuum. It is placed between the two dimensions (impersonality and invisibility). All the instances in which an existential *there* appears carry with it a sense of impersonalisation. When this pattern only accounts for the neutral existence of a particular element, there is no connection to authorial invisibility. However, when there is information which is empirically obtained through procedures, authorial invisibility is clearly portrayed.

Lastly, the last two other patterns, *-ing* and *-ed* reduced relative clauses, are dependent on others and because of this, they are represented as adjacent to the typical patterns they tend to accompany. They are supplementary structures that may be considered to operate at a lower level if compared to the rest of the structures. In any case, they show a hint of impersonality but not all of them relate to authorial visibility. Again, when the verb is connected to material procedures, they portray some researchers' invisibility.

In sum, many of these linguistic resources convey authorial manifestation when the verb is related to procedural processes, and this really connects the semantics, the lexis, the grammar, and the pragmatics of the language to the notion of impersonality and authorial visibility in the specialised and conventionalised register of medical writing.

Figure 27. Continuum of all the lexicogrammatical structures analysed from more personal to more impersonal



Promotion of Agent	Partial demotion of Agent through backgrounding	Almost total demotion of Agent	Total demotion of Agent
Maximal negotiability	Minor degree of negotiability	Minimal negotiability	No negotiability

Conclusions

The more you know,
the more you know you do not know.
Aristotle

The core aim of this thesis has been, since its original conception, to shed some light on the complexity associated with the notion of impersonality and its multi-dimensional representation in English scientific medical writing. The notion of impersonality can be considered a truly abstract (and pragmatic) concept which encompasses many distinct interpretations derived from a wide array of nuances that make it complex to operationalise and materialise. This thesis considers the existence of a fuzzy boundary between the notion of impersonality and that of authorial visibility, which are often seen as synonymic and interdependent. In fact, this dissertation sees them as two distinct but interrelated concepts.

As impersonality, at least in medical writing, has traditionally been related to the extensive use of the passive voice, taking grammatical voice as a starting point to operationalise the notion of impersonality seems well-justified. By operationalising the notion of impersonality through the choice of grammatical voice, a partial account of impersonality has been explored in the corpus-based study presented in Chapter 5. It is considered as a partial description because it has only considered active and passive verb forms, which may be categorised as one of the main means by which the notion of (im)personality is transmitted, yet not the only one.

Indeed, to take a more complete picture of impersonality, a close textual analysis has been designed in Chapter 6, to enable the exploration of this abstract notion in a restricted set of representative medical texts (RCTs). For the completion of the study, a different type of operationalisation has been considered. The notion of authorial (in)visibility has been foregrounded as one of the main distinctive parameters, as opposed to that of grammatical voice, which has been relegated to a second-tier status. In sum, impersonality has served as a super-ordinate concept, encompassing two distinct ordinates: grammatical voice, in the corpus-based study (despite the inclusion of some hints about authorial (in)visibility); and authorial (in)visibility, in the close textual study (despite specific reference to active and passive grammatical voice in verbal choices).

The main reason to shift the parameters from one study to the other relies on the adjustment one needs to make when exploring real instances in their textual context. As the transmission of knowledge in medical writing has shown to heavily rely on nominalisations and stacked nominal phrases, among many other resources, the notion of grammatical voice could not serve as a primary distinctive parameter and had to be shifted to authorial (in)visibility. Indeed, with the conjunction of these three parameters (grammatical voice, (im)personalisation, and authorial (in)visibility) a more representative approach could be made.

In this section, and to conclude this doctoral dissertation, I will refer to the Research Questions posed in the Introduction section to assess whether they have been answered throughout the various chapters included in this thesis, while including a summary of the main findings. Once the reasoning of the answers is succinctly interpreted, I will briefly highlight the major contributions that this dissertation, to my view, provide to the development of the notion of impersonality in medical register. A final section on future research directions is also provided.

I) Referring to the Research Questions Underlying this Dissertation

In this section, I will refer to the Research Questions that led to the completion of this dissertation.

1) Is a medical Research Article (in particular, a Randomised Controlled Trial) impersonal because of its extensive use of the passive voice?

As seen in the corpus-based study and in the close textual analysis, the passive voice is extensively used in medical writing, despite the efforts being made, particularly by the Plain Language Movement and some medical journals' houses of style, to counteract its omnipresence in an attempt to simplify the often so-called 'obscure' and 'verbose'

scientific language. In fact, to state that the pervasive use of the passive voice is what makes medical texts be perceived as impersonal may be considered an oversimplification. Obviously, passive instances contribute to the impersonalisation of medical research articles, but the existence of a wide range of verbal and verbless linguistic strategies (referred to in Research Question 2) which also contribute to the impersonal dimension of medical texts cannot be underestimated. These instances are seen to impersonalise medical writing as much as passive voice constructions do.

The impression that the passive voice is what makes medical writing impersonal may rely on the interpretation of verbs as being more hierarchically prominent than nouns and other different lexical categories. Medical language tends to lexically overload arguments with many linguistic resources, mainly aimed at providing descriptive information on the objects under study. Through the use of verb forms, these heavily loaded arguments are connected in specific ways. The passive plays a crucial role in the explicit connection made between these arguments, as it foregrounds information on the objects under study while backgrounding and demoting the presence of the agent. Apart from its high frequency of occurrence, this powerful function of enabling the ordering of clausal arguments to attain a specific focus in the message is what may make the passive voice be ranked in a superior hierarchical position within the whole set of linguistic choices available to transmit this sense of detachment and objectivity, inherent in medical writing. As said, however, a wide repertoire of strategies also serves this purpose.

2) **Does there exist any other lexicogrammatical resources which contribute to the (im)personalisation of medical discourse? If so, how do these patterns relate to (im)personalisation, grammatical voice, and authorial (in)visibility?**

There is a wide array of linguistic strategies that serve to either personalise or impersonalise medical writing. On the one hand, personal active structures (*we* + active verb) and non-*we* personal active structures are the ones which contribute the most to

the personalisation of texts. Both are active structures which include a human agent, performer of the action expressed by the verb in the active form; however, only personal active structures refer to authorial manifestation, as they promote authors to their full extent.

Another structure which seems to carry some kind of personalisation is the agentive prototypical passive voice. Despite being a passive structure and despite the partial demotion of the agent by means of backgrounding, its mere presence contributes to the recognition of agency within the process being described. The classification of this pattern as relatively personal opposes the conception of the passive as related to impersonality. Apart from this, although the agent is explicitly mentioned, this pattern never contributes to the representation of writers in the text.

From now on, all the structures that I will summarise will contribute to the impersonal characterisation of medical writing. It is true, however, that each of them will contribute to its own extent. Grading structures towards a more impersonal dimension, one can find agentless passives, stacked nominal phrases, impersonal active constructions or abstract rhetors, and nominalisations. These four linguistic lexicogrammatical constructions show almost total demotion of the agent, and they do so differently.

Agentless passives totally demote the agent, as it is not overtly expressed. However, it can be inferred from the context and no explicitness is required in these cases, particularly when describing procedural activities carried out by the researchers themselves. As the implicit agents are the writers themselves, authorial visibility is also implied. This pattern then would be half-way on the continuum as far as authorial manifestation is concerned.

Nominalisations and stacked nominal phrases do not really relate to active or passive voice but may be considered a bit more impersonal than agentless passives, insofar as they hide authors' involvement in the research carried out through the objectification of verbal process into nominal entities, allowing a more unbiased perception of what is being transmitted. When the noun head of the nominalisation is related to procedural

research activities carried out by the researchers themselves, partial authorial visibility can be identified. This is not the case of stacked nominal phrases, which do not relate to authorial manifestation.

Impersonal active structures or abstract rhetors are considered impersonal insofar as they foreground objects which are given human characteristics by means of metonymy. These impersonal subjects are followed by an active verb, so here, again, the notion of grammatical voice fails to relate personality and impersonality to active and passive voice, respectively. Abstract rhetors tend to interplay with nominalisations and stacked nominal phrases. If a nominalisation refers to procedural activities, authorial visibility is partially manifested; if a nominalisation refers to other different types of actions or if the syntactic function of the subject is realised by a stacked nominal phrase, authorial invisibility is clearly shown.

The non-finite *as*-passive can be considered as an impersonal form, and it can even be classified as more impersonal than the agentless passive, simply because of the more abstract perception due to its non-finiteness. It can be considered as an alternative to the agentless passive. Its nature as a guiding discourse indicator makes it be a good resource for writers to point at spatial and local information, together with the use of adverbs. Because of the types of verbs appearing in this pattern (essentially verbs of procedure and reporting verbs), authorial manifestation can be retrieved from the context if the structure is presented as agentless. If presented as agentive, reference to other agents different from the authors is provided.

The reporting passive structure includes the impersonal pronoun *it*, followed by a passive voice structure including a reporting verb, and followed by a *that*-clause. These reporting verbs contribute to the possible inference of agency, and thus, authorial visibility is presented as demoted.

The existential *there* structure is considered the most impersonal of all. Despite its unclear classification as far as voice is concerned, I will consider it as a special active type of structure, because of the presence of an expletive grammatical subject,

recognising the agreement in number the verb needs to show with the argument that follows. When an existential *there* neutrally describes the existence or absence of a given item, no authorial visibility is shown, whereas when depicting information objectively obtained through experimentation, authorial visibility is clearly manifested.

Past participial reduced relative clauses are non-finite subordinate clauses that can be embedded in any other types of clausal arguments. They operate at a lower level insofar as they have been seen to essentially depend on agentless passives, stacked nominal phrases, abstract rhetors, and nominalisations. If they were transformed into clausal constructions, they would contribute to an extreme oversimplification of medical writing, which would make the resulting language totally different from the existing one. These linguistic resources highly contribute to the impersonalisation of medical texts, and they rarely imply authorial visibility.

Present participial reduced relative clauses are also non-finite subordinate clauses which can be embedded in non-*we* personal active structures, abstract rhetors, and nominalisations. Their effect is that of slightly personalising the pattern they co-occur with, as they can be easily paraphrased into active constructions. Authorial visibility is never represented through present participial reduced relative clauses.

Apart from the choice and use of one linguistic strategy, what heavily contributes to the impersonalisation of medical research articles is the amalgam of structures which have been shown to interplay. The summative effect they exert over medical writing is ubiquitous and appears in all the sections of the research paper.

3) **Is there any correlation between the dichotomies *personal* vs. *impersonal* and *active voice* vs. *passive voice*? Is there any correlation between these two concepts and *authorial (in)visibility*?**

As seen in Research Question 2, in fact, there is no straightforward correlation between the dichotomies *personal* vs. *impersonal*, *active voice* vs. *passive voice*,

and *authorial visibility* and *authorial invisibility*. They are three very interrelated concepts, but each pattern, either active or passive, represents and entails a different degree of authorial representation and (im)personality by means of specific elements in the patterns, specific choices of lexical verbs, and specific functions within the overall textual configuration. As seen, an active voice sentence can be considered as more impersonal than a passive voice one, and a nominalisation can be considered as more personal than an existential *there*.

4) **Is the Cartesian coordinate system the best way to frame and categorise all possible lexicogrammatical resources to delve into the notions of (im)personality, grammatical voice, and authorial (in)visibility in medical writing, or is a continuum better?**

The Cartesian coordinate system has been the one used to conclude the corpus-based study, whereas the continuum has served to portray all the most frequent lexicogrammatical instances encountered in the close textual study. Even though they serve the same purpose, they are thought to be used in different ways.

The conception of the Cartesian coordinate system departs from the premise of connecting the notion of grammatical voice, which has been the original classification of structures, to that of (im)personality. By adding these two parameters, I considered that it was worth illustrating the spatial representation of the structures and the best way to do so was within a two-axial system, assigning the notion of Personality to the x-axis, because of its horizontally visual gradation, and the notion of Voice to the y-axis, presented vertically. In this sense, the x-axis serves as a gradual cline from impersonality (-x) to personality (x), whereas the y-axis serves as a cline from active (y) to passive voice (-y). By bearing these two main parameters in mind, different combinations were possible, and these combinations were spatially related to one of the four Quadrants in the plane.

As for the active structures, I Quadrant would include personal structures and II Quadrant would include impersonal ones. Passive structures would be included in III and IV Quadrants and would distribute passives according to their degree of (im)personality: to III Quadrant if impersonal, and to IV if personal. Therefore, I Quadrant will be represented by (x, y), II Quadrant by (-x +y), III Quadrant by (-x, -y), and IV Quadrant by (x, -y).

This system tries to scientifically categorise most of the verbal structures analysed in this dissertation. The spatial position of each of the constructions within the Quadrant plane has been individually assessed, so I take full responsibility for any errors in interpretation.

As this is a bi-dimensional system, the limit number of parameters is two. That is why authorial (in)visibility is left aside in this case. It would be possible to include it by highlighting structures that demonstrate visibility, contrasting them with those indicating invisibility, while considering authorial (in)visibility as a secondary-tier parameter. I will refer to this when dealing with future research directions.

As for the use of the continuum, grading structures according to their level of impersonality has been previously done. In this case, however, the continuum is also thought to have a bi-dimensional structure, which is presented as a new contribution to the study of impersonality. This bi-dimensional cline represents the parameters (im)personality and authorial (in)visibility. Both dimensions have been divided through a black discontinuous line to mark the interrelatedness of both concepts. Bearing in mind these parameters, all the structures have been positioned within the spatial representation of the continuum, according to their levels of (im)personality and authorial (in)visibility. If only showing one of these parameters, the linguistic structure has been placed within the one representing that notion. If showing both, the linguistic structure has been placed on the discontinuous line. Again, the spatial position of each of the structures within the continuum has been individually assessed, so I take full responsibility for any errors in interpretation.

The continuum is also a bi-dimensional system, so only two parameters can be represented. For this reason, the notion of grammatical voice has been disregarded as a main distinctive parameter and has been conceived as a secondary-tier parameter with the inclusion of a legend of colours, to represent either grammatical voice existing in the given patterns.

Both the Cartesian coordinate system and the continuum make use of impersonality as a super-ordinate element because it always appears as a primary distinctive parameter. The former serves to scientifically represent the interaction between the notions of (im)personality and grammatical voice, whereas the latter serves to relate the notions of (im)personality and authorial (in)visibility.

The Cartesian coordinate system may be considered as more ordered insofar as all the structures belong to one of the four Quadrants in the plane, that is why I would consider it a more empirical representation than the continuum. It is true, however, that the position each pattern occupies within the spatial representation of the Quadrant is at the analyst's discretion. One disadvantage of using the Cartesian coordinate system is that all lexicogrammatical patterns that do not exhibit grammatical voice, because they are non-verbal forms, like nominalisations or stacked nominal phrases, cannot be represented, and therefore, they need to be excluded from the graphic representation. As it has been observed that these linguistic forms are highly prevalent and contribute significantly to the impersonalisation of medical discourse, the representation on a continuum appears to be a better option for the inclusion of all the patterns described in this doctoral thesis, demoting the importance of grammatical voice as a main distinctive criterion to delve into the impersonal dimension of medical writing.

While the ordering of elements in the continuum is also based on the linguist's analysis criterion, and therefore, it represents an approximate interpretation of reality, it is true that the continuum offers a clearer depiction of the gradual classification of various linguistic choices by visually arranging them from left to right. In contrast, the Cartesian coordinate system presents this gradation counterclockwise, aligning with the position of the Quadrants and the combination of their positive and/or negative polarity. While

this approach provides a more empirical and rationalised representation, it also introduces a certain degree of intricacy in mental processing due to its complex spatial configuration.

For all these reasons, Chapter 5 explicitly makes use of the Cartesian coordinate system, whereas Chapter 6, which includes highly recurrent non-verbal forms in medical discourse, opts to make use of the continuum as the final representation of the analysis of all the linguistic forms conducted in this thesis.

II) Major Contributions

This thesis has endeavoured to tackle the abstraction of the notion of impersonality through two different study models, one from a corpus-based approach and the other one from a text analysis-based approach. The parameters applied have always considered the notion of (im)personality and the concepts of grammatical voice and authorial (in)visibility. In this respect, it has been found that, despite being highly related, their connection is not actually straightforward.

The literature shows examples of studies which have contributed to the analysis of impersonality in academic writing; however, these studies usually include partial representations of linguistic choices, limiting thereby the presence or absence of impersonality to a very restricted set of items and not elucidating the overall picture of such a notion.

It is true that the pervasiveness of certain elements contributes to the (im)personal perception of a given textual body; however, the more structures one analyses as far as these three notions are concerned, the more general picture one can take. For example, non-finite forms tend to be disregarded when dealing with impersonality and its relationship to the passive voice, despite the high frequency of these non-finite forms and their clear connection as alternatives to finite passives. The same occurs with the relationship shown between nominalisations and impersonal active structures, which have demonstrated to show a high level of interconnection. The inclusion in this study

of all these lexicogrammatical items has provided, to my judgement, a more realistic representation if compared to partial analyses which shed light on the use of a very limited set of particular linguistic units.

This thesis emphasises the interplay of all these lexicogrammatical patterns as a means to characterise a text as being impersonal. It is not just through particular choices that the text becomes more impersonal but through the amalgamation of structures and their interrelatedness.

Another important aspect worth noting is the inclusion of two distinct visual representations: the Cartesian coordinate system to represent the notions of grammatical voice and (im)personality, and the creation of a continuum dealing with the notions of (im)personality and authorial (in)visibility. Continuums tend to only represent one dimension because only one distinctive parameter is used. The fact of enlarging the continuum to acquire the two aforementioned dimensions is worth considering.

III) Future Research Directions

Corpus-based analysis has proved to be a very useful tool to analyse impersonality by means of active and passive voice structures. However, it has posed some problems when retrieving some of the specific structures searched for due to the impossibility of recognising specific particularities typically related to the abstract notion of (im)personality, such as the retrieval of agents. Tagging the corpus semantically may help in the automatic detection of certain specificities in the text and this will allow a more refined material to be analysed by the linguist. A more pedagogical approach to the use of specific syntaxes would also allow the active participation of other linguists onto the analysis of complex colligational structures.

Another interesting research line would be to expand the continuum by means of the inclusion of other lexicogrammatical patterns which may not be very frequent, but which may contribute to the impersonal dimension of the text. These other linguistic

realisations can shed light on the relationship between the notion of (im)personality, authorial (in)visibility and, whenever possible, grammatical voice.

Indeed, the pedagogical application of this continuum is something that has not been considered in this thesis, but how the visual representation of all these forms may help writers decide how to transmit their messages is something yet unexplored. Together with general criteria of language use for the writing of articles, this continuum can serve as a primary source to the development and enhancement of the writing skills of authors who need to publish a scientific medical paper. Exploring the pedagogical applications underlying the notions dealt with in this thesis may contribute to the design of pedagogical and teaching material as an attempt to facilitate linguistic transmission among members of the same community of practice who may not share the same linguistic background or who may find it arduous to express linguistic information in the way expected by their discourse community.

Another research line would consist of comparing if the results obtained in this dissertation match the reality of the linguistic realisations used by other scientific disciplines. Establishing commonalities and main differences between disciplines can enable the classification of primary forms used in all disciplines, and secondary forms, restricted to more specific types of disciplines.

To conclude, another research line worth highlighting would be the development of a Cartesian coordinate system for a three-dimensional space, including a z-axis, which would represent the notion of authorial (in)visibility. This would allow a three-dimensional representation of the notion of impersonality related to its two ordinates, which are grammatical voice and authorial visibility.

The 3-D Cartesian coordinate system would consist of an ordered triplet of lines which intersect at a common point, the origin (O). Each pair of axes would define a coordinate hyperplane, which would divide the space not into Quadrants, but into Octants.

The Octants are formed of three different axes, namely, the x-axis (Personality; the abscissa), the y-axis (Voice; the ordinate), and the z-axis (Authorial Visibility; the applicate), as shown in Figure 28. With these different axes, three coordinate hyperplanes would be created: the XY-plane (Personality and Voice, like the plane shown in Chapter 5), the YZ-plane (Voice and Authorial Visibility), and the XZ-plane (Personality and Authorial Visibility), as shown in Figure 29.

Figure 28. Three-dimensional Cartesian coordinate system

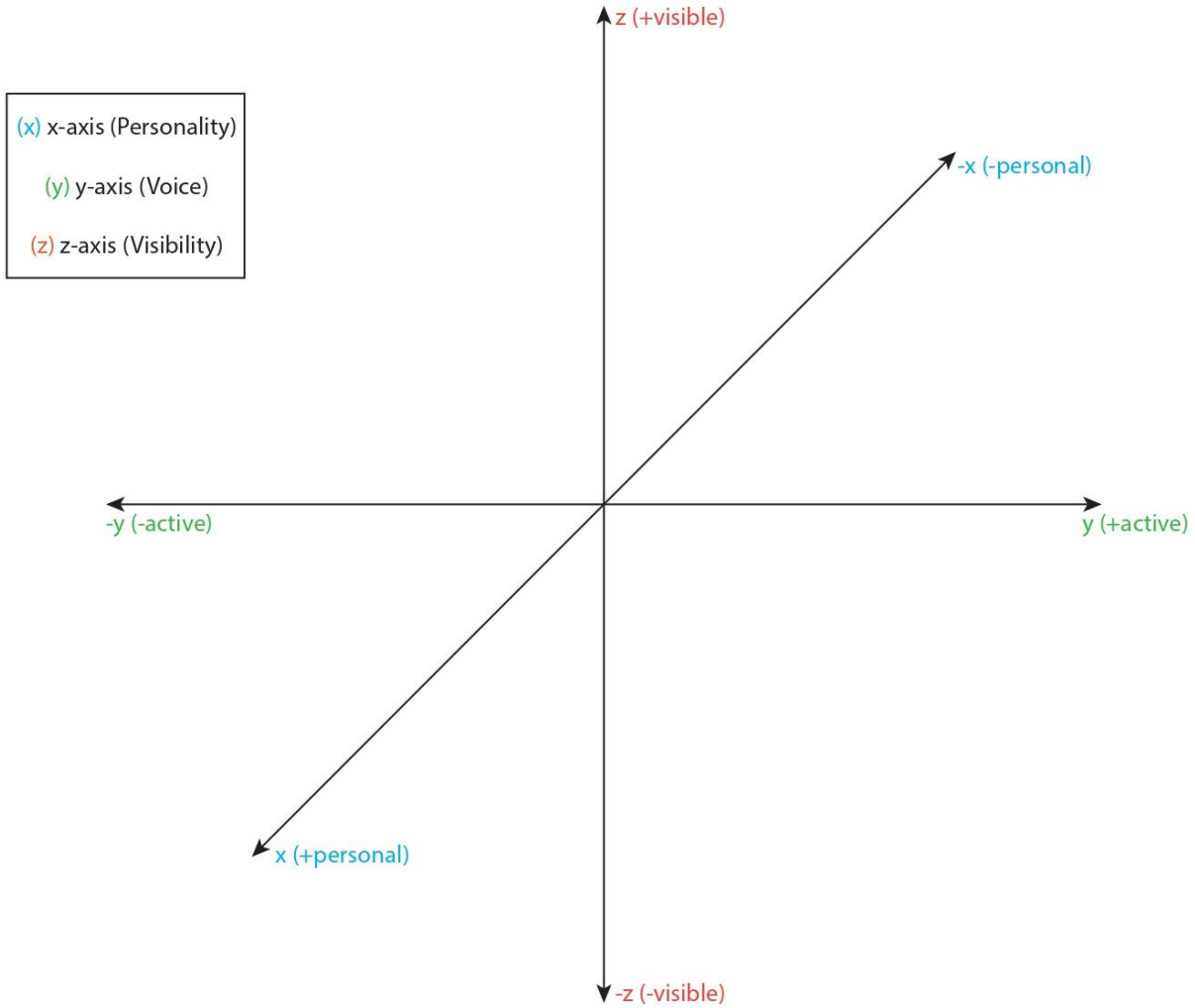
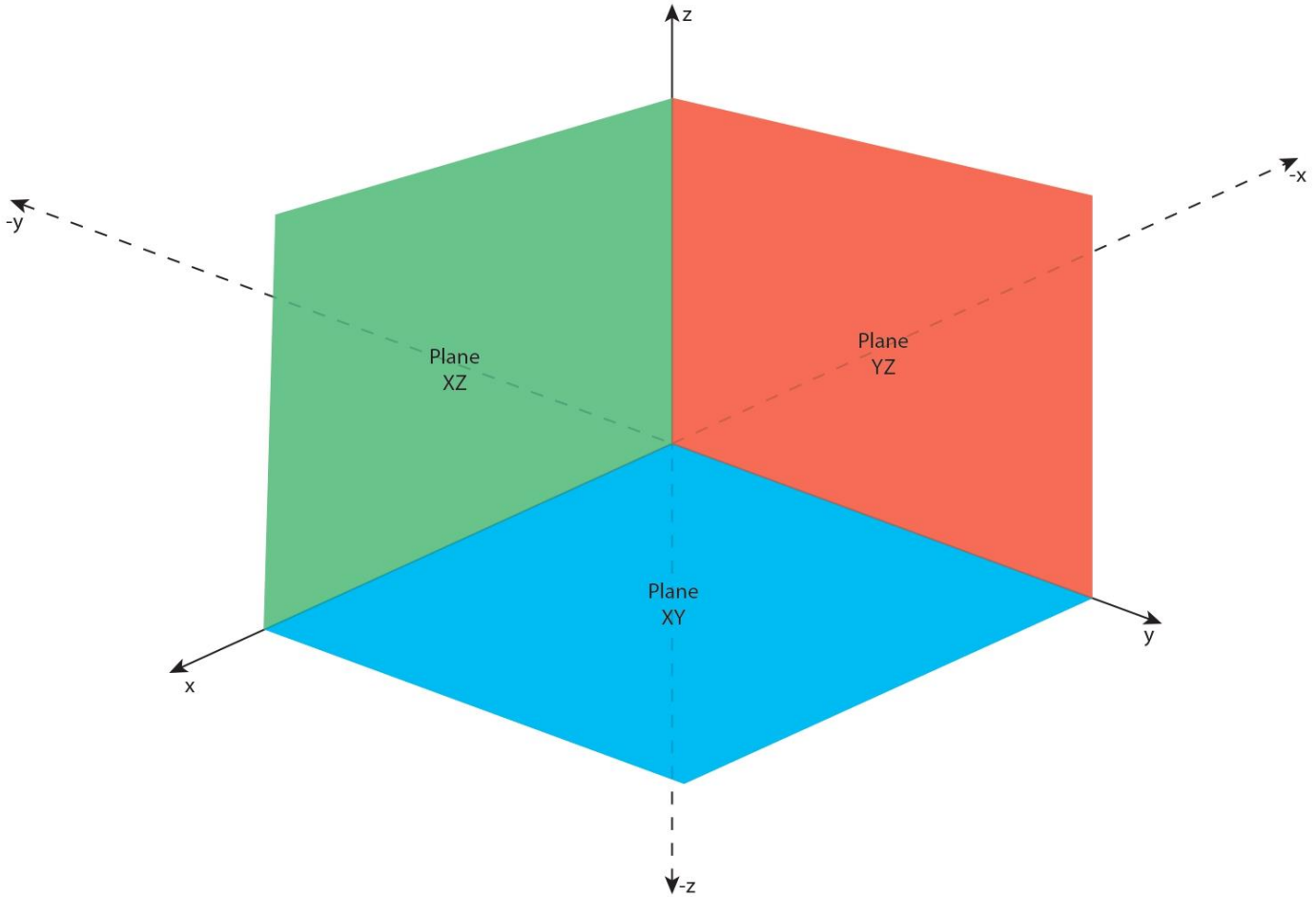


Figure 29. Representation of the Coordinate Hyperplanes [Plane XZ Personality and Authorial Visibility, Plane YZ (Voice and Authorial Visibility), and Plane XY (Personality and Voice)]



Considering these three different axes, the resulting Octants would be as follows:

Table 87. Octants of the Cartesian coordinate system, and their corresponding axes described through literal and used parameters – [(x) corresponds to Personality; (y) corresponds to Voice; (z) corresponds to Authorial Visibility]

Octants	Literal parameters	Used Parameters
(+x, +y, +z)	(+personal, +active, +visible)	(personal, active, visible)
(+x, -y, -z)	(+personal, -active, -visible)	(personal, passive, invisible)
(-x, +y, +z)	(-personal, +active, +visible)	(impersonal, active, visible)
(-x, +y, -z)	(-personal, +active, -visible)	(impersonal, active, invisible)
(+x, -y, +z)	(+personal, -active, +visible)	(personal, passive, visible)
(-x, -y, +z)	(-personal, -active, +visible)	(impersonal, passive, visible)
(+x, +y, -z)	(+personal, +active, -visible)	(personal, active, invisible)
(-x, -y, -z)	(-personal, -active, -visible)	(impersonal, passive, invisible)

In today's technological world, where artificial intelligence and supercomputing are prevalent, this project would assist in analysing the abstract notion of impersonality through a three-dimensional medium. Developing this conceptualisation further by attempting to physically represent impersonality through the assignment of linguistic items to each of the Octants (also assessing the inclusion of those showing no Voice) may provide insights into how such a notion could be better understood by and taught to the medical scientific community.

References

Ädel, A. (2014). “What I want you to remember is...” Audience orientation in monologic academic discourse. In L. Brems, L. Ghesquière, & F. Van de Velde (Eds.), *Intersubjectivity and Intersubjectification in Grammar and Discourse* (Vol. 65, pp. 101–127).

Aissen, J. (1999). Markedness and Subject Choice in Optimality Theory. *Natural Language & Linguistic Theory*, *17*(4), 673–711. https://doi.org/10.1023/A:1006335629372

Aston, G., Bernardini, S., & Stewart, D. (2004). *Corpora and language learners.* (G. Aston, S. Bernardini, & D. Stewart, Eds.). John Benjamins Publishing Company.

Atkins, S. B. T., & Rundell, M. (2008). *The Oxford guide to practical lexicography.* Oxford University Press.

Baker, M. (1993). Corpus linguistics and translation studies: implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: in honour of John Sinclair.* (pp. 233–352). John Benjamins Publishing Company.

Banks, D. (1994). *Writ in Water, Aspects of the Scientific Journal Article*. ERLA, Université de Bretagne Occidentale.

Banks, D. (1996). The passive and metaphor in scientific writing. *Cuadernos de Filología Inglesa*, *5*(2), 13–22.

Banks, D. (2005). On the historical origins of nominalized process in scientific text. *English for Specific Purposes*, *24*(3), 347–357. https://doi.org/10.1016/j.esp.2004.08.002

Banks, D. (2008). *The development of scientific writing: Linguistic features and historical context.* Equinox.

Baratta, A. M. (2009). Revealing stance through passive voice. *Journal of Pragmatics*, *41*(7), 1406–1421. https://doi.org/10.1016/j.pragma.2008.09.010

Barlow, M. (2000). Parallel texts and language teaching. In S. Botley, T. McEnery, & A. Wilson (Eds.), *Multilingual Corpora in Teaching and Research.* (pp. 106–115). Rodopi.

Barnbrook, G. (1996). *Language and computers.* Edinburgh University Press.

Bartolic, L. (1978). Nominal Compounds in Technical English. In M. T. Trimble, L. Trimble, & K. Drobnic (Eds.), *English for Specific Purposes: Science and Technology*. English Language Institute: Oregon State University.

Basilico, D. (1997). The topic is 'there.' *Studia Linguistica*, *51*(3), 278–316. https://doi.org/10.1111/1467-9582.00023

Bathia, V. K. (1993). *Analysing genre: Language use in professional settings*. Longman.

Bazerman, C. (1988). *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science.* University of Wisconsin Press.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, *8*(4), 243–257.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* John Benjamins Publishing Company.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*, 371–405.

Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics*, *15*(2), 223–250. https://doi.org/10.1017/S1360674311000025

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Pearson Education Ltd.

Biber, D., & Reppen, R. (2015). *The Cambridge handbook of English corpus linguistics.* Cambridge University Press.

Biddle, L., Donovan, J., Owen-Smith, A., Potokar, J., Longson, D., Hawton, K., Kapur, N., & Gunnell, D. (2010). Factors influencing the decision to use hanging as a method of suicide: Qualitative study. *British Journal of Psychiatry*, *197*(4), 320–325. https://doi.org/10.1192/bjp.bp.109.076349

Bizzell, P. (1992). *Academic discourse and critical consciousness.* University of Pittsburgh Press.

Black, M. (1955). Metaphor. *Proceedings of the Aristotelian Society. New Series.*, 273–294. https://www-jstor-org.sire.ub.edu/stable/4544549

Booth, V. (1993). *Communicating in Science. Writing a Scientific Paper and Speaking at Scientific Meetings.* (2nd edn.). Cambridge University Press.

Bordet, G. (2013a). The rhetorical role of "collocational chains" and their implications in the building of scientific discourse strategies. *European Journal of English Studies, 17*(3), 235–250.

Boskovic, Z. (2014). Now I'm a Phase, Now I'm Not a Phase: On the Variability of Phases with Extraction and Ellipsis [Article]. *Linguistic Inquiry, 45*(1), 27–89. https://doi.org/10.1162/LING_a_00148

Botley, S., McEnery, T., & Wilson, A. (2000). *Multilinguial corpora in teaching and research.* (S. Botley, T. McEnery, & A. Wilson, Eds.). Rodopi.

Brants, S., Dippper, S., Hansen, S., Lezius, W., & Smith, G. (2002). The TIGER treebank. In *Proceedings of TLT-02.*

Breeze, R. (2010). They say, we do: Writers' strategic positioning in the discourses of political communication research. In R. Lorés-Sanz, P. Mur-Dueñas, & E. Lafuente-Millán (Eds.), *Constructing Interpersonality: Multiple Perspectives on Written Academic Genres.* (pp. 163–179). Cambridge Scholars Publishing.

Breivega, K. R., Dahl, T., & Fløttum, K. (2002a). Traces of self and others in research articles. A comparative pilot study of English, French and Norwegian research articles in medicine, economics and linguistics. *International Journal of Applied Linguistics, 12*(2), 218–239. https://doi.org/10.1111/1473-4192.00032

Breivega, K. R., Dahl, T., & Fløttum, K. (2002b). Traces of self and others in research articles. A comparative pilot study of English, French and Norwegian research articles in medicine, economics and linguistics. *International Journal of Applied Linguistics, 12*(2), 218–239. https://doi.org/10.1111/1473-4192.00032

Brno, O. D. (2005). *Authorial presence in academic discourse: functions of author-reference pronouns*.

Bruce, I. (2008). *Academic writing and genre: A systematic analysis*. Continuum.

Brutt-Griffler, J. (2002). *World English: A Study of its Development.* (C. Baker & N. H. Hornberger, Eds.). Bilingual Education and Bilingualism: 34.

Butler, J. (2006). *Gender trouble feminism and the subversion of identity* [Book]. Routledge.

Canale, M., & Swain, M. (1981). A Theoretical Framework for Communicative Competence. In A. Palmer, P. Groot, & G. Trosper (Eds.), *The construct validation of test of communicative competence.* (pp. 31–36).

Carrió-Pastor, M. L. (2008). English Complex Noun Phrase Interpretation. *Revista Española de Lingüística Aplicada*, *21*(June), 27–44.

Castelló, M., Iñesta, A., Pardo, M., Liesa, E., & Martínez-Fernández, R. (2012). Tutoring the end-of-studies dissertation: helping psychology students find their academic voice when revising academic texts [Article]. *Higher Education*, *63*(1), 97–115. https://doi.org/10.1007/s10734-011-9428-9

Celce-Murcia, M. (1991). Discourse analysis and grammar instruction. *Annual Review of Applied Linguistics1*, *11*, 459–480.

Chaopricha, Sangrawee. (1997). *Coauthoring as learning and enculturation: a study of writing in biochemistry*.

Charles, M. (1999). Aspects of the author's voice in three academic research articles. In P. Thompson (Ed.), *Issues in EAP Writing Research and Instructions.* (pp. 71–83). CALS Publication.

Charles, M. (2004). *The construction of stance: a corpus-based investigation of two contrasting disciplines.* University of Birmingham (unpublished PhD).

Cherry, R. D. (1988). Ethos Versus Persona: Self-Representation in Written Discourse. *Written Communication*, *5*(3), 251–276. https://doi.org/10.1177/0741088388005003001

Chomsky, N. (1957). *Syntactic Structures*. De Gruyter Mouton. https://doi.org/doi:10.1515/9783112316009

Clark, R., & Ivanic, R. (1997). *The Politics of Writing* (1st Edition). Routledge.

Connor, U. (2003). Changing currents in contrastive rhetoric: Implications for teaching and research. In B. Kroll (Ed.), *Exploring the Dynamics of Second Language Writing* (pp. 218–241). Cambridge University Press. https://doi.org/DOI:10.1017/CBO9781139524810.015

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *Tesol Quarterly*, *34*(3), 548–560. https://doi.org/10.2307/3587743

Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In J. McH. Sinclair (Ed.), *How to Use Corpora in Language Teaching 2* (pp. 67–85). John Benjamins Publishing Company.

Cooper, M. M. (1989). Why are we talking about Discourse Communities? Or Foundationalism rears its ugly head once more. In *Writing as Social Action* (pp. 202–220). Boynton-Cook.

Coxhead, A. (1998). An academic word list. In *English Language Institute Occasional Publication, number 18*. Victoria University of Wellington.

Coxhead, A. (2000). A new academic word list. *Tesol Quarterly*, *34*(2), 213–238.

Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, *XII*, 65–78.

Crompton, P. (1997). Hedging in academic writing: Some theoretical problems. *English for Specific Purposes*, *16*(4), 271–287. https://doi.org/10.1016/S0889-4906(97)00007-0

Crystal, D. (2003a). *English as a Global Language* (2nd edn.). Cambridge University Press.

Cutting, J. (2000). *Analysing the Language of Discourse Communities.* Elsevier Science.

Dalen, J. E., & Devries, S. (2014). Diets to prevent coronary heart disease 1957-2013: What have we learned? *American Journal of Medicine*, *127*(5), 364–369. https://doi.org/10.1016/j.amjmed.2013.12.014

David Bartholomae. (1995). Writing with Teachers: A Conversation with Peter Elbow. *National Council of Teachers of English*, *46*(1), 62–71.

Davis, M., Davis, K. J., & Dunagan, M. (2012). *Scientific Papers and Presentations.* (3rd edn.). Academic Press (Elsevier).

Day, R. A. (1998). *How to Write and Publish a Scientific Paper* (5th ed.). The Oryx Press.

Day, R. A., & Gastel, B. (2011). *How to Write and Publish a Scientific Paper*. (7th edn.). Greenwood.

Devitt, A. J. (2004). *Writing Genres*. Southern Illinois University Press, Project MUSE.

Dickey, R. J. (2004). Content (adj) or Content (n) With Your English Classes? In *Education International* (Vol. 1, Issue 3). http://www.content-english.org

Ding, D. (2002). The Passive Voice and Social Values in Science. *Journal of Technical Writing and Communication*, *32*, 137–154. https://doi.org/10.2190/EFMR-BJF3-CE41-84KK

Dixon, A., Clarkin, C., Barrowman, N., Correll, R., Osmond, M. H., & Plint, A. C. (2014). Reduction of radial-head subluxation in children by triage nurses in the emergency department: A cluster-randomized controlled trial. *CMAJ*, *186*(9), 317–323. https://doi.org/10.1503/cmaj.131101

Dontcheva-Navrátilová, O. (2013). Authorial Presence in Academic Discourse: Functions of Author-Reference Pronouns. *Linguistica Pragensia*, *1*, 9–30.

Dressen-Hammouda, D. (2013). The role of indexicality in the emergence of disciplinary writing expertise. In J. Bamford (Ed.), *Space, place and the discursive construction of identity*. Peter Lang.

Dressen-Hammouda, D. (2014). Measuring the voice of disciplinarity in scientific writing: A longitudinal exploration of experienced writers in geology. *English for Specific Purposes*, *34*(1), 14–25. https://doi.org/10.1016/j.esp.2013.10.001

Du Bois, B. Lou. (1981). The construction of noun phrases in biomedical journal articles. In J. Hoedt (Ed.), *Proceedings of the 3rd European Symposium on LSP: Pragmatics and LSP*. (pp. 46–67). The LSP Centre.

Duboviciene, T., & Skorupa, P. (2017). Passive Voice as a Means of Impersonal Presentation of Facts in English Quality Press. *Man and the Word / Foreign Languages*, *19*(3), 38–47. https://doi.org/http://dx.doi.org/10.15823/zz.2017.24

Dzuganová, B. (2013). English medical terminology – different ways of forming medical terms. *JAHR European Journal of Bioethics*, *4*(7), 55–69.

Eggins, S. (1994). *An Introduction to Systemic Functional Linguistics*. Pinter.

Eggins, S. (2004). *An Introduction to Systemic Functional Linguistics.* (2nd ed.). Continuum.

Elbow, P. (1994). What do we mean when we talk about voice in texts? In K. Blake Yancey (Ed.), *Voices on Voice: Perspectives, Definitions, Inquiry.* (pp. 1–35). National Council of Teachers of English.

Elbow, P. (1995). Being a Writer vs. Being an Academic: A Conflict in Goals [Article]. *College Composition and Communication*, *46*(1), 72–83. https://doi.org/10.2307/358871

Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: psycholinguistics, corpus linguistics and TESOL. *Tesol Quarterly*, *42*(3), 375–396.

Ellis, R. (1998). Teaching and research: Options in grammar teaching. *Tesol Quarterly*, *32*, 39–60.

Elsness, J. (2014). Clausal modifiers in noun phrases: A comparison of English and Norwegian based on the Oslo Multilingual Corpus. *Oslo Studies in Language*, *6*(1). https://doi.org/10.5617/osla.718

Espinoza, A. M. (1997). Contrastive analysis of the Spanish and English passive voice in scientific prose. *English for Specific Purposes*, *16*(3), 229–243. https://doi.org/10.1016/S0889-4906(96)00034-8

Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook* (pp. 1212–1248). Mouton de Gruyter. https://doi.org/10.1515/9783110213881.2.1212

Fairclough, N. (1993). *Discourse and social change* [Book]. Polity Press.

Fairclough, N. (1995). *Critical discourse analysis: the critical study of language* [Book]. Longman.

Ferguson, G., Pérez-Llantada, C., & Plo, R. (2011). English as an international language of scientific publication: a study of attitudes. *World Englishes*, *30*(1), 41–59.

Firth, A. (1996). The discursiveness accomplishment of "normality": On conversation analysis and "lingua franca." *Journal of Pragmatics*, *26*, 237–259.

Fløttum, K. (2003). Bibliographical References and Polyphony in Research Articles. In K. Fløttum & F. Rastier (Eds.), *Academic discourse. Multidisciplinary approaches* (pp. 97–119). Novus Press.

Fløttum, K. (2009). *Academic voices in the research article*. 109–122. https://www.jbe-platform.com/content/books/9789027288738-pbns.193.07flo

Fløttum, K., Dahl, T., & Kinn, T. (2006a). *Academic Voices: Across languages and disciplines*. https://doi.org/10.1075/pbns.148

Fløttum, K., Dahl, T., & Kinn, T. (2006b). *Academic Voices: Across languages and disciplines*. https://doi.org/10.1075/pbns.148

Flowerdew, J. (2014). Introduction: Approaches to the Analysis of Academic Discourse in English. In J. Flowerdew (Ed.), *Academic Discourse* (2nd Editio). Routledge. Taylor & Francis Group.

Flowerdew, L. (2004). *The argument for using English specialized corpora to understand academic and professional language* (pp. 11–33). https://doi.org/10.1075/scl.16.02flo

Foucault, M. (1972). *The Archaeology of knowledge and The dicourse on language* [Book]. Pantheon books.

Francis, N. W., & Kucera, H. (1964). *The Standard Corpus of Present-Day English Edited American English (the Brown Corpus)*. www.helsinki.fi/varieng/CoRD/corpora/BROWN

Freed, R. C., & Broadhead, G. J. (1987). Discourse Communities, Sacred Texts, and Institutional Norms. *College Composition and Communication*, *38*(2), 154–165. https://www.jstor.org/stable/357716

Fromkin, V., & Rodman, R. (1993). *An Introduction to Language.* (5th ed.). Harcourt Brace Jovanovich.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning. A Journal of Research in Language Studies*, *67*(S1), 155–179. https://doi.org/10.1111/lang.12225

Ganavadiya, R., Chandra Shekar, B. R., Goel, P., Hongal, S. G., Jain, M., & Gupta, R. (2014). Comparison of anti-plaque efficacy between a low and high cost dentifrice: A short term randomized double-blind trial. *European Journal of Dentistry*, *8*(3), 381–388. https://doi.org/10.4103/1305-7456.137652

Gardner, D., & Davies, M. (2016). A response to "To what extent is the Academic Vocabulary List relevant to university student writing?" *English for Specific Purposes*, *43*, 62–68. https://doi.org/10.1016/j.esp.2016.03.001

Gardner, D. E. E., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Geertz, C. (1983). *Local knowledge: further essays in interpretive anthropology*. BasicBooks.

Gellerstam, M. (1996). Translations as a source for cross-linguistic studies. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in Contrast.* (pp. 53–62). Lund University Press.

Giere, R. N. (1991). Explaining Science: A Cognitive Approach. *Journal of Philosophy*, *88*(3), 163–167. https://doi.org/10.2307/2026988

Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, *51*, 1–17. https://doi.org/10.1016/j.esp.2018.02.002

Givón, T. (1994). *Voice and Inversion* (T. Givón, Ed.). John Benjamins Publishing Company.

Givón, T. (2017). *On Understanding Grammar Revised Edition.* Amsterdam: J. Benjamins.

Glasman-Deal, H. (2009). *Science Research Writing for Non-Native Speakers of English.* Imperial College Press.

Gnutzmann, C. (1999). *Teaching and Learning English as a Global Language.* (C. Gnutzmann, Ed.). Stauffenburg.

Goffman, E. (1981). Footings. In E. Goffman (Ed.), *Forms of talk.* University of Pennsylvania Press.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure.* The University of Chicago Press.

Goldberg, A. (2006). *Constructions at Work.* Oxford University Press.

Gosden, H. (1992). Research writing and NNSs: From the editors. *Journal of Second Language Writing*, *1*(2), 123–139. https://doi.org/10.1016/1060-3743(92)90012-E

Gosden, H. (1993). Discourse Functions of Subject in Scientific Research Articles. *Applied Linguistics*, *14*(1), 56–75. https://doi.org/10.1093/applin/14.1.56

Granger, S., Hung, J., & Peych-Tyson, S. (2002). *Computer learner corpora, second language acquisition and foreign language teaching.* John Benjamins Publishing Company.

Granger, S., & Paquot, M. (2012). *Electronic lexicography.* Oxford University Press.

Gries, S. Th. (2008). Phraseology and linguistic theory. A brief survey. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective.* (pp. 3–25). John Benjamins Publishing Company.

Gries, S. Th. (2009a). *Quantitative corpus linguistics with R: A practical introduction.* Routledge.

Gries, S. Th. (2009b). *Statistics for linguists with R: A practical introduction.* Mouton de Gruyter.

Gries, S. Th. (2013). 50-something years of work on collocations: What is or should be next...? *International Journal of Corpus Linguistics*, *18*, 137–166. https://doi.org/10.1075/ijcl.13.4.02gri

Gries, S. Th., & Stefanowitsch, A. (2004). Extending collostructional analysis: a corpus-based perspective on “alternations”. *International Journal of Corpus Linguistics*, *9*(1), 97–129.

Grisham, R. (1986). *Computational linguistics. An introduction.* Cambridge University Press.

Gross, A. G., & Chesley, P. (2012). Hedging, Stance, and Voice in Medical Research Articles. In K. Hyland & C. S. Guinda (Eds.), *Stance and Voice in Written Academic Genres* (pp. 85–100). Palgrave Macmillan UK. https://doi.org/10.1057/9781137030825_6

Guillén, I. (1998a). The textual interplay of grammatical metaphor on the nominalizations occurring in written medical English. *Journal of Pragmatics*, *30*(3), 363–385. https://doi.org/10.1016/s0378-2166(98)00002-2

Guinda, C. S. (2014). Abridged Abstracts: Rushing the Research Race? *Revista Canaria de Estudios Ingleses*, *69*, 15–33.

Halliday, M. A. K. (1973). *Explorations in the functions of language*. Edward Arnold.

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar* (1st ed.). Edward Arnold.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. Edward Arnold.

Halliday, M. A. K. (1998). Things and Relations: Regrammaticizing Experience as Technical Knowledge. In J. Martin & R. Veel (Eds.), *Reading Science: Critical and Functional Perspectives on Discourse of Science.* (pp. 185–235). Routledge.

Halliday, M. A. K., & Martin, J. R. (1993a). *Writing Science: Literacy and Discursive Power*. The Falmer Press.

Halliday, M. A. K., & Matthiessen, C. (2013). *Halliday's Introduction to Functional Grammar*. https://doi.org/10.4324/9780203431269

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar.* (3rd ed.). Routledge. https://doi.org/10.4324/9780203431269

Halliday, M. A. K., & Matthiessen, M. I. M. (2014). *Halliday's introduction to functional grammar* (4th edn.). Routledge, Taylor and Francis Group. https://doi.org/10.4324/9780203431269

Hammond, M. (2003). *Programming for linguists: Perl for language researchers.* Blackwell.

Hanauer, D. I. (2008). *Scientific Discourse: Multiliteracy in the Classroom* (2nd edn.). Continuum.

Hao, J. (2020). Nominalisations in scientific English. *Functions of Language*, *27*(2), 143–173. https://doi.org/10.1075/fol.16055.hao

Hardie, A. (2012). CQPweb — combining power, flexibility, and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, *17*(3), 380–409. https://doi.org/10.1075/ijcl.17.3.04har

Hardjanto, T. D., & Roselani, N. G. A. (2022). Objectifying Science: Impersonalization in English Research Articles from Different Disciplines. *Theory and Practice in Language Studies*, *12*(10), 2072–2083. https://doi.org/10.17507/tpls.1210.14

Hare, M., Tanenhaus, M. K., & McRae, K. (2007). Understanding and producing the reduced relative construction: Evidence from ratings, editing and corpora. *Journal of Memory and Language*, *56*(3), 410–435. https://doi.org/10.1016/j.jml.2006.08.007

Harman, G. (1975). Language, Thought, and Communication. *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science.*, *7*, 270–298.

Harwood, N. (2005a). 'I hoped to counteract the memory problem, but I made no impact whatsoever': discussing methods in computing science using I [Article]. *English for Specific Purposes (New York, N.Y.)*, *24*(3), 243–267. https://doi.org/10.1016/j.esp.2004.10.002

Harwood, N. (2005b). 'Nowhere has anyone attempted … In this article I aim to do just that': A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics*, *37*(8), 1207–1231. https://doi.org/https://doi.org/10.1016/j.pragma.2005.01.012

Harwood, N. (2005d). "We do not seem to have a theory ... The theory I present here attempts to fill this gap": Inclusive and exclusive pronouns in academic writing. *Applied Linguistics*, *26*(3), 343–375. https://doi.org/10.1093/applin/ami012

Harwood, N. (2005e). What do we want EAP teaching materials for? [Article]. *Journal of English for Academic Purposes*, *4*(2), 149–161. https://doi.org/10.1016/j.jeap.2004.07.008

Harwood, N., & Hadley, G. (2004). Demystifying institutional practices: critical pragmatism and the teaching of academic writing. *English for Specific Purposes*, *23*(4), 355–377. https://doi.org/10.1016/J.ESP.2003.08.001

Harwood, W. (2013). *Being progressive is just a phase: dividing the functional hierarchy*. Department of Linguistics.

Harwood, W. (2015). Being Progressive is just a Phase: celebrating the uniqueness of progressive aspect under a phase-based analysis. *Natural Language & Linguistic Theory*, *33*(2), 523–573.

Harwood, W. (2018a). Reduced Relatives and Extended Phases: A Phase-Based Analysis of the Inflectional Restrictions on English Reduced Relative Clauses. *Studia Linguistica*, *72*(2), 428–471. https://doi.org/10.1111/stul.12069

Heard, S. B. (2016a). *The Scientist's Guide to Writing.* Princeton University Press.

Helms-Park, R., & Stapleton, P. (2003). Questioning the importance of individualized voice in undergraduate L2 argumentative writing: An empirical study with pedagogical implications [Article]. *Journal of Second Language Writing*, *12*(3), 245–265. https://doi.org/10.1016/j.jslw.2003.08.001

Herrando-Rodrigo, I. (2019). Raising awareness around writers' voice in academic discourse: An Analysis of Writers' (In)visibility. *Brno Studies in English*, *45*(2), 53–76. https://doi.org/10.5817/BSE2019-2-3

Herrando-Rodrigo, I. (2022). 'Is voice inevitable in written texts?' Animate agents followed by active verbs as writers' (de) voicing mechanisms in knowledge dissemination practices. *Circulo de Linguistica Aplicada a La Comunicacion*, *90*, 81–94. https://doi.org/10.5209/clac.81309

Herrando-Rodrigo, I. (2014). *A contribution to the Study of Writers' Self-representation: Visible Researchers, Invisible Writers, or How to Make Medical Electronic Popularizations Trustworthy.* Universidad de Zaragoza.

Hersh, W., & Zhang, M. S. (1999). Teaching English Medical Terminology Using the UMLS Metathesaurus and World Wide Web. *Proceedings of the 1999 AMIA Annual Symposium*, 1078.

Hewings, A. (2012). *Stance and Voice in Academic Discourse across Channels* (pp. 187–201). https://doi.org/10.1057/9781137030825_12

Hewings, A., & Coffin, C. (2007a). Writing in multi-party computer conferences and single authored assignments: Exploring the role of writer as thinker [Article]. *Journal of English for Academic Purposes*, *6*(2), 126–142. https://doi.org/10.1016/j.jeap.2007.04.002

Hiltunen, T. (2016). Passives in academic writing: Comparing research articles and student essays across four disciplines. *Corpus Linguistics on the Move: Exploring and*

Understanding English through Corpora, 132–157. https://doi.org/10.1163/9789004321342_008

Hoadley, C. (2012). What is a Community of Practice and how can we support it? In D. H. Jonassen & S. M. Land (Eds.), *Theoretical Foundations of Learning Environments* (2nd edn., pp. 287–300). Routledge.

Hockey, S. (2004). History of humanities computing. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A Companion to Digital Humanities.* (pp. 3–19). Blackwell.

Hoey, M. P. (2004). Lexical priming and the properties of text. In A. Partington, J. Morley, & L. Haarman (Eds.), *Corpora and Discourse.* (pp. 385–412). Peter Lang.

Hoey, M. P. (2005). *Lexical Priming. A new theory of words and language.* Routledge.

Hoffmann, S., & Evert, S. (2006). BNCweb (CQP edition) - the marriage of two corpus tools. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy?: new resources, new tools, new methods.* (pp. 177–195). Peter Lang.

Hoffmann, S., Evert, S., Smith, N., Lee, D. Y. W., & Berglund, P. Y. (2008). *Corpus Linguistics with BNCWeb - a Practical Guide.* Peter Lang.

Horsella, M., & Pérez, F. (1991). Nominal Compounds in Chemical English Literature: Towards an Approach to Text Typology. *English for Specific Purposes*, *19*(2), 125–139.

House, J. (1999). Misunderstanding in intercultural communication: Interactions in English as a lignau franca and the myth of mutual intelligibility. In C. Gnutzmann (Ed.), *Teaching and Learning English as a Global Language* (pp. 77–89). Stauffenburg.

Huckin, T. N., & Olsen, L. A. (1983). *English for Science and Technology: A Handbook for Non-native Speakers.* McGraw-Hill.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language* (1st ed.) [Book]. Cambridge University Press. https://doi.org/10.1017/9781316423530

Hughes, J., Jewson, N., & Unwin, L. (2007). *Communities of Practice. Critical Perspectives* (J. Hughes, N. Jewson, & L. Unwin, Eds.). Routledge.

Hundt, M., Röthlisberger, M., & Seoane, E. (2021). Predicting voice alternation across academic Englishes. *Corpus Linguistics and Linguistic Theory*, *17*(1), 189–222. https://doi.org/10.1515/cllt-2017-0050

Hundt, M., Schneider, G., & Seoane, E. (2016). The use of the be-passive in academic Englishes: local versus global usage in an international language. *Corpora*, *11*(1), 29–61. https://doi.org/10.3366/cor.2016.0084

Hunston, S. (2002). Corpora in Applied Linguistics. In *Cambridge Applied Linguistics*. Cambridge University Press. https://doi.org/DOI:10.1017/CBO9781139524773

Hunston, S. (2006). Corpus Linguistics. *Linguistics*, *7*(2), 215–244.

Hunston, S., & Francis, G. (1999). *Pattern Grammar. A corpus-driven approach to the lexical grammar of English.* John Benjamins Publishing Company.

Hunston, & Thompson. (2000). *Evaluation in text: authorial stance and the construction of discourse / edited by Susan Hunston and Geoff Thompson.* (S. Hunston & G. Thompson, Eds.) [Book]. Oxford University Press.

Hyland, K. (1994). Hedging in academic writing and EAF textbooks. *English for Specific Purposes*, *13*(3), 239–256. https://doi.org/10.1016/0889-4906(94)90004-3

Hyland, K. (1996). Writing Without Conviction? Hedging in Science Research Articles [Article]. *Applied Linguistics*, *17*(4), 433–454. https://doi.org/10.1093/applin/17.4.433

Hyland, K. (1998a). *Hedging in Scientific Research Articles* (Vol. 54). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.54

Hyland, K. (1999a). Academic attribution: citation and the construction of disciplinary knowledge [Article]. *Applied Linguistics*, *20*(3), 341–367. https://doi.org/10.1093/applin/20.3.341

Hyland, K. (1999b). Talking to Students: Metadiscourse in Introductory Coursebooks. *English for Specific Purposes*, *18*(1), 3–26. https://doi.org/10.1016/S0889-4906(97)00025-2

Hyland, K. (2000). *Disciplinary Discourses. Social Interaction in Academic Writing.* . Longman.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, *20*(3), 207–226. https://doi.org/10.1016/S0889-4906(00)00012-0

Hyland, K. (2002a). Authority and invisibility: authorial identity in academic writing [Article]. *Journal of Pragmatics*, *34*(8), 1091–1112. https://doi.org/10.1016/S0378-2166(02)00035-8

Hyland, K. (2002b). Directives: Argument and Engagement in Academic Writing [Article]. *Applied Linguistics*, *23*(2), 215–239. https://doi.org/10.1093/applin/23.2.215

Hyland, K. (2004). Disciplinary interactions: metadiscourse in L2 postgraduate writing [Article]. *Journal of Second Language Writing*, *13*(2), 133–151. https://doi.org/10.1016/j.jslw.2004.02.001

Hyland, K. (2005). *Metadiscourse: exploring interaction in writing* [Book]. Continuum.

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2008b). Genre and academic writing in the disciplines [Article]. *Language Teaching*, *41*(4), 543–562. https://doi.org/10.1017/S0261444808005235

Hyland, K. (2008c). Persuasion, Interaction and the Construction of Knowledge: Representing Self and others in Research Writing. *International Journal of English Studies*, *8*(2), 1–23. http://revistas.um.es/ijes/article/view/49151

Hyland, K. (2008d). "Small bits of textual material": A discourse analysis of Swales' writing. *English for Specific Purposes*, *27*(2), 143–160. https://doi.org/10.1016/j.esp.2006.10.005

Hyland, K. (2010). Community and Individuality: Performing Identity in Applied Linguistics [Article]. *Written Communication*, *27*(2), 159–188. https://doi.org/10.1177/0741088309357846

Hyland, K. (2011a). The presentation of self in scholarly life: Identity and marginalization in academic homepages. *English for Specific Purposes*, *30*(4), 286–297. https://doi.org/10.1016/j.esp.2011.04.004

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes*, *24*(2), 123–139. https://doi.org/10.1016/j.esp.2004.02.002

Hyland, Ken. (2009). *Academic discourse: English in a global context* [Book]. Continuum.

International Committee of Medical Journal Editors (ICMJE). (2016). *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. Citeseer. https://doi.org/10.1080/08941920.2016.1150542

Hyland, Ken. (2012). *Disciplinary Identities: Individuality and Community in Academic Discourse*. Cambridge University Press.

Ivanič, R. (1994). I is for interpersonal: Discoursal construction of writer identities and the teaching of writing [Article]. *Linguistics and Education*, *6*(1), 3–15. https://doi.org/10.1016/0898-5898(94)90018-3

Ivanič, R. (1995). Writer Identity. *Prospect*, *10*(1), 8–31.

Ivanič, R. (1998). *Writing and identity: the discoursal construction of identity in academic writing. (vol.5)*. John Benjamins Publishing Company.

Ivanič, R., & Camps, D. (2001). I am how I sound: Voice as self-representation in L2 writing. *Journal of Second Language Writing*, *10*(1–2), 3–33. https://doi.org/10.1016/S1060-3743(01)00034-0

Ivanič, R., & Roach, D. (1990). Academic Writing, Power and Disguise. In N. Clark, N. Fairclough, R. Ivanič, N. Mcleod, J. Thomas, & P. Meara (Eds.), *British Studies in Applied Linguistics 5: Language and Power* (pp. 103–121). C.I.L.T. for B.A.A.L.

Ivanič, R., & Simpson, J. (1992). Who's Who in Academic Writing? In N. Fairclough (Ed.), *Critical Language Awareness* (pp. 141–173). Longman.

Jairam, A. P., Timmermans, L., Eker, H. H., Pierik, R. E. G. J. M., van Klaveren, D., Steyerberg, E. W., Timman, R., van der Ham, A. C., Dawson, I., Charbon, J. A., Schuhmacher, C., Mihaljevic, A., Izbicki, J. R., Fikatas, P., Knebel, P., Fortelny, R. H., Kleinrensink, G. J., Lange, J. F., Jeekel, H. J., … Golling, M. (2017). Prevention of incisional hernia with prophylactic onlay and sublay mesh reinforcement versus primary suture only in midline laparotomies (PRIMA): 2-year follow-up of a multicentre, double-blind, randomised controlled trial. *The Lancet*, *390*(10094), 567–576. https://doi.org/10.1016/S0140-6736(17)31332-6

Jenkins, J. (2000). *The Phonology of English as an International Language: New Models, New Norms, New Goals.* Oxford University Press.

Jenkins, J. (2003). *World Englishes: A resource book for students.* Routledge.

Jenkins, J. (2006). Current Perspectives on Teaching World Englishes and English as a Lingua Franca. *Tesol Quarterly*, *40*(1), 157–181.

Jenkins, J. (2007). *English as a Lingua Franca: Attitude and Identity*. Oxford University Press.

Jenkins L. (1975). *The English Existential*. Niemeyer.

Jensen, K. E. (2014). Linguistics and the digital humanities. *Mediekultur: Journal of Media and Communication Research*, *30*(57), 117–136. https://doi.org/10.7146/mediekultur.v30i57.15968

Joe, H. B., Lee, S. Y., Kim, J.-S., Chang, H. S., Jeong, Y., Jeong, H., & Park, S. Y. (2016). Effect of total intravenous anaesthesia and prophylactic ramosetron on postoperative nausea and vomiting after thyroidectomy: A prospective, randomized controlled study. *Journal of International Medical Research*, *44*(1), 81–88. https://doi.org/10.1177/0300060515607384

John, S. (2005). *The Writing Process and Writer Identity: Investigating the Influence of Revision on Linguistic and Textual Features of Writer Identity in Dissertations.* [PhD Thesis (unpublished)]. University of Birmingham.

John, S. (2007). Meeting the challenge of developing an academic identity: a textual approach. In P. Teo & C. Ho (Eds.), *Discourse in the Modern World: Perspectives and Challenges* (pp. 28–52). McGraw-Hill.

John, S. (2009). Using the revision process to help international students understand the linguistic construction of the academic identity. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic Writing: At the Interface of Corpus and Discourse.* (pp. 272–290). Continuum.

Johns, Tim. (2001). From evidence to conclusion: The case of 'indicate that.' In H. Martin (Ed.), *Academic writing in contexts: Implications and applications* (pp. 55–62). Bloomsbury Academic.

Joubert, P. H., & Rogers, S. M. (2015). *Strategic scientific and medical writing - The road to success*. Springer. https://doi.org/10.1007/978-3-662-48316-9

Kachru, B. B. (1992). *The Other Tongue: English across Cultures (English in the Global Context)* (B. B. Kachru, Ed.; 2nd edn.). University of Illinois Press.

Kawaguchi, Y., Minegishi, M., & Viereck, W. (2011). *Corpus-based analysis and diachronic linguistics.* John Benjamins Publishing Company. https://doi.org/https://doi.org/10.1075/tufs.3

Kesäniemi, J., Vartiainen, T., Säily, T., & Nevalainen, T. (2018). Open Science for English historical corpus linguistics: Introducing the language change database. *CEUR Workshop Proceedings*, *2084*, 51–62.

Khedri, M. (2016). *Are we visible? An interdisciplinary data-based study of self-mention in research articles*. *52*(3), 403–430. https://doi.org/doi:10.1515/psicl-2016-0017

Khedri, M., & Kritsis, K. (2020). How do we make ourselves heard in the writing of a research article? A study of authorial references in four disciplines [Article]. *Australian Journal of Linguistics*, *40*(2), 194–217. https://doi.org/10.1080/07268602.2020.1753011

Kieran O'Halloran. (2003). *Critical Discourse Analysis and Language Cognition* (1st ed.) [Book]. Edinburgh University Press. https://doi.org/10.3366/j.ctvxcrwcs

Kilgarriff, A. (1993). Dictionary word-sense distinctions: An enquiry into their nature. *Computers and the Humanities*, *26*(1–2), 365–387.

Kilgarriff, A. (1997). The hard parts of lexicography. *International Jounral of Lexicography*, *11*(1), 51–54.

Kilgarriff, A. (2009). Simple maths for keywords. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings, Corpus Linguistics*.

Kilgarriff, A., & Kosem, I. (2012). Corpus tools for lexicographers. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography*. Oxford University Press.

Kilgarriff, A., & Rundell, M. (2002). Lexical profiling software and its lexicographic applications: A case study. In A. Braasch & C. Povlsen (Eds.), *Proceedings of the Tenth Euralex Congress.* (pp. 807–818). University of Copenhagen.

Killingsworth, M. J. (1992). Discourse Communities. Local and global. *Rhetoric Review*, *11*(1), 110–122. https://www.jstor.org/stable/465883

Kuhi, D., & Behnam, B. (2011). Generic variations and metadiscourse use in the writing of applied linguists: A comparative study and preliminary framework. *Written Communication*, *28*(1), 97–141. https://doi.org/10.1177/0741088310387259

Kuhn, T. S. (1970). The Structure of Scientific Revolutions. In *International Encyclopedia of Unified Science* (2nd Editio, Vol. 2). The University of Chicago Press.

Kulikov, L. (2004). T. Givón, Syntax: an introduction (2 vols.). Amsterdam & Philadelphia: John Benjamins, 2001. Pp. xvii+500, x+406 [Article]. *Journal of Linguistics*, *40*(2), 419–425. https://doi.org/10.1017/S002222670429274X

Kulkarni, A. D., Adashi, E. Y., Jamieson, D. J., Crawford, S. B., Sunderam, S., & Kissin, D. M. (2017). Fertility Treatments and Multiple Births in the United States. *Paediatric and Perinatal Epidemiology*, *31*(5), 438–448. https://doi.org/10.1111/ppe.12383

Kuo, C.-H. (1999). The Use of Personal Pronouns: Role Relationships in Scientific Journal Articles [Article]. *English for Specific Purposes (New York, N.Y.)*, *18*(2), 121–138. https://doi.org/10.1016/S0889-4906(97)00058-6

Lafuente Millán, E. (2008). Epistemic and approximative meaning revisited: the use of hedges boosters and approximators when writing research in different disciplines. *English as an Additional Language in Research Publication and Communication*, 65–82.

Lafuente-Millán, Mur-Dueñas, Lorés-Sanz, & Vázquez-Orta. (2010). Interpersonality in written academic discourse: three analytical perspectives. In *Constructing Interpersonality. Multiple Perspectives on Written Academic Genres.*

Lauridsen, K. M. (1996). Text corpora and contrastive linguistics: Which type of corpus for which type of analysis? In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in Contrast.* (pp. 63–71). Lund University Press.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Startvik (Ed.), *Directions in corpus linguistics.* (pp. 105–122). Mouton de Gruyter.

Leech, G., Johansson, S., Hofland, K., & Garside, R. (1976). *The Lancaster-Oslo/Bergen Corpus (LOB Corpus)*. www.helsinki.fi/varieng/CoRD/corpora/LOB

Lehman, I. (2018). Authorial Presence in English Academic Texts. In *A Comparative Study of Student Writing across Cultures and Disciplines*. Peter Lang Verlag. https://doi.org/10.3726/b14175

Lehman, I., de Celis, K. C.-G., & Sułkowski, Ł. (2022). Writing to make a difference: Discursive analysis of writer identity in research articles on management [Article]. *Ibérica (Castellón de La Plana, Spain)*, *44*, 155–178. https://doi.org/10.17398/2340-2784.44.155

Lehman, I. M., & Anderson, R. (2021a). Inviting individual voice to second language academic writing [Article]. *International Review of Pragmatics*, *13*(1), 61–85. https://doi.org/10.1163/18773109-01301002

Lehman, I. M., & Anderson, R. (2021b). Inviting individual voice to second language academic writing [Article]. *International Review of Pragmatics*, *13*(1), 61–85. https://doi.org/10.1163/18773109-01301002

Lehmann, H. M., Schneider, P., & Hoffmann, S. (2000). BNCweb. In *Corpora galore.* (pp. 259–266). Rodopi.

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, *22*, 42–53. https://doi.org/10.1016/j.jeap.2016.01.008

Levin, B. (1991). Building a lexicon: The contribution of linguistics. *International Journal of Lexicography*, *4*(3), 205–226.

Liardét, C. L. (2016a). Grammatical metaphor: Distinguishing success. *Journal of English for Academic Purposes*, *22*, 109–118. https://doi.org/10.1016/j.jeap.2016.01.009

Liardét, C. L. (2016b). Grammatical metaphor: Distinguishing success. *Journal of English for Academic Purposes*, *22*, 109–118. https://doi.org/https://doi.org/10.1016/j.jeap.2016.01.009

Liardét, C. L. (2016c). Nominalization and grammatical metaphor: Elaborating the theory. *English for Specific Purposes*, *44*, 16–29. https://doi.org/10.1016/j.esp.2016.04.004

Loock, R. (2013). Close Encounters of the Third Code: Quantitative vs. Qualitative Analyses in Corpus-Based Translation Studies [Article]. *Belgian Journal of Linguistics*, *27*, 61–86. https://doi.org/10.1075/bjl.27.04loo

Lorés-Sanz, R. (2008). Genres in Contrast: The Exploration of Writers' Visibility in Research Articles and Research Article Abstracts. In S. Burgess & P. Martín-Martín (Eds.), *English as an Additional Language in Research Publication and Communication* (pp. 105–122). Peter Lang.

Lores-Sanz, R. (2011). The construction of the author's voice in academic writing: the interplay of cultural and disciplinary factors [Article]. *Text & Talk*, *31*(2), 173–193. https://doi.org/10.1515/text.2011.008

Lorés-Sanz, R. (2011). The study of authorial voice: using a Spanish-English corpus to explore linguistic transference. *Corpora*, *6*(1), 1–24.

Lorés-Sanz, R., & Herrando-Rodrigo, I. (2020). New concepts, different approaches: Tackling e-visibility in research project websites. *Revista de Linguistica y Lenguas Aplicadas*, *15*(1), 83–98. https://doi.org/10.4995/RLYLA.2020.12782

Lorés-Sanz, R., Mur-Dueñas, P., & Lafuente-Millán, E. (2010a). *Constructing Interpersonality: Multiple Perspectives on Written Academic Genres.* (R. Lorés-Sanz, P. Mur-Dueñas, & E. Lafuente-Millán, Eds.). Cambridge Scholars Publishing.

Lorés-Sanz, R., Mur-Dueñas, P., & Lafuente-Millán, E. (2010b). Constructing Interpersonality: Multiple Perspectives on Written Academic Genres. In *Constructing Interpersonality*. Cambridge Scholars Publishing.

Lorés-Sanz, R., Mur-Dueñas, P., & Lafuente-Millán, E. (2010c). *Constructing Interpersonality: Multiple Perspectives on Written Academic Genres.* (R. Lorés-Sanz, P. Mur-Dueñas, & E. Lafuente-Millán, Eds.). Cambridge Scholars Publishing.

Lorés-Sanz, R., & Murillo, S. (2007). Authorial identity and reader involvement in academic writing: a contrastive study of the use of pronouns in RA abstracts. *Aprendizaje de Lenguas, Uso Del Lenguaje y Modelación Cognitiva. Perspectivas Aplicadas Entre Disciplinas: Actas Del XXIV Congreso Internacional de AESLA.*

Lourenco, T., Pickard, R., Vale, L., Grant, A., Fraser, C., MacLennan, G., & N'Dow, J. (2008). Alternative approaches to endoscopic ablation for benign enlargement of the prostate: Systematic review of randomised controlled trials. *BMJ*, *337*(7660), 36–39. https://doi.org/10.1136/bmj.39575.517674.BE

Luke, A. (1993). Series Editor's Introduction. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing Science: Literacy and Discursive Power.* (pp. xi–xv). The Falmer Press.

MacDonald, S. P. (1992). A method for analysing sentence-level differences in disciplinary knowledge making. *Written Communication*, *9*(4), 533–569.

Maglie, R. (2009). *Understanding the Language of Medicine* (1st edn.). Aracne.

Mair, C. (2003). *The Politics of English as a World Language.* (C. Mair, Ed.). Rodopi.

Majumder, K. (2015). *6 Article types that journals publish: A guide for early career researchers.* http://www.editage.com/insights/6-article-types-that-journals-publish-a-guide-for-early-career-researchers

Malmfors, B., Garnsworthy, P., & Grossman, M. (2004a). *Writing and Presenting Scientific Papers.* (2n ed.). Nottingham University Press.

Maria Carciu, O. (2009). An intercultural study of first-person plural references in biomedical writing [Article]. *Ibérica (Castellón de La Plana, Spain)*, *18*, 71–92.

Marín Arrese, J. (2002). Mystification of agency in passive, impersonal and spontaneous situation types. In J. Marín Arrese (Ed.), *Conceptualization of Events in Newspaper Discourse: Mystification of agency and degree of implication in news reports.* (pp. 31–54). Universidad Complutense.

Marín-Arrese, J., Gómez, M., Martínez Caro, E., Neff, J., & Becerril, S. (2003). *Mystification of agency in newspaper discourse and writer's relative degree of implication in reported events: grammatical strategies of impersonalisation* (pp. 859–866).

Marín-Arrese, J. I. (1993). *La pasiva en inglés: un estudio funcional-tipológico*. Universidad Complutense de Madrid.

Marín-Arrese, J. I. (2008). Construal and the Use of Impersonalization Strategies in English and Spanish in an FLL Context. *General & Theoretical Papers*, *730*.

Marin-Arrese, J., Martínez Caro, E., & Becerril, S. (2001a). *A corpus study of impersonalisation strategies in newspaper discourse in English and Spanish* (pp. 369–377).

Marin-Arrese, J., Martínez Caro, E., & Becerril, S. (2001b). *A corpus study of impersonalisation strategies in newspaper discourse in English and Spanish* (pp. 369–377).

Martin, J. R., & White, P. R. R. (n.d.). *The Language of Evaluation Appraisal in English*. London: Palgrave MacMillan https://doi.org/10.1057/9780230511910

Martínez, I. A. (2001). Impersonality in the research article as revealed by analysis of the transitivity structure. *English for Specific Purposes*, *20*(3), 227–247. https://doi.org/10.1016/S0889-4906(00)00013-2

Martínez, I. A. (2005a). Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing*, *14*(3), 174–190. https://doi.org/10.1016/j.jslw.2005.06.001

Martín-Martín, P. (2004). Genre and discourse community. *English Studies*, *25*, 153–167.

Master, P. (1991). Active verbs with inanimate subjects in scientific prose. *English for Specific Purposes*, *10*(1), 15–33. https://doi.org/https://doi.org/10.1016/0889-4906(91)90013-M

Master, P. (2001). Active verbs with inanimate subjects in scientific research articles. *Academic Writing in Context. Implications and Applications*, 169–181.

Matsuda, P. K. (2001). Voice in Japanese written discourse: Implications for second language writing. *Journal of Second Language Writing*, *10*(1), 35–53. https://doi.org/https://doi.org/10.1016/S1060-3743(00)00036-9

Matsuda, P. K., & Tardy, C. M. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes*, *26*(2), 235–249. https://doi.org/10.1016/j.esp.2006.10.001

Matsuda, P. K., & Tardy, C. M. (2008). Continuing the conversation on voice in academic writing. *English for Specific Purposes*, *27*(1), 100–105. https://doi.org/10.1016/j.esp.2007.04.002

Matthews, J. R., & Matthews, R. W. (2008a). *Successful Scientific Writing: A Step-by-step Guide for the Biological and Medical Sciences.* (3rd edn.). Cambridge University Press.

Matthews, J. R., & Matthews, R. W. (2014). *Successful Scientific Writing: A Step-by-Step Guide for the Biological and Medical Sciences* (Fourth edition) [Book]. Cambridge University Press. https://doi.org/10.1017/CBO9781107587915

Matthiessen, C. M. I. M. (2004a). *Descriptive motifs and generalizations.* (M. Caffarel & C. M. I. M. Matthiessen, Eds.; pp. 537–673). de Gruyter.

Matthiessen, C. M. I. M. (2004b). The evolution of language: a systemic functional exploration of phylogenetic phases. In G. Williams & A. Lukin (Eds.), *Language development: functional perspectives on evolution and ontogenesis.* (pp. 45–90). Continuum.

Matthiessen, C. M. I. M. (2012). Systemic Functional Linguistics as appliable linguistics: social accountability and critical approaches [Article]. *Revista de Documentação de Estudos Em Lingüística Teórica e Aplicada*, *28*(spe), 435–471. https://doi.org/10.1590/S0102-44502012000300002

Mauranen, A. (1993). *Cultural Differences in Academic Rhetoric: A Text Linguistic Study.* Peter Lang.

Mauranen, A. (2008). *ELFA 2008. The Corpus of English as a Lingua Franca in Academic Settings.* http://www.helsinki.fi/elfa/elfacorpus

Mauranen, A. (2009). English as a Lingua Franca. In A. Mauranen & E. Ranta (Eds.), *English as a Lingua Franca: Studies and Findings*. Cambridge Scholars Publishing.

Mauranen, A. (2015). English as a global Lingua Franca: changing language in changing global academia. *Exploring ELF in Japanese Academic and Business Contexts*, *33*(January 2015), 29–46. https://doi.org/10.4324/9780203929353

Mauranen, A., Hynninen, N., & Ranta, E. (2010a). English as an academic lingua franca: The ELFA project. *English for Specific Purposes*, *29*(3), 183–190. https://doi.org/10.1016/j.esp.2009.10.001

Mauranen, A., Hynninen, N., & Ranta, E. (2010b). English as an academic lingua franca: The ELFA project. *English for Specific Purposes*, *29*(3), 183–190. https://doi.org/10.1016/j.esp.2009.10.001

McArthur, T. (1998). *The English Languages*. Cambridge University Press.

McEnery, T., & Wilson, A. (2001). *Corpus Linguistics.* Edinburgh University Press.

McEnery, T., & Xiao, R. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning. (Vol. 2)* (pp. 364–380). Routledge.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-Based Language Studies. An Advanced Rosource Book.* Routledge.

McGrath, I. (2016). *Materials evaluation and design for language teaching* (2nd ed.). Edinburgh University Press.

McKoon, G., & Ratcliff, R. (2003). Meaning through syntax: Language comprehension and the reduced relative clause construction. *Psychological Review*, *110*(3), 490–525.

Meierkord, C. (1996). *Englisch als Medium der interkulturellen Kommunikation: Unterschungen zum non-native-/non-native-Speaker-Discurs*. Lang.

Melchers, G., & Shaw, P. (2011). *World Englishes: An Introduction.* (2nd edn.). Hodder Education.

Merlo, P., & Stevenson, S. (1998). What grammars tell us about corpora: the case of reduced relative clauses. *Sixth Workshop on Very Large Corpora*, 134–142.

Mesthrie, R., & Bhatt, R. M. (2008). *World Englishes: The study of new linguistic varieties.* Cambridge University Press.

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

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

Miettinen, O. S. (2015). *Medicine as a Scholarly Field: An Introduction*. Springer. https://doi.org/10.1007/978-3-319-19012-9

Millar, N., Budgell, B., & Fuller, K. (2013). 'Use the active voice whenever possible': The Impact of Style Guidelines in Medical Journals. *Applied Linguistics*, *34*(4), 393–414. https://doi.org/10.1093/applin/ams059

Millward, C. M., & Hayes, M. (2011). *A Biography of the English Language* (3rd edn.). Wadsworth CENGAGE Learning.

Molino, A. (2010). Personal and impersonal authorial references: A contrastive study of English and Italian Linguistics research articles. *Journal of English for Academic Purposes*, *9*(2), 86–101. https://doi.org/10.1016/j.jeap.2010.02.007

Mozaffari, A., & Moini, R. (2014). Academic words in education research articles: A corpus study. *Procedia - Social and Behavioral Sciences*, *98*, 1290–1296. https://doi.org/10.1016/j.sbspro.2014.03.545

Mur-Dueñas, P., Lorés-Sanz, R., & Lafuente-Millán, E. (2010). Editorial. *Journal of English for Academic Purposes*, *9*(2), 83–85. https://doi.org/10.1016/j.jeap.2010.02.006

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223–242.

Nesselhauf, N. (2005). *Collocations in a learner corpus.* John Benjamins Publishing Company.

Nguyen, T. M. H., & Webb, S. (2016). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 1–23. https://doi.org/10.1177/S0267190512000098

Nickols, F. (2003). Communities of Practice: An Overview. *Distance Consulting*, 6.

Nunn, R. (2008). Do academic reviewers readily accept a first-person voice? *Asian EFL Journal*, *10*(2), 210–229.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge University Press.

Oliver del Olmo, S. (2004). *Análisis contrastivo español/inglés de la atenuación retórica en el discurso médico. El artículo de investigación: el caso clínico.* Universitat Pompeu Fabra.

Orchard, T. J., Temprosa, M., Goldberg, R., Haffner, S., Ratner, R., Marcovina, S., & Fowler, S. (2005). The Effect of Metformin and Intensive Lifestyle Intervention on the Metabolic Syndrome: The Diabetes Prevention Program. *Annals of Internal Medicine*, *142*(8), 611–619. https://doi.org/142/8/611 [pii]

Osborne, J. (2002). Science Without Literacy: A ship without a sail? *Cambridge Journal of Education*, *32*(2), 203–218. https://doi.org/10.1080/03057640220147559

Owens, T. (2011). Defining data for humanists: Text, artifact, information or evidence? *Journal of Digital Humanities*, *1*(1), 6–8.

Pahta, P., Taavitsainen, I., Nevalainen, T., & Tyrkkö, J. (2007). *Towards multimedia in corpus studies.* Research Unit for Variation, Contacts and Change in English (VARIENG). University of Helsinki.

Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching.* John Benjamins Publishing Company.

Pérez-Iglesias, E. (2004). Modern English Classical Scientific Terminology. *Odisea*, *5*, 133–141.

Pérez-Llantada, C. (2009). Native and non-native English scholars publishing research internationally: A small-scale study on authorial (in)visibility. *Journal of Applied Linguistics*, *4*(2), 217–238. https://doi.org/10.1558/japl.v4i2.217

Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance [Article]. *Discourse Studies*, *10*(2), 231–250. https://doi.org/10.1177/1461445607087010

Piper, R. J. (2013). *How to write a systematic literature review: a guide for medical students*.

Pogner, K.-H. (2005). Discourse Communities and Communities of Practice: On the social context of text and knowledge production in the workplace. In *Intercultural Communication*.

Porter, J. E. (1986). Intertextuality and the Discourse Community. *Rhetoric Review*, *5*(1), 34–47. http://www.jstor.org/stable/466015

Poulsen, C. (2015). Student writing: gender and visibility; then and now [Article]. *English Teaching: Practice and Critique*, *14*(3), 404–418. https://doi.org/10.1108/ETPC-06-2015-0048

Prior, P. (2001). Voices in text, mind, and society: Sociohistoric accounts of discourse acquisition and use [Article]. *Journal of Second Language Writing*, *10*(1–2), 55–81. https://doi.org/10.1016/S1060-3743(00)00037-0

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* Longman.

Rafoth, B. A. (1988). Discourse Community: Where Writers, Readers, and Texts Come Together. In B. A. Rafoth & D. L. Rubin (Eds.), *The Social Construction of Written Communication*. Ablex Publications.

Rafoth, B. A. (1990). The concept of discourse community: Descriptive and explanatory adequacy. In G. Kirsch & D. H. Roen (Eds.), *A sense of audience in written communication* (pp. 140–152). Sage Publications.

Rahimivand, M., & Kuhi, D. (2014). An Exploration of Discoursal Construction of Identity in Academic Writing. *Procedia - Social and Behavioral Sciences*, *98*, 1492–1501. https://doi.org/10.1016/j.sbspro.2014.03.570

Railsback, B. (2018). *What is Science?* GEOL 1122. www.gly.uga.edu/railsback/1122science2.html

Ravelli, L. J. (1988a). Grammatical metaphor: An initial analysis. In E. H. Steiner & R. Veltman (Eds.), *Pragmatics, discourse and text: Some systematically-oriented approaches* (pp. 133–147). Pinter.

Ravelli, L. J. (1988b). Grammatical metaphor: An initial analysis. *Pragmatics, Discourse and Text: Some Systemically-Oriented Approaches*, 133–147.

Ravelli, L. J. (1996). Making language accessible: Successful text writing for museum visitors [Article]. *Linguistics and Education*, *8*(4), 367–387. https://doi.org/10.1016/S0898-5898(96)90017-0

Reilly, J., Zamora, A., & McGivern, R. (2005). Acquiring perspective in English: The development of stance. *Journal of Pragmatics*, *37*, 185–208. https://doi.org/10.1016/j.pragma.2004.08.010

Rodman, L. (1994). The Active Voice in Scientific Articles: Frequency and Discourse Functions. *Journal of Technical Writing and Communication*, *24*(3), 309–331. https://doi.org/10.2190/D9BR-CAP2-LW5N-LCRP

Rodríguez-Vergara, D. (2017). A systemic functional approach to the passive voice in English into Spanish translation: Thematic development in a medical research article. *Open Linguistics*, *3*(1), 1–17. https://doi.org/10.1515/opli-2017-0001

Rogers, S. M. (2014a). *Mastering Scientific and Medical Writing: A Self-Help Guide* (2nd edn.). Springer.

Rogers, S. M. (2014b). *Mastering Scientific and Medical Writing: A Self-Help Guide* (2nd edn.). Springer.

Römer, U. (2005). *Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts, and didactics.* John Benjamins Publishing Company.

Römer, U. (2009). The inseparability of lexis and grammar Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, *7*(1), 140–162. https://doi.org/10.1075/arcl.7.06rom

Römer, U., & Schulze, R. (2009). *Exploring the lexis-grammar interface.* (U. Römer & R. Schulze, Eds.). John Benjamins Publishing Company.

Ross, J. R. (1967). On the cyclic nature of English pronominalization. In *In To honor Roman Jakobson: Essays on the occasion of his seventieth birthday* (Vol. 3, pp. 1669–1982). Mouton.

Ross, J. R. (1972). Doubl-ing. *Linguistic Inquiry*, *3*(1), 61–86. http://www.jstor.org.sire.ub.edu/stable/4177686

Rundblad, G. (2007a). Impersonal, General, and Social: The Use of Metonymy Versus Passive Voice in Medical Discourse. *Written Communication*, *24*, 250–277. https://doi.org/10.1177/0741088307302946

Rundblad, G. (2007b). Impersonal, general, and social: The use of metonymy versus passive voice in medical discourse. *Written Communication*, *24*(3), 250–277. https://doi.org/10.1177/0741088307302946

Rundblad, G. (2007c). Impersonal, general, and social: The use of metonymy versus passive voice in medical discourse. *Written Communication*, *24*(3), 250–277. https://doi.org/10.1177/0741088307302946

Rutkove, S. B. (2016). *Biomedical Research: An Insider's Guide*. Springer.

Salager-Meyer, F. (1983). Compound Nominal Phrases in Scientific-Technical Literature: Proportion and Rationale. Reading for Professional Purposes. *Studies and Practices in Native and Foreign Languages*, 136–144.

Salager-Meyer, F. (1985). Syntax and Semantics of Compound Nominal Phrases in Medical English Literature: A Comparative Study with Spanish. *English for Specific Purposes, Special Issue: ESP in South America*.

Salager-Meyer, F. (1993). Imprecision and vagueness (hedging) in today's medical discourse: courtesy, coyness or necessity? *The ESPecialist*, *14*(1), 1–15.

Salager-Meyer, F. (1994a). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*. https://doi.org/10.1016/0889-4906(94)90013-2

Salager-Meyer, F. (1994b). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*. https://doi.org/10.1016/0889-4906(94)90013-2

Salazar, D., Ventura, A., & Verdaguer, I. (2013). A cross-disciplinary analysis of personal and impersonal features in English and Spanish scientific writing. In I. Verdaguer, N. J. Laso, & D. Salazar (Eds.), *Biomedical English. A Corpus-Based Approach* (pp. 121–143). John Benjamins Publishing Company.

Saraceni, M. (2015). *World Englishes: A critical analysis.* Bloomsbury.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education.* John Benjamins Publishing Company.

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *Journal of Applied Linguistics*, *11*(2), 133–158.

Seidlhofer, B. (2004). Research Perspectives on Teaching English As a Lingua Franca. *Annual Review of Applied Linguistics*, *24*, 209–239. https://doi.org/10.1017/S0267190504000145

Seidlhofer, B. (2005). English as a Lingua Franca. *ELT Journal*, *59*(4), 339–341. https://doi.org/https://doi.org/10.1093/elt/cci064

Seoane, E., & Hundt, M. (2018). Voice Alternation and Authorial Presence: Variation across Disciplinary Areas in Academic English. *Journal of English Linguistics*, *46*(1), 3-22. https://doi.org/10.1177/0075424217740938

Seronde, M. F., Geha, R., Puymirat, E., Chaib, A., Simon, T., Berard, L., Drouet, E., Bataille, V., Danchin, N., & Schiele, F. (2014). Discharge heart rate and mortality after acute myocardial infarction. *American Journal of Medicine*, *127*(10), 954–962. https://doi.org/10.1016/j.amjmed.2014.06.034

Seskauskiene, I. (2022). *Metaphor in Legal Discourse*. Cambridge Scholar Publishing.

Shaw, P. (2003). Evaluation and promotion across languages [Article]. *Journal of English for Academic Purposes*, *2*(4), 343–357. https://doi.org/10.1016/S1475-1585(03)00050-X

Shaw, P., & Vassileva, I. (2009). Co-evolving academic rhetoric across culture; Britain, Bulgaria, Denmark, Germany in the 20th century. *Journal of Pragmatics*, *41*(2), 290–305. https://doi.org/10.1016/j.pragma.2008.07.009

Silver, M. (2012). Voice and stance across disciplines in academic discourse. In *Stance and Voice in Written Academic Genres* (pp. 202–217). https://doi.org/10.1057/9781137030825

Silyn-Roberts, H. (2012). *Writing for Science and Engineering. Papers, Presentations and Reports.* (2nd edn.). Elsevier.

Sinclair, J. McH. (1991). *Corpus, concordance, collocation.* Oxford University Press.

Sinclair, J. McH. (2004). *How to use corpora in language teaching.* (J. McH. Sinclair, Ed.). John Benjamins Publishing Company.

Sinclair, J. McH. (2008). The phrase, the whole phrase and nothing but the phrase. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective.* (pp. 407–410). John Benjamins Publishing Company.

Smith, C. S. (1961). A class of complex modifiers in English. *Language, 37*, 342–365.

Smith, C. S. (1964). Determiners and Relative Clauses in a Generative Grammer of English. *Language*, *40*(1), 37–52. https://doi.org/10.2307/411923

Smitterberg, E., & Kytö, M. (2015). English genres in diachronic corpus linguistics. In P. Shaw, B. Erman, G. Melchers, & P. Sundkvist (Eds.), *From clerks to corpora: Essays on the English language yesterday and today.* (pp. 117–133). Stockholm University Press. https://doi.org/http://dx.doi.org/10.16993/bab.g

Solstad, T., & Lyngfelt, B. (2006). Perspectives on Demotion: Introduction to the Volume. In T. Solstad & B. Lyngfelt (Eds.), *Demoting the Agent: Passive, Middle and Other Voice Phenomena* (pp. 1–20). John Benjamins B.V.

Solstad, T., Lyngfelt, B., & Krave, M. (2004). Pre-Conference Proceedings of the Oslo Workshop "Demoting the Agent: Passive and Other Voice-Related Phenomena". In T. Solstad, B. Lyngfelt, & M. Krave (Eds.), *Demoting the Agent: Passive and Other Voice-Related Phenomena'*. University of Oslo.

Sperling, M., & Appleman, D. (2011). Voice in the Context of Literacy Studies [Article]. *Reading Research Quarterly*, *46*(1), 70–84. https://doi.org/10.1598/RRQ.46.1.4

Stapleton, P. (2002). Critiquing voice as a viable pedagogical tool in L2 writing: Returning the spotlight to ideas [Article]. *Journal of Second Language Writing*, *11*(3), 177–190. https://doi.org/10.1016/S1060-3743(02)00070-X

Stapleton, P., & Helms-Park, R. (2008). A response to Matsuda and Tardy's "Voice in academic writing: The rhetorical construction of author identity in blind manuscript review." *English for Specific Purposes*, *27*(1), 94–99. https://doi.org/10.1016/j.esp.2007.04.001

Stefanowitsch, A., & Gries, S. Th. (2003). Collostructions. Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, *8*(2), 209–243.

Stock, I., & Eik-Nes, N. L. (2016). Voice features in academic texts – A review of empirical studies. *Journal of English for Academic Purposes*, *24*, 89–99. https://doi.org/10.1016/j.jeap.2015.12.006

Stotesbury, H. (2006). Gaps and False Conclusions: Criticism in Research Article Abstracts across the Disciplines. In K. Hyland & M. Bondi (Eds.), *Academic Discourse Across Disciplines* (Peter Lang, Vol. 42, pp. 123–148).

Stuart, Mark. C. (2007a). *The Complete Guide to Medical Writing* (M. C. Stuart, Ed.). Pharmaceutical Press.

Stuart, Mark. C. (2007b). *The Complete Guide to Medical Writing* (M. C. Stuart, Ed.). Pharmaceutical Press.

Stubbs, M., & Barth, J. (2003). Using recurrent phrases as text-type discriminators. *Functions of Corpus Linguistics*, *10*(1), 61–104.

Sun, Y., & Cheng, L. (2015). There-existential Sentences in Trial: A Corpus-based Approach. *Proceedings of the 2015 International Conference on Social Science, Education Management and Sports Education*, 1853–1856. https://doi.org/10.2991/ssemse-15.2015.473

Svartvik, J. (1990). *The London-Lund Corpus of Spoken English*. www.helsinki.fi/varieng/CoRD/corpora/LLC/background.html

Svensson, P. (2010). The landscape of digital humanities. *DHQ: Digital Humanities Quartely*, *4*(1).

Swales. (1981). *Aspects of article introductions*. The Unviersity of Aston, Language Studies Unit.

Swales, J. (1991). ESP development worldwide. In Luzares et al. (Ed.), *Proceedings of the ESP State-of-the-Art Conference.* (pp. 11–19).

Swales, J. D. (1990). *Genre analysis: English in academic and research settings* [Book]. Cambridge University Press.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

Swales, J. M. (1993). Genre and Engagement. *Revue Belge de Philologie et d'Histoire*, *73*, 687–698.

Swales, J., & Najjar, H. (1987). The writing of research article introductions. *Written Communication*, *4*, 175–191.

Tan, M. (2002). *Corpus studies in language education.* IELE Press.

Tang, R. (2004). *An approach to written academic voice: exploring the interpersonal negotiations in student academic writing through APPRAISAL.* University of Birmingham.

Tang, R., & John, S. (1999a). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun [Article]. *English for Specific Purposes (New York, N.Y.)*, *18*(supplement), S23–S39. https://doi.org/10.1016/S0889-4906(99)00009-5

Tang, R., & John, S. (1999b). The "I" in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes*, *18*(SUPPL. 1). https://doi.org/10.1016/S0889-4906(99)00009-5

Tardy, C. M. (2012a). Current Conceptions of Voice. In K. Hyland & C. Sancho Guinda (Eds.), *Stance and Voice in Written Academic Discourse.* (pp. 34–46). Palgrave Macmillan.

Tardy, C. M. (2012b). Voice Construction, Assessment , and Extra-Textual Identity Voice Construction in Academic Writing. *Research in the Teaching of English*, *47*(1), 64–99.

Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1998). On the use of the passive and active voice in astrophysics journal papers: With extensions to other languages and other fields. *English for Specific Purposes*, *17*(1), 113–132. https://doi.org/10.1016/S0889-4906(97)00032-X

Taylor, C. A. (1996). *Defining Science: A Rhetoric of Demarcation*. The University of Wisconsin Press.

The Science Council. (2018). *Definition of Science.* http://www.sciencecouncil.org

Thompson, G. (2012a). Intersubjectivity in newspaper editorials Construing the reader-in-the-text [Article]. *English Text Construction*, *5*(1), 77–100. https://doi.org/10.1075/etc.5.1.05tho

Thompson, G. (2012b). Intersubjectivity in newspaper editorials: Construing the reader-in-the-text. *English Text Construction*, *5*. https://doi.org/10.1075/etc.5.1.05tho

Tognini-Bonelli, E. (2001). *Corpus linguistics at work.* John Benjamins Publishing Company.

Toulmin, S. (1958). *The uses of argument.* Cambridge University Press.

Tribble, C. (2000). Practical uses for language corpora in ELT. In P. Brett & G. Motteram (Eds.), *A special interest in computers: Learning and teaching with information and communication technologies.* (pp. 31–41). Peter Lang.

Trowler, P. (2001). *Becher, T. and Trowler, P. (2001) Academic Tribes and Territories: intellectual enquiry and the cultures of disciplines (2nd edition). Buckingham: Open University Press/SRHE.*

Tsui, A. B. M. (2004). What teachers have always wanted to know - and how corpora can help. In J. McH. Sinclair (Ed.), *How to Use Corpora in Language Teaching.* (pp. 39–61). John Benjamins Publishing Company.

Tyrkkö, J., & Hiltunen, T. (2009). *Frequency of nominalization in Early Modern English medical writing* (pp. 297–320). https://doi.org/10.1163/9789042029101_016

Tyrkko, J., & Hiltunen, T. (2009). Frequency of nominalization in Early Modern English medical writing [Proceeding]. *Corpora: Pragmatics and Discourse*, *68*, 297–320.

Ulijn, J., & Salager-Meyer, F. (1998). The professional reader and the text: insights from L2 research. *Journal of Research in Reading*, *21*(2), 79–95. https://doi.org/https://doi.org/10.1111/1467-9817.00046

Vafaie, M., Slagman, A., Möckel, M., Hamm, C., Huber, K., Müller, C., Vollert, J. O., Blankenberg, S., Katus, H. A., Liebetrau, C., Giannitsis, E., & Searle, J. (2016). Prognostic Value of Undetectable hs Troponin T in Suspected Acute Coronary Syndrome. *American Journal of Medicine*, *129*(3), 274–282. https://doi.org/10.1016/j.amjmed.2015.10.016

Vande Kopple, W. (1994). Some Characteristics and Functions of Grammatical Subjects in Scientific Discourse. *Written Communication*, *11*, 534–564. https://doi.org/10.1177/0741088394011004004

Vassileva, I. (1998). Who am I/who are we in academic writing?L: A contrastive analysis of authorial presence in English, German, French, Russian and Bulgarian [Article]. *International Journal of Applied Linguistics*, *8*(2), 163–185. https://doi.org/10.1111/j.1473-4192.1998.tb00128.x

Vassileva, I. (2000). *Who is the author? (A contrastive analysis of authorial presence in English, German, French, Russian and Bulgarian academic discourse).*

Verdaguer, I., Laso, N. J., & Salazar, D. (2013). *Biomedical English. A Corpus-Based Approach.* (I. Verdaguer, N. J. Laso, & D. Salazar, Eds.). John Benjamins Publishing Company.

Vergaro, C. (2011). Shades of impersonality: Rhetorical positioning in the academic writing of Italian students of English. *Linguistics and Education*, *22*(2), 118–132. https://doi.org/10.1016/j.linged.2010.11.001

Vetvicka, V., & Vetvickova, J. (2016). Addition of Selenium Improves Immunomodulative Effects of Glucan. *North American Journal of Medical Sciences*, *8*(2), 88–92. https://doi.org/10.4103/1947-2714.177311

Vihla, M. (1999). *Medical writing. Modality in focus (Language and Computers: Studies in Practical Linguistics 28)*. Rodopi, B. V.

Vincent, B. (2013). Investigating academic phraseology through combinations of very frequent words : A methodological exploration. *Journal of English for Academic Purposes*, *12*(1), 44–56. https://doi.org/10.1016/j.jeap.2012.11.007

Vladimirou, D. (2007). I Suggest That We Need More Research: Personal Reference in Linguistics Journal Articles. In C. Gabrielatos, R. Slesser, & J. Unger (Eds.), *Lancaster University Postgraduate Conference in Linguistics and Language Teaching* (pp. 139–157). Lancaster University Department of Linguistics and English Language.

von Gottberg, A., de Gouveia, L., Tempia, S., Quan, V., Meiring, S., von Mollendorf, C., Madhi, S. A., Zell, E. R., Verani, J. R., O'Brien, K. L., Whitney, C. G., Klugman, K. P., & Cohen, C. (2014). Effects of Vaccination on Invasive Pneumococcal Disease in South Africa. *New England Journal of Medicine*, *371*(20), 1889–1899. https://doi.org/10.1056/NEJMoa1401914

Walker, N., Howe, C., Glover, M., McRobbie, H., Barnes, J., Nosa, V., Parag, V., Bassett, B., & Bullen, C. (2015). Cytisine versus nicotine for smoking cessation. *Obstetrical and Gynecological Survey*, *70*(4), 255–256. https://doi.org/10.1097/01.ogx.0000464934.19513.05

Wanner, Anja. (2009). *Deconstructing the English passive* (1st ed.) [Book]. Mouton de Gruyter. https://doi.org/10.1515/9783110199215

Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business School Press.

Widdowson, H. G. (1994). The Ownership of English. *Tesol Quarterly*, *28*(2), 377–389. https://doi.org/10.2307/3587438

Widdowson, H. G. (2002). Discourse Analysis: A Critical View. In M. Toolan (Ed.), *Critical Discourse Analysis: Critical Concepts in Linguistics.* (pp. 3131–3147). Routledge.

Wilson, E. O. (2004). *On Human Nature* (2nd edn.). Harvard University Press.

Wiraszka, L. (2023). Active verbs with inanimate, text-denoting subjects in Polish and English abstracts of research articles in linguistics [Article]. *Poznan Studies in Contemporary Linguistics*, *59*(1), 217–237. https://doi.org/10.1515/psicl-2022-2003

Wiraszka, Ł. (2023). Active verbs with inanimate, text-denoting subjects in Polish and English abstracts of research articles in linguistics. *Poznan Studies in Contemporary Linguistics*, *59*(1), 217–237. https://doi.org/10.1515/psicl-2022-2003

Wise, L. (2005). Blogs versus discussion forums in postgraduate on line continuing medical education. *BlogTalk Conference Paper*.

Xiao, R. (2003). Use of parallel and comparable corpora in language study. *English Education in China*, *1*.

Yakhontova, T. (2006). Cultural and disciplinary variation in academic discourse: The issue of influencing factors. *Journal of English for Academic Purposes*, *5*, 153–167. https://doi.org/10.1016/j.jeap.2006.03.002

Yamamoto, M. (2006). *Agency and Impersonality* (Vol. 78). John Benjamins Publishing Company. https://doi.org/10.1075/slcs.78

Yore, L. D., Florence, M. K., Pearson, T. W., & Weaver, A. J. (2006). Written Discourse in Scientific Communities: A conversation with two scientists about their views of science, use of language, role of writing in doing science, and compatibility between

their epistemic views and language. *International Journal of Science Education*, *28*(2–3), 109–141. https://doi.org/10.1080/09500690500336601

Zhao, C. G. (2013). Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric [Article]. *Language Testing*, *30*(2), 201–230. https://doi.org/10.1177/0265532212456965