A CORPUS-BASED STUDY OF THE PHRASEOLOGICAL BEHAVIOUR OF ABSTRACT NOUNS IN MEDICAL ENGLISH –A NEEDS ANALYSIS OF A SPANISH MEDICAL COMMUNITY

Tesi doctoral presentada per

Natàlia Judith Laso Martín

com a requeriment per a l'obtenció del títol de

Doctora per la Universitat de Barcelona

Programa de Doctorat: Nivells d'anàlisi a la llengua espanyola
Bienni: 2000-2002
Departament de Filologia Anglesa i Alemanya
Universitat de Barcelona
2009

Directores: **Dra. Anna Poch Higueras** (juliol 2005-juny 2008)
Dra. Isabel Verdaguer Clavera (juny 2008-octubre 2008)

A mi padre, por ser el patrón de cualquier viaje
A mi madre, por enseñarme la vida secreta de las palabras
A mi hermana, por ser la compañera de todas mis travesías

In Memoriam

O Captain! my Captain! our fearful trip is done,
The ship has weather'd every rack, the prize we sought is won,
The port is near, the bells I hear, the people all exulting,
While follow eyes the steady keel, the vessel grim and daring:
But O heart! heart! heart!
O the bleeding drops of red,
Where on the deck my Captain lies,
Fallen cold and dead.

O Captain! my Captain! rise up and hear the bells;
Rise up –for you the flag is flung- for you the bugle trills,
For you bouquets and ribbon'd wreaths –for you the shores a-crowding,
For you they call, the swaying mass, their eager faces turning;
Here Captain! dear father!
This arm beneath your head!
It is some dream that on the deck,
You've fallen cold and dead.

My Captain does not answer, his lips are pale and still,
My father does not feel my arm, he has no pulse nor will,
The ship is anchor'd safe and sound, its voyage closed and done,
From fearful trip the victor ship comes in with object won;
Exult O shores, and ring O bells!
But I with mournful tread,
Walk the deck my Captain lies,
Fallen cold and dead.

(Walt Whitman [1855] *Leaves of grass*)

Conocí a la Dra. Anna Poch en el aňo 1998. Fue mi profesora de *Lexicología y Morfología inglesa* y de *Sintaxis inglesa*. Había finalizado mis estudios de Filología Hispánica y dediqué mi último año de carrera a acabar la mención en Filología Inglesa, que tan cautivada me tenía. Me había quedado prendada de las clases de *Lengua española* y *Sintaxis española* de la profesora Marisa Santiago y conocer a la Dra. Poch y tener el privilegio de asistir a sus clases de *Lexicología* y *Sintaxis* me adentraron en un campo de investigación que empezó a apasionarme enormemente.

Mis planes tras finalizar la carrera iban encaminados hacia la docencia de español para extranjeros, de ahí mi determinación de hacer un máster y continuar, a ser posible, con mi beca en la EIM, lugar en el que me sentía como en casa. La Dra. Poch fue la artífice de que mis planes iniciales se esfumasen. Me propuso presentarme a una beca del Ministerio e iniciar un doctorado. Casi sin darme cuenta me vi envuelta en la rueda del mundo académico. La beca nunca se materializó pero ella, que siempre contaba con un plan B (o un plan C: "*tu no et preocupis, Judith, que tinc moltes alternatives per a tu*"), me animó a que me presentase a una plaza del Departamento de Filología Inglesa y Alemana y desde entonces he tenido el privilegio y la satisfacción de ir siempre de su mano.

La recuerdo firme y determinada en el ámbito profesional ("*a mi em toca fer de poli dolent, Judith, pero ja entendràs que ho faig pel teu bé*", me decía) y a la vez maternal y preocupada por velar por los suyos en su faceta más personal. Siempre sabía qué hacer en cada momento, no daba puntada sin hilo y, como siempre ha comentado la Dra. Verdaguer, "*hem de fer allò que diu l'Anna, perquè l'Anna sempre dóna bons consells*", Y es cierto, los daba. Por eso, hoy me siento totalmente huérfana porque ahora no la tengo para preguntarle tantas y tantas cuestiones que siempre le consultaba y para las que siempre tenía el teléfono de su casa y su propio hogar disponible. Me recibió muchísimas veces en su salón para animarme cuando flaqueaba y me desanimaba pensando que no acabaría nunca esta tesis, o cuando mis problemas personales, tan acuciantes desde noviembre de 2006, me hacían tirar la toalla. Su casa era un remanso de paz entre tanta zozobra. Solía decir que la casa de uno tenía que ser un espacio vivido y que invitase a los demás a serenarse. También era cierto: siempre salí de esa casa con ilusiones renovadas.

Hoy, pues, estoy triste porque no puedo volver a ese puerto donde atracaban todas mis mareas y porque siento que no le cumplí a tiempo la promesa que le hice de acabar pronto esta tesis. Lo siento, Anna. Lo siento tanto. El 25 de febrero de 2008 te prometí que trabajaría duro en la redacción final de mi trabajo y, meses después, el 5 de junio, tras ese abrazo que me diste al despedirnos, te dije que en julio te traería toda la tesis y que lo celebraríamos juntas. Guardo ese abrazo en mi corazón y también la pena honda de no haberte acabado esto a tiempo, pero espero que puedas perdonarme y que allá donde estés puedas ver que llegué (llegamos) finalmente a puerto.

Gracias por los consejos, por las reprimendas a tiempo, por los silencios, por las conversaciones sobre Harry Potter y los *whodunnits* de H. Mankel, por la paciencia, por enviarme a ese “retiro espiritual” (Birmingham) que resultó ser tan fructífero, por permitirme formar parte de un grupo de investigación tan humano, por darme la oportunidad de conocer a gente tan maravillosa y tan importante en mi vida como la Prof. Eva Giménez y la Dra. Isabel Verdaguer, por aceptar dirigir mi tesis y por la honradez con la que siempre me obsequiaste. GRACIAS, en definitiva, por capitanear todas mis travesías profesionales. Te estoy echando tanto de menos… Nos quedaron pendientes una infinidad de proyectos: modificar y publicar un *Text-Guia* de *Lexicología y Morfología Inglesa*, visitar Birmingham juntas y alquilar un coche para explorar sus alrededores, seguir analizando *adjuncts*, materializar esa base de datos léxica que estabas ideando, probar esa famosa crema de zanahorias con agua de Vichy y … un largo etcétera. Nos quedan pendientes, Anna. Espero que algún día podamos llevarlos a cabo juntas.

Te quiero mucho, Capitán, Mi Capitán

Birmingham, 13 de junio de 2008

Acknowledgements

A number of people have contributed to the development of this thesis project, which is the result of a seven-year research period. I would like to take this opportunity to thank each of them for being most understanding and supportive with my ups and downs on finalising this dissertation. Needless to say that this thesis would not have come to light without their invaluable help and encouragement. I do hope life will give me the chance someday to “balance the scales”.

First and foremost, I owe a debt of gratitude to my former supervisor, Dr. Anna Poch, who has been my travelling companion all through these years. I have been extremely fortunate to benefit from her research guidance, valuable advice and unconditional support. Dr. Poch was responsible for embarking me on this thesis journey and for guiding me to this shore. Special thanks must also go to her daughters: Patrícia and Sílvia Bofill, whose assistance on technical matters (gràcies mil, Pat i Henri, pel recolzament tècnic i pels dissenys gràfics) and moral support in difficult times has come as a breeze of air. Gràcies per ser com sou i per donar-me l’oportunitat de compartir temps amb l’Anna fins l’últim moment.

Dr. Isabel Verdaguer’s role in making this project come true has also been decisive. It would be unfair to say that she came to my rescue last June. Of course she did, but she has been in the backstage as far as I can remember. Dr. Verdaguer’s research guidance has been indispensable since I first joined the English Department of the University of Barcelona in 2000. I now recall a conversation with Dr. Guàrdia at the photocopy machine of the English section. I was photocopying some handouts for my *English Lexicology and Morphology* course and told him I was about to participate in an International Conference for the first time. To my relief, Dr. Verdaguer was coming with me. “*No need to worry about anything*”, he said. And then he added: “*tens moltíssima sort perquè la Isabel Verdaguer és la Joia de la Corona d’aquest departament*”. He was absolutely right: Dr. Verdaguer is not only an excellent researcher but also, and most importantly, one of the kindest and most honest people I have ever

met. Gràcies, Isabel, per incorporar-me al teu grup de recerca, per compartir viatges, per ser un exemple constant d'honestedat dins i fora de la feina i, sobretot, per endolcir amb poesia la prosa del dia a dia. Gràcies també merescudes a en Martí per ser el company ideal dels nostres viatges de congressos i de sopars de celebració. Gràcies, doncs, a tots dos per fer-m'ho tot tan fàcil.

I would also like to express my heartfelt thanks to Dr. Pilar Zozaya for encouraging me to apply for a post in the English Department at the University of Barcelona; Dr. Pere Guàrdia for his kind words and wise advice; Dr. Mercè Puig for her clear frame of mind in difficult times (gràcies de tot cor per recordar-me què hauria volgut l'Anna quan jo era incapaç de mantenir el cap fred); Dr. Coloma Lleal, Dr. Emília Anglada, Dr. Teresa Espanyol and Dr. Emma Martinell for the doctoral courses I was lucky enough to take as part of the programme "*Nivells d'anàlisi a la llengua espanyola*"; Dr. Roser Estapà for her classes on Spanish syntax and for helping me out with various PhD issues; Prof. Marisa Santiago, whose lectures on syntax awakened my interest in linguistics in general and syntax in particular; Dr. Mar Cruz, whose classes on lexicology and morphology were highly influential in my future research on lexical studies; Dr. Mª Luz Celaya for being most supportive (gracias por las clases de *Descriptive Grammar of English* y *Second Language Acquisition*, por aceptar ser mi tutora del Máster de Profesorado Novel, y, sobre todo, gracias por el apoyo prestado en los últimos meses: gracias por los cálidos correos electrónicos enviados a la gélida Inglaterra y por estar tan pendiente de mí en momentos realmente duros. Gracias de corazón); Dr. Joan Carles Mora for encouraging me to carry on with my thesis. Many thanks too for being so thoughtful of assistant professors' needs.

My gratitude is further extended to the Head of the English Department, Dr. Bill Phillips; the Head of the English Language Section, Dr. Brian Mott, and my colleagues at the English Department, who have been a great help too along the building up of my thesis. Special thanks to Dr. Joseph Hilferty, who kindly accepted to read through the very first draft of my work and persuaded me not to give up; to Dr. Rodrigo Andrés, Dr. Cristina Alsina and Dr. Teresa Requena for encouraging me to go abroad and make the most of that opportunity, and to my officemates, Prof. Eva Giménez, Prof. Mònica Miravet and Dr. John Stone, for coping with me.

I am also grateful to the University of Barcelona and, most specifically, to the School of Philology and Dr. Montserrat Camps, former Dean of that School, for granting me some funding to complete this project at the University of Birmingham. Here, I would also like to thank the *Lexicom* team, Sue Atkins, Michael Rundell and Adam Kilgarriff, who suggested Birmingham and, particularly, the *Centre for Advanced Research in English* (*CARE*) as a suitable destination to conduct my research. To this respect, my attendance to the *Lexicom Masterclass* held in Brno (Masaryk University) in 2007 was also highly beneficial. Many thanks are due to Michael Rundell for putting me in contact with Dr. Christopher Gledhill from the Université Marc Bloch in Strasbourg. Dr. Gledhill was immensely helpful. His work on medical discourse in English has been hugely influential in this dissertation, as the copious references to his research throughout this thesis reveal.

My academic visit to the University of Birmingham has been absolutely crucial for the completion of this dissertation. To the English Department and the *Centre for Advanced Research in English* (*CARE*) of the University of Birmingham, I owe the greatest debt. I must particularly mention Dr. Pernilla Danielsson, Director of the *Centre for Corpus Research*, who read and commented on early versions of this thesis. Her constructive comments and advice have been extremely helpful. Dr. Danielsson was my supervisor in Birmingham and I must say that not only did she supervise the contents and editing of my work but also provided moral support when it was most needed. Thank you ever so much for allowing me to benefit from your talented mind, generosity of spirit, precious time and challenging comments. On a more personal note, I would also like to thank Dr. Danielsson for welcoming me on Spring Avon Croft and, most specifically, for her sharing of the heart. Thanks for the Taste Festival, Amelie’s birthday party, the barbeque at the Kings Head, popcorn and bedtime stories with the little princess, Cannon Hill Park, the Waitrose and an endless list of unforgettable warm moments in chilly

Birmingham. Thanks for making my pain and sorrow bearable as well. Here, my thanks are certainly extended to Amelie. “*Oh dear*”, as she would say, no doubt that her coffee parties and small talk made my thesis confinement more endurable and entertaining.

During my time as a research associate at the *CARE* (University of Birmingham) I was allowed to make use of the 450 million-word Bank of English as well as other corpus tools such as *WordSmith*. Besides, I was given the opportunity to audit postgraduate classes and attend weekly staff and postgraduate seminars.

Invaluable advice on how to present and analyse some of the data in this dissertation was also provided by Dr. Suganthi John, professor of the English Department at the University of Birmingham. I am deeply indebted to Dr. John for taking an interest in my thesis and reading several sections of it, hence providing extremely valuable guidance and useful remarks. Thank you for your time, your generosity, for giving me the opportunity to share your knowledge and for putting me in touch with Dr. Crayton Walker, whose work on collocations has also been very profitable.

I am also extremely grateful to the following people at the University of Birmingham for their assistance in providing bibliographical references and for their interest in my work. I must particularly mention Dr. Susan Hunston, Head of the Department of English, at the University of Birmingham, for her useful suggestions on Chapters 4 and 5 and for her revision of my references. It was a sheer pleasure and an honour to meet someone whose work I have greatly admired for so long. Dr. Anna Poch used to say that outstanding people are always very humble and welcoming. I can tell Dr. Hunston is a good example of both qualities. Dr. Rosamund Moon was also very helpful in providing me offprints as well as recently published articles on phraseology in English. All faults are my own responsibility.

I also have to acknowledge Michelle Devereaux from the *Centre for Advanced Research in English* for her continuing support with everyday matters during my stay in Birmingham. She is an absolute star. Thanks for making my academic life in Birmingham so straightforward.

I would also like to say a big thank you to my landlord there, Charles Owen, and his wife, Yvonne Owen. They are some of the kindest souls anybody could ever meet. Thanks for opening the doors of Barnsley Road and for giving me the opportunity to experience family life in Birmingham. You two offered me a cosy place to stay in, “peaceful evenings”, as Charles would say, to concentrate on my work and a warm atmosphere that made me feel at home. Many thanks for being my British family, for the barbeque, the theatre evening and the thought-provoking chat on multiculturalism on our way back home, the lovely picnic in Compton Verney, the delicious ‘Costorphine Cream’ (among other superb recipes by Yvonne), the small talk about Birmingham over a cup of coffee, Scotland, blackbirds, monkey puzzles (I’ll never forget that tree, I swear), *Dr. Who*, David Lodge, gardening, joists and other issues that made my time in England an unforgettable ‘been to’ experience. Thanks for introducing me to some of your distinguished colleagues, Wolfgang Teubert, Michael Toolan, Malcolm Coulthard and Carmen Rosa Caldas-Coulthard.

Back to Barcelona. Many special thanks must go to the Escola d’Idiomes Moderns (EIM) of the University of Barcelona. Many kind souls there have helped me a great deal to fulfil this academic task. I must particularly mention Dr. Philip Banks, former Academic Director of the EIM, whose work has served as an example to me along the way. Thanks for giving me the opportunity to work in the EIM, which has turned out to be a great learning, as well as extremely rewarding, experience in my academic career. First, I was lucky enough to hold a two-year grant, which gave me lots of teaching and administrative practice. I certainly have learned a lot from Dr. Banks’ expertise and would like to express my great admiration for him. I am also deeply grateful to the current Academic Director of the EIM, Paul Ambrose, as well as to the English Co-ordinators, Jonathan Gregg and Richard Blair, for helping me out with the *Servei Extern d’Idiomes* (*SEI*) while being abroad. To this respect, I must also thank SEI teachers for their patience and understanding, in particular, in the last few months. Your good work has certainly made my life much easier.

Thanks for your encouraging emails as well. Many thanks too to EIM teachers and, in particular, to Rosemary Hancock for her interest and support all through the years.

As the current co-ordinator of the Servei Extern d'Idiomes (SEI), I would like to express my gratitude to late Patrick Moss, who introduced me to the SEI and let me teach some classes for the Personal d'Administració i Serveis (PAS) of the University of Barcelona. PAS students are a delight to teach. Thanks are also extended to the former co-ordinator of the SEI, Peter Loveday, who recommended me for the post.

My colleagues in the EIM administration also deserve special mention. Roser Nicolás is one of the most understanding people I have ever met. Without her help and encouragement to carry on with my research, I would not have been able to complete this thesis. Gràcies per facilitar-me tant les coses en els darrers mesos: et respecto i t'estimo molt. The following people have also been fundamental, especially when I felt overburdened: Yolanda Murcia, Maite Portabales, Pilar Fernández, Xavier Navarro, David Canal, Marta Costoya, Laura Chaparro and Merche Zamora. Què seria de mi sense vosaltres? Gràcies per fer-me costat, per animar-me en moments tristos, pel gran hit nadalenc de desembre de 2006, per les festes, els esmorzars a Gràcia i a la Central, el support al SEI i, sobretot, per la paciència! I must also mention Dolors Buetas, Carme Serra and Alicia Paris (us trobo molt a faltar a l'EIM!) Us estimo molt i sou part de la meva família acadèmica.

I am equally grateful to the following people for their assistance in administrative tasks: Montse Bartomeus (gràcies per fer fàcil allò que sembla tan difícil i pel teu somriure), Gemma Barea (gràcies per l'eficiència i per la teva delicadesa en situacions difícils), Mercedes Padial (gracias por estar ahí desde el principio y por convertir en arte la prosa diaria), Olga Borderas (gracias por los correos electrónicos que contribuían a hacerme más llevadera la distancia y por todas las gestiones), Josefina Cordero (gracias por tu trabajo y por compartir complicidades) and Lali Colom (gràcies per tota la feinada i per la paciència amb la reserva d'aules). Within the administration of the University of

Barcelona, there are hundreds of people that have been extremely helpful in the last seven years. Among them, I would like to mention Marisa Royo, Dolors Güell and their colleagues at the Formació Corporativa of the University of Barcelona. Thanks for your patience with my "absence" last year and for your help with the organisation of the PAS courses.

I believe that many things in my life, such as my job as an assistant teacher in the School of Philology (UB), have come about quite by chance, and when you do not expect something to happen, you do not really know how to react in front of such situations. I do think this is one of the best parts of my teaching job: I find it really challenging! I love teaching. As B. Shaw would say, "*Those who can, do and those who can't, teach!*" I must definitely be a member of the latter because, honestly, I cannot see myself *doing* anything better than teaching. Thus, I would also like to thank the various groups of students I have had all through these years at the English Department of the University of Barcelona (*English Language I*, *English Lexicology and Morphology*, *Descriptive Grammar of English III*), at the EIM (levels 3 and 4, and most especially, my beloved Saturday morning groups!), at the SEI and my online courses at the Universitat Oberta de Catalunya (UOC) for contributing to this thesis with their comments and questions, which have been extremely beneficial in my current research. They are certainly a constant source of inspiration. Many thanks must also go to the twenty-four informants from three Spanish hospitals (i.e. *Hospital Universitario Puerta del Hierro*, *Complejo Hospitalario Universitario de Albacete* and *Hospital de Navarra*) that willingly participated in this study and showed an interest in my thesis. Without their involvement this research would not have been possible.

On a more personal level, I must express my eternal gratitude to Eva Giménez, Elisa Rosado, Eli Comelles, Mercè González and Sílvia Miquel for providing me invaluable help and support in addition to their friendship. Eva Giménez has been my officemate all through this period and one of the most delightful friends I have ever had. L'Eva té un talent especial per a la docència i un cap especialment dotat per a les idees brillants. És un plaer poder treballar amb ella, compartir complicitats, tasques, congressos i projectes i veure créixer als

meus estimats nebots (l'Aleix, el Joel i la Zoe, que està en camí en el moment d'escriure aquestes línies). Gràcies per ser la meva germana del cor i la meva companya incondicional en aquesta carrera d'obstacles.

Dr. Elisa Rosado has also made every endeavour to help me out with this thesis. She was actually the first person who read the very first draft of some parts of this thesis and gave me all sorts of useful, constructive feedback on my research. We were colleagues for a while and I must say that I am now enjoying a post that should have been hers in the first place (sabes que llevo esa espinita clavada). Anyway, life has luckily given me the opportunity to witness her academic success, which fills me with tremendous pride. Gracias, lady Eliza, por tanto… tú sabes… por los mensajes de "*martes: sol y nubes, nubes rosa y gris claro y blanco*", por ser y estar, por compartir y, sobre todo, por acompañarme en la conquista diaria de la propia ración de aire.

Agraeixo també haver conegut a l'Elisabet Comelles. N'he après tant d'ella en aquests anys! Té una fortalesa que mou muntanyes i és un d'aquests esperits dotats amb el talent de convertir en or allò que passa per les seves mans. Ha estat un plaer compartir beca a l'EIM amb ella i amb la Sílvia Miquel, *Miss Lunatic*. Totes dues són art en estat pur i em sento molt afortunada d'haver-les conegut. Espero continuar gaudint de les nostres trobades culturals i lúdico-festives i dono gràcies també per tornar a coincidir amb l'Eli Comelles com a companyes de feina: n'haig d'aprendre molt del seu coneixement lexicogràfic. Serà, doncs, un altre honor professional. Gràcies mil també a la Mercè González pel seu suport incondicional no només en la confecció de la meva tesi sinó en moments personals de "forta marejada". It has also been a pleasure to meet such a gifted mind. Thanks for teaching me the important things in life and for sharing your metaphors: qui s'afanya, com tu, a buscar la perla amagada en l'interior de cada individu és senzillament una gran persona.

Last, but by no means least, I must thank my other close friends and family for being next to me through thick and thin. Agradezco públicamente a la vida el haberme concedido el lujo de contar con un "círculo de confianza" que ha sabido darme un consejo, una palabra de aliento en momentos de

desesperación, una opinión, una crítica constructiva, un saludo, una mirada de complicidad, un llanto y una risa compartidas, y en definitiva, todos los ingredientes necesarios para que estos folios pudiesen ser presentados algún día.

Gratitud inmensa para Ana Belén Palomino por sobrevivir a mi ignorancia estadística y ayudarme con los gráficos y los análisis matemáticos. Gracias, tú sabes, por ser el apoyo de mi madre en momentos que tanto te hemos necesitado. Eres amiga, hermana y compañera de viajes. Gracias, pues, también por ser y estar. Mi otra hermana del alma, Sara Sánchez, también ha sido un apoyo imprescindible. Gracias a ti y a Sara Sonsoles y Lucas Brito (sobrinos del corazón) por traernos siempre la brisa del Mediterráneo y por regalarnos alegrías con las que olvidar los estragos del oleaje. Gracias igualmente a Miguel Brito por animarme con la tesis, por interesarse por mi trabajo y, cómo no, por proporcionarme a un gran número de informantes del Hospital Universitario Puerta del Hierro y del Complejo Hospitalario Universitario de Albacete. Eres sencillamente un sol.

Mis amigos Mònica Grau, Marc Masa, Sonia López y Miguel López también me han ayudado a sobrellevar mi "reclusión" en Birmingham y los momentos de crisis con sus correos, con *El juego del ángel* el 23 de abril de 2008, con nuestras escapadas "rurales" y con conversaciones en múltiples cenas. Gracias también por el soporte técnico (Mònica & Marc) y por vuestro apoyo, especialmente necesario desde noviembre de 2006 (gracias, Marc, por tus palabras en El Serrat antes de mi partida a UK, las cuales me animaron a seguir con el proyecto).

Y, ahora sí, finalmente quiero dar las gracias en mi lengua materna a mi familia. Les he dejado para el final pero ellos saben que son principio y fin y el motor de todo cuanto hago. Gracias a mi cuñado, Fer, el "*boss*", por apoyarme en mi trabajo y por creer ciegamente en que esta tesis se acabaría pronto. Sabes que tú y yo trabajamos conjuntamente en la realización de muchos proyectos: nos propusimos dejar el tabaco (tú) y la coca-cola (yo) y lo logramos; prometimos que esta tesis se terminaba en 2008 y lo hemos conseguido. Ahora tenemos otro proyecto por delante. Estoy convencida de que los dos, junto con nuestro "círculo de confianza", "*volaremos tan alto, tan alto, que le daremos a la caza alcance*".

Gracias a mi Curro porque él sabe que esta tesis está dedicada a todo su esfuerzo de una vida entera. Gracias por ser padre y maestro y por compartir complicidades durante tanto tiempo.

Gracias a mi madre por su fe constante en mi trabajo y por seguir acompañándome en esta etapa de mi vida. Gracias por ser madre y vecina y por enseñarme la importancia de la vida interior.

Gracias a mi tatah porque juntas somos "hogar". Decía Mónica en *La Reina de las Nieves* de C. M. Gaite que "*vive uno como montado en una noria que cada vez gira más deprisa. Da miedo. Y vértigo. Sobre todo, vértigo*". Es verdad, pero yo soy muy afortunada por haber contado con tu compañía en este y todos los viajes (reales e imaginarios) que hemos realizado juntas. Gracias, sister, por ser lugar de partida y lugar de llegada.

Os quiero mucho. A big THANK YOU to you all. This thesis owes EVERYTHING to each of you.

Barcelona, August 22 2008

CONTENTS

CHAPTER 1: INTRODUCTION 1

- 1.1 Research questions underlying this thesis project 3
- 1.2 Outline of the thesis structure 4

CHAPTER 2: LITERATURE REVIEW AND THEORETICAL FRAMEWORK 7

- 2.1 Corpus linguistics as a methodological tool to explore language 7
  - 2.1.1 Corpus-based research in language description 11
  - 2.1.2 Corpus-based research in dictionary-making 11
  - 2.1.3 Corpus-based research for the production of grammar reference books 12
  - 2.1.4 Corpus-based research in language acquisition and language instruction 13
- 2.2 Lexico-grammar and the association between lexis and syntax in descriptive linguistics 15
- 2.3 Sinclair's notion of *collocation* 19
- 2.4 Phraseology in English Research Writing 28
  - 2.4.1 Phraseology in specialised registers 32
  - 2.4.2 Phraseology in Foreign Language Learning 37

CHAPTER 3: METHODOLOGY 44

- 3.1 Corpus linguistics as a methodology. The rationale for using a corpus-based approach 44
- 3.2 Corpus data 46
- 3.3 The Survey 55
- 3.4 Informants 56
- 3.5 Worksheet of exercises 73
- 3.6 Summing up 76

CHAPTER 4: DATA ANALYSIS: THE COLLOCATIONAL PATTERNING OF ABSTRACT NOUNS IN THE *HEALTH SCIENCE CORPUS* (*HSC*) 78
4.1 Introduction to the Chapter 78
4.2 The collocational patterning of abstract nouns in the *HSC*. Some preliminary considerations 79
4.3 The patterns of the noun *conclusion* 82
a) The noun *conclusion* as an introductory item 83
b) verb + *conclusion* 83
c) *conclusion* + preposition 91
d) preposition + *conclusion* 93
e) adjective + *conclusion* 94
4.4 The patterns of the noun *agreement* 98
a) verb + *agreement* 98
b) adjective + *agreement* 104
c) *agreement* + preposition 106
d) preposition + *agreement* 108
e) The lexical bundle *there is* + (adjective) *agreement* 110
4.5 The patterns of the noun *comparison* 110
a) verb + *comparison* 111
b) adjective + *comparison* 114
c) *comparison* + preposition 118
d) preposition + comparison 122
4.6 The patterns of the noun *contribution* 129
a) verb + *contribution* (*to*) 130
b) adjective + *contribution* 137
c) noun + *contribution* 140
d) *contribution* + preposition 141
4.7 The patterns of the noun *decision* 144
a) verb + decision 144
b) *decision* + noun 147
c) *decision* + preposition 152
d) *decision* + *to*-infinitive 154
e) adjective + decision 155
4.8 Discussion of overall results 156

CHAPTER 5: INFORMANTS' DATA ANALYSIS 163

5.1 Results obtained from the Worksheet of exercises 164
  5.1.1 Fill in the Gaps Exercise 164
    a) Group 1: restricted verb collocates 165
    b) Group 2: free verb collocates 174
    c) Group 3: lexical bundles 178
  5.1.2 Translation Exercise (Spanish into English) 185
    a) Informants' command of the lexical bundles *there* + *be* + (adjective) *agreement* and *in conclusion* 186
    b) Informants' use of the pattern restricted verb + *comparison* 192
    c) Informants' use of the verb *contribute* vs. the noun *contribution* 195
    d) Informants' command of the patterns associated with the noun *conclusion* 200
    e) Informants' knowledge of the compound *decision making* 206
  5.1.3 Matching Exercise (adjective + abstract noun) 209
  5.1.4 Matching Exercise (verb + abstract noun) 221
    a) Group A: restricted verb + abstract noun collocations 224
    b) Group B: restricted verb + nouns characteristic of the medical discourse 225
    c) Group C: restricted verb + general English nouns 230
5.2 Summing up of findings 236
5.3 Informants' self-assessment (Survey Content-Based Questions) 240

CHAPTER 6: CONCLUDING REMARKS 250
  6.1 Referring back to the research questions underlying this thesis project 250
  6.2 Summing up of main findings 251
  6.3 Implications of the study 259
  6.4 New research directions 266

REFERENCES 269

APPENDICES 289
- Appendix 1 Survey 289
- Appendix 2 Survey Question 10: Informants’ time estimates in completing the questionnaire 292
- Appendix 3 Consent form 293
- Appendix 4 Worksheet of Exercises 294
- Appendix 5 Typology of Exercises and Aims 297
- Appendix 6 Informants’ translations to sentences in Exercise 3 299
- Appendix 7 Informants’ verb-noun collocations in Exercise 4: Matching Exercise (verb + abstract noun) 313
- Appendix 8 Survey Content-Based Question 8: Informants’ self-evaluation of their main problem areas when writing in English 319

LIST OF FIGURES

Figure 1 WordList of the *HSC* provided by *WordSmith Tools* 48

Figure 2 Concordance lines for the node word *contribution* provided by *WordSmith Tools* 50

Figure 3 Source text of a concordance line of the node word *contribution* displayed by *WordSmith Tools* 51

Figure 4 Clusters of the node word *agreement* as displayed by *WordSmith Tools* 52

Figure 5 Collocates of the node word *comparison** 53

Figure 6 Sex group of the informants 60

Figure 7 Age group of the informants 60

Figure 8 Informants’ centre of studies 61

Figure 9 Number of years studying English 63

Figure 10 Informants’ experience working in their current hospital 65

Figure 11 List of concordances of the noun *conclusion* (extracted from *WordSmith Tools*) 82

Figure 12 **Verb + *conclusion*** collocates in the *HSC* 84

Figure 13 Percentages of occurrences of restricted collocates of the type **Verb + *conclusion*** 85

Figure 14 Percentages of occurrences of free collocates of the type **Verb + *conclusion*** 86

Figure 15 Overall number of occurrences of **draw + *conclusion*** in the *HSC* 87

Figure 16 Overall number of occurrences of *arrive at* / *come to* + *conclusion* in the *HSC* 89

Figure 17 Random selection of the search word *conclusion* from the *British National Corpus* 89

Figure 18 Randomly selected occurrences of the construction *in conclusion* in the *HSC* 94

Figure 19 Number of occurrences of restricted and free collocates in combination with the noun *agreement* 99

Figure 20 *agree* vs. *be in agreement* 100

Figure 21 Semantic connotations of *agree* vs. *be in agreement* 104

Figure 22 Frequency of occurrence of **adjective + *agreement*** collocations in the *HSC* 106

Figure 23 Collocational patterns of the noun *comparison* 112
Figure 24 Semantic fields of free collocates of the type **Verb + *comparison*** 113
Figure 25 Distribution of evaluative premodifiers of the abstract noun *comparison* 115
Figure 26 Distribution of time / descriptive premodifiers of the abstract noun *comparison* 115
Figure 27 Distribution of relational premodifiers of the abstract noun *comparison* 116
Figure 28 Distribution of topical premodifiers of the abstract noun *comparison* 116
Figure 29 Concordance lines to the node word *comparison* 118
Figure 30 Prepositions following the noun *comparison* 119
Figure 31 Overall number of occurrences of the pattern **preposition + *comparison*** (lexical bundle) 125
Figure 32 Realisations of the lexical bundle *in comparison* 125
Figure 33 Realisations of the lexical bundle *for comparison* 126
Figure 34 Realisations of the lexical bundle *by comparison* 126
Figure 35 Overview of the node word *contribute* (extracted from *WordNet 2.1*) 132
Figure 36 Active and passive constructions with the pattern **restricted verb + *contribution*** 134
Figure 37 Active and passive constructions with the pattern **free verb + *contribution*** 134
Figure 38 Restricted and free collocates of the pattern **V + *contribution*** grouped into semantic fields and described as part of a logical sequence of events 136
Figure 39 Concordance lines of the node word *contribution* 138
Figure 40 Concordance lines of the node word *contribute* 139
Figure 41 Prepositions following the noun *contribution* 141
Figure 42 Syntactic structures related to restricted collocates of the type **V + *decision*** 145
Figure 43 Syntactic and semantic environment of **free verb collocates + *decision*** found in the *HSC* (Group 1) 146

Figure 44 Syntactic and semantic environment of **free verb collocates + *decision*** found in the *HSC* (Group 2) 147
Figure 45 Tree diagram for the compound *decision models* 149
Figure 46 Tree diagram for the compound *decision making* 150
Figure 47 Tree diagram for the compound decision makers 151
Figure 48 Examples of *to*-infinitive clauses following *decision* and its corresponding verb *decide* 155
Figure 49 Frequency of abstract nouns in comparison with their corresponding verbs 159
Figure 50 Collocations provided by informants in Sentence 1 (Exercise 2) 166
Figure 51 Collocations provided by informants in Sentence 2 (Exercise 2) 167
Figure 52 Collocations provided by informants in Sentence 5 (Exercise 2) 168
Figure 53 Collocations provided by informants in Sentence 6 (Exercise 2) 170
Figure 54 Collocations provided by informants in Sentence 7 (Exercise 2) 171
Figure 55 Collocations provided by informants in Sentence 8 (Exercise 2) 171
Figure 56 Collocations provided by informants in Sentence 9 (Exercise 2) 172
Figure 57 Collocations provided by informants in Sentence 15 (Exercise 2) 173
Figure 58 Collocations provided by informants in Sentence 3 (Exercise 2) 175
Figure 59 Collocations provided by informants in Sentence 11 (Exercise 2) 175
Figure 60 Collocations provided by informants in Sentence 13 (Exercise 2) 177
Figure 61 Collocations provided by informants in Sentence 14 (Exercise 2) 178
Figure 62 Collocations provided by informants in Sentence 4 (Exercise 2) 179
Figure 63 Collocations provided by informants in Sentence 12 (Exercise 2) 181
Figure 64 Collocations provided by informants in Sentence 10 (Exercise 2) 182
Figure 65 Informants' writing process 242
Figure 66 Survey Question 10: Informants' time estimates in completing the questionnaire (Appendix 2) 292
Figure 67 ***make* + noun collocations** provided by informants 313
Figure 68 ***draw* + noun collocations** provided by informants 314
Figure 69 ***reach* + noun collocations** provided by informants 315
Figure 70 ***lead to* + noun collocations** provided by informants 316
Figure 71 ***take* + noun collocations** provided by informants 317
Figure 72 ***do* + noun collocations** provided by informants 318
Figure 73 Use of specific (medical) terminology when writing in English 319

Figure 74 Use of general terms when writing in English 320
Figure 75 Distinction between similar pairs of words when writing in English 321
Figure 76 Paragraph distribution when writing in English 322
Figure 77 Grammatical rules when constructing sentences in English 323

LIST OF TABLES

Table 1 Informants' Identification codes and hospital 56

Table 2 Informants' centre of studies and workplace 62

Table 3 Informants' stays abroad 64

Table 4 Informants' medical specialities 66

Table 5 Informants' training courses 67

Table 6 Frequency of English journal reading 69

Table 7 Frequent NPs as Prepositional complements of the ***conclusion*** **+ preposition** collocational pattern 91

Table 8 Predicative adjectives in combination with the abstract noun *conclusion* 95

Table 9 Main colligational and collocational patterns of the abstract noun *conclusion* found in the *HSC* 97

Table 10 Delexical and primary uses of the verbs *reach*, *find*, *lead to* and *make* 99

Table 11 Semantic relations expressed by the different prepositions postmodifying *agreement* 107

Table 12 Classification of modifiers of the abstract noun *comparison* (number of occurrence sin the *HSC*) 115

Table 13 Random selection of right-sorted concordance lines of the pattern ***comparison*** **+ preposition** 119-120

Table 14 List of 2, 3 and 4-word lexical bundles with the structure **preposition + noun phrase fragment (*comparison*)** 123

Table 15 Discourse functions of the pattern **preposition +** ***comparison*** 124

Table 16 Context distribution of the *in comparison* bundle in the *HSC* 127

Table 17 Context distribution of the *for comparison* bundle in the *HSC* 128

Table 18 Context distribution of the *by comparison* bundle in the *HSC* 128-129

Table 19 Frequency rates of the pattern **free verb +** ***contribution*** 135

Table 20 Frequency rates of **descriptors** and **classifiers +** ***contribution*** 137

Table 21 **Noun +** ***contribution*** collocation pattern 140-141

Table 22 Noun Phrases as Prepositional complements of the sequence *contribution of / to / from* 143

Table 23 Inflected forms of ***decision*** **+ noun** compounds 149

Table 24 *HSC* examples of the pattern ***decision*** **+ preposition** vs.

***decide* + preposition** 153
Table 25 Semantic groupings of the attributive adjectives that collocate with the abstract noun *decision* 156
Table 26 Lexical bundles used to introduce conclusions and summaries in the *HSC* and the informants' data 191
Table 27 Deviant formal realisations of the noun *comparison* provided by informants in Sentences 2 and 6 192
Table 28 **Restricted verb collocates + *conclusion*** in the *HSC* and the informants' data (Sentence 5) 200
Table 29 Semantic classification of the adjectives selected in Exercise 1 211
Table 30 Adjective collocates in combination with *agreement* / *conclusion* / *comparison* / *contribution* / *decision* extracted from the *BNC* 213
Table 31 Adjective collocates in combination with *agreement* / *conclusion* / *comparison* / *contribution* / *decision* extracted from the *HSC* 213
Table 32 Adjective collocates in combination with *agreement* / *conclusion* / *comparison* / *contribution* / *decision* extracted from the informants' sample 214
Table 33 Highest frequency adjective collocates in the *BNC*, the *HSC* and the informants' sample 215
Table 34 Adjective collocates only found in the informants' sample 220
Table 35 Restricted verb-noun collocations (Group A: abstract nouns; Group B: nouns characteristic of the medical discourse; Group C: general English nouns not found in the *HSC*) 223
Table 36 Group A: **restricted verb + abstract noun collocations** 224
Table 37 Most frequent collocates of **verb + *conclusion***, **verb + *agreement*** and **verb + *comparison*** in the *HSC* and the informants' data 224
Table 38 Group B: **restricted verb + noun** characteristic of the medical discourse 226
Table 39 Most frequent collocates of **verb + noun** characteristic of the medical discourse in the *HSC* and the informants' data 226-228
Table 40 Group C: **restricted verb + general English nouns** 231
Table 41 Most frequent **verb + general English nouns** in the *HSC* and the informants' data 232-233
Table 42 Suggested tools and resources 245

CHAPTER 1: INTRODUCTION

Corpus-based studies have drawn the attention to the study of the lexicon as the central principle in language and have also emphasised the interconnections between lexis and syntax (Francis 1993; Hunston & Francis 2000; Wray 2002, among others). Linguistic investigation of naturally-occurring data has revealed that language is organised in terms of a lexico-grammar and, thus, it consists of recurrent patterns of words (Renouf & Sinclair 1991; Sinclair 1991; Altenberg & Tapper 1998; Stubbs 2001). The study of how words are used to make meanings; in other words, how meaning maps onto use, is one of the key concerns in current research in phraseology.

Phraseological empirical studies have confirmed the important role of prefabricated expressions in the textual development of meaning (Gledhill 2000b; Kaszubski 2000) and have also highlighted the need for further research on the phraseological conventions characteristic of specialist genres. As Kaszubski (2000) points out:

> “Word combinations are inextricably related to the layer of style –the appropriateness and/or naturalness of selection and co-occurrence of items, subject to genre-sensitive restrictions and conventions. Thus, in order to compare aspects of lexical use, one is bound to focus attention on phraseology.” (Kaszubski 2000:2)

Specifically, in the field of Language for Specific Purposes (LSP), corpus linguistics has established itself as a fundamental methodological tool for determining the defining linguistic features of different discourse communities (Swales 1990; Lee 2001; Gledhill 2000; Lee & Swales 2006).

It has been long acknowledged (Carter 1998, Williams 1998; Biber 2006; Hyland 2008) that writing a text not only entails the accurate selection of correct terms and grammatical constructions but also a good command of appropriate lexical combinations and phraseological expressions. This assumption becomes especially apparent in scientific discourse, where a precise expression of ideas and description of results is expected. Several scholars (Gledhill 2000; Flowerdew 2003; Hyland 2008) have pointed to the importance of mastering the prototypical formulaic patterns of scientific discourse so as to produce phraseologically competent scientific texts.

Research on specific-domain phraseology has demonstrated that acquiring the appropriate phraseological knowledge (i.e. mastering the prototypical lexico-grammatical patterns in which multiword units occur) is particularly difficult for non-native speakers, who must gain control of the conventions of native-like discourse (Howarth 1996/1998; Wray 1999; Oakey 2002; Williams 2005; Granger & Meunier 2008).

The present dissertation aims to analyse native and non-native speakers' usage of five abstract nouns, i.e. *agreement*, *comparison*, *conclusion*, *contribution* and *decision*, in medical English. In order to explore non-native speakers' lexical and phraseological command of the five selected nouns, a worksheet of four exercises was distributed among a community of twenty-four Spanish doctors, who were practising medicine in three different Spanish hospitals; namely, *Complejo Hospitalario Universitario de Albacete*, *Hospital Universitario Puerta de Hierro* and *Hospital de Navarra* (cf. chapter 3).

With the aim of assessing non-native speakers' phraseological competence regarding the use of abstract nouns, their written production will be compared against a native corpus. Thus, it can be claimed that this investigation also attempts to contribute to the linguistic characterisation of the discourse of medical science. More precisely, this thesis project intends to explore native speakers' prototypical lexico-grammatical patterns around the five abstract nouns under investigation. This analysis is based entirely on corpus evidence, since all collocational patterns discussed have been extracted from the *Health Science Corpus* (*HSC*), which consists of a 4 million word collection of health science (i.e. medicine, biomedicine, biology and biochemistry) texts, specifically compiled for the current research study (cf. section 3.2). The exploration of the collocational behaviour of abstract nouns in medical English will serve as a benchmark against which to measure non-native speakers' production.

1.1 Research questions underlying this thesis project

As already stated, this study mainly aims at identifying non-native speakers' problem areas when using medical English and, in particular, their phraseological competence as far as the usage of abstract nouns is concerned. Throughout this investigation, three basic research questions are pursued, which the planned study will aim to answer:

(i) Are Spanish doctors aware of the particularities of the medical genre?
(ii) What are the main difficulties the phraseological patterns of abstract nouns in medical English pose for the non-native speakers' medical community?
(iii) What are the prototypical usage patterns around the following abstract nouns: *agreement*, *comparison*, *conclusion*, *contribution* and *decision* in medical English?

The first research question addresses the issue of whether Spanish doctors are competent enough in English so as to express themselves in medical literature. With the aim of exploring the way in which non-native speakers make particular collocations for scientific writing, a worksheet of exercises, purposely designed for this thesis project, was distributed among 24 Spanish doctors. The findings obtained from their written production will be compared against native data. This comparison will provide information about their knowledge of the collocational patterns associated with the abstract nouns under study.

The second research question investigates the difficulties posed by the phraseological patterns of abstract nouns in medical genre. The analysis of non-native speakers' production of abstract noun collocations in the exercises will provide information about the problem areas (e.g. overuse and/or underuse of collocational patterns, lexical bundles, and formulaic expressions) Spanish doctors encounter when using abstract nouns, such as *agreement*, *comparison*, *conclusion*, *contribution* and *decision* in scientific articles.

The third research question looks at the collocational patterns of the selected five nouns that typically occur in the health science discourse. In order to answer the first research question; that is, non-native speakers' command of the phraseological behaviour of abstract nouns in medical papers, a native corpus will be explored as a comparison measure. All the observations made concerning native phraseological and collocational usage of abstract nouns will thus serve to determine Spanish doctors' phraseological competence in scientific English.

1.2 Outline of the thesis structure

This thesis consists of six chapters; namely:

a) Chapter 1: "Introduction"
b) Chapter 2: "Literature review and theoretical framework"
c) Chapter 3: "Methodology"
d) Chapter 4: "Data Analysis: the collocational patterning of abstract nouns in the *Health Science Corpus* (*HSC*)"
e) Chapter 5: "Informants' Data Analysis"
f) Chapter 6: "Conclusions, implications and further research"

Chapter 1, the present one, is concerned with three research questions that motivated this thesis project, explaining why the investigation is undertaken and what the expected outcome will be. As mentioned earlier, the planned study is an attempt at comparing the lexical and phraseological command of non-native speakers in the use of five selected English abstract nouns. The expected outcome is a needs analysis of a Spanish medical community as far as their phraseological competence in scientific English is concerned. Finally, the present section (1.2) offers an outline of the structure of the actual research study.

Chapter 2, "Literature review and theoretical framework", focuses on earlier studies in the field of corpus linguistics as a methodological tool and describes the theoretical foundation of the conducted research. The literature review in chapter 2 starts by describing the wide range of disciplines that can benefit from a corpus-based approach to the study of language. Section 2.2 presents a lexical approach to linguistic analysis adopted in this research. An overview of the notion of *lexico-grammar* and the interrelationship of lexis and syntax in descriptive linguistics is followed by a summary of the different definitions applied to the term *collocation* and Sinclair's "idiom principle" (section 2.3). Section 2.4 closes this chapter by exploring the role of phraseology in writing in general and, in specialised registers, in particular (2.4.1) and by summarising the current research studies on non-native speakers' phraseological competence (2.4.2).

Chapter 3, "Methodology", describes the methods employed in the research study. Section 3.1 establishes the rationale for using a corpus-based approach. Sections 3.2-3.5 present the procedures followed in the native corpus[1] compilation (3.2) and the materials designed to elicit informants' data; namely, a Survey (3.3) and a worksheet of exercises (3.5). A brief summary of the inquiry methods employed is included in section 3.6.

Chapter 4, "Data Analysis: the collocational patterning of abstract nouns in the *Health Science Corpus* (*HSC*)", seeks to explore the collocational behaviour of five abstract nouns (i.e. *agreement*, *comparison*, *conclusion*, *contribution* and *decision*) in the *HSC*. Following an introduction to the chapter (4.1), section 4.2 offers an overview of the lexico-grammatical patterning of abstract nouns in the health science discourse. Two main linguistic processes; i.e. delexicalisation and grammaticalisation, will be considered thoroughly as the analysis of the patterns of abstract nouns in the *HSC* has revealed that both phenomena play an important role in the characterisation of such patterns. Sections 4.3-4.7 concern the native corpus analyses with reference to the particular phraseology of each of the five abstract nouns in question. This lexico-grammatical description is followed by section 4.8, which reports on the results obtained and focuses on four linguistic processes observed:

a) delexicalisation in restricted verb-noun collocations
b) grammaticalisation in lexical bundles
c) high use of periphrastic structures in the passive voice
d) preference for adjective collocates used in attributive position

Chapter 5, "Informants' Data Analysis", aims to provide answers to research questions 1 and 2[2], formulated in section 1.1, by means of discussing the results obtained in the worksheet of exercises distributed among informants. Section 5.2 provides a general summary of all the findings of the study, which, on the whole, seem to indicate that despite being aware of certain domain-specific collocations, Spanish doctors lack knowledge of how to use them correctly. Section 5.3 closes this chapter by discussing informants' opinions of their linguistic skills and needs, as stated in the Survey. While the problem areas identified in section 5.2 are based on the analysis of non-native speakers' performance in the worksheet of exercises, the difficulties discussed in the final section of chapter 5 have been extracted from informants' own view of the kind of linguistic help that might contribute to improve their scientific writing in English.

---

[1] The *Health Science Corpus* (*HSC*)

[2] Are Spanish doctors aware of the particularities of the medical genre? and what are the main difficulties the phraseological patterns of abstract nouns in medical English pose for the non-native speakers' medical community?, respectively.

In chapter 6, "Conclusions, implications and further research", section 6.1 relates the obtained results to the research questions formulated in the introduction (cf. section 1.1). A summary of the main findings of the analysis of both the native and non-native data (section 6.2) is then followed by a number of implications of the results for the teaching and learning of phraseological units in specialist registers (section 6.3). Section 6.4 concludes the present dissertation by suggesting future research areas in the field of phraseology in EFL and ESP that need to be further explored.

Finally, a list of bibliographical references and eight appendices[3] conclude this thesis dissertation. The next chapter will set up the literature review that informed the present study.

---

[3] (a) Appendix 1: Survey; (b) Appendix 2 Survey Question 10: Informants' time estimates in completing the questionnaire; (c) Appendix 3: Consent form; (d) Appendix 4: worksheet of Exercises; (e) Appendix 5: Typology of Exercises and Aims; (f) Appendix 6: Informants' translations to sentences provided in Exercise 3; (g) Appendix 7: Informants' verb-noun collocations in Exercise 4: Matching Exercise (verb + abstract noun) and (h) Appendix 8 Survey Content-Based Question 8: Informants' self-evaluation of their main problem areas when writing in English.

CHAPTER 2: LITERATURE REVIEW AND THEORETICAL FRAMEWORK

In this chapter, a literature review is presented with the aim of establishing the theoretical and methodological framework within which this thesis project was carried out. In section 2.1, the role of corpus linguistics as a methodological tool in linguistic inquiry is explored, by providing a general account of the numerous linguistic disciplines which can benefit from corpus-based research. The following section (2.2) will serve as the theoretical foundation of the present study, which is framed within the lexico-grammar approach to language description and, thus, emphasises the association between pattern and meaning. Section 2.3 addresses the notion of *collocation* in the light of Sinclair's 'idiom principle'. A general overview of the different terms this concept has been associated with as well as its various classificatory attempts is provided. The chapter ends with a discussion of the role of phraseology in English research writing, paying special attention to specialised registers (2.4.1), followed by an examination of current empirical research on phraseology in foreign language learning (2.4.2).

2.1 Corpus linguistics as a methodological tool to explore language

Over the past decades, there has been an increasing interest in data-oriented language research. Corpus linguistics is currently seen as a way of doing linguistics, rather than a separate linguistic discipline (Meyer 2002: xi), whose object of inquiry is a collection of texts upon which different linguistic analyses can be conducted. To this respect, it is also worth mentioning that corpus linguists rely their corpus-based studies on actual instances of language instead of self-created examples. Therefore, the observation of naturally occurring data allows empirical contextualised analyses, which constitutes a valuable methodological tool in linguistic description. In this thesis project, corpus linguistics is also viewed as a methodology rather than as a strand of linguistics on its own.

As McEnery et al. (2006) state, empirical studies based on corpus data collection date back to the pre-Chomskyan period. Field linguists (e.g. Boas 1940) and structuralists (e.g. Sapir 1949; Bloomfield 1933) made use of the corpus methodology in their linguistic studies. However, such methodology was largely pushed aside in the 1950s by generative linguistics, partly due to the fact that language was considered to be productive and generativists (e.g. Chomsky 1957/1965) claimed that corpus linguistics could only inform about the language that had already been produced.

Although generativists and corpus linguists may have different goals in mind: "*while the generative grammarian strives for explanatory adequacy (…), the corpus linguist aims for descriptive adequacy*" (Meyer 2002:1), it has been demonstrated that corpus research can make an enormous impact on the creation and development of linguistic theories:

> "corpora can be invaluable resources for testing out linguistic hypotheses based on more functionally based theories of grammar, i.e. theories of language more interested in exploring language as a tool of communication." (Meyer 2002:2)

Meyer (2002) further discusses the role of corpus studies in linguistics. He claims that there is no sharp dividing line between descriptive linguists, whose linguistic studies are often based on corpus observation, and theoretical linguists, who are more concerned with building up linguistic theories. On the contrary, despite their differences, Meyer states that both types of linguists can benefit from each other:

> "while the goals of the corpus linguist and the generative grammarian are often different, there is an overlap between the two disciplines and, in many cases, the findings of the corpus linguist have much to offer to the theoretical linguist." (Meyer 2002: xiv)

As advocated by McEnery et al. (2006), "*the key to using corpus data is to find the balance between the use of corpus data and the use of one's intuition*" (McEnery et al. 2006:7). In this view, the more traditional intuition-based approach (Chomsky), which relied on native speakers' introspection of what was considered to be grammatically correct/acceptable in their language, and the corpus-based approach (Sinclair 1992a; Francis, Hunston & Manning 1996; Singleton 2000; Reppen et al. 2002; Halliday et al. 2004; among others) must be considered as mutually inclusive (Leech 1991; McEnery et al. 2006):

> “Neither the corpus linguist of the 50s, who rejected intuition, nor the general linguist of the 60’s, who rejected corpora data, was able to achieve the interaction of data coverage and the insight that characterise the many successful corpus analyses of recent years.” (Leech 1991:14)

It would certainly be unwise to claim that a compiled collection of texts, regardless of its size, can be representative of all the possible instances of a language. Nevertheless, the use of computer-based corpus material is currently seen as a very useful methodological tool in linguistics. With the possibility of storing large electronic corpora that computers provide with and the help of concordance techniques and other computer-assisted methods (i.e. frequency analyses, study of the distribution of lexical items throughout a text and calculation of the keyword sequences for particular words, to name but a few), linguists can now observe characteristics and verify results which are not apparent to a native speaker’s intuition (Francis, Hunston & Manning 1996; Chief et al. 2000)

Both approaches, i.e. the use of the linguist’s intuition and empirical data, can enrich language description if understood as complementary. As Fillmore (1992) points out:

> “I don’t think there can be any corpora, however large, that contain information about all of the areas of English lexicon and grammar that I want to explore (…) [but] every corpus I have had the chance to examine, however small, has taught me facts I couldn’t imagine finding out any other way. My conclusion is that the two types of linguists need one another.” (Fillmore 1992: 35)

Due to the growing awareness of the importance of analysing real language in use, corpus linguistics has lately had a great impact on language description theories. According to Gledhill (2000b), this rapidly expanding area of research has given rise to three clearly distinct corpus linguistics schools; namely, a) corpus-based studies in computational linguistics (Butler 1985; Oakes 1998); b) corpus-based research on corpus tagging, parsing and information retrieval (McEnery & Wilson 1996; Biber et al. 1998; Danielsson 2003) and c) the use of corpus analysis to investigate issues that have different applications in modern linguistics, such as language acquisition and language learning (Barnbrook & Sinclair 1995; Barnbrook 1996; Granger 1998; Altenberg & Tapper 1998), contrastive and translation studies (Baker 1993, 1996; Johansson & Hofland 1994; Schmied & Schäffler 1996; Bernardini & Zanettin 2000), historical linguistics (Nevalainen & Raumolin-Brunberg 1996; Rissanen 2000) and the creation of lexicographical works (Renouf & Sinclair 1991; Bogaards 1996; Cowie 1999; Moon 2007a/b/c) as well as grammar reference books (Quirk et al. 1985; Biber et al. 1999).

All the above corpus-based research areas seem to corroborate that the use of corpora is a very effective methodological tool from which numerous linguistic disciplines can benefit. Meyer (2002) highlights the various applications of the corpus linguistics methodology as a new means of exploring language which contributes successfully to the development of linguistic studies:

> “[The use of] corpora [has] numerous uses, ranging from the theoretical to the practical, making them valuable resources for descriptive, theoretical, and applied discussions of language. Because corpus linguistics is a methodology, all linguists -even generativists- could in principle use corpora in their studies of language (…) corpora (…) are used for creating dictionaries, studying language change and variation, understanding the process of language acquisition, and improving foreign -and second- language instruction.” (Meyer 2002:28)

Nowadays, much corpus-based research (i.e. the creation of lexicographical works and reference grammars, historical analyses of the evolution of language over a period of time and studies on the phraseology characteristic of specialised discourse communities, among others) is grounded on the analysis of large bodies of texts. The extensive use of both spoken and written corpora and the development of computer-aided tools to exploit such corpora have favoured the production of research studies, based on corpus evidence, on the already mentioned linguistic disciplines, which will be further discussed in the following pages.

2.1.1 Corpus-based research in language description

Data-oriented language research has proven that linguistic analyses can greatly benefit from the observation of real language in use; that is, naturally-occurring data. As a result, the past decades have witnessed an explosion of corpus studies on the usage of different grammatical patterns and multiword units, such as lexical bundles in freshman compositions (Cortes 2002) and their distribution across registers (Hyland 2008); prototypical collocations in science writing (Gledhill 2000b); patterns and meaning (Hunston & Francis 2000); verb complement constructions (Altenberg 1993); delexical chunks (May Fan 1999; Tognini-Bonelli 2001); delexicalised restricted verb-noun collocations in scientific English (Verdaguer & Laso 2006), introductory *it* patterns (Groom 2005); transitive and intransitive uses of *eat* and *drink* in English (Newman & Rice 2006); modal verbs (Coates 1983), reporting clauses used in citation (Charles 2006); the phraseology characteristic of the verb *make* (Granger, Paquot & Rayson 2006); the phraseological patterns of *make* and *take* in EFL textbook materials (Gouverneur 2008); among others.

2.1.2 Corpus-based research in dictionary-making

The creation of dictionaries based on corpus evidence stands out as one of the most prominent applications of corpus linguistics: "*corpora are now commonplace as lexicographical resources*" (Moon 2007c:165). Since the first edition of the *Collins Cobuild*[4] *English Dictionary* (*CCELD*)[5], first published in 1987, many dictionary-making projects have been based on corpus data. As Cowie (1999) highlights, the influence of Sinclair et al.'s Cobuild project in lexicography is thus undeniable: "*Sinclair and the Cobuild project have*

---

[4] *Collins Birmingham University International Language Database*.

[5] Based on the Bank of English, which today amounts to 2,5 billion words.

contributed enormously to the development of EFL lexicography, in particular, and dictionary-making in general" (Cowie 1999:128). Similarly, Carter (1998) claims that the corpus-based research carried out at the University of Birmingham (Cobuild project) has contributed enormously to the development of many lexicographically relevant works:

> "[the influence of the Cobuild project] can be measured by the fact that by the late 1990s all major English-language learner dictionary projects have incorporated reference to extensive language corpora and developed computational techniques for extracting lexicographically significant information from such corpora." (Carter 1998:167)

Without a doubt, the use of corpus data along with the software tools employed in its analysis have both speeded the process of elaborating a lexicographical work: "*because so much text is now available in computer-readable form, many stages of dictionary creation can be automated*" (Meyer 2002:16). The advances made in corpus technology have thus been a determining factor in the recent proliferation of corpus-based dictionaries, such as the *Cambridge International Dictionary of English* (1995), the *Longman Dictionary of Contemporary English* (2005) and the *Macmillan English Dictionary* (2007), which have relied upon the following repositories of corpus evidence: the Cambridge International Corpus and the Cambridge Learners' Corpus, the British National Corpus (*BNC*) and the World English Corpus[6], respectively.

2.1.3 Corpus-based research for the production of grammar reference books

By analysing the lexico-grammar of grammatical constructions, the information obtained from corpus analysis can also be used as the basis for the creation of reference grammar books that provide a comprehensive description of language. Some examples can be found in the following corpus-based modern grammar works: Quirk et al.'s *A Grammar of Contemporary English* (1972) and

---

[6] This corpus was created in the 1990s and it consists of a collection of 200 million words. It includes a general corpus, a learner corpus and an ELT corpus. Its main particularity lies in the fact that it constitutes the first ELT dictionary to be informed by electronic collections of authentic foreign language data.

A Comprehensive Grammar of the English Language (1985), some analyses of which were based on corpus data extracted from the London Corpus; Greenbaum's *Oxford English Grammar* (1996), based on the analyses of the British examples of the International Corpus of English; Biber et al.'s *Longman Grammar of Spoken and Written English* (1999), informed by the Longman Spoken and Written English Corpus, as well as a whole series of grammar works that developed from the *Collins Cobuild English Grammar* (1990), which contain actual examples from the Bank of English Corpus.

It is worth noting that the above corpus-based grammar books do not display grammatical constructions in isolation. On the contrary, pedagogical grammars of this kind aim to help users associate meanings with structures by providing information not only on the form but also on the usage of actual words and phrases. As Meyer points out, "*one of the principles underlying these grammars is that a complete description of English entails information not just on the form of grammatical constructions but on their use as well.*" (Meyer 2002:14).

2.1.4 Corpus-based research in language acquisition and language instruction

In the last few years a growing interest in the creation of learner corpora has been especially noticeable. This kind of corpora has given rise to many corpus-based studies (Altenberg & Tapper 1998; Aston 2001; Cortes 2002; Granger 1998/2003; Conrad 2005; Meunier & Granger 2008; among others) which focus mainly on the analysis of non-native speakers' performance in a second or foreign language.

The information obtained from learner corpora like the International Corpus of Learner English (*ICLE*)[7], the Longman's Learner Corpus[8], the Learner Corpus (Milton & Freeman 1996) and the Hong Kong University of Science and Technology (*HKUST*) has the potential to enhance the teaching and learning of both second and foreign languages since its analysis can help researchers design more effective Second Language (SL) and Foreign Language (FL) pedagogical materials.

---

[7] The *ICLE* Corpus consists of 2 million words and contains EFL learner essays.

[8] http://www.longman-elt.com/dictionaries/corpus/lclearn.html

In addition, Meyer (2002) advocates that non-native speakers' language learning process can also benefit from corpus analysis ("data-driven learning", as labelled by Johns 1991/1994 and Hadley 1997). By observing a corpus of native-speaker production, learners get access to examples of real language rather than artificially created instances. Then, the information generated by concordancing programmes will enable "*the student [to] conduct inductive explorations of grammatical constructions*" (Meyer 2002:27-28) which can contribute to a successful learning experience.

Retrieving information from a corpus has thus proven to be a fundamental resource for any language user, as it provides them with new insights into language structure and use. According to Tsui (2004: 42), EFL learners are not sufficiently exposed to the target language so as to be able to acquire that language efficiently on their own. Consequently, the use of corpora in an EFL context contributes to fill that gap since it allows learners to focus on naturally-occurring utterances as well as highly-frequent combinations of words. Among its other notable contributions to the teaching of EFL, it must be stressed that corpus linguistics offers "*objective quantitative evidence of the distribution of lexical items on which the contents of the curriculum can be based.*" (Tsui 2004: 41).

As already discussed, prior to the emergence of corpora, linguistic descriptions used to rely on native speakers' intuition regarding language use; that is, it was merely based on the way they perceive language to be rather than on the way language is naturally used. The intuition learners may have (especially being non-native speakers) should be backed by corpus evidence. Needless to say that EFL learners have less intuition about second language use than native speakers (Carter 1998); however, the use of corpus linguistics techniques in the classroom enables the former to become language researchers, and, as a result, they gain more autonomy as language learners since they realise they have means of acquiring language knowledge from their observation of real language, rather than merely depending on teacher's instructions.

Another worth mentioning contribution of corpus linguistics to the field of second and foreign language acquisition lies in the creation of newly designed teaching materials and linguistic resources, such as lexical databases, which incorporate recurrent combinations of words and trigger class discussions as regards the analysis of language behaviour.

The corpus research conducted in this dissertation describes the language characteristic of medical discourse from an empirical perspective and thus makes use of corpus linguistics as a methodological tool rather than a mere technology. The analysis carried out in this thesis will highlight the benefits of such a methodology not only in language description but also in language learning and language teaching.

2.2 Lexico-grammar and the association between lexis and syntax in descriptive linguistics

Over the last few decades, work on corpus linguistics has stressed the interaction between lexis and grammar in language as well as the importance of collocational and colligational patterning in the construction of meaning[9]. This issue concerning the fluidity between lexis and syntax seems to have strayed a long way from traditional descriptions of language, which showed a sharp separation between the already mentioned linguistic domains:

> "Throughout the history of language study there has also been an unchallenged acceptance of the individual, independent word as the repository of meaning." (Carter 1998:62)

On the contrary, theoretical and applied linguistics have lately focused their attention on the fact that lexis and grammar should not be viewed as separate levels of language. Within this framework, grammar is regarded as a "*heavily constrained and abstract form of vocabulary rather than a separate linguistic level*" (Gledhill 2000b:74); that is, both grammar and vocabulary are understood as two ends of the same spectrum. The following quotations from various linguists illustrate this approach to linguistic description:

---

[9] This interconnection between lexis and grammar has been extensively documented in the literature. The interested reader will find further discussion of this linguistic approach in the following studies: Sinclair 1991/2004; Halliday 1992; Francis et al. 1996/1997/1998; Hunston, Francis & Manning 1997; Hunston and Francis 1998/2000; Singleton 2000; Tognini-Bonelli 2001; Hunston 2002/2007; Stubbs, 2001; Wray 2002; Deignan 2005; Hoey 2005; Groom 2005 and Hyland 2008.

> "**Grammar and vocabulary** are not two different things; they **are the same thing seen by different observers**." [Emphasis added] (Halliday 1992:63)

> "Particular syntactic structures tend to co-occur with particular lexical items, and -the other side of the coin- lexical items seem to occur in a limited range of structures. **The interdependence of syntax and lexis is such that they are ultimately inseparable, and it becomes merely a methodological convenience to regard them as different perspectives from which to view language use.**" [Emphasis added] (Francis 1993:143)

> "the relationship between syntax and lexicon may therefore be more fluid than is usually supposed: under some circumstances an expression may be retrieved from the lexicon as a single unit; under others it may constructed from partially assembled pieces in the lexicon, requiring somewhat more syntactic processing (…) **Syntax and lexicon are** thus seen to be **complementary in a dynamic and redundant way**." [Emphasis added] (Peters 1983:90)

At the heart of this view of the intimate connection between grammar and meaning is the notion of *lexico-grammar* (Halliday 1992) / *pattern grammar* (Hornby 1954[10]), which gives account of the patterned nature of language. Within this approach to language description, the concept of *pattern* becomes central to the way discourse is construed (Halliday & Martin 1993), in the sense that it represents a link between lexis, grammar and meaning:

---

[10] Further studies on *pattern grammar* can also be found in Hasan 1987/1996; Francis, Manning & Hunston 1997 and Hunston & Francis 2000.

> "There are two main points about patterns to be made: firstly, that all words can be described in terms of patterns; secondly, that words which share patterns, share meanings." (Hunston, Francis & Manning 1997:209)

As Hunston (2002) points out, this linguistic theory dates back to Hornby's (1954) work but it has its theoretical roots in Sinclair's (1991) observations that words tend to occur in typical phraseologies and meanings are often expressed in a variety of combinatorial patternings. This neo-Firthian conception of language analysis understands that grammar and lexis must be treated as intertwined very tightly in linguistic description. Hunston & Francis' (2000) *pattern grammar* gives priority not only to the behaviour of individual lexical items but also to the patterns they form part of (Francis et al. 1996/1997/1998; Hunston & Francis 1998/2000), which implies that the "*semantics of a word [is seen] as textually distributed, and syntax as intimately linked with lexical knowledge*. " (Gledhill 2000a:131).

Bearing in mind that "*any identification of a recurring pattern rests on observation*" (Hunston 2007:259), work on corpora has recently become a commonplace for lexico-grammatical studies, since the use of naturally-occurring data allows descriptions of meaning that give account of lexico-grammatical patternings. As already stated (cf. section 2.1), growing research tradition focusing upon corpus linguistics has largely contributed to emphasising the unity of lexis and grammar, which extends the boundaries of a given initial unit, giving rise to the creation of extended units of meaning. (Tognini-Bonelli 2001:24).

It is today widely assumed that one of the areas in which the corpus (r)evolution has had greater impact is on the study of lexis. Singleton (2000) notes that Chomskyan approach to syntax (i.e. generative grammar) since the early 1950s has gradually evolved into a new linguistic perspective which prioritises the study of the lexis as the central principle of language, to the extent that a general trend towards lexical approaches to linguistic analysis can be traced (Willis 1990; Lewis 1993; Singleton 2000; Hunston 2002). The lexicon which had been neglected for a long time (i.e. Saussure, the structuralists of the Prague School and the Chomskyan modern linguists were more interested in the study of the structure of language, which was regarded as a system of rules), as it was assumed to be a list of irregularities, has recently achieved a salient position in linguistic studies:

> "an increasing number of applied linguists are beginning to address such questions. Vocabulary is now on the agenda in several domains of applied language study." (Carter 1998:282)

It is thus important to note that the present revival of the lexicon is closely related to the advent of corpus linguistics, which, as mentioned earlier, has played a fundamental role in the development of linguistic frameworks which bring to the forefront the syntagmatic relations of lexical items as well as the notion that lexical choice is prior to the selection of grammatical patterns (Sinclair 1991; Francis 1993; Hunston & Francis 2000):

> "lexis is communicatively prior to syntax. As communicators we do not proceed by selecting syntactic structures and independently choosing lexical items to slot into them. Instead, we have concepts to convey and communicative choices to make which require central lexical items, and these choices find themselves syntactic structures in which they can be said comfortably and grammatically." (Francis 1993:142)

The pervasiveness of this lexical approach to linguistic description has brought about a wide range of studies that understand lexis as "*systematically structured through patterns of use*" (Hyland 2008:6). Firth, founder of the "London School of Linguistics", and his followers (e.g. Sinclair, Halliday) took the view that the meaning of a lexical item could equal the sum of its linguistic environments; that is, the analysis of the collocates of a word would provide a great deal of information about its meaning. As Sinclair (1991) claims:

> "By far the majority of text is made of the occurrence of common words in common patterns, or in slight variants of those common patterns. Most everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up a text. This is totally obscured by the procedures of conventional grammar." (Sinclair 1991:108)

In Sinclair's view, meanings are not independent selections. On the contrary, the close interrelationship between lexical items and their context of occurrence, often referred to as *collocation*, is considered to play an important role in the interpretation of texts. This thesis is built upon the notion of lexico-grammar, which approaches language as a unitary system where grammar and the lexicon are two different perspectives of the same continuum rather than two different systems. The following section provides an overview of the notion of *collocation* and summarises its different uses in the literature.

2.3 Sinclair's notion of *collocation*

In the previous section, some of the theoretical issues surrounding the notion of *collocation* were set out. It is now turn to focus on what was meant by Firthian linguists when they first coined that term and the different uses it embraces nowadays in modern linguistics.

As discussed in the related literature, the "London School of Linguistics" (Firth, Sinclair, Halliday) used, in broad terms, the notion of *collocation* to refer to the environments in which a word occurs[11]; that is, it is based on the assumption that certain words tend to appear in combination and that, consequently, they should not be described in isolation, but rather in terms of the patterns they belong to (Francis, Manning & Hunston 1997). This intersection between grammar and lexis, embodied in the concept of *collocation*, has received increasing attention in modern linguistics in recent years. The analysis of the collocational properties of words is now considered to be fundamental in descriptive linguistics (cf. section 2.2) since it provides substantial information about the phraseologies characteristic of different genres and registers (Howarth 1996; Gledhill 2000a/b; Conrad 2000; Lee 2001; Groom 2005; Hyland 2008). As Groom (2005) states:

---

[11] "*you shall know a word by the company it keeps*" (Firth 1957:179) .

> “It is reasonable to suppose that different genres and disciplines might make differential use of these phraseological resources.” (Groom 2005:263)

Although the concept of *collocation* has been used in many different and rather vague ways, it is generally agreed that it refers to a habitual combination of words: “*a meaningful relationship between two words in each other’s environment*” (Tognini-Bonelli 2000:209). Sinclair’s definition of the term *collocation*: “*the occurrence of two or more words within a short space of each other in a text*” (Sinclair 1991: 170) emphasises the fact that the use of one word in a text affects the co-occurrence of its collocates (Barnbrook 1996:88). He claims that:

> “The choice of one word conditions the choice of the next, and of the next again. The item and the environment are ultimately not separable, or certainly not separable by present techniques.” (Sinclair 2004:19)

Thus, in this context, word co-occurrence is viewed as crucial in the interpretation of texts. Sinclair’s description of *collocation* led to his well-known “idiom principle”:

> “The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. To some extent, this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort or it may be motivated in part by the exigencies of real-time conversation.” (Sinclair 1987:320)

He suggests that there are two governing principles of language organisation: the “open-choice principle” and the “idiom principle”. While the former approaches lexis and syntax as two separate language domains, the latter refers to the tendency of language towards the creation of phrases and idioms:

> “a model of language which divides grammar and lexis, and which uses the grammar to provide a string of lexical choice points, is a secondary model (…) It has an abstract relevance, in the sense that much of the text shows a potential for being analysed as the result of open choices, but the other principle, the idiom principle, dominates. The open choice analysis could be imagined as an analytical process which goes on in principle all the time, but whose results are only intermittently called for.” (Sinclair 1991:114)

Within this frame, Sinclair’s “idiom principle” draws attention to the inadequacy of the “open-choice principle” to account for meaning in language:

> “It is clear that words do not occur at random in a text, and that the open-choice principle does not provide for substantial enough restraints on consecutive choices. We would not produce normal text simply by operating the open-choice principle (…) [T]he principle of idiom is put forward to account for the restraints that are not captured by the open-choice model.” (Sinclair 1991:110)

According to Barnbrook (2007), the grammatical models of language suggested by the “open-choice (“slot-filler”) principle”, on which most grammars are based, entail a biased view of language since they fail to identify the phraseologies typically associated with words. Within this “slot-filler” model, words are seen as being selected individually in order to occupy (fill) a position (slot) in the syntactic structure rather than as co-selected in extended units of meanings (i.e. the “idiom principle”).

Sinclair's "idiom principle" and his notion of *collocation* have been hugely influential in corpus linguistics. Highly topical corpus-based studies (Renouf & Sinclair 1991; Hunston & Francis 1998; Gledhill 2000b; to name but a few) have stressed the close link between the grammatical form of a string of words (pattern) and its meaning. Current phraseological work has demonstrated that collocations that share specific particular patterns may also convey a closely related meaning (Sinclair 1987/1991; Hunston & Francis 2000; Gledhill 2000b; Hoey 2005; Groom 2005; Hunston 2007).

From this perspective, lexis and grammar are regarded as equally necessary in describing how meaning is construed (Sinclair 1991), and thus the notion of *collocation* and the analysis of how some uses of words tend to occur in certain lexico-grammatical patterns have become central in mainstream corpus linguistics.

The study of multiword units and their collocational patternings has been extensively discussed in the literature. Yet while most studies of *collocation* are grounded, in general terms, on Firth's and Sinclair's notion of *collocation* as the way words tend to be associated in language, this "togetherness factor" (Singleton 2000) has been referred to with different terms and has also increasingly embraced different uses.

There have been many classificatory attempts of multiword expressions which have given rise to a variety of defining terms such as idioms or idiomatic expressions (Moon 1992; Schmitt 2004); prefabs or prefabricated expressions/routines (Wray & Perkins 2000); clusters (Scott 1996); chunks (Nattinger& DeCarrico 1992); formulae or formulaic sequences (Moon 1992); ready-made expressions (Cowie 1988); lexical bundles (Biber et al. 1999/2003/2004; Cortes 2002/2004; Partington & Morley 2004); word strings, clichés, composite units (Howarth 1996) and lexical phrases (Nattinger & DeCarrico 1992; Oakey 2002a/b).

As that wide range of labels reflects, the treatment of collocations may be approached from a number of perspectives (e.g. lexicographic, semiotic, sociolinguistic, psycholinguistic and semantic) and with different goals (e.g. pedagogic, lexicographic) in mind, which will, inevitably, determine the choice of each of the above phraseological items.

Research on the phraseological mechanisms of language has focused not only on the descriptive analysis of word combinations (Becker 1975; Altenberg 1993; Biber et al. 1999; Moon 1998) but also on the role of phraseology on language teaching and learning (e.g. Fillmore's [1979] and Pawley & Syder's [1983] work on native-like fluency; DeCarrico & Nattinger's [1988] and Nattinger & DeCarrico's [1989] studies on lexical phrases in academic lectures and speech acts; Howarth's [1996/1998] analysis of non-native speakers' written academic performance in English and Granger & Meunier's [2008] study of the role of phraseology in foreign language acquisition and instruction). Both issues will be more thoroughly considered in section 2.4.

Taking into account that, as has been observed, the concept of *collocation* may be defined in various ways to serve different linguistic purposes, it seems now worth discussing how different traditions have approached the phenomenon of multiword expressions. Howarth (1996) presents a synthesis of how the heading of *collocation* has been viewed by two[12] main traditions in which this concept originated; the "Firthian London School of Linguistics", on the one hand and Russian lexicology, on the other. He observes that the main difference between both views is that whereas the former focuses mainly on the concept of *collocation* on its own, the latter is more concerned with the phraseological classification of word combinations (Howarth 1996:25).

Following on from Firth's notion of *collocation*, linguists such as Halliday, Sinclair and Hoey have adopted what has been referred to as a "frequency-based, statistical or textual" (Herbst 1996; Nesselhauf 2004; Gledhill 2000b) approach to the study of collocations. In their view, collocations are framed in terms of the frequent co-occurrence of words, rather than on their various classifying categories. The following definitions illustrate this statistical perspective:

---

[12] Howarth (1996) also mentions a third tradition which has lately gained much attention. This new approach to the notion of *collocation*, introduced by linguists such as Palmer (1933) & Hornby (1954, tackles the difficulties that phraseological units pose for language learners (Howarth 1996:25). Further discussion on the role of phraseology in language learning and teaching will be put forward in section 2.4.

> "Collocation is the syntagmatic association of lexical items, quantifiable, textually, as the probability that there will occur at *n* removes (a distance of *n* lexical items) from an item *x*, the items *a*, *b*, *c* (…) Any given item thus enters into a range of collocation, the items with which it is collocated being ranged from more to less probable." (Halliday 1961:276)

> "a cover term for the kind of cohesion that results from the co-occurrence of lexical items that are in some way or other typically associated with one another, because they tend to occur in similar environments." (Halliday & Hasan 1976:287)

> "the occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is a maximum of four words intervening. Collocations can be dramatic and interesting because unexpected, or they can be important in the lexical structure of the language because of being frequently repeated." (Sinclair 1991:170)

> "[The term *collocation* refers to] the relationship that a lexical item has with items that appear with greater than random probability in its textual context." (Hoey 1991:6)

The above quotes highlight the statistical probability of certain words to occur in combination, which has proven to be of great value, particularly in computational studies of syntagmatic relations of multiword expressions (Nesselhauf 2005:12). However, Howarth (1996) argues that:

> "collocation is not purely a matter of probability of linear co-occurrence, since there are grammatical and semantic or purely lexical factors that constrain lexical co-occurrence in a large number of cases. One of the tasks facing the lexicologist constructing an integrated model for composite units is to

> make general descriptive statements about lexical relationships between items which are highly diverse in nature, extremely numerous, very hard to measure quantitatively and about which speakers of the language have differing opinions.”
> (Howarth 1996:29)

In this context, he takes account of the treatment of *collocation* in the “semantic/syntactic” (Gledhill 2000b) tradition, also known as “significance-oriented” (Herbst 1996) or “phraseological” (Nesselhauf 2004) approach, which is rooted in Russian lexicology. Unlike the “statistical/textual” approach, the “semantic/syntactic” model defines *collocation* as a type of word combination in a certain grammatical context. The emphasis is not on the frequency of occurrence of expressions but on their combinatory possibilities since, in Gledhill’s (2000b) view, collocations are “*typically seen either as units of meaning (lexical items or idioms) or units of grammar (phrases).*” (Gledhill 2000b:9).

As referred to in Howarth (1996) and Nesselhauf (2005), the work of linguists such as A. P. Cowie (1992/1994/1998), Greenbaum (1970), Mel’čuk (1998) and Hausmann (1989) can be included in this phraseological approach, whose main aim consists of describing phraseological expressions as well as analysing their different typologies. According to Howarth (1996), the most salient defining trait of this approach is that

> “it sees no watertight division between the various types of collocation and idiom, rather a continuum from, at one extreme, the most freely co-occurring lexical items and transparent combinations to, at the other, the most cast-iron and opaque idiomatic expressions.” (Howarth 1996:32)

From this phraseological perspective, thus, the term *collocation* is being used to refer to different phraseological categories, ranging from highly fixed (idiomatic) expressions[13] (e.g. *spill the beans, blow the guff, kick the bucket, foot the bill*)[14] to unrestricted ("free combinations") word sequences[15] (e.g. *run a business, explode a bomb, drink tea*)[16].

Regardless of the fact that the distinction between the various types of word combinations suggested by Cowie (1981) is sometimes difficult to delimit from a purely semantic point of view (Van der Wouden 1997; Gledhill 2000b),

> "you cannot predict that the meaning of *sleep like a log* will denote an intense form of sleeping, but after you have learned what it means, you see that *like a log* is an intensifier. The essence of collocation is that the assignment of *like a log* to the meaning 'very' does not feed any other combinations. So even though we have a meaning for it, that meaning is only valid in a certain collocation." (Van der Wouden 1997:54-55)

it cannot be denied that Cowie's (1994) approach to *collocation* as the "*association of two or more lexemes (or roots) recognized in and defined by their occurrence in a specific range of grammatical constructions*" (Cowie 1994:3169) has proven remarkably influential in collocational studies, since it has underlined the assumption that the elements that constitute sequences of words/word associations (understood as abstract combinations) are syntactically related (Nesselhauf 2005).

Although fuller attention has been accorded to the two approaches discussed so far; namely, the "frequency-based approach" and the "phraseological approach", it is important, nonetheless, to note that some other perspectives have defined the term *collocation* in other ways, highlighting numerous other collocational properties. It seems worth mentioning, for instance, Gledhill's (2000b) "discoursal/rhetorical approach" to the definition of the scope of the notion of *collocation*. He argues that apart from analysing recurrent strings of words and examining their syntactic and semantic structure, collocations can also be observed from a discoursal point of view; that is, by examining their rhetorical function in a given discourse:

---

[13] In Howarth's (1996) words, idiomatic expressions are: "*combinations that have a unitary meaning that cannot be derived from the meanings of the components*." (Howarth 1996:47)

[14] These examples have been taken from Cowie (1981).

[15] Unrestricted or free combinations are defined as "*combinations of two or more words in which the elements are used in their literal sense. Each component may be substituted without affecting the meaning of the other*." (Howarth 1996:47)

[16] Cf. Footnote 14

> "From this perspective, idioms such as *to get the sack*, *to be fired* can be contrasted stylistically with less marked expressions: *to be dismissed*, *to lose one's job*. The difference between these expressions lives in their emphasis or rhetorical effect (…)" (Gledhill 2000b:13)

Gledhill's model is therefore more concerned with the pragmatic role of phraseological units, which has been broadly discussed in the field of discourse analysis[17]. As he notes, this approach goes beyond Halliday's "syntagmatic association of lexical items" (Halliday 1961) and Howarth's "collocational continuum" (Howarth 1996/1998) and focuses on the pragmatic textual function performed by collocations and idioms in the construction of discourse.

Within this model, the emphasis is shifted to the role of lexical phrases in discourse:

> "Lexical phrases are parts of language that often have clearly defined roles in guiding the overall discourse. In particular, they are the primary markers which signal the direction of discourse, whether spoken or written. When they serve as discourse devices, their function is to signal, for instance, whether the information to follow is *in contrast to*, *in addition to* or is *an example of* information that it to proceed." (Nattinger & DeCarrico 1992:60)

All in all, although each of the abovementioned approaches (i.e. statistical, phraseological and rhetorical) to the phenomenon of *collocation* foregrounds different properties of multiword sequences, they should not be considered as inevitably exclusive but rather as complementary, since the three perspectives contribute to a better understanding of the associations established between the different constituents of multiword expressions. As Gledhill (200b) asserts:

---

[17] Studies in this area are, for example, Pawley & Syder 1983; Fillmore, Kay & O'Connor 1988; Nattinger & DeCarrico 1992; Moon 1994; Hyland 2008.

> "Despite differences of methods, each approach leads us to reconsider the relationship between words within the collocational expression and to revise the traditional notion of phraseology (…) a lexico-grammatical analysis of a specific discourse can be supplemented by an analysis of phraseology." (Gledhill 2000b: 17)

In general terms, this thesis will use the notion of *collocation* as a syntagmatic recurrent combination of words. More specifically, following Gledhill's view, the three collocational models presented in this section have been adopted at different stages of the present research. First, from a statistical perspective, the word combinations examined in this corpus-based study have been analysed in terms of their frequency of occurrence. Secondly, from a phraseological approach, the lexico-grammatical patterning characteristic of the abstract nouns considered has been explored in a corpus of medical English[18]. Finally, the role that the use of the analysed expressions play in the construction of phraseology in medical discourse has also been discussed ("rhetorical approach").

The following section addresses the issue of phraseology in language, paying special attention to its treatment in scientific English (2.4.1) and in foreign language learning (2.4.2).

2.4 Phraseology in English Research Writing

> "*Phraseological analyses* demonstrate that much of communication makes use of fixed expressions memorized as formulaic chunks, that language is rich in collocational and colligation restrictions and semantic prosodies, that the phrase is the basic level of language representation where form and

---

[18] *The Health Science Corpus (HSC)*; cf. chapter 3.

> meaning meet with greatest reliability (…) The phrase is at the centre of language, and thus calls the attention of the broad range of language sciences." (Ellis 2008:6)

As noted by Ellis (2008), phraseology plays an important part in language production. Thus, the description of the characteristics, lexico-grammatical patterning, usage and discourse function of prefabricated expressions has been extensively researched and discussed in the literature over the last decades. Recent phraseological studies[19], mainly focused on the analysis of academic English, have pointed to the fact that phrases (also referred to as formulaic sequences) are central to the grammar. In this context, phrases are viewed as the main focus of linguistic analysis since they contribute to characterise the language of different discourse communities. In Oakey's (2002) view, a lexical phrase can be defined as:

> "a discontinuous formulaic sequence which has been assigned a pragmatic function, and thus each occurrence of a pattern needs to be examined in context in order to identify the function it appears to have in the text." (Oakey 2002a:116)

According to Singleton (2000), several linguistic schools such as Lexical-Functional Linguistics, Construction Grammar, Valency Grammar and the various forms of Dependency Grammar have enormously contributed to placing the study of the lexicon at the forefront of much linguistic research being conducted nowadays. Within this lexical perspective, it is argued that there is a continuum between lexis and grammar and, therefore, lexical items are described in terms of the constructions they take part in:

> "Not only do constructionists see as a continuum the properties of syntactic, phraseological and lexical structures, but they also are convinced that phraseological patterns make up the vast majority of structures that enter into everyday discourse." (Fillmore 1989:34)

---

[19] See Howarth 1996/1998; Moon 1998; Hunston & Francis 2000; Wray 2002; Oakey 2002a/b; Biber 2006; Hyland 2008; Granger & Meunier 2008.

Phraseological studies have greatly benefited from all the above linguistic theories which have promoted the study of meaning across different patterns of language. This attested interaction between meaning and form has already been discussed (cf. section 2.2), but it seems worth noting that this symbiotic relationship becomes particularly apparent in the analysis of lexical phrases. Nattinger and DeCarrico (1989) comment that:

> "multi-word lexical phenomena (…) exist somewhere between the traditional poles of lexicon and syntax. They are similar to lexicon in being treated as a unit, yet most of them consist of more than one word, and many of them can at the same time be derived from the regular rules of syntax, just like other sentences." (Nattinger & DeCarrico 1989:118)

Likewise, the advent of corpus linguistics has also been fundamental in the development of empirical studies of phraseology, since it has given the computational means by which automatic analyses of large bodies of naturally-occurring data can be carried out (Kjellmer 1984/1990; Sinclair 1987b/1991). The close observation of real language in context has confirmed the fact that "*natural language makes considerable use of recurrent patterns of words and constructions*" (Ellis 2008:4), so the lexical context is considered to provide the linguist with an invaluable wealth of information regarding meaning and grammatical function.

Consequently, the potential of corpus research should not be underestimated in phraseological studies because it constitutes a step further in the methodology employed in linguistic description and "*leads to a deeper understanding of the obligatory nature of much that is written in academic discourse*." (Gledhill 2000a:133).

According to Krishnamurthy (2000), traditional grammar used to rely mostly on native speakers' intuitions with respect to word combinations:

> “These [native speakers’ intuitions] were often incorrect, or at least inexact, because each of us has only a partial knowledge of the language, we have prejudices and preferences, our memory is weak, our imagination is powerful (so we can conceive of possible contexts for the most implausible utterances), and we tend to notice unusual words or structures but often overlook ordinary ones.” (Krishnamurthy 2000:32-33)

In contrast, corpus linguistics is based on the actual use of such combinations in real language as being produced by speakers, which, in Oakey’s (2000) words, certainly entails “*a clearer, less intuitive insight into these units, and illuminate how their form and use differs in different academic contexts*.” (Oakey 2002a:111).

Another area which is receiving phraseologists’ attention (Glaser 1988; Cowie 1988; Howarth 1996; Partington 1998; Wray 1999; Gledhill 2000b; Hyland 2008; Granger & Meunier 2008) is the key role of multiword expressions in the textual development of meaning. Present-day phraseological studies are concerned not only with the lexico-grammatical analysis of formulaic sequences of specific discourses, but also with their rhetorical uses (Gledhill 2000b:17) since, as will be further discussed in subsection 2.4.1, lexical phrases fulfil varied discourse signalling and cohesive functions across disciplines (Oakey 2002a:111). The following assertions traced in the literature will serve to illustrate this point:

> “[Multi-word expressions] are extended collocations which appear more frequently than expected by chance, helping to shape meanings in specific contexts and contributing to our sense of coherence in a text (…) [They] are not only central to the creation of academic discourse, but (…) they offer an important means of differentiating written texts by discipline.” (Hyland 2008:4)

> “The notion of phraseology implies much more than inventories of idioms and systems of lexical patterns. Phraseology is a dimension of language use in which patterns

> of wording (lexico-grammatical patterns) encode semantic views of the world, and at a higher level idioms and lexical phrases have rhetorical and textual roles within a specific discourse. Phraseology is at once a pragmatic dimension of linguistic analysis, and a system of organisation which encompasses more local lexical relationships, namely collocation and the lexico-grammar." (Gledhill 2000b:202)

All in all, research on the field of phraseology seems to indicate that mastering formulaic constructions is a significant factor in successful linguistic production. In that same line, this thesis stresses the importance of gaining control of the phraseology characteristic of medical English so as to show linguistic competence. As will be further discussed, this is especially relevant for the medical community as they are committed to disseminate their research internationally (Flowerdew 2007) and, thus, must be familiar with the phraseological conventions of the discourse community they are part of.

2.4.1 Phraseology in specialised registers

> "[T]he picture we have of the research article is far from complete. That picture suggests that there are certain characteristics of RAs [research articles] which, by and large, tend to occur and recur in samples drawn from an extensive range of disciplines (…) However, it remains the case that RAs [research articles] are rarely simple narratives of investigations. Instead, they are complexly distanced reconstructions of research activities, at least part of this reconstructive process deriving from a need to anticipate and discountenance negative reactions to the knowledge claims being made." (Swales 1990:174-175)

Swales' account of the genre of the scientific article points to the existence of a number of conventions which contribute to define and characterise the scientific discourse. Textual analyses of the genre are hence extremely important so as to identify the phraseological structures characteristic of scientific English.

Genre analysis is concerned with a particular subset of language; that is, a specific language practice, characterised by a number of linguistic features and phraseological conventions. It can be therefore claimed that genres make use of different ways of expressing meaning (Hunston 2002:178). This assumption is intimately linked with the concept of *local grammar* (Gross 1993; Barnbrook & Sinclair 1995; Hunston & Sinclair 2000), which consists of a description of particular areas of language (e.g. the analysis of the phraseology characteristic of medical discourse), rather than the language as a whole (Bednarek 2007).

The current treatment of phraseology in specialised registers acknowledges the need for corpus-based studies of the prototypical lexico-grammatical patternings and discourse functions of lexical phrases across disciplines (cf. Carter 1998; Oakey 2002a/b; Hyland 2008)[20]. According to Hyland (2008),

> "Gaining control of a new language or register requires a sensitivity to expert users' preferences for certain sequences of words over others that might seem equally possible." (Hyland 2008:5)

Thus, it seems that getting familiar with the specific phraseology of the register of a discourse community will imply not only a better knowledge of the genre but also an enhanced competence in the process of writing and reading in specialised registers. As Williams (2002) claims:

> "In order to understand texts, we must look at them closely to find the lexico-grammatical strategies that they adopt to assist communication within a specialised community." (Williams 2002:60)

Studies in genre analysis, such as Swales' (1990) investigation of academic and research settings; Bhatia's (1993) exploration of genre in professional contexts and Gledhill's (2000b) work on collocations in science writing, suggest that there are significant textual variations in different specific-domains and genres[21]. These findings have underlined the convenience of analysing specialist corpora so as to find out the "*kinds of language data which particular communities of users might encounter and which will inform their use*." (Hyland 2008:8).

---

[20] Other linguists who have stressed the importance of analysing the prototypical phraseological expressions in different genres and registers are Lewis 1996; Williams 1998/1999/2002; Biber 2006 and Scott & Tribble 2006.

As already stated, close observations of real language in use by means of corpus-based data constitute a useful method for the textual and rhetorical study of phraseological units on the one hand, and for the study of how those expressions serve to characterise the discourse within which they have been produced, on the other.

Despite the abovementioned growing interest in the formulaic aspects of language knowledge in specialised registers, Gledhill (2000b) observes that in comparison with linguistic analyses based on general English corpora, less work has been conducted on specialised language to date. More specifically, he claims that there is a noticeable shortage of linguistic corpus-based studies in the field of phraseology in scientific discourse.

There are, however, some remarkable exceptions. Several worth mentioning studies focusing on scientific articles as a whole are, for example, Myers' (1989) account of the pragmatics of politeness involved in scientific papers; Master's[22] (1991) study of active verbs with inanimate subjects in scientific English and Banks' (1994) analysis of the organisation of different clause types in the scientific journal article.

Other studies on the phraseology characteristic of scientific discourse, on the contrary, have centred their investigation either on a specific domain within the field of scientific English or, to a lesser extent, on the different (sub)sections of the scientific article. Examples of the former can be traced in the following literature that explores the textual properties of health science (i.e. medical and biological) discourse:

---

[21] Further genre analyses can be found in Biber et al. 1999; Lee 2001; Lee & Swales 2006 and Scott & Tribble 2006.

[22] In his analysis of scientific prose, Master (1991) pays special attention to writers with an Asian background.

a) Adams' analysis of authorial comments (1984) and variation (1987) in medical English.
b) Malcolm's (1987) study of the rules governing tense usage in medicine.
c) West's (1980) exploration of "*that*-clauses" in biological science.
d) Thomas' & Hawes' (1994) classification of a list of reporting verbs (e.g. *demonstrate*, *indicate*, *show*, *suggest*) in eleven medical journals.
e) Salager-Meyer's (1994) study of hedging in medical discourse.
f) Gledhill's (1995/1996) phraseological analysis of cancer research articles.
g) Williams' (1996) work on lexical verbs in clinical and experimental research reports.

With respect to research on, more specifically, the different (sub)sections of the scientific article, Salager-Meyer's (1992) exploration of medical English abstracts (with a special emphasis on verb tense and distribution of modality) and Gledhill's (2000a/b) work on the rhetorical function of collocation in research article introductions are also remarkable examples.

All the above literature has proven extremely useful in the characterisation of science writing, and has also confirmed Swales' (1990) assertion of the complexity of the scientific research article: "*the RA [research article] is anything but a simple genre*" (Swales 1990:128). Several textual properties of the scientific discourse such as modality (Huddleston 1971; Widdowson 1979; Adams-Smith 1984; Salager-Meyer 1992; Banks 1994; Gledhill 2000a/b), hedging (Myers 1989; Swales 1990; Salager-Meyer 1994; Banks 1994; Varttala 1999; Gledhill 2000), the use of the passive and the anticipatory *it*-pattern so as to disguise authorial interpretations (Huddleston 1971; Swales 1990; Banks 1994; Biber et al. 1998/1999; Hyland 2008), an attested tendency to use grammatical metaphor (Salager-Meyer 1992; Banks 1994; Halliday 1998; Gledhill 2000a/b) and a high use of abstract nouns in the expression of processes and methods (Halliday 1993; Flowerdew 2003) have been identified as defining rhetorical devices which contribute to a great extent to the development of scientific discourse (Luzón 2000; Gledhill 2000b; Noguchi 2006; Hyland 2008).

If, as already discussed in the reported literature, much of the language involved in scientific discourse is "*highly stereotypical in nature"* (Gledhill 2000a:116), it seems of paramount importance that members of that discourse community become familiar with the collocational expressions considered to be "good scientific style":

> "collocations in science writing are undoubtedly selected as the best ways of expressing certain ideas (…) [T]he selection is largely a feature of convention and acceptability within the discourse community." (Gledhill 2000a:133)

In Gledhill's (2000a) view, conforming to those conventions will provide scientists with the phraseological competence necessary for effective and accurate communication. Following Gledhill (2000), this dissertation focuses on the linguistic problems Spanish doctors encounter when writing their research articles in English. Bearing in mind that they are expected to spread their results around the medical community, the present research will highlight the fact that non-native writers are faced with the further challenge of producing native-like publications for international journals.

The difficulties that mastering the formulaic patterns of specialised registers may pose for non-native speakers as well as the role of phraseology in English as a Second Language (ESL) and English as a Foreign Language (EFL) Writing for Specific Purposes will now be examined in the light of the reviewed literature.

2.4.2 Phraseology in Foreign Language Learning

> "The notion that native-like proficiency in a language depends crucially on a stock of prefabricated units -or 'prefabs'- varying in complexity and internal stability is now set in critical opposition to the atomistic view, rooted in generative theory, that the workings of a language can be explained by a system of rules of general applicability, a lexicon largely made up of minimal units and a set of basic principles of semantic interpretation." (Cowie 1998:1-2)

As Cowie (1998) points out, the role of prefabricated expressions as a significant indicator of communicative excellence in a language is now acquiring increasing relevance in phraseological studies. Although the existence of fixed forms in language had long been observed (Bloomfield 1933; Firth 1957), it is with modern linguistic theories (in particular, with the neo-Firthian's pioneering work of, for instance, Halliday and Sinclair) that the issue of formulaicity in language has been brought to the fore of current linguistic discussion.

After the predominance of the Chomskyan generative model to language, more recent linguistic approaches such as Cognitive Grammar (Langacker 1991/2000; Croft & Cruise 2004), Construction Grammar (Fillmore, Kay & O'Connor 1988), Lexical-Functional Grammar (Bresnan 1982), Valency Grammar (Allerton 1982) and the various forms of Dependency Grammar, and Pattern Grammar (Hunston & Francis 2000) have emphasised the mutual dependence between the lexicon and syntax and have addressed the issue of the pervasiveness of phraseology in language:

> "[I]f lexicon comprises the fixed expressions of a language, it subsumes not only single morphemes and polymorphemic stems and words, but also thousands of multiword conventional expressions (clichés, idioms, standard collocations, etc.) representing usual ways of expressing oneself in a language." (Langacker 1991:3)

Recent corpus-based studies of multiword expressions (Sinclair 1991; Gledhill 2000; Oakey 2002a/b; Stubbs 2001; Nesselhauf 2005; Hyland 2008; Granger & Meunier 2008) have confirmed that discourse communities tend to make use of recurrent word combinations and prefabs. As pointed out in subsection 2.4.1, effective communication in a language is highly dependent on the user's knowledge of the expected fixed expressions in a given context. In other words, lexical phrases are regarded as the means by which speakers are able to process language effectively and improve fluidity when encoding messages (Wray 2002). This, inevitably, holds a number of implications since non-native speakers must conform to the conventions of native-like discourse, which entails, as has been observed, not only selecting the right lexical items, but also full communicative mastery of the lexico-grammatical patterns and discourse functions of prototypical collocations (Howarth 1998a:186).

As research on non-native speakers' phraseological competence has demonstrated, the processing and use of phraseology pose many difficulties for learners (Carter 1998; Howarth 1996/1998; Wray 1999; Oakey 2002a/b; Williams 2005; Nesselhauf 2005; Ellis 2007; Hyland 2008; Granger & Meunier 2008; Paquot 2008, among others):

> "knowing which subset of grammatically possible utterances is actually commonly used by native speakers is an immense problem for even the most proficient of non-natives." (Wray 1999:468)

According to Howarth (1996), non-native speakers' lack of phraseological competence is partly due to the fact that they are not taught collocations explicitly. As he comments:

> "few learners are taught collocations and they are perhaps unaware of the collocations that they do know and have acquired accidentally. The evidence from their errors in producing overlapping collocations suggests that their competence is full of gaps and highly variable between

> learners and that this is a particularly unstable area of interlanguage." (Howarth 1996: 192)

Howarth (1996) thus claims that the field of phraseology should play a predominant role in foreign language instruction as empirical studies have revealed that deviations occurring in collocations are particularly frequent in non-native speakers' oral and written production (Howarth 1996/1998; Granger 1998; Wray 1999/2002; Altenberg & Granger 2001; Nesselhauf 2005):

> "The problems for learners are, firstly, to recognize that the phenomenon of conventionally restricted collocability exists and is widespread and, secondly, to acquire knowledge of the particular facts of how the phenomenon is realized in practice." (Howarth 1996:136)

and then

> "many learners fail to understand the existence of the central area of the phraseological spectrum between free combinations and idioms. It is in handling restricted collocations that errors of both a lexical and grammatical nature constantly occur." (Howarth 1998a:186)

As can be inferred from Howarth's (1996/1998) assertions, it is also of vital importance to draw learners' attention to the phenomenon of phraseology in language and raise their awareness of the collocations and prefabricated expressions which are characteristic of native linguistic production.

Research into pedagogical tools and materials that help learners of English as a Foreign Language (EFL) to improve their phraseological competence is also beginning (Granger 1998; Walker 2008; Gouverneur 2008; Pecman 2008). Walker (2008) stresses the need to design teaching materials that lay greater emphasis on collocations, understood as motivated and, therefore, susceptible to be explained, word strings. He argues that the right approach to the teaching of collocation should challenge learners to reflect on why some words are often associated with some others, so as to make their collocational learning process more meaningful. As he claims:

> “There needs to be a movement away from the current emphasis on techniques for the rote memorisation of lists of collocations. Instead, teaching material needs to contain exercises which focus upon one aspect of collocation, such as the semantics of individual items or the use of metaphor (…) Once collocations become items which can be explained, they are immediately more meaningful to the learner and therefore more memorable (…) With the right kind of materials and the right approach to the teaching of collocation, the whole process of learning collocations should become more meaningful for the learner, as well as more enjoyable.” (Walker 2008: 307)

However, phraseologists (Cowie 1992/1998; Howarth 1996; Granger 1998; Kaszubski 1998; Wray 2002; Nesselhauf & Tschichold 2002; Schmitt 2004; Walker 2008) are also aware of the fact that more empirical analyses studies on lexical phrases are needed. As Granger states:

> “prefabs certainly need to play a greater role in EFL than they have in the past (…) [L]earners’ phraseological skills are severely limited: they use too few native-like prefabs and too many foreign-sounding ones. But if we are to devise the ‘ideal’ pedagogical tools, a great deal of more empirical data on prefabs is required.” (Granger 1998:158)

Another important issue in the creation of useful tools aimed at improving non-native speakers’ control of phraseology is, as suggested by Granger (1998), the need for an in-depth analysis of the phraseology of the learners’ mother tongue. To this respect, she claims that further research on the phraseology characteristic of the non-native speakers’ L1 would be desirable because transfer of structures from their mother tongue has been found to affect L2 phraseology (Gabrys-Biskup 1992; Bahns & Eldaw 1993; Kellerman 1995; Granger 1998; Ellis 2007; Paquot 2008):

> “L1 plays an important role in the acquisition and use of prefabs in the L2. For obvious commercial reasons, most EFL material is aimed at all learners, irrespective of their mother tongue. Given the essentially language-specific nature of prefabs, this is a major issue that must be addressed if we are serious about giving learners the most effective learning aids.”
> (Granger 1998: 159)

Regarding L1 influence in the learners’ phrasicon, Gabrys-Biskup’s (1992) and Bahns’ & Eldaw’s (1993) analyses of L1 transfer and its various implications in L2 phraseology are also worth recalling. Gabrys-Biskup (1992) examines deeply the influence of L1 in translations produced by Polish and German learners of English. In particular, this study focuses upon the strategies employed by learners when translating L1 word combinations into English. Similarly, Bahns & Eldaw (1993) investigate the mechanisms used by German subjects in translating a text into English. The results of both investigations show that non-native speakers lack knowledge of English collocations.

Other collocational studies in learner language stress non-native speakers’ misuse of native-like word combinations. Granger’s (1998) analysis of the production of adverb-adjective constructions shows that learners make less use of prefabricated expressions in comparison with native speakers. In the same line, Howarth (1996) and Kaszubski (2000) explore the use of verb-noun combinations by non-native speakers. The former compares English verb-noun combinations produced by speakers of different mother tongues against native verb-noun associations. Howarth’s findings suggest that non-native speakers tend to use fewer collocations and that their level of proficiency, contrary to Zhang’s (1993) observations, does not seem to have any remarkable influence in their phraseological competence. Likewise, Kaszubski’s (2000) comparison of the collocational uses of five supporting verbs (i.e. *do*, *get*, *have*, *make* and *take*) both in native and non-native corpora yields similar results concerning the limited number of collocations used by non-native speakers.

Nesselhauf's (2005) work on non-native speakers' production of multiword units also constitutes a thorough investigation of the difficulties German learners face when producing verb-noun combinations. Her data reveal that the collocations provided by advanced learners of English show a high degree of deviation and, therefore, her study corroborates Howarth's (1996) conclusion that non-native speakers' advanced level of English does not have any remarkable positive effect on their collocational competence.

As observed in the reviewed literature, the influence of learners' L1 as well as the problems that the use of multiword units pose for non-native speakers is gaining considerable attention in the field of EFL learning and teaching. There is now growing evidence that phraseological units should be acknowledged as a particularly interesting research area, not only for linguistic description, but also for foreign language instruction (Nuccorini 2002; Granger & Meunier 2008).

All in all, numerous empirical studies of L2 phraseology, based on learners written production (Gabrys-Biskup 1992; Bahns & Eldaw 1993; Howarth 1996; Granger 1998; Kaszubski 2000), have evidenced that non-native speakers lack awareness of the most typical collocations in English, both in the general language and specialised registers. As the processing and production of such prefabricated units have proven particularly problematic for learners, a phraseologically-oriented approach to FL instruction seems to be justified (Cowie 1998; Lewis 2000; Oakey 2002b; Ellis 2008). As Cowie (1998) observes:

> "[S]ensitizing advanced learners to the crucial importance of this central area of the phraseological spectrum must be a major priority of EFL teachers." (Cowie 1998:15)

This priority and the pervasive role of phraseological units in language are, as already noted, at the centre of much of the current work devoted to phraseology in foreign language learning and teaching. In this dissertation the important role played by phraseological units in foreign language learning as well as the difficulties non-native speakers are confronted with in writing their medical articles in English will also be explored. In the next chapter, the methodology used in the research study will be presented.

CHAPTER 3: METHODOLOGY

The following chapter describes in detail the methodological issues considered in the present study. In section 3.1 the notion of corpus linguistics as a methodology as well as the criteria followed for applying a corpus-based approach to this research are examined. Section 3.2 gives account of the construction and processing of the corpus compiled in this investigation and section 3.3 presents the survey conducted among a community of twenty-four Spanish doctors whose profile will be outlined in section 3.4. Following the characterisation of the participants in the sample, section 3.5 will provide a detailed description of the worksheet of exercises administered among them. The final section of this chapter (3.6) will serve as a summary of the methodology applied.

3.1 Corpus linguistics as a methodology. The rationale for using a corpus-based approach

> “Neither the corpus linguists of the 1950s, who rejected intuition, nor the general linguists of the 1960s, who rejected corpus data, was able to achieve the interaction of data coverage and the insight that characterise the many successful corpus analyses of recent years.” (Leech 1991:14)

This section is devoted to the role of corpus linguistics as a methodology in linguistic research. Nowadays corpus methodology enjoys widespread recognition. As many scholars (Leech 1992; Hunston & Francis 2000; Tognini-Bonelli 2001; Meyer 2002; McEnery et al. 2006; Moon 2007) have already pointed out, corpus linguistics should be regarded as a system of methods applied to the analysis of real language data, “*a methodological basis for pursuing linguistic research*” (Leech 1992:105), rather than a theory of language of its own: “*corpus linguistics provides a methodology, not an end or theory in itself*” (Moon 2007b:1058). According to McEnery et al. (2006), corpus linguistics does have a theoretical basis but it is not a linguistic theory as such: “*as (…) a system of methods and principles of how to apply corpora in language* *studies and teaching/learning, [corpus linguistics] certainly has a theoretical status. Yet theoretical status is not theory itself*' (McEnery et al. 2006:7-8).

In line with Meyer's (2002) and McEnery et al.'s (2006) statement, I understand corpus linguistics as a methodology on its own rather than as an independent linguistic theory. Unlike some other branches of linguistics such as lexicology, morphology and semantics, corpus linguistics is not merely limited to the analysis of a single aspect of language. Quite on the contrary, it can be applied to the study of many linguistic areas.

The work presented in this thesis follows thus the corpus methodology and highlights how linguistic analyses can greatly benefit from a corpus-based approach[23]. Corpus-based approaches to language studies emphasise how the use of computers to manipulate and analyse large bodies of language data has tremendously affected "*the methodological frame of linguistic enquiry*". (Tognini-Bonelli 2001:210). Traditionally, language has been described in language classrooms by means of setting rules and offering examples to reinforce and back up such pre-established rules. As Tognini-Bonelli (2001:17) advocates, set rules hardly reflect real instances of language use. Thus, a new approach that takes into account the close interrelationship between syntax and the lexicon (Hunston & Sinclair 2000; Hunston & Francis 2000; Hunston 2002) and provides examples of real language in context seems extremely beneficial for both language teaching and learning.

In this view, I maintain that corpus linguistics as a methodology can help a great deal to compensate the existing mismatch between traditional descriptions and real language instances. As a corpus-based research study, this thesis examines the information drawn from corpus evidence, from which some generalisations regarding the recurrent lexico-grammatical patterning of abstract nouns in the medical academic texts under study will be made.

---

[23] This term is used following McEnery et al.'s (2006) perspective; that is, in a general sense; without making a sharp distinction between corpus-based and corpus-driven approaches.

It has been long acknowledged (Sinclair 1991; Hunston & Francis 2000; Gledhill 2000; Verdaguer 2003; among others) that when encoding a message it is necessary not only to use the correct terms and grammatical structures, but also the appropriate lexical combinations and collocations. The same can also be applied to the description and analysis of a given word. As has already been noted, the use of corpora enables both the researcher and the teacher to have access to a large amount of naturally occurring language: language in use. Within this framework, words can be analysed as part of a chain, which allows the direct observation, identification and even generalization of the most frequent collocates of node words. Besides, the analysis of a large body of real language in context also provides useful information with regard to the different grammatical structures and semantic categories in which a given sense might occur.

Having discussed the concept of corpus linguistics as a methodology as well as the rationale of the proposed corpus-based approach for linguistic description and analysis, we now turn to section 3.2 where the compiled corpus is presented.

3.2 Corpus data

This section presents the corpus on which this research is based as well as the principles of corpus design adopted in its compilation. Likewise, it explores the software employed in retrieving data from the selected compilation of texts. The main corpus analysed in this study is the *Health Science Corpus* (*HSC*), which is a representative sample of texts specifically assembled for the current investigation of the use of abstract nouns by the health science community.

> "[…] corpus analysis presents considerable methodological advantages for a description of languages for specific purposes. In the first instance, the rhetorical aims of the writers are known and can be prioritised in the analysis: this is not an anonymous collection of texts. In addition, we have seen that while there are many studies of phraseology and lexico-grammar in the general

> language, few specialist varieties have benefited from a large-scale corpus analysis of this kind." (Gledhill 2000:90)

Following Gledhill's view, I strongly believe in the methodological potentiality of corpus analysis for the description of language for specific purposes. Interested as we were and still are in the lexico-grammatical patterns of non-technical terms in scientific English and the conventionalized phraseological characteristics of that genre, and bearing in mind that there was no corpus of scientific English publicly available[24], our research group[25] decided to compile our own micro-corpus, now consisting of approximately 4 million words (see Figure 1 below) of scientific research articles from prestige online journals[26] that cover different disciplines such as medicine, biology, biochemistry and biomedicine.[27]

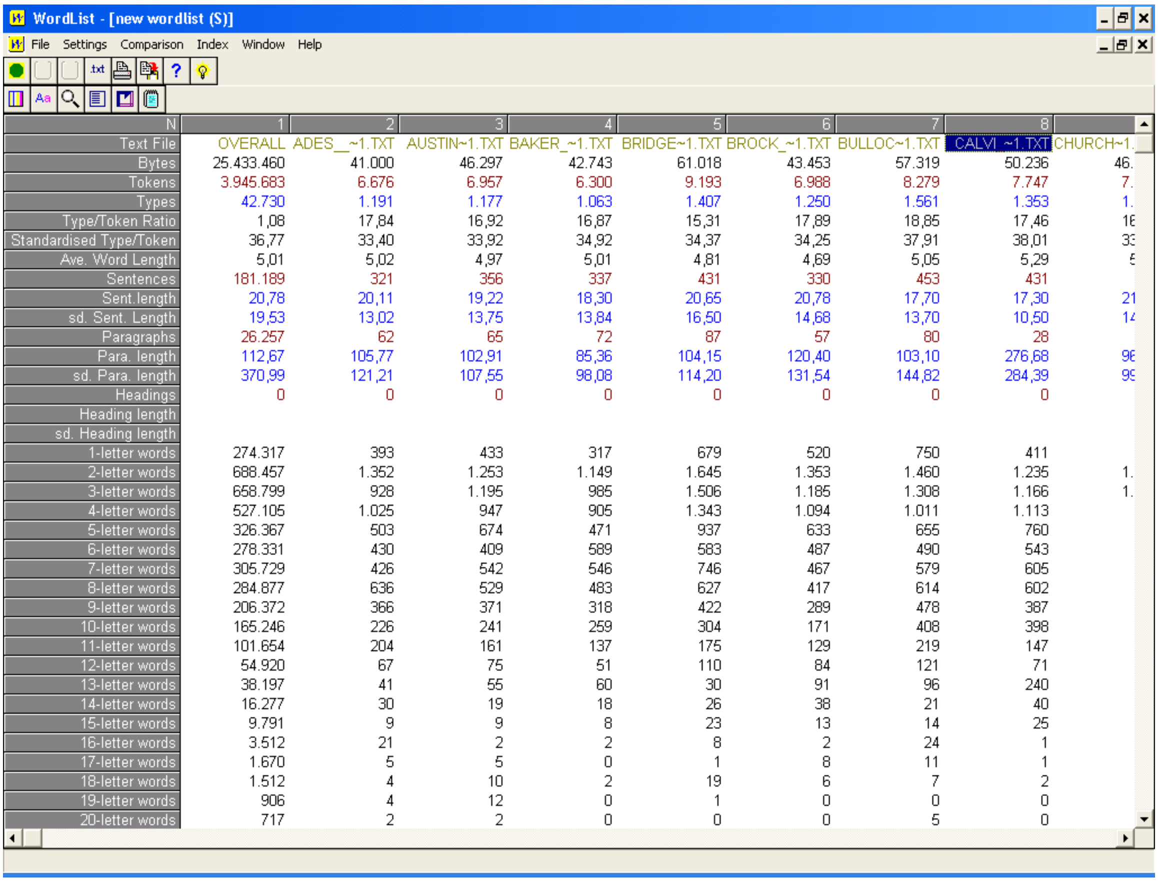
---

[24] The Professional English Research Consortium (PERC) has developed a 100 million word corpus of Professional English, which has been recently released.

[25] The data corpus was compiled as part of a research project, 'Creation of a Database of Lexical Combinations in Scientific English", financed by the Spanish Ministry of Science and Education and FEDER (BFF2001-2988). Within this project, co-ordinated by Dr. Isabel Verdaguer (University of Barcelona), the research team developed *SciE-Lex*, a dictionary that provides contextual information on usage as well as the combinatorial potential of words commonly used in scientific registers.

[26] All these articles correspond to the years 1998 and 1999.

[27] Biology: *Genes and Development* (40 articles), *Genetics* (54 articles), *Journal of Cell Biology* (26 articles), *The American Journal of Primatology* (20 articles), *Biological Control* (97 articles), *The Journal of Experimental Zoology* (32 articles), *BioEssays* (99 articles), *Integrative Biology* (18 articles), *Zoo Biology* (65 articles); biochemistry: *Biochemical Journal* (53 articles), *The Embo* Journal (64 articles); medicine: *Journal of Clinical Investigation* (53 articles), *British Medical Journal* (58 articles), *Journal of Bacteriology* (39 articles). It is worth mentioning that the informants in our study (cf. section 3.3.2) regarded these journals as very relevant to the scientific community.



N	1	2	3	4	5	6	7	8	
Text File	OVERALL	ADES__~1.TXT	AUSTIN~1.TXT	BAKER_~1.TXT	BRIDGE~1.TXT	BROCK_~1.TXT	BULLOC~1.TXT	CALVI ~1.TXT	CHURCH~1.
Bytes	25.433.460	41.000	46.297	42.743	61.018	43.453	57.319	50.236	46.
Tokens	3.945.683	6.676	6.957	6.300	9.193	6.988	8.279	7.747	7.
Types	42.730	1.191	1.177	1.063	1.407	1.250	1.561	1.353	1.
Type/Token Ratio	1,08	17,84	16,92	16,87	15,31	17,89	18,85	17,46	1
Standardised Type/Token	36,77	33,40	33,92	34,92	34,37	34,25	37,91	38,01	3
Ave. Word Length	5,01	5,02	4,97	5,01	4,81	4,69	5,05	5,29	
Sentences	181.189	321	356	337	431	330	453	431	
Sent.length	20,78	20,11	19,22	18,30	20,65	20,78	17,70	17,30	21
sd. Sent. Length	19,53	13,02	13,75	13,84	16,50	14,68	13,70	10,50	1
Paragraphs	26.257	62	65	72	87	57	80	28	
Para. length	112,67	105,77	102,91	85,36	104,15	120,40	103,10	276,68	9
sd. Para. length	370,99	121,21	107,55	98,08	114,20	131,54	144,82	284,39	9
Headings	0	0	0	0	0	0	0	0	
Heading length									
sd. Heading length									
1-letter words	274.317	393	433	317	679	520	750	411	
2-letter words	688.457	1.352	1.253	1.149	1.645	1.353	1.460	1.235	1.
3-letter words	658.799	928	1.195	985	1.506	1.185	1.308	1.166	1.
4-letter words	527.105	1.025	947	905	1.343	1.094	1.011	1.113	
5-letter words	326.367	503	674	471	937	633	655	760	
6-letter words	278.331	430	409	589	583	487	490	543	
7-letter words	305.729	426	542	546	746	467	579	605	
8-letter words	284.877	636	529	483	627	417	614	602	
9-letter words	206.372	366	371	318	422	289	478	387	
10-letter words	165.246	226	241	259	304	171	408	398	
11-letter words	101.654	204	161	137	175	129	219	147	
12-letter words	54.920	67	75	51	110	84	121	71	
13-letter words	38.197	41	55	60	30	91	96	240	
14-letter words	16.277	30	19	18	26	38	21	40	
15-letter words	9.791	9	9	8	23	13	14	25	
16-letter words	3.512	21	2	2	8	2	24	1	
17-letter words	1.670	5	5	0	1	8	11	1	
18-letter words	1.512	4	10	2	19	6	7	2	
19-letter words	906	4	12	0	1	0	0	0	
20-letter words	717	2	2	0	0	0	5	0	

Figure 1 WordList of the *HSC* provided by *WordSmith Tools*[28]

Taking into account the fact that all the disciplines it embraces fall into the category of health sciences, it came to be called the *Health Science Corpus* (*HSC*). As will be seen, the *HSC* comprises a collection of written material produced by native speakers of English put together according to a common corpus design. It is thus necessary at this stage to set out the criteria followed in its building up. Gries & Stefanowitsch (2006:4) propose some principles of corpus design which were adopted in the compilation of the *HSC*:

a) The present study is based on naturally-occurring language in machine-readable form.
b) The retrieval of search patterns was made by means of linguistic software tools.
c) The *HSC* is intended to be a representative sample of the research article in the health science register this research is focused on.

---

[28] © Mike Scott & Oxford University Press, 1998 (http://www.lexically.net/wordsmith/)

d) The whole collection of running text (rather than a random selection of examples) was considered.
e) The linguistic analysis undertaken, which will be thoroughly described in section 4.1, proceeds on the basis of frequency lists of the abstract nouns under study, concordance lines that display those nouns in their natural context of appearance, and collocations: that is, the actual words that co-occur with the nouns of interest to the study.

Once the corpus was compiled, all scientific articles, a total of 718, were fully edited to smooth out problems concerning the typographical form of texts. Every downloaded text presented problems with capital letters, paragraphing, diagrams, numbers, photographs, columnar layouts, etc. Thus, they were edited manually, which took us a great deal of time and effort, converted into plain text files and then stored into different folders and subfolders. All these steps were prior to the data processing stage.

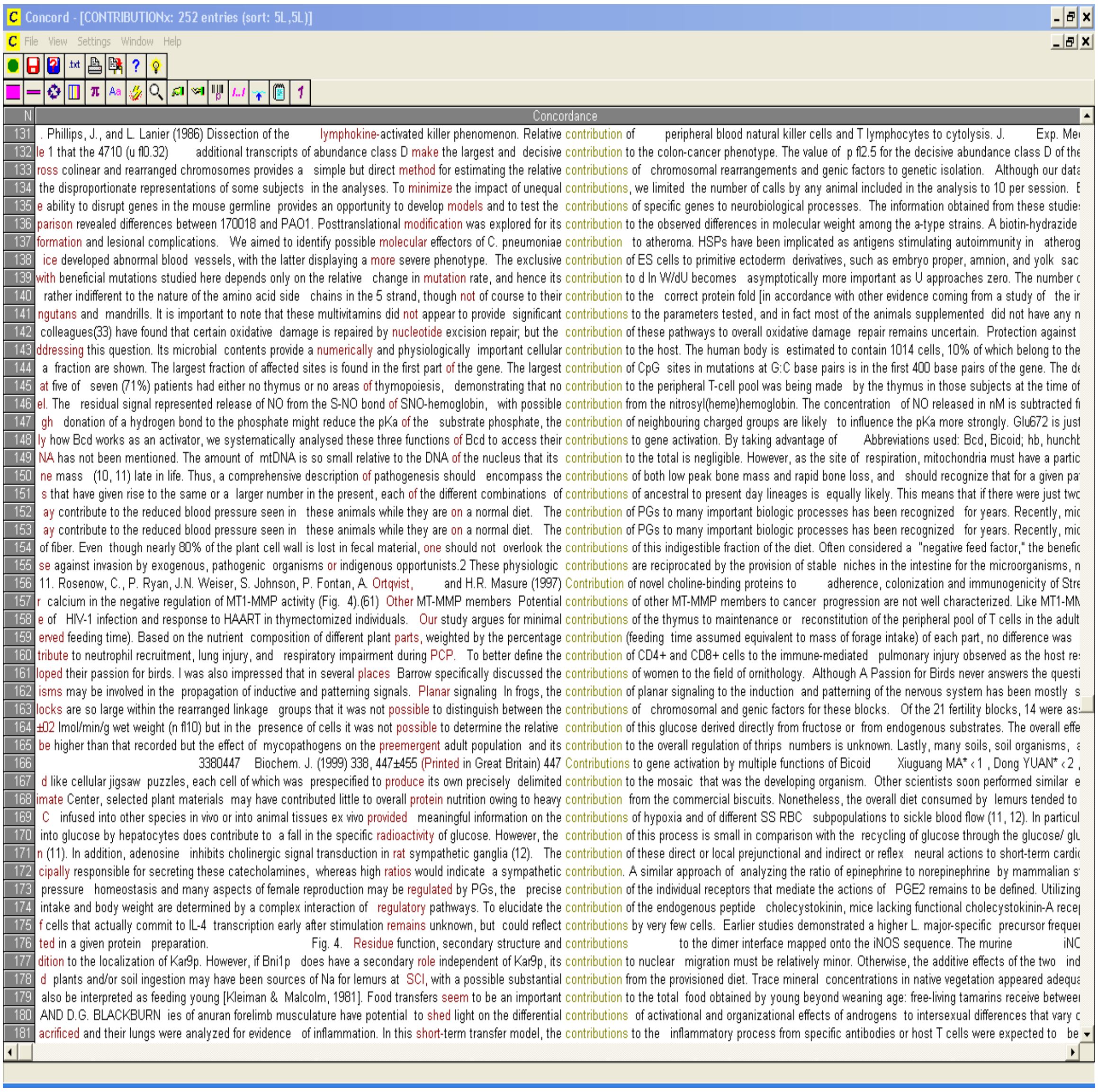
Computerised corpora and linguistic software tools are essential for linguistic data management. The processing of a corpus by means of computerised methods has proven to be a very useful tool for the researcher to process in real time large quantities of texts, which had been otherwise completely impossible. There are several software programmes at the lexicographer's disposal to store a corpus electronically. The software used in retrieving data from the *HSC* and refining the results further was version 3.0 of the concordancing program *WordSmith Tools*[29]. This program provided us with a list of words, which allowed us to find out what general terms are most frequently used in scientific English. Taking such a list as the starting point, I selected five abstract nouns, based on frequency counts, for the present study.

WordSmith software also supplied a list of concordance lines of the chosen words that made it possible to identify the patterns of word co-occurrence and their frequency of appearance. As Figure 2 shows below, this software displays

---

[29] Ibid.

KWIC[30] concordance lines within a collocation span of 5 words to the left and 5 words to the right. As Clear (1987:41) observes, the value of the information provided by this tool should not be underestimated since it saves the researcher a great deal of time and brings out in a few seconds patterns of word co-occurrence that could have easily passed unnoticed. Such a tool also enables to have a quick access to the source text each line comes from by right clicking on “View Text” (cf. Figure 3).

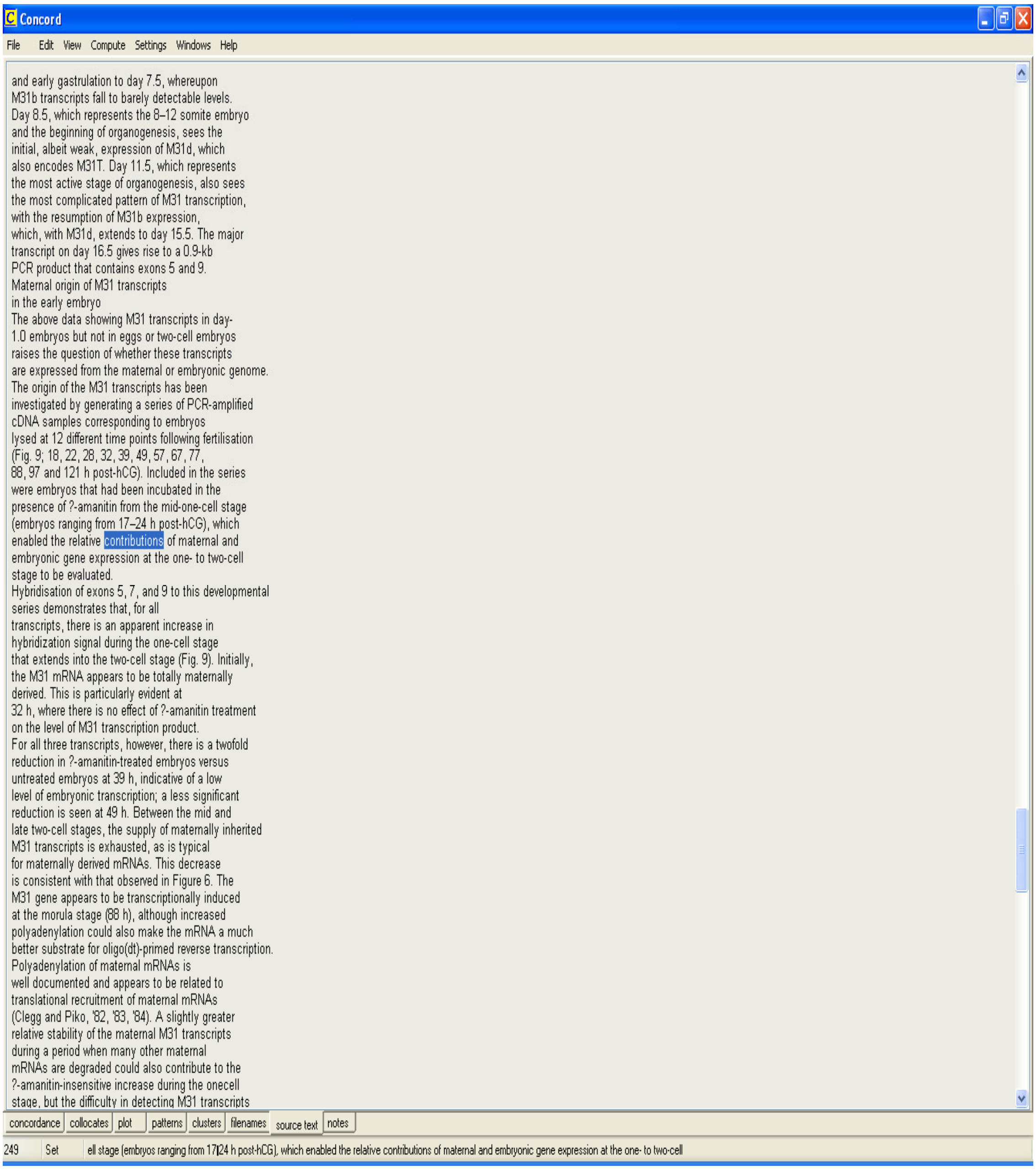


```
131 . Phillips, J., and L. Lanier (1986) Dissection of the      lymphokine-activated killer phenomenon. Relative contribution of      peripheral blood natural killer cells and T lymphocytes to cytolysis. J.      Exp. Me
132 le 1 that the 4710 (u fl0.32)      additional transcripts of abundance class D make the largest and  decisive contribution to the colon-cancer phenotype. The value of  p fl2.5 for the decisive abundance class D of the
133 ross colinear and rearranged chromosomes provides a   simple but direct method for estimating the relative contributions of   chromosomal rearrangements and genic factors to genetic isolation.   Although our data
134 the disproportionate representations of some subjects  in the analyses. To minimize the impact of unequal contributions, we limited  the number of calls by any animal included in the analysis to 10 per session.  E
135 e ability to disrupt genes in the mouse germline  provides an opportunity to develop models and to test the  contributions of specific genes to neurobiological processes.  The information obtained from these studie
136 parison revealed differences between 170018 and PAO1. Posttranslational modification was explored for its contribution to the observed differences in molecular weight among the a-type strains. A biotin-hydrazide
137 formation and lesional complications.   We aimed to identify possible molecular effectors of C. pneumoniae contribution   to atheroma. HSPs have been implicated as antigens stimulating autoimmunity in   atherog
138  ice developed abnormal blood  vessels, with the latter displaying a more severe phenotype.  The exclusive contribution of ES cells to primitive ectoderm  derivatives, such as embryo proper, amnion, and yolk  sac
139 with beneficial mutations studied here depends only on the relative   change in mutation rate, and hence its contribution to d In W/dU becomes   asymptotically more important as U approaches zero. The number c
140  rather indifferent to the nature of the amino acid side   chains in the 5 strand, though not of course to their contribution to the   correct protein fold [in accordance with other evidence coming from a study of   the ir
141 ngutans and  mandrills. It is important to note that these multivitamins did not appear to provide  significant contributions to the parameters tested, and in fact most of the animals supplemented  did not have any n
142 colleagues(33) have found that certain oxidative  damage is repaired by nucleotide excision repair; but the  contribution of these pathways to overall oxidative damage  repair remains uncertain.  Protection against
143 ddressing this question. Its microbial  contents provide a numerically and physiologically  important cellular contribution to the host. The human body is  estimated to contain 1014 cells, 10% of which belong to the
144  a  fraction are shown. The largest fraction of affected sites is found in the first part of the gene. The largest contribution of CpG  sites in mutations at G:C base pairs is in the first 400 base pairs of the gene. The de
145  at five of   seven (71%) patients had either no thymus or no areas of thymopoiesis,   demonstrating that no contribution to the peripheral T-cell pool was being made   by the thymus in those subjects at the time of
146 el. The   residual signal represented release of NO from the S-NO bond of SNO-hemoglobin,   with possible contribution from the nitrosyl(heme)hemoglobin. The concentration   of NO released in nM is subtracted fr
147  gh   donation of a hydrogen bond to the phosphate might reduce the pKa of the   substrate phosphate, the contribution of neighbouring charged groups are likely   to influence the pKa more strongly. Glu672 is just
148 ly how Bcd works as an activator, we systematically analysed these three functions of Bcd to access their contributions to gene activation. By taking advantage of      Abbreviations used: Bcd, Bicoid; hb, hunchb
149  NA has not been mentioned. The amount of  mtDNA is so small relative to the DNA of the nucleus that its  contribution to the total is negligible. However, as the site of  respiration, mitochondria must have a partic
150  ne mass   (10, 11) late in life. Thus, a comprehensive description of pathogenesis should   encompass the contributions of both low peak bone mass and rapid bone loss, and   should recognize that for a given pa
151  s that have given rise to the same or a  larger number in the present, each of the different combinations  of contributions of ancestral to present day lineages is  equally likely. This means that if there were just two
152  ay contribute to the reduced blood pressure seen in   these animals while they are on a normal diet.   The contribution of PGs to many important biologic processes has been recognized   for years. Recently, mic
153  ay contribute to the reduced blood pressure seen in   these animals while they are on a normal diet.   The contribution of PGs to many important biologic processes has been recognized   for years. Recently, mic
154 of fiber. Even  though nearly 80% of the plant cell wall is lost in fecal material, one should not  overlook the contributions of this indigestible fraction of the diet. Often considered a  "negative feed factor," the benefic
155  se against invasion by exogenous, pathogenic  organisms or indigenous opportunists.2 These physiologic  contributions are reciprocated by the provision of stable  niches in the intestine for the microorganisms, n
156 11. Rosenow, C., P. Ryan, J.N. Weiser, S. Johnson, P. Fontan, A. Ortqvist,      and H.R. Masure (1997) Contribution of novel choline-binding proteins to      adherence, colonization and immunogenicity of Stre
157 r  calcium in the negative regulation of MT1-MMP activity (Fig.  4).(61)  Other MT-MMP members  Potential contributions of other MT-MMP members to cancer  progression are not well characterized. Like MT1-MM
158 e of   HIV-1 infection and response to HAART in thymectomized individuals.   Our study argues for minimal contributions of the thymus to maintenance or   reconstitution of the peripheral pool of T cells in the adult
159  erved feeding time). Based on the nutrient  composition of different plant parts, weighted by the percentage contribution (feeding  time assumed equivalent to mass of forage intake) of each part, no difference was
160 tribute to neutrophil recruitment, lung injury, and   respiratory impairment during PCP.   To better define the contribution of CD4+ and CD8+ cells to the immune-mediated   pulmonary injury observed as the host re
161 loped their passion for birds. I was also impressed that in several places  Barrow specifically discussed the contributions of women to the field of ornithology.   Although A Passion for Birds never answers the questi
162  isms may be involved in the  propagation of inductive and patterning signals.  Planar signaling  In frogs, the contribution of planar signaling to the induction  and patterning of the nervous system has been mostly  s
163 locks are so large within the rearranged linkage   groups that it was not possible to distinguish between the contributions of   chromosomal and genic factors for these blocks.   Of the 21 fertility blocks, 14 were as
164 ±02 Imol/min/g wet weight (n fl10) but in the  presence of cells it was not possible to determine the relative  contribution of this glucose derived directly from fructose or  from endogenous substrates. The overall effe
165  be higher than that recorded but the effect of  mycopathogens on the preemergent adult population  and its contribution to the overall regulation of thrips  numbers is unknown. Lastly, many soils, soil organisms,  a
166                               3380447    Biochem. J. (1999) 338, 447±455 (Printed in Great Britain) 447 Contributions to gene activation by multiple functions of Bicoid      Xiuguang MA* ‹ 1 , Dong YUAN* ‹ 2 ,
167  d like cellular jigsaw  puzzles, each cell of which was  prespecified to produce its own precisely  delimited contribution to the mosaic  that was the developing organism.  Other scientists soon performed similar  e
168 imate Center, selected plant materials  may have contributed little to overall protein nutrition owing to heavy contribution  from the commercial biscuits. Nonetheless, the overall diet consumed by  lemurs tended to
169  C   infused into other species in vivo or into animal tissues ex vivo provided   meaningful information on the contributions of hypoxia and of different SS RBC   subpopulations to sickle blood flow (11, 12). In particul
170  into glucose by hepatocytes does contribute to  a fall in the specific radioactivity of glucose. However, the  contribution of this process is small in comparison with the  recycling of glucose through the glucose/ glu
171 n (11). In addition, adenosine   inhibits cholinergic signal transduction in rat sympathetic ganglia (12).   The contribution of these direct or local prejunctional and indirect or reflex   neural actions to short-term cardio
172  cipally responsible for secreting these catecholamines,  whereas high ratios would indicate  a sympathetic contribution. A similar approach of  analyzing the ratio of epinephrine to norepinephrine  by mammalian s
173  pressure   homeostasis and many aspects of female reproduction may be regulated by PGs, the   precise contribution of the individual receptors that mediate the actions of   PGE2 remains to be defined. Utilizing
174  intake and body weight are determined by a complex interaction of   regulatory pathways. To elucidate the contribution of the endogenous peptide   cholecystokinin, mice lacking functional cholecystokinin-A recep
175  f cells that actually commit to IL-4  transcription early after stimulation remains unknown, but  could reflect contributions by very few cells.  Earlier studies demonstrated a higher L. major-specific  precursor frequen
176  ted in a given protein   preparation.               Fig. 4.   Residue function, secondary structure and contributions         to the dimer interface mapped onto the iNOS sequence. The murine         iNO
177 dition to the localization of Kar9p. However, if Bni1p   does have a secondary role independent of Kar9p, its contribution to nuclear   migration must be relatively minor. Otherwise, the additive effects of the two   ind
178  d  plants and/or soil ingestion may have been sources of Na for lemurs at  SCI, with a possible substantial contribution from the provisioned diet. Trace mineral  concentrations in native vegetation appeared adequa
179  also be interpreted as feeding young [Kleiman &  Malcolm, 1981]. Food transfers seem to be an important contribution to the total  food obtained by young beyond weaning age: free-living tamarins receive betwee
180  AND D.G. BLACKBURN  ies of anuran forelimb musculature have potential  to shed light on the differential contributions  of activational and organizational effects of androgens  to intersexual differences that vary c
181  acrificed and their lungs were analyzed for evidence   of inflammation. In this short-term transfer model, the contributions to the   inflammatory process from specific antibodies or host T cells were expected to   be
```

Figure 2 Concordance lines for the node word *contribution* provided by *WordSmith Tools*

---

[30] *Key Word in Context*; that is, a computer-generated set of concordance lines in which the node word is centred in each line.



and early gastrulation to day 7.5, whereupon
M31b transcripts fall to barely detectable levels.
Day 8.5, which represents the 8–12 somite embryo
and the beginning of organogenesis, sees the
initial, albeit weak, expression of M31d, which
also encodes M31T. Day 11.5, which represents
the most active stage of organogenesis, also sees
the most complicated pattern of M31 transcription,
with the resumption of M31b expression,
which, with M31d, extends to day 15.5. The major
transcript on day 16.5 gives rise to a 0.9-kb
PCR product that contains exons 5 and 9.
Maternal origin of M31 transcripts
in the early embryo
The above data showing M31 transcripts in day-
1.0 embryos but not in eggs or two-cell embryos
raises the question of whether these transcripts
are expressed from the maternal or embryonic genome.
The origin of the M31 transcripts has been
investigated by generating a series of PCR-amplified
cDNA samples corresponding to embryos
lysed at 12 different time points following fertilisation
(Fig. 9; 18, 22, 28, 32, 39, 49, 57, 67, 77,
88, 97 and 121 h post-hCG). Included in the series
were embryos that had been incubated in the
presence of ?-amanitin from the mid-one-cell stage
(embryos ranging from 17–24 h post-hCG), which
enabled the relative contributions of maternal and
embryonic gene expression at the one- to two-cell
stage to be evaluated.
Hybridisation of exons 5, 7, and 9 to this developmental
series demonstrates that, for all
transcripts, there is an apparent increase in
hybridization signal during the one-cell stage
that extends into the two-cell stage (Fig. 9). Initially,
the M31 mRNA appears to be totally maternally
derived. This is particularly evident at
32 h, where there is no effect of ?-amanitin treatment
on the level of M31 transcription product.
For all three transcripts, however, there is a twofold
reduction in ?-amanitin-treated embryos versus
untreated embryos at 39 h, indicative of a low
level of embryonic transcription; a less significant
reduction is seen at 49 h. Between the mid and
late two-cell stages, the supply of maternally inherited
M31 transcripts is exhausted, as is typical
for maternally derived mRNAs. This decrease
is consistent with that observed in Figure 6. The
M31 gene appears to be transcriptionally induced
at the morula stage (88 h), although increased
polyadenylation could also make the mRNA a much
better substrate for oligo(dt)-primed reverse transcription.
Polyadenylation of maternal mRNAs is
well documented and appears to be related to
translational recruitment of maternal mRNAs
(Clegg and Piko, '82, '83, '84). A slightly greater
relative stability of the maternal M31 transcripts
during a period when many other maternal
mRNAs are degraded could also contribute to the
?-amanitin-insensitive increase during the onecell
stage, but the difficulty in detecting M31 transcripts

249 Set ell stage (embryos ranging from 17|24 h post-hCG), which enabled the relative contributions of maternal and embryonic gene expression at the one- to two-cell

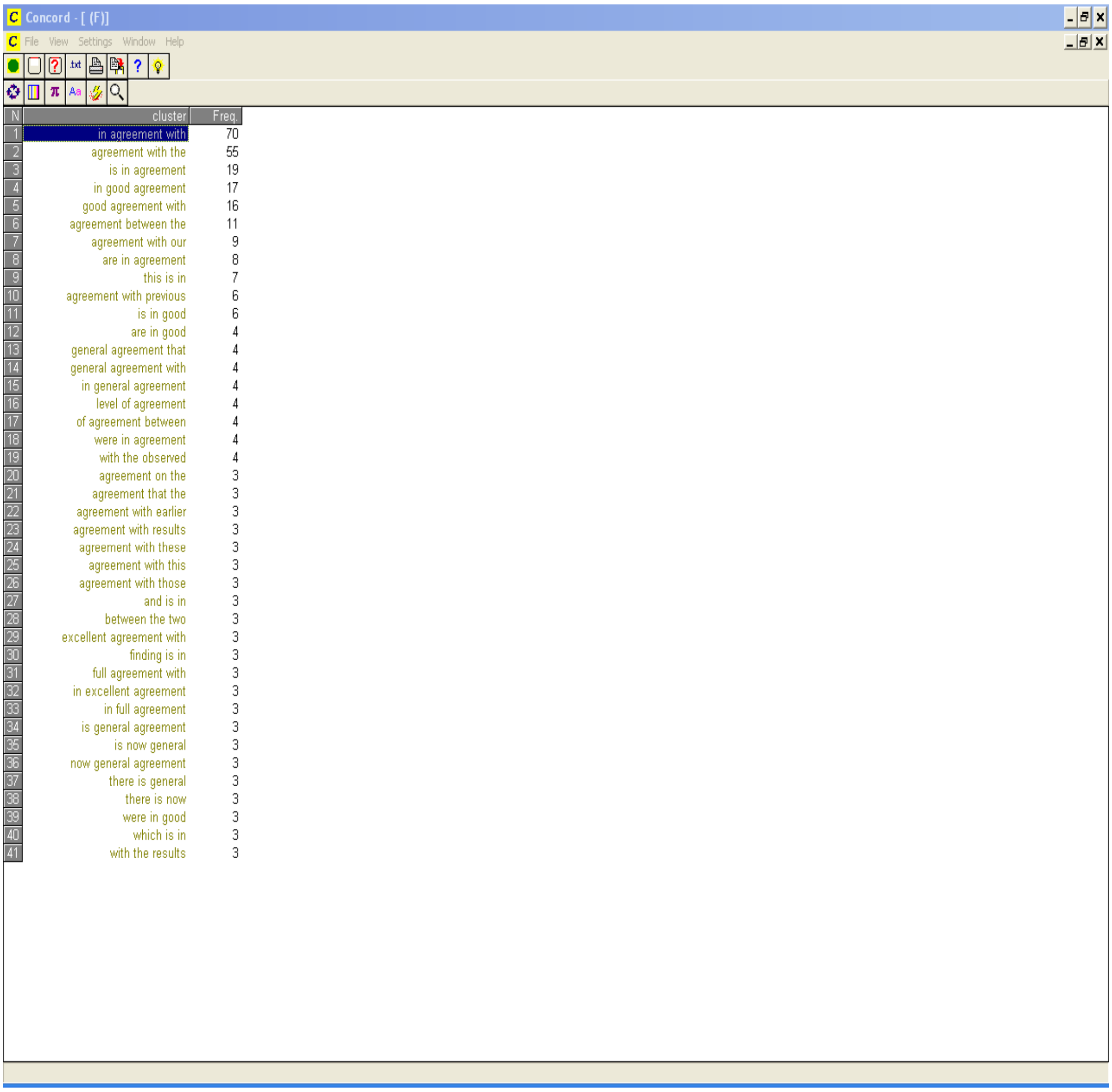
Figure 3 Source text of a concordance line of the node word *contribution* displayed by *WordSmith Tools*

Among the various tools available, it should be mentioned the fact that concordance lines can be sorted, edited, exported, saved as plain text format and printed:

> “The fact that the concordance will be read vertically, looking for the patterns of word co-selection entertained by the node with its context, shifts the focus from a model which is primarily paradigmatic and emphasises the freedom of choice to one where the syntagmatic units established in the

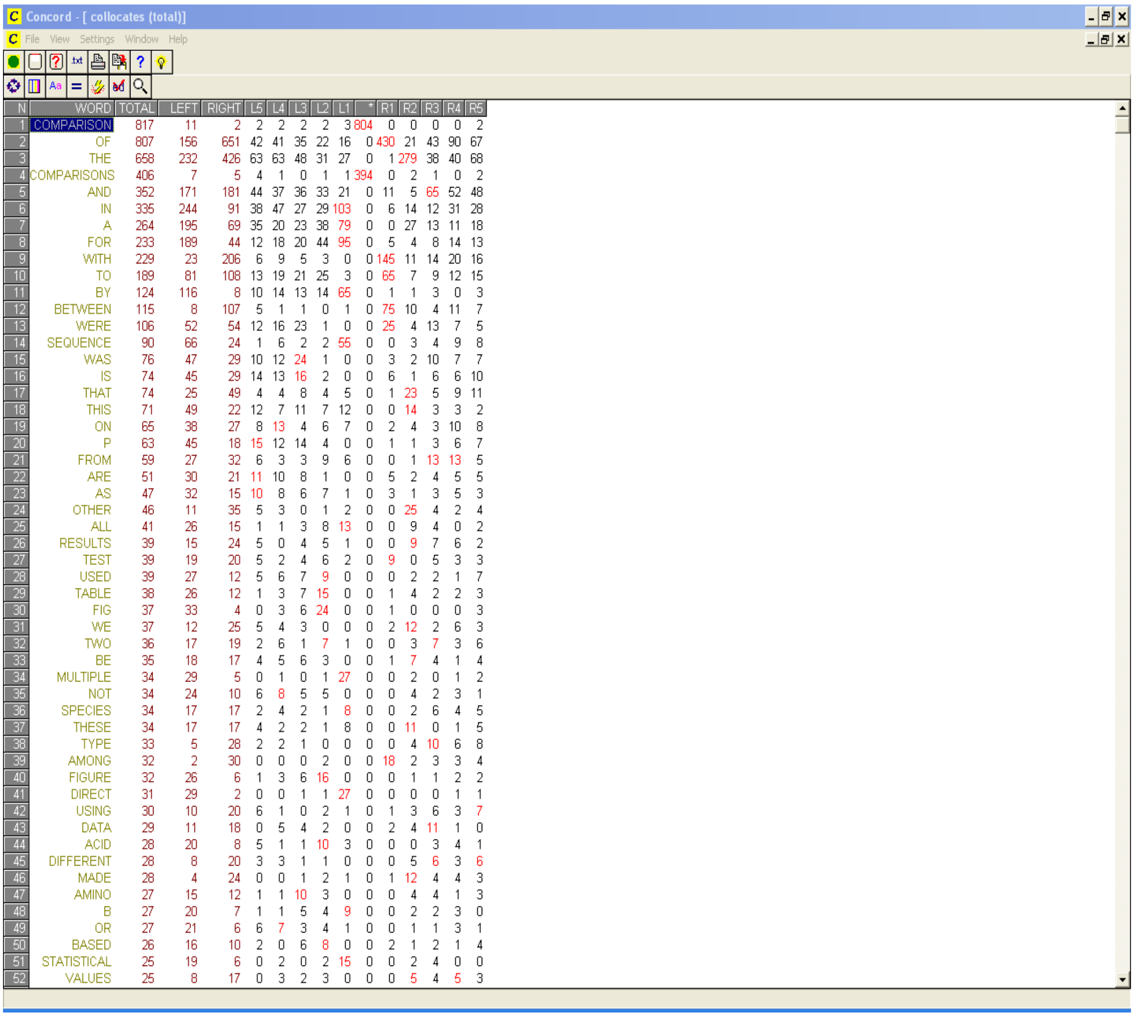
> light of repetition on the vertical axis of the concordance take precedence.” (Tognini-Bonelli 2001:18)

Furthermore, this software was extremely useful not only to identify collocates and their frequency, but also to classify such collocates in terms of grammatical position, word class, semantic category, etc., to analyse word clusters so as to see the patterns of repeated phraseology in the concordance lines analysed and to make generalisations from the observation of repeated language events. The following figures show illustrative screenshots of these already mentioned tools:



N	cluster	Freq.
1	in agreement with	70
2	agreement with the	55
3	is in agreement	19
4	in good agreement	17
5	good agreement with	16
6	agreement between the	11
7	agreement with our	9
8	are in agreement	8
9	this is in	7
10	agreement with previous	6
11	is in good	6
12	are in good	4
13	general agreement that	4
14	general agreement with	4
15	in general agreement	4
16	level of agreement	4
17	of agreement between	4
18	were in agreement	4
19	with the observed	4
20	agreement on the	3
21	agreement that the	3
22	agreement with earlier	3
23	agreement with results	3
24	agreement with these	3
25	agreement with this	3
26	agreement with those	3
27	and is in	3
28	between the two	3
29	excellent agreement with	3
30	finding is in	3
31	full agreement with	3
32	in excellent agreement	3
33	in full agreement	3
34	is general agreement	3
35	is now general	3
36	now general agreement	3
37	there is general	3
38	there is now	3
39	were in good	3
40	which is in	3
41	with the results	3

Figure 4 Clusters of the node word *agreement* as displayed by *WordSmith Tools*



N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	COMPARISON	817	11	2	2	2	2	2	3	804	0	0	0	0	2
2	OF	807	156	651	42	41	35	22	16	0	430	21	43	90	67
3	THE	658	232	426	63	63	48	31	27	0	1	279	38	40	68
4	COMPARISONS	406	7	5	4	1	0	1	1	394	0	2	1	0	2
5	AND	352	171	181	44	37	36	33	21	0	11	5	65	52	48
6	IN	335	244	91	38	47	27	29	103	0	6	14	12	31	28
7	A	264	195	69	35	20	23	38	79	0	0	27	13	11	18
8	FOR	233	189	44	12	18	20	44	95	0	5	4	8	14	13
9	WITH	229	23	206	6	9	5	3	0	0	145	11	14	20	16
10	TO	189	81	108	13	19	21	25	3	0	65	7	9	12	15
11	BY	124	116	8	10	14	13	14	65	0	1	1	3	0	3
12	BETWEEN	115	8	107	5	1	1	0	1	0	75	10	4	11	7
13	WERE	106	52	54	12	16	23	1	0	0	25	4	13	7	5
14	SEQUENCE	90	66	24	1	6	2	2	55	0	0	3	4	9	8
15	WAS	76	47	29	10	12	24	1	0	0	3	2	10	7	7
16	IS	74	45	29	14	13	16	2	0	0	6	1	6	6	10
17	THAT	74	25	49	4	4	8	4	5	0	1	23	5	9	11
18	THIS	71	49	22	12	7	11	7	12	0	0	14	3	3	2
19	ON	65	38	27	8	13	4	6	7	0	2	4	3	10	8
20	P	63	45	18	15	12	14	4	0	0	1	1	3	6	7
21	FROM	59	27	32	6	3	3	9	6	0	0	1	13	13	5
22	ARE	51	30	21	11	10	8	1	0	0	5	2	4	5	5
23	AS	47	32	15	10	8	6	7	1	0	3	1	3	5	3
24	OTHER	46	11	35	5	3	0	1	2	0	0	25	4	2	4
25	ALL	41	26	15	1	1	3	8	13	0	0	9	4	0	2
26	RESULTS	39	15	24	5	0	4	5	1	0	0	9	7	6	2
27	TEST	39	19	20	5	2	4	6	2	0	9	0	5	3	3
28	USED	39	27	12	5	6	7	9	0	0	0	2	2	1	7
29	TABLE	38	26	12	1	3	7	15	0	0	1	4	2	2	3
30	FIG	37	33	4	0	3	6	24	0	0	1	0	0	0	3
31	WE	37	12	25	5	4	3	0	0	0	2	12	2	6	3
32	TWO	36	17	19	2	6	1	7	1	0	0	3	7	3	6
33	BE	35	18	17	4	5	6	3	0	0	1	7	4	1	4
34	MULTIPLE	34	29	5	0	1	0	1	27	0	0	2	0	1	2
35	NOT	34	24	10	6	8	5	5	0	0	0	4	2	3	1
36	SPECIES	34	17	17	2	4	2	1	8	0	0	2	6	4	5
37	THESE	34	17	17	4	2	2	1	8	0	0	11	0	1	5
38	TYPE	33	5	28	2	2	1	0	0	0	0	4	10	6	8
39	AMONG	32	2	30	0	0	0	2	0	0	18	2	3	3	4
40	FIGURE	32	26	6	1	3	6	16	0	0	0	1	1	2	2
41	DIRECT	31	29	2	0	0	1	1	27	0	0	0	0	1	1
42	USING	30	10	20	6	1	0	2	1	0	1	3	6	3	7
43	DATA	29	11	18	0	5	4	2	0	0	2	4	11	1	0
44	ACID	28	20	8	5	1	1	10	3	0	0	0	3	4	1
45	DIFFERENT	28	8	20	3	3	1	1	0	0	0	5	6	3	6
46	MADE	28	4	24	0	0	1	2	1	0	1	12	4	4	3
47	AMINO	27	15	12	1	1	10	3	0	0	0	4	4	1	3
48	B	27	20	7	1	1	5	4	9	0	0	2	2	3	0
49	OR	27	21	6	6	7	3	4	1	0	0	1	1	3	1
50	BASED	26	16	10	2	0	6	8	0	0	2	1	2	1	4
51	STATISTICAL	25	19	6	0	2	0	2	15	0	0	2	4	0	0
52	VALUES	25	8	17	0	3	2	3	0	0	0	5	4	5	3

Figure 5 Collocates of the node word *comparison**

When dealing with the data extracted from the worksheet of exercises (section 5.1) and comparing them against the *HSC*, some methodological decisions were taken. Due to the fact that in some cases[31] very modest conclusions could be drawn from comparing the *HSC* and the informants' data, I decided to use the *British National Corpus* (*BNC*)[32] as a third comparison measure. Being a 100 million word collection of samples of written and spoken general language from a wide range of sources, the *BNC* provided a higher context of co-occurrence and enabled me, whenever needed, to supplement my selected sample of contemporary medical texts. Some word combinations produced by non-native speakers in the worksheet of exercises were also compared against the Birmingham's *Bank of English* (*BoE*), currently a 450 million word corpus of present-day English.

---

[31] This was especially noticeable in Exercise 1, where data revealed a low percentage of occurrence of some evaluative adjectives in combination with abstract nouns in the *HSC* (see Table 29 in section 5.1.3)

[32] http://www.natcorp.ox.ac.uk/

It should also be stressed that while this analysis is based on a large dataset, my study is still more qualitative in nature than quantitative. It was never the main aim of my investigation to present a purely quantitative account of the use of abstract nouns in both the *HSC* (cf. sections 4.3-4.7) and the informants' production (cf. section 5.1), but instead to focus mainly on their similarities and differences of use from a more qualitative perspective.

As pointed out above that one of the main aims when compiling the *HSC* was trying to make a representative selection of naturally-occurring language in a very specific type of genre, the health science discourse, so as to analyse the collocations and syntagmatic structures associated with abstract nouns in that particular register. To this respect, it seems worth-recalling the notion of "local grammar" (Barnbrook & Sinclair 1995; Hunston & Sinclair 2000; Hunston 2002), which refers to descriptions of particular areas of language (rather than the language as a whole): "*the connection between pattern and meaning opens the possibility of quantifying ways of expressing meanings in different registers via the concept of 'local grammar.'*" (Hunston 2002:178).

The compilation of the *HSC* thus understood as "*an authoritative body of linguistic evidence which can support generalizations and against which hypotheses can be tested*" (Sinclair 1987:2) has facilitated the exploration of the phraseological behaviour of abstract nouns in medical English, with respect to patterning by means of corpus evidence. For many research and pedagogical purposes, I do believe that the larger the corpus is, the more reliable conclusions can be drawn from the careful examination of the language shown. Nevertheless, it must be stressed that I am fully aware of the fact that the *HSC* constitutes a sample[33], a cross-section of the health science discourse[34], so claims will be just based on the results obtained from the in-depth analysis of the *HSC* data. As Partington (1998) observes: "*a corpus, no matter how large and varied, is only representative of itself and claims made about the behaviour of linguistic items after studying corpus data should bear this in mind*" (Partington 1998:146). Following Partington's consideration, Hunston (2002) notes that all observations made from a particular collection of texts "*must be dealt with as deductions rather than as facts*". (Hunston 2002:23)

---

[33] This is a central idea in corpus linguists

[34] In Teubert's & Cěrmáková's words: "*we are only justified in claiming that a given corpus is representative of a discourse, however we have defined it, if we have, at least in principle, access to all the texts the discourse consists of.*" (Teubert & Cěrmáková 2004: 117).

3.3 The Survey

This section concerns the data collection method used in this investigation, which consists of a survey conducted among the chosen medical community. A total of twenty-four Spanish doctors from three Spanish hospitals were asked to fill in a survey of ten open questions in which they were required to provide different sorts of information. The main aim of that survey was to gather information on the participants' profile, henceforth referred to as the Survey Profile Questions, as well as to find out their writing process and problem areas when producing their articles in English, henceforth referred to as the Survey Content-Based Questions. The Survey Profile Questions (i.e. informants' personal details along with questions 1-6) contained information about the participants' nationality, their sex and age group, the number of years they had been learning English, whether they ever stayed in an English-speaking country (and, if so, for how long), and how frequent they read and wrote research medical papers in English. The Survey Content-Based Questions (questions 7-9), on the other hand, were concerned with the informants' writing process, the main problems they encountered when writing scientific articles in this foreign language and the tools as well as resources they found to be useful so as to overcome such problem areas. Each survey question will be analysed in detail in sections 3.4 (Survey Profile Questions) and 5.3 (Survey Content-Based Questions).

In the original survey there was also another question (question 10) which attempted to obtain information about the time participants needed to complete the whole questionnaire; that is, not only the time spent in answering the initial survey but also the worksheet of exercises distributed among them.[35] The length of the initial questionnaire as well as the workload involved in the completion of the exercises indicated that both activities would take the participants much time. However, in light of their responses[36], I decided to exclude that question from the data analysis since some informants misunderstood the inquiry as the question itself turned out to be ambiguous. It was placed at the end of the initial questionnaire and some participants thought they were being asked about the time spent in completing the survey (e.g. less than five minutes, between five and ten minutes, fifteen minutes), whereas some other responses seem to include the time devoted to doing the exercises as well (e.g. forty minutes, one hour, two hours). The richness and detailed information of the data drawn from each survey along with the worksheet of exercises, though, undoubtedly indicate that completing the whole questionnaire must have been time-consuming for all informants. In any case, the statistical information in Appendix 2 should not be taken into account due to its unreliability. The full questionnaire is included in Appendix 1.

3.4 Informants

A total of twenty-four Spanish doctors constitute the community of informants considered in this research study. As Table 1 shows, each participant in the sample was assigned a code number while analysing their data and, from now onwards, in order to differentiate their opinions they will be referred to by means of that identification number[37].

INFORMANTS' WORKPLACE	
ALB1-ALB3	Complejo Hospitalario Universitario de Albacete
PH1- PH13	Hospital Universitario Puerta de Hierro
HN1- HN8	Hospital de Navarra

Table 1 Informants' Identification codes and hospital

Informants belong to a wide range of specialities; most participants from the *Complejo Hospitalario Universitario de Albacete* (**ALB1**, **ALB2** and **ALB3**) and the *Hospital Universitario Puerta de Hierro* (**PH1**-**PH13**) were working for the *Servicio de Endocrinología y Nutrición*, whereas the sample from the *Hospital de Navarra* (**HN1**-**HN8**) embraces different departments: "*Unidad de Anatomía Patológica*" (**HN3**, **HN6**), "*Unidad de Neurología*" (**HN1**, **HN4**), "*Unidad de Radiología*" (**HN7**), "*Medicina Interna*" (**HN2**, **HN5**) and "*Servicio de Oncología*" (**HN8**).

---

[35] This worksheet of exercises will be presented in section 3.5.

[36] See Appendix 2.

[37] Identification codes consist of the acronym of the hospital each informant was working at; namely, **ALB** (*Complejo Hospitalario Universitario de Albacete*), **PH** (*Hospital Universitario Puerta de Hierro*) and **HN** (*Hospital de Navarra*), together with a number as there is more than one subject from each hospital.

The choice of the abovementioned hospitals was determined by the fact that one endocrinologist from the *Hospital Universitario Puerta de Hierro* (**PH2**) and one pathologist from the *Hospital de Navarra* (**HN3**) were aware of my research work on phraseology in general and, more specifically, on the syntactic and collocational patterns of non-technical words in scientific English, and had already expressed their interest in collaborating as informants of my research whenever needed. Both doctors put me in contact with the other informants who were either working in the same hospitals as **PH2** and **HN3**, (**PH1**-**PH13**) and (**HN1**-**HN8**) respectively, or had qualified in the same year as **PH2** (**ALB1**-**ALB3**) but were practising medicine in a different medical centre.

The fact that the candidates were working in so many different places and devoted themselves to several professional activities (i.e. dealing with patients, practising surgery and doing research) made the gathering of the data a time-consuming process. As already pointed out, it was also really difficult to restrict the sample to doctors from the same hospital and the same research area; so the Survey was extended to also include those doctors, however their speciality, provided by **PH2** and **HN3** in order to get a broader view of different areas of expertise.

The samples from the *Complejo Hospitalario Universitario de Albacete* and the *Hospital Universitario Puerta de Hierro* have proven to be quite homogeneous in terms of their research discipline as most of them work on endocrinology and nutrition. Doctors from the *Hospital de Navarra*, on the contrary, constitute a more heterogeneous group as far as the main field to which they belong to is concerned, including internal medicine, medical oncology, (neuro)pathology and radiology.

Regardless of their different specialisms, it was likely that I would work with a very similar sample in terms of the age group of our doctors. On the one hand, it was naturally expected that all of them would be older than twenty-five years old since graduates in Medicine are not younger than twenty-four years old. It was also quite likely that most of them would be of similar age as they had qualified in either the same academic year or rather close in time.

It seems important to bear in mind that when finishing their degree in Medicine, or Medicine and Surgery, as it used to be called in Spain not long ago, all doctors-to-be must follow the same path within the Spanish health system. In short, a graduate in Medicine takes, on average, between six months and two years to become a *MIR*[38]. Depending on their results in the *MIR* examination, they are allowed to choose the speciality they want and where to practise it. This period tends to last, in broad terms, a minimum of three years and a maximum of five. Then they can be properly referred to as “specialists” and are entitled to work in both private and public health centres. With much effort, time and a “bit of luck” they become what is referred to as “*personal estatutario*”, which in **PH2**’s view means some substantial improvement in a doctor’s career. All our informants fitted into the Spanish health system in a similar way:

> “con la nota [del MIR] eliges qué especialidad hacer y dónde, según quieras puedes irte al fin del mundo a hacer lo que te gusta o quedarte en donde quieres y hacer lo que quede… dura entre tres (cuatro recientemente) y cinco años, al terminar a la calle con título de especialista, puedes trabajar en la privada o en la pública si te cogen, habitualmente con contratos basura… pasa el tiempo y puedes aprobar una oposición y entonces ‘voilà’, eres personal ‘estatutario’ (semejante a funcionario pero no igual). […] Lo de estudiar en un sitio y formarte en otro es habitual […] Antes la gente tendía a formarse en los ‘mejores sitios’ y luego volver a su tierra a trabajar; hoy probablemente es más fácil que te cojan si te conocen,

---

[38] *MIR* stands for “*médico interno residente*”; it refers to a public medical examination used to judge progress of medical students towards becoming doctors and being allowed to practise medicine in Spain.

> es decir, formado en tu tierra." (personal communication with **PH2**, 08/17/2007)[39]

A Consent form (cf. Appendix 3) was also prepared and distributed among the twenty-four participants in the study so that they could be well-informed of their involvement in this research. In order to avoid possible misunderstandings of their rights and duties as informants, the Consent form was written in their mother tongue (Spanish) and provided them with very detailed information concerning the preservation of their anonymity.

Furthermore, informants were acknowledged of their right to have access to the results obtained from the present study as well as to withdraw from it at any time they wanted or even refuse to answer any question without risk of prejudice. Only data from those informants who signed the Consent form were included in this investigation[40].

As was stated in section 3.3, part of the data considered in this empirical investigation come from a survey administered to the informant sample, which consists of twenty-four Spanish doctors who were working in three different hospitals when this research was carried out. In the pages to follow, the defining traits of the discourse community under study, based on the information provided by the Survey Profile Questions (questions 1-6), will be carefully analysed.

First of all, participants were asked to fill in some information regarding their sex, age group and contact with the English language. Evidence from the sample reveals that the female participants who provided the data outnumbered male ones with 15 to 9 (62,5% as opposed to 37,5%). With regard to the age group the informants belonged to, most of them (70,8%) were between 30 and 40 years old, which makes the sample quite homogeneous as far as the age span is concerned (cf. Figures 6 and 7 below).

---

[39] This quotation comes from a personal communication on August 17 2007: "*depending on the results obtained in the [MIR] examination, one can choose what speciality and where to practise it; you may go far away and do what you really want to or stay where you prefer and do what is left…; this period lasts between 3-5 years. Afterwards, you are a proper specialist and can work in either private or state medical centres (very often with low-paid contracts). Time goes by and if you pass an official examination, you become what is called 'personal estatutario', which is a kind of civil servant. Studying in a place and being trained somewhere else is a very commonplace among doctors. In the past, people used to be trained in the 'best hospitals' and then came back to their home place to practise medicine. Nowadays things have changed: it is more likely to get a post in a place where they know you, so it is better to get your training where you studied.*"

[40] There were originally twenty-five questionnaires but one of them was disregarded as the informant had not filled in that form.

SEX

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MALE	9	37,5	37,5	37,5
	FEMALE	15	62,5	62,5	100,0
	Total	24	100,0	100,0	

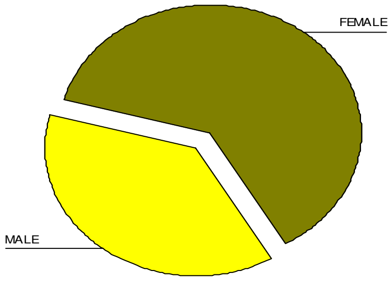


Figure 6 Sex group of the informants

AGE

		Frequency	Percent	Cumulative percent
Valid	25-30	4	16,6	16,6
	30-40	17	70,8	87,4
	40-50	2	8,3	95,7
	+50	1	4,2	100,0
	Total	24	100,0	100,0

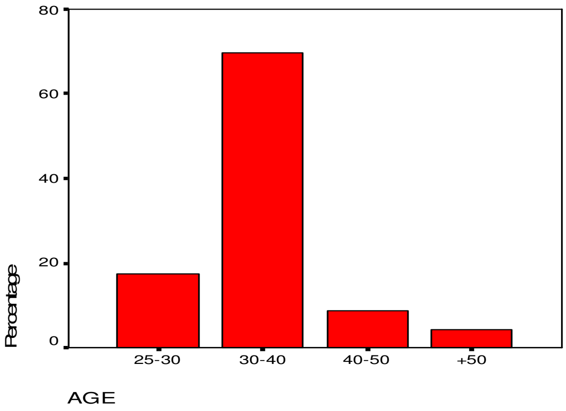


Figure 7 Age group of the informants

All informants, with no exception, held a bachelor's degree in Medicine. All graduated from different Spanish universities (see Figure 8 below), apart from one being **PH7,** who graduated from the *Universidad Central de Ecuador*.

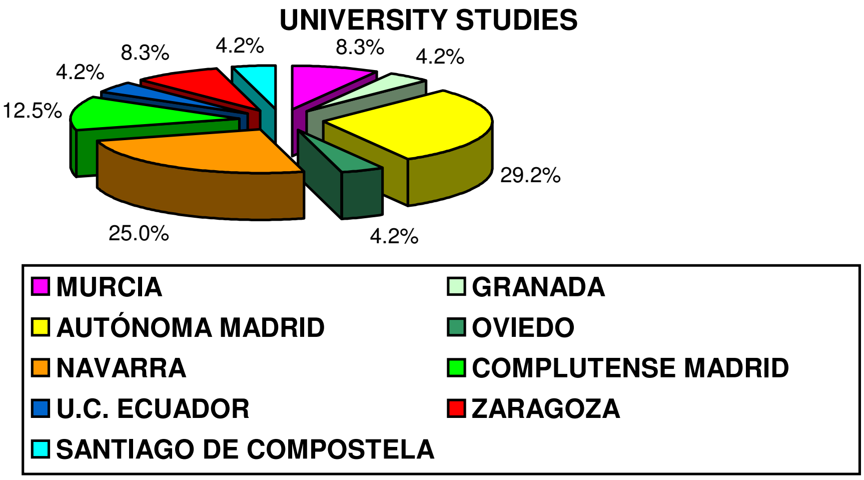


Figure 8 Informants' centre of studies

Most participants (54,2%) graduated from either the *Universidad de Navarra* (**HN1**, **HN2**, **HN3**, **HN5**, **HN6**, **PH2**) or the *Universidad Autónoma de Madrid* (**ALB3**, **PH3**, **PH4**, **PH5**, **PH8**, **PH9**, **PH11**). The rest studied at the *Universidad de Murcia* (**ALB1**, **PH10**), *Universidad Complutense de Madrid* (**PH6**, **PH12**, **PH13**), *Universidad de Granada* (**ALB2**), *Universidad de Oviedo* (**PH1**), *Universidad Central de Ecuador* (**PH7**), *Universidad de Zaragoza* (**HN4, HN8**) and *Universidad de Santiago de Compostela* (**HN7**).

As it is common practice among doctors, they may not necessarily practise medicine in the same town for a long time, as they very often participate in selection processes involving competitive examinations[41], which makes them work in different hospitals until they get a more permanent destination. Table 2 gives up-to-date information about the University each informant studied at and the hospital where they were at the time both practising medicine and doing their research in their various specialities.

---

[41] Cf. Footnote 38

IDENTIFICATION NUMBER	GRADUATED FROM THE UNIVERSITY OF	CURRENTLY WORKING AT
ALB1	MURCIA	COMPLEJO HOSP. UNIVERSITARIO DE ALBACETE
ALB2	GRANADA	COMPLEJO HOSP. UNIVERSITARIO DE ALBACETE
ALB3	AUTÓNOMA MADRID	COMPLEJO HOSP. UNIVERSITARIO DE ALBACETE
PH1	OVIEDO	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH2	NAVARRA	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH3	AUTÓNOMA MADRID	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH4	AUTÓNOMA MADRID	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH5	AUTÓNOMA MADRID	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH 6	COMPLUTENSE MADRID	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH7	U. C. ECUADOR	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH8	?	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH9	AUTÓNOMA MADRID	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH10	MURCIA	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH11	AUTÓNOMA MADRID	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH12	COMPLUTENSE MADRID	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
PH13	COMPLUTENSE MADRID	HOSPITAL UNIVERSITARIO PUERTA DE HIERRO
HN1	NAVARRA	HOSPITAL DE NAVARRA
HN2	NAVARRA	HOSPITAL DE NAVARRA
HN3	NAVARRA	HOSPITAL DE NAVARRA
HN4	ZARAGOZA	HOSPITAL DE NAVARRA
HN5	NAVARRA	HOSPITAL DE NAVARRA
HN6	NAVARRA	HOSPITAL DE NAVARRA
HN7	SANTIAGO DE COMPOSTELA	HOSPITAL DE NAVARRA
HN8	ZARAGOZA	HOSPITAL DE NAVARRA

Table 2 Informants' centre of studies and workplace

Once informants had provided some general information about their degree, field of research and professional career, they were asked to answer ten open-ended questions. As already mentioned, the first six questions explored in this section aimed, more specifically, at gathering details about the respondents' academic and professional background in English. The analysis of these questions helped enormously to characterise the sample. Each Survey Profile Question will now be exhaustively examined.

Survey Question 1: How long have you been studying English?

Survey Question 2: Have you ever lived abroad, in an English speaking country? If so, how long?

These two questions were intended to elicit the participants' overall exposure to the English language. The data shown in Figure 9 reveal that our informants had been learning English for a number of years so far. Only four participants in the survey (**ALB1**, **PH4**, **HN2,** and **HN8**) stated that they had studied English for less than five years, whereas the vast majority of the sample (18 informants) had been learning English for more than five years.

Despite the fact that on average our doctors had spent about 13,16[42] years studying English, the standard deviation (9,74) should be carefully looked at in this case: while a 54,5% (12 informants) claimed that they had spent between four and ten years learning English, a remarkable 31,8% (7 informants) revealed a minimum exposure of fifteen years to learning this foreign language. Figure 9 below presents the exact percentages[43]:

INFORMANTS' EXPOSURE TO ENGLISH

STUD.ENG

N	Valid	22
	Missing	2
Mean		13,159
Median		10,000
Mode		4,0(a)
Std. Deviation		9,7460

a Multiple modes exist. The smallest value is shown

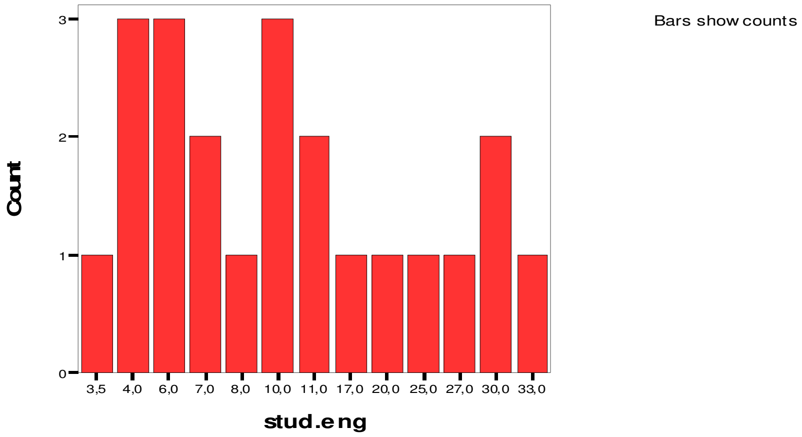


Figure 9 Number of years studying English

---

[42] This figure should be taken cautiously as informants had just some (unspecified) hours of English instruction per year.

[43] Two informants (**ALB3**, **PH8**) were disregarded as they did not provide any information concerning this issue.

With regard to participants living in an English-speaking country for a period of time, data reveal that 62,5% of our informants had spent some time abroad with the aim of improving their proficiency in English. Most (73,3%) referred to one, two or three-month stays (**ALB3**, **PH1**, **PH5**, **PH8**, **PH10**, **PH11**, **PH12**, **PH13**, **HN1**, **HN4**, **HN6**), whilst the rest (26,7%) referred to longer periods, ranging from four months (**ALB1**) to two academic years (**PH7**). Although they were not asked about the place they had stayed in, some participants included that piece of information in the surveys. The following Table displays all data found in the questionnaires:

ID	LENGTH OF THE STAY	DESTINATION
ALB1	3 months 1 months	USA Australia
ALB3	3 months	?
PH1	1 month	England
PH2	3 summers & a whole academic year	?
PH5	2 months 3,5 – 4 months	Ireland Sweden[44]
PH7	2 years	England
PH8	3 months	Philadelphia
PH9	10 months	London
PH10	3 months	USA
PH11	3 months	?
PH12	2 months	?
PH13	3 months	Canada
HN1	1 month	?
HN4	1 month	?
HN6	1 month	Ireland

Table 3 Informants' stays abroad

Survey Questions 1 and 2 turned out to be very useful as they enabled me to verify that eighteen subjects had devoted a minimum of five years to the continuous study of English. Indeed, seven out of those eighteen informants had been learning English for more than 13,16[45] years on average. As will be discussed in Survey Question 4, this actually means classroom instruction of English for Specific Purposes (ESP), on the one hand, and content-based courses in English, on the other. Besides, fifteen informants reported to have lived in an English-speaking country for a minimum of one month. These two factors cannot claim to prove (nor did they seek to prove) that all the informants can be placed within the same level of proficiency in English. However, it can certainly be stated that these twenty-four doctors had been attending English lessons for a number of years and a 62,5% of them had also been in real language communicative settings either for professional (conferences, work practices) or academic (*COU*[46], scientific / English courses) reasons.

---

[44] Although question 2 explicitly asked informants about their stays in an English-speaking country, PH5 included her time spent in Sweden. This information was disregarded.

[45] Cf. Footnote 42

Survey Question 3: How long have you been working in this hospital? What is your *MIR*[47] speciality (field/research work)?

Informants were also required to specify their length of service in their current hospital in order to find out whether doctors from different hospitals had a comparable medical practice background and, given the case, whether they felt as part of their medical community.

Figure 10 reveals that participants can be easily grouped into three different categories as far as their experience working in their hospital is concerned: a) "new" doctors (less than five years working in that hospital), b) "settled" doctors (between five and ten years practising medicine in that hospital) and c) "senior" doctors (more than ten years in their current hospital).

Statistics

N	Valid	24
	Missing	0
Mean		6,375
Median		5,000
Std. Deviation		4,9788
Minimum		1,0
Maximum		25,0

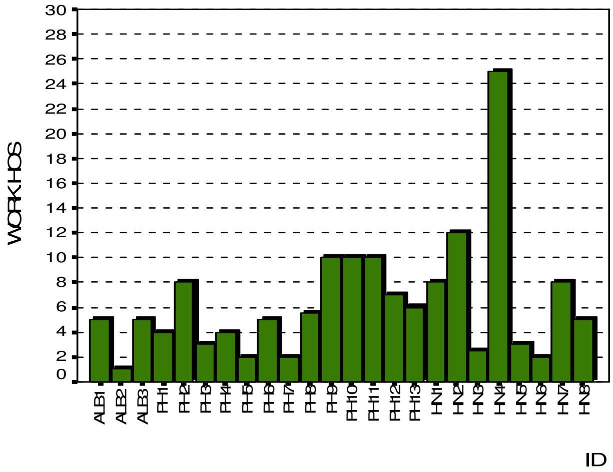


Figure 10 Informants' experience working in their current hospital

---

[46] *COU* stands for "*Curso de Orientación Universitaria*" and refers to a course students must take so as to be admitted to university studies at any Spanish university.

[47] "*médico interno residente*" (cf. Footnote 38)

In addition, this question sought to gather information about the research areas our participants were more closely involved in so as to identify their main fields of expertise. A close look at their answers (cf. Table 4 below) suggests that these scientists fitted into different research areas according to their various *MIR*[48] specialities, namely Endocrinology and Nutrition (8 informants), Internal Medicine (7 informants), (Neuro)pathology (3 informants), Clinical Pharmacology (2 informants), Medical Oncology (2 informants), Neurology (1 informant) and Radiology (1 informant).

Endocrinology & Nutrition	Internal Medicine	Clinical Pharmacology	Medical Oncology	Neurology	(Neuro)Pathology	Radiology
ALB1	PH4	PH5	PH10	HN1	HN3	HN7
ALB2	HN2	PH6	HN8		HN4	
ALB3	HN5*				HN6	
PH1	PH9**					
PH2	PH11***					
PH3	PH12					
PH7	PH13***					
PH8						

Table 4 Informants' medical specialities

*currently working at an Emergency Unit
** doing research work on *Lupus*
*** currently working at an HIV Unit

With reference to their various specialities, data have shown that informants fell into seven disciplines (cf. Table 4) and that had been working in their current hospital for about 6,4 years on average (cf. Figure 10). The survey also suggests that this factor plays an important role in the doctors' engagement with the institution. In the informants' view, being a doctor does not only imply dealing with patients, but also doing research. Unlike other disciplines or professional activities, research in medicine is understood as a team work rather than as an individual activity (Gledhill 2000:55). Thus, the longer they have been working for the same hospital, the more involved they will feel in the hospital's research lines. In this respect, it seems worth mentioning that this is the main reason why I refer to these group participants as a discourse community. Following Swales' (1990:9) approach, a discourse community can be defined as a group that has goals and purposes in common, and use communication to achieve these goals. This close relationship will be more evident when it comes to the production of articles, which will be more extensively developed later on in section 5.3.

---

[48] Ibid.

Survey Question 4: Have you ever taken any course related to your speciality in English? When? What was the course about?

With the aim of obtaining additional information with respect to informants' exposure to English, they were asked about the training courses, if any, they used to take in English. Only ten participants declared that they had taken some courses related to their speciality in English. As Table 5 illustrates, their answers refer to two clearly distinct kinds of courses: content-based courses in English, on the one hand, and English for Specific Purposes (ESP) courses, on the other.

Three informants (**PH5**, **HN1**, **HN4**) stressed the fact that they used to attend conferences and congresses in English all through the year since English is the general language of communication in their discourse community:

> "the vast majority of our courses are in English. The last one took place two weeks ago in Paris. It was about 'Clinical Investigation in children'." (**PH5**)

ID	TYPE OF COURSE	WHEN?	LENGTH
ALB2	Using scientific English in Medicine	?	**1 month**
ALB3	"EFES[49] course on Clinical Endocrinology"	September 2003	**?**
PH2	Training speakers	2005	**1 week**
PH5	Most courses in English (i.e. "Clinical investigation in children")	2007	**1week**
PH10	Medical presentations in English	2006	**?**
PH11	"Advances in Antiretroviral treatment"	2006	**?**
PH12	Medical English"	2003	**?**
HN1	Conferences, congresses & training courses in English	all through the year	**?**
HN4	Conferences, congresses (i.e. "Brain tumours")	all through the year	**?**
HN7	Neuroradiology	2005-2006	**4 one-week courses**

ESP courses
Content-based courses in English

Table 5 Informants' training courses

---

[49] European Federation of Endocrine Societies.

Data (cf. Table 5) suggest that training is also a key issue among doctors. 41,7% of the informants (10) stated that they used to attend courses and seminars related to their speciality in English. It seems, though, that they interpreted that survey question in different ways; some of them (50%) referred in their answers to the medical English language courses (i.e. writing medical presentations and/or reports in English, speech-making, using scientific English in medicine, to name but a few) whilst others (60%) mentioned content-based courses related to their medical speciality taught in English (i.e. specialised courses organised by the *European Federation of Endocrine Societies* [*EFES*[50]], international seminars on neuroradiology, oncology, and the like).

Whatever the type of course, it seems clear that these informants were used to practising both oral and written skills in English for their everyday medical praxis. What is more, they all underlined the fact that English was essential in their field and thus they needed to be familiar with the English system in general and the kind of English characteristic of the discourse community they belonged to, in particular.

Survey Question 5: Do you ever read academic papers/scientific articles in English? If so, how often? Which books/journals, etc. do they come from?

Question 5 attempted at verifying whether participants were used to reading scientific papers in English as well as finding out what kind of publications they highlighted as their primary information sources; that is, those books and/or journals they considered essential to their own field work.

All informants, with no exception, declared themselves as regular readers of English medical journals, mainly (cf. Table 6). The *British Medical Journal* (*BMJ*), the *New England Journal of Medicine* (*NEJM*) and the *Journal of Clinical Endocrinology and Metabolism* (*JCEM*) stood out as the most frequently read journals by eight doctors. Other journals like *AIDS*, *Nature*, the *Lancet British Medical Journal*, the *American Journal of Medicine*, the *American Journal of*

---

[50] The *EFES* is aimed at promoting endocrinology in Europe.

Neuroradiology, the *American Journal of Radiology* and *the Journal of Bacteriology* also constituted the informants' reading list.

The answers also suggested that both American and British journals used to be given much more priority than manuals and English textbooks: "*I rarely read books in English*" (**PH2**). However, the following manuals were mentioned once: *Williams Textbook of Endocrinology*, DeGroot's *Endocrinology* and *Harrison's Endocrinology*. In addition to journals and textbooks, **PH5** mentioned reports and international dossiers as part of her everyday reading in English:

> "I'm working for the European Medicine Agency and the reports and
> dossiers I have to deal with are all in English. So everyday I have to
> read and write in English." (**PH5**)

When being questioned about how often they read scientific articles in English, a remarkable 58,4% (14 doctors) asserted that they read academic papers related to their field work on a daily basis. 25% consulted English journals once or twice a week, whereas the remaining 16,6% did so with less frequency. See more detailed figures in Table 6:

How often do you read academic papers/scientific articles in English?		
Daily	14	58,4%
Weekly	6	25%
Every month (one article a month)	2	8,3%
Every 3 months (one article; four articles a year)	2	8,3%

Table 6 Frequency of English journal reading

Reading scientific papers in English has thus also proven to be part of a doctor's almost everyday activity. As was expected, the range of sources was wide and clearly determined by the speciality each researcher was most interested in. Nevertheless, within the same speciality, doctors mentioned the same journals and, to a lesser extent, manuals as part of their daily/monthly reading list. What must be stressed, though, is that this question has revealed some very interesting methodological information. The following quotes appear to suggest that reading in English is seen as a professional need since most medical journals are written in English:

> "Yes, I do read scientific papers in English everyday. This is part of my job as a doctor." (**PH4**)
>
> "Yes, everyday. Pathology books and journals. They are from USA, UK and Europe. I must be up-to-date with what is going on in neuropathology." (**HN4**)

Interestingly, two of the journals selected for the compilation of the *HSC*, the *British Medical Journal* (*BMJ*) and the *Journal of Bacteriology* were mentioned by thirteen informants. As was explained in section 3.2, the criteria followed for the selection of the medical journals compiled in the *HSC* was determined by two main factors: on-line accessibility and relevance. Even though the *BMJ* and the *Journal of Bacteriology* are addressed to the whole medical community rather than specifically to endocrinologists, pathologists or oncologists, participants underlined the impact of the medical journals selected for the *HSC* as central sources of current research on different medical fields.

Survey Question 6: Do you usually write scientific papers in English? Which journals do you usually publish in?

This question provided details concerning the informants' written production in English. They were asked to comment on whether they used to write scientific papers in English on the one hand and, given the case, they were required to specify the kind of journals in which they published their papers. According to their information, participants can be divided into three main categories:

a) Group 1 (**12,5%**): they published articles and/or reports in English **on a regular basis** (**PH5**, **PH6**, **HN1**).
b) Group 2 (**37,5%**): they **occasionally** published articles in English (**ALB2**, **ALB3**, **PH1**, **PH2**, **PH9**, **PH10**, **HN3**, **HN4, HN8**).
c) Group 3 (**50%**): they **hardly ever** published articles in English, but some other written productions were mentioned. (**ALB1**, **PH3**, **PH4**, **PH7**, **PH8**, **PH11**, **PH12**, **PH13**, **HN2**, **HN5**, **HN6**, **HN7**).

Only three participants in the survey (Group 1) acknowledged that they were used to writing up papers and/or reports in English to be published in different formats. **HN1** mentioned she was a regular contributor to the *European Journal of Neurology*, *Cardiovascular Diseases* and *Experimental Brain Research*. **PH5** commented on her frequent written contributions to the different events organised by the *European Medicine Agency* and, finally, **PH6** highlighted her usual assessment reports for medicinal product agencies and expert reports for pharmaceutical companies.

Group 2 is formed by a total of eight doctors (37,5%) who declared themselves as occasional English writers. With the only exception of **ALB2**, who had only published in Conference Proceedings[51], the other members of this group named the following journals to which they had contributed: the *European Journal of Endocrinology* (**ALB3**), *Journal of Endocrinological Investigation* (**PH2**), *Journal of the American Society of Nephrology* (**PH2**), *Journal of Rheumatology* (**PH9**), *AIDS* (**PH9**), *Annals of Rheumatic Diseases* (**PH9**), *Annals of Oncology* (**PH10**), *Clinical Cancer Research* (**PH10**) and various (unspecified) journals on neurology and neuropathology (**HN3**, **HN4**) and medical oncology (**HN8**).

Finally, the number of informants who pointed out that they hardly ever published articles in English (Group 3) is remarkably high (50%). But for **PH13** who mentioned that she occasionally wrote summaries for International Conferences in English, the rest provided no further information in this respect.

Bearing in mind the importance of reading and writing to the dissemination of research, it was expected that data related to this question would reveal informants as regular writers of scientific papers as well. Thus, a 12,5% seemed to be a very low percentage. By contrast, 50% of informants claimed that they hardly ever published articles in English. Such an assertion came as a surprise since I knew they all had publications in English[52], so with the aim of shedding light on this apparent inconsistency, I decided to contact **PH2**, who very kindly gave me details of their usual writing up process. He stated that research articles tend to be jointly written in a team, which means that different people are responsible for the different tasks and steps involved in writing up a scientific paper.

---

[51] "*I write scientific papers in English but not very often. I have only published in Conference Proceedings. Indeed, I still haven't published in any English journal.*"

[52] I was aware of the fact that all the participants were included as authors in *Pubmed*, which is a service of the U.S. National Library of Medicine that includes over 17 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s.

> "el trabajo suele ser conjunto; los residentes hacen el trabajo de campo, los adjuntos guiamos la publicación: qué tema abordar, qué datos publicar y la orientación general. La redacción suele hacerse en español y después traducirse al inglés para intentar 'colarlo' en alguna revista internacional. Nadie es totalmente responsable del artículo y suele ser el jefe o el que ha planeado el estudio el que figura como persona de contacto para correspondencia. (…) pero cuando se va a mandar a alguna revista 'top' sueles pasar el artículo a algún revisor, por ejemplo, profesores de inglés que suelen dar clases a médicos incluso en el mismo hospital. Controlar cómo escribir nuestro tipo de artículos en inglés, como ves, es esencial para nosotros." (personal communication with **PH2**, 08/17/2007)

Most importantly, in my view, is that the writing up of a paper is understood as a process where different tasks are assigned to different agents: housemen (*MIR*) are responsible for the field work and are also in charge of checking for specific information in the lab, whereas assistant doctors guide the publication and take decisions with regard to the topic, the data analysis and the overall orientation of the research question(s). The actual writing up of the article is usually done in Spanish and then is translated into English (see further details in the discussion of the writing process in section 5.3). Thus, any written production is also seen as a team work, which seems to be the reason why they feel they have not published in English.

Having explored the defining traits of the medical community under study, the following section presents a description of the worksheet of exercises distributed among participants in the sample as well as a detailed account of the various steps taken in their design process.

3.5 Worksheet of exercises

Along with the initial Survey[53] and the Consent form[54], informants were handed out a worksheet consisting of four different exercises especially designed for the current research. The analysis of these exercises supplied an incredible amount of data with respect to the participants' competence in the use of abstract nouns in medical English. The comparison between the data extracted from the *HSC* (cf. chapter 4) and the worksheet of exercises yielded some interesting results that will further be discussed in chapter 5.

As pointed out earlier, the main purpose of these exercises was exploring the way in which participants (i.e. non-native speakers of English) made particular collocations with the abstract nouns selected. So as to determine the contents and plan the design of these exercises, I focused on the recurrent lexico-grammatical patterns of the nouns under study across the *HSC* (see sections 4.3-4.7). All the examples included in the designed worksheet were therefore extracted from the compiled corpus.

The careful examination of the recurrent patterning associated with abstract nouns in the *HSC* revealed very interesting findings with reference to their behaviour in the health science discourse (cf. section 4.8) and it also raised the following issues: a general tendency for the use of classifying adjectives, delexicalised uses of verb collocates, grammaticalized multiword units and a preference for the use of periphrastic structures in passive forms. With the aim of determining whether the above processes posed problems to the non-native informants of my study, I designed four exercises, each of which attempted at providing reliable data concerning the participants' competence with respect to the colligational and collocational patterns of the selected abstract nouns.

Originally, the initial repertoire of exercises consisted of six different activities and the number of items per task was considerably higher. When producing the first draft, I decided to shorten the worksheet so as to prevent informants from dropping out of the study. The final worksheet is made up of two "Matching" exercises (Exercises 1 and 4), a “Fill-in the gaps” question (Exercise 2) and a “Sentence-translation” activity (Exercise 3).[55]

---

[53] Cf. Appendix 1
[54] Cf. Appendix 3

In Exercise 1, informants were asked to associate nineteen adjectives with the abstract nouns considered in the research (i.e. *agreement*, *comparison*, *conclusion*, *contribution* and *decision*). It was clearly stated in the instructions that several combinations were possible; that is, it was stressed that each noun could be paired with more than one adjective from the list provided and, similarly, each adjective could be applied to more than one noun. Respondents were also warned that some adjectives could be left out.

It must also be pointed out that fifteen out of those nineteen adjectives had been found to be frequent collocates of at least one of the abstract nouns considered. In contrast, the other four adjectives included in the exercise (i.e. *contradictory*, *fashionable*, *short* and *unforeseeable*) could not be found in combination with those nouns in the *HSC*. But for the size descriptor, *short*, the other opinion descriptors (i.e. *contradictory*, *fashionable* and *unforeseeable*) were randomly selected. The choice of *short*, on the contrary, was determined by the fact that it was expected that some informants would have difficulties in distinguishing between *little*, *small* and *short*, since they are similar in meaning and may have similar translations into their L1 (Spanish).[56]

With the purpose of examining what collocates the informants regarded as the most usual ones in combination with the nouns of interest, in Exercise 2 respondents were given fifteen real sentences extracted from the *HSC*. They were asked to fill in the gaps provided with a suitable word, which depending on the context could be a preposition or a verb[57]. As will be illustrated and discussed in section 5.1.1, this activity revealed very interesting findings regarding the participants’ level of linguistic competence, as far as the collocational patterns associated with abstract nouns is concerned.

---

[55] The actual worksheet of exercises are included in Appendix 4 and a Summary of the typology of exercises along with the aims pursued in each activity can be seen in Appendix 5.

[56] (*poco ~ little*; *pequeño ~ little*, *small*, *short*).

[57] Informants, however, were not given any clue about the word class they were expected to use.

The "Sentence-translation" activity (Exercise 3) consisted of a total of eight Spanish sentences for which informants were required to provide suitable translations into English. For the designing of this exercise, eight real examples from the *HSC* were randomly selected from the concordance lines provided by *WordSmith Tools*[58] and, subsequently, translated into Spanish. The main purpose of this task was to get participants to face a real scenario of linguistic production and find out the kind of constructions they used when being asked to translate into English the multiword units analysed in this research. The wide range of structures observed in the collected data offered highly valuable information with regard to their command of English morphology, syntax and semantics (cf. section 5.1.2).

Like Exercise 1, the last activity of the worksheet corresponded to a "Matching exercise" (cf. section 5.1.4). This time respondents were asked to associate certain highly-frequent verbs in English; namely, *make*, *draw*, *reach*, *lead to*, *take* and *do* with a list of thirty-six nouns, which had proven to appear in combination with the above mentioned verbs. Despite the fact that all the noun phrases included in Exercise 4 were collocates of at least one of the verbs from the list, some of them (i.e. *experience*, *a guess*, *a verdict*, *a place*, *a picture*, *an exam*, *a goal*, *a profit*, *a speech*, *knowledge*, *skills*, *some writing*, *notes* and *[financial] success*) were not traced in the *HSC* in combination with the above verbs.

By comparing informants' collocational preferences against the actual word combinations found in the *HSC*, I aimed at examining the possible mismatches between them and discussing their implications. As results will be measured against native production, I shall refer to the collocations provided by the informants as either morphologically, syntactically or semantically inappropriate/deviant combinations or (un)likely to occur in the native corpus data. All the observations made and the reported findings for each exercise will be discussed in chapter 5 (cf. sections 5.1.1-5.1.4).

---

[58] Cf. Footnote 28

3.6 Summing Up

This chapter has focused on the inquiry methods used in this corpus-based research. The *HSC* corpus that forms the basis of my analysis as well as the corpus tools used in retrieving data from it were presented in section 3.2. Section 3.3 informed about a survey distributed among a group of twenty-four Spanish doctors, which has provided very useful data with reference to the participants' profile as well as their main difficulties when writing their scientific papers in English.

The participants' profile, based on the information obtained from the Survey Profile Questions, was carefully examined in section 3.4. These questions revealed that participants believe they must be familiar with the kind of English used in the medical community they are part of. This necessarily entails, as will be further analysed, that they need to develop reading and writing skills which will help them identify, understand and produce structures and word combinations typical of the style of the scientific discourse.

Another remarkable finding from the Survey Profile Questions pointed to the informants' publications in English. The fact that half of the participants claimed that they did not use to publish articles in English contributed to the characterisation of their writing up process. This investigation has demonstrated that one of the main reasons why informants sounded extremely cautious in declaring themselves as regular writers in English lies in the fact that, as **PH2** reported, the main writing up tends to be done by a translator and/or an English teacher who provides expert advice on the draft produced by the research team. Each researcher sees themselves as a contributor to the research itself but not as the actual writer of the final product. Many people contribute towards the building up of that final product, so that seems to be the reason why the writing of an article is interpreted as an ongoing activity: a paper is produced and redrafted by several participants.

In section 3.5 the materials, in the form of a worksheet of four exercises administered among informants, specifically designed for this investigation, were presented. In the next chapter, I set out a comprehensive examination of the collocational patterns in which the selected abstract nouns can be found in the *HSC*.

CHAPTER 4: NATIVE CORPUS DATA ANALYSIS

4.1 Introduction to the chapter

This chapter is devoted to the analysis and discussion around the findings from the *HSC* corpus as well as further comparison from the *British National Corpus* (*BNC*). Firstly, in section 4.2 some considerations regarding the collocational patterning of abstract nouns in the *HSC* will be presented. Secondly, sections 4.3-4.7 will describe the collocational behaviour of the following abstract nouns; *conclusion*, *agreement*, *comparison*, *contribution* and *decision*. This is followed by a discussion of the overall results (section 4.8), with a view to exploring the particular phraseology of abstract nouns in the health science discourse as well as considering the different functions they can acquire because of the different patterns of co-selection they may entertain. Section 5.2 (cf. chapter 5) will offer a comparison between these specific phraseological and collocational properties and the data obtained from the informants' performance in the worksheet of exercises. This comparison will outline the main difficulties the phraseology of abstract nouns poses for the selected discourse community when writing their scientific articles in English.

It must be noted that the decision to analyse abstract nouns was based on the findings from a study[59] conducted on the behaviour of the noun *conclusion* and its restricted collocations in scientific register. While focusing on that abstract noun, it became apparent that there was a frequent list of comparable nouns, etymologically related to a verb (*conclusion* ~ *conclude*; *agreement* ~ *agree*; *comparison* ~ *compare*; *contribution* ~ *contribute* and *decision* ~ *decide,* to name but a few), which needed more thorough investigation.

The selection of five abstract nouns (i.e. *conclusion*, *agreement*, *comparison*, *contribution*, *decision*) set out in sections 4.3-4.7 was motivated by the fact that they were frequently used in combination with verbs. Other significant collocates (prepositions, adjectives, clauses and related verbs: *conclude*, *agree*, *compare*, *contribute* and *decide*) have also been included in the analysis, so as to provide a comprehensive account of the syntactic and semantic behaviour of the above-named abstract nouns. In order to present this study more systematically, the most salient lexico-grammatical patterns of each noun will be described in turn[60].

---

[59] See Laso, N. J. & Verdaguer, I. 2005. "An invitation to explore conclusions and its restricted collocations", *Specific*, 2: 47-54.

4.2 The collocational patterning of abstract nouns in the *HSC*. Some preliminary considerations

Abstract nouns in combination with other parts of speech are frequently used in the academic register to refer to scientific processes, methods, evidence and findings. A detailed study of the patterns of abstract nouns in the *HSC* has revealed a vast amount of phraseological units whose overall meaning is the result of the interaction among its various elements. As will be shown later, some of the collocates these frequent abstract nouns co-occur with have undergone a process of **delexicalisation** and **semantic bleaching**, which means that these nouns mostly provide the semantic content to the whole unit. In most cases, such abstract nouns collocate with delexicalised or support verbs which contribute very little to the meaning of the whole unit. It is thus the larger unit that is the complete unit of meaning, rather than the individual lexical items.

This phenomenon has already been pointed out in innumerable corpus studies which highlight the importance of multiword units in linguistic production (Sinclair 1991; Gledhill 2000; Altenberg & Granger 2001, Oakey 2002; Stubbs 2001; Simpson 2004). In line with Sinclair's idea that meanings are clustered into lexico-grammatical patternings and not in isolated terms, the following five sections focus on the recurrent sequences of words[61] in combination with the abstract nouns under study that commonly co-occur in the health science discourse.

In Partington's words, the process of **delexicalisation** involves "*the reduction of the independent lexical content of a word, or group of words, so that it comes to*

---

[60] The main features of their syntactic environment were identified with the assistance of the concordancing program *WordSmith Tools* (cf. chapter 3, section 3.2 for further reference on this programme.)

[61] These sequences have been referred to with different labels such as "speech formulas" (Pawley 1985), "multiword units" (Cowie 1988), "prefabricated routines" (Cowie 1988), "collocations" (Sinclair 1991), "formulae" (Moon 1992) and lexical bundles (Biber et al. 1999; Cortes 2002), among others (cf. chapter 2).

fulfil a particular function but it has no meaning apart from this to contribute to the phrase in which it occurs." (Partington 1993:183). In this respect, expressions such as *make a conclusion*, *reach an agreement*, *make a comparison*, *make a contribution* and *make a decision* must be analysed as multiword units, where the abstract noun provides the semantic content to the extended lexical string of words to the detriment of the semantic content of the lexical verb. Such restricted verbs have adopted a more grammatical role and simply perform a verbal function.

As Partington (1993) observes, the notion of **delexicalisation** is closely related to Sinclair's concept of **shared meaning**, "*a distribution of meaning across a number of words*" (Sinclair 1987b:110), which accounts for the fact that single words and their context of appearance are mutually co-selected: "*words in English do not normally constitute independent selections* (…) *The item and the environment are ultimately not separable.*" (Sinclair 1992:15).

According to Sinclair (1997:323), **semantic depletion** is common with high frequency words, which tend to lose their independent meaning and adopt a more grammatical role. The analysis of **V + Noun** periphrastic structures in the corpus data reinforces this statement. In this view, the boundaries between a lexical item and its environment become fuzzy. Evidence from the *HSC* has revealed that delexical uses of verbs co-occurring with abstract nouns are usual. The conclusion to be drawn from this fact is that the more delexicalised a unit is, the more widely it collocates (Partington 1993:183).

When stressing some of the benefits of corpus work, Tognini-Bonelli (2001:24) highlights the fact that it shows the strict correlation between lexical and grammatical choices which extends the boundaries of a given initial unit, giving rise to the creation of extended units of meaning. The exploration of abstract noun patterns in terms of this interdependence of lexis and syntax has also led this study to the consideration of another linguistic phenomenon referred to as **grammaticalisation**.

Hopper & Traugott (1993:2) define such a linguistic process as a "*subject of linguistic changes whereby a lexical item or construction in certain uses takes on grammatical characteristics, or through which a grammatical item becomes more grammatical*". This notion of **grammaticalisation** points to the fact that certain lexical items and constructions in certain contexts serve different grammatical functions.

Since first introduced by the German neogrammarian Georg von der Gabelentz[62] (1891), many scholars (Sweetser 1988; Sinclair 1992/1997; Hopper & Traugott 1993; Partington 1993; Howarth 1996; May Fan 1999; Tognini-Bonelli 2001; Aitchison & Lewis 2003, among others) have made use of several locutions, such as semantic change, desemanticization, fading meaning, depletion[63], to name but a few, to refer to **semantic bleaching** as a key notion that "*pertains almost exclusively to late stages of grammaticalization*" (Hopper & Traugott 1993:98)[64].

The concept of **grammaticalisation** thus should be regarded as closely related to the aforementioned process of **delexicalisation**. In Tognini-Bonelli's view,

> *"delexicalisation very rarely is an end in itself. In most cases it signals a new role for the original unit. More studies are needed to understand this phenomenon thoroughly; from the evidence available, however, it would seem that there is a tendency for delexicalisation to coincide with grammaticalisation; i.e. the process by which, with the passage of time, certain items lose their lexical meaning and acquire a grammatical function. It would also seem that it is very common words, which are used in a large number of contexts that are more exposed to the phenomenon of delexicalisation and are more likely to lose their independence to 'mean' with other words."* (Tognini-Bonelli 2001:123).

---

[62] "(…) *linguistic forms 'fade or grow pale' (**verblassen**), their colors 'bleach' (**verbleichen**), and must be covered with fresh paint*". In Hopper & Traugott (1993:20)

[63] See Aitchison & Lewis (2003:254).

[64] The term **semantic bleaching** was coined by the German neogrammarian Georg von der Gabelentz in 1891 (cf. References). Some other linguists (Sweetser 1988; Aitchinson and Lewis 2003) refer to that concept when discussing polysemy and grammaticalisation.

However, this continuum between **delexicalisation** and **grammaticalisation** cannot be understood as an automatic association of fading meaning (semantic bleaching or *verbleichen* in Gabelentz's[65] terminology) and sudden semantic loss with grammaticalisation. As Hopper & Traugott (1993:95) highlight, "*when a form undergoes grammaticalization from a lexical to a grammatical item, some traces of its original lexical meanings tend to adhere to it*"; that is, at an initial stage of grammaticalisation, there is a redistribution of meaning, but not a complete loss of semantic content. On the contrary, meanings tend to be weakened and former lexical items gradually become empty syntactic elements.

All these processes along with some other observations made with reference to the lexico-grammatical patterning of abstract nouns in medical writing will be analysed in the pages to follow.

4.3 The patterns of the noun *conclusion*

In order to discover the lexical items that usually co-occur with the abstract noun *conclusion*, concordance lines from the *HSC* will be studied. Figure 11 below shows some of the concordance lines for this node word:

```
WordSmith Tools -- 07/05/2008 15:16:17

N      Concordance
382    there were no interferences with the    conclusions drawn  because of alkanes
383    by N17  Rac, and therefore support the  conclusion drawn from the use of  the V
384    velopment to term was much better. The  conclusion drawn from this elegant
385    etween vital and nonvital genes, so the conclusions drawn here should apply to
386     the Results section) also supports the conclusion for maize. It is clear that
387     H. virescens were consistent with  the conclusions from 40 T. ni; M2R addition
388    mplete nuclear division, confirming the conclusion from the earlier experiment
389    e  precursor tissue of the fetus.29 The conclusion from examining  both coat
390    ng to ensure adequate accuracy of the   conclusions; in our case 10 000
391    odological quality did not affect   the conclusions. Interestingly, the two tri
392     are thus completely independent. The   conclusion is therefore reinforced that
393    mechanism  is postulated to explain the conclusions,  it is difficult to know
394     and Wiese [1997] further criticize the conclusions of Templeton and Read  [198
395    s have been completed   since 1992, the conclusions of the original review may
396    ctomyosin-based contractile ring at the conclusion of   mitosis has long been a
397    tonic swelling, are consistent with the conclusions of our analysis of 2D
398    as judged at P < 0.05.  RESULTS  At the conclusion of the freezing trials, the
399    his assertion, however, contradicts the conclusions of Hatini et al. (1996).
400     a numbered 10 cm wooden stake. At the  conclusion of the count, one researcher
401    eing present in equal frequency at the  conclusion of both meiotic divisions
402    rgets children aged under 1 year.   The conclusions of this meta-analysis
403    the leader peptidase. This supports the conclusions of Strom   and Lory (41) an
404    stify her statement about refuting  the conclusion of elk scarcity. If any
405     lobster NHE is regulated in vivo.  The conclusions of this investigation
```

Figure 11 List of concordances of the noun *conclusion* (extracted from *WordSmith Tools*)

---

[65] Cf. Footnote 62

The lexico-grammatical patterning of this noun is straightforward. Associated with it in colligation we find verbs, prepositions, *that*-clauses and adjectives. These combinations must be analysed in much more detail so as to provide a motivated description of its various meanings and functions.

a) The noun *conclusion* as an introductory item

One of the most prominent uses of *conclusion* is as a heading/subheading introducing a section of the scientific article. 184 occurrences (out of 578) reveal that this can appear in isolation or in combination with other fixed expressions such as the following: *summary and conclusion*, *conclusion and future directions*, *conclusions and prospects*, *conclusions and unresolved issues*, *conclusions and speculation*. As can be seen in the above examples, these expressions are all related to further research on the *conclusion* stated in the article they appear.

In regard to its morphology, it should be stressed that there does not seem to be any syntactic and/or semantic difference between the base form (*conclusion*) and the inflected form for the plural (*conclusions*). However, there is a preference for the latter (145 occurrences) in comparison with the former (39 occurrences).

The initial expectation was to find this structure more frequently in the data as the different sections of an article are clearly distinguished by the use of subheadings. 394 occurrences of such an abstract noun give evidence of the fact that the *conclusions* drawn in the *HSC* tend to be introduced by means of more elaborated constructions, as will be seen in the following pages.

b) verb + *conclusion*

In the study of this abstract noun, another remarkable pattern that can be traced is the **Verb + noun** colligation. The following Figure charts the numerous verbs that collocate with the noun *conclusion*:

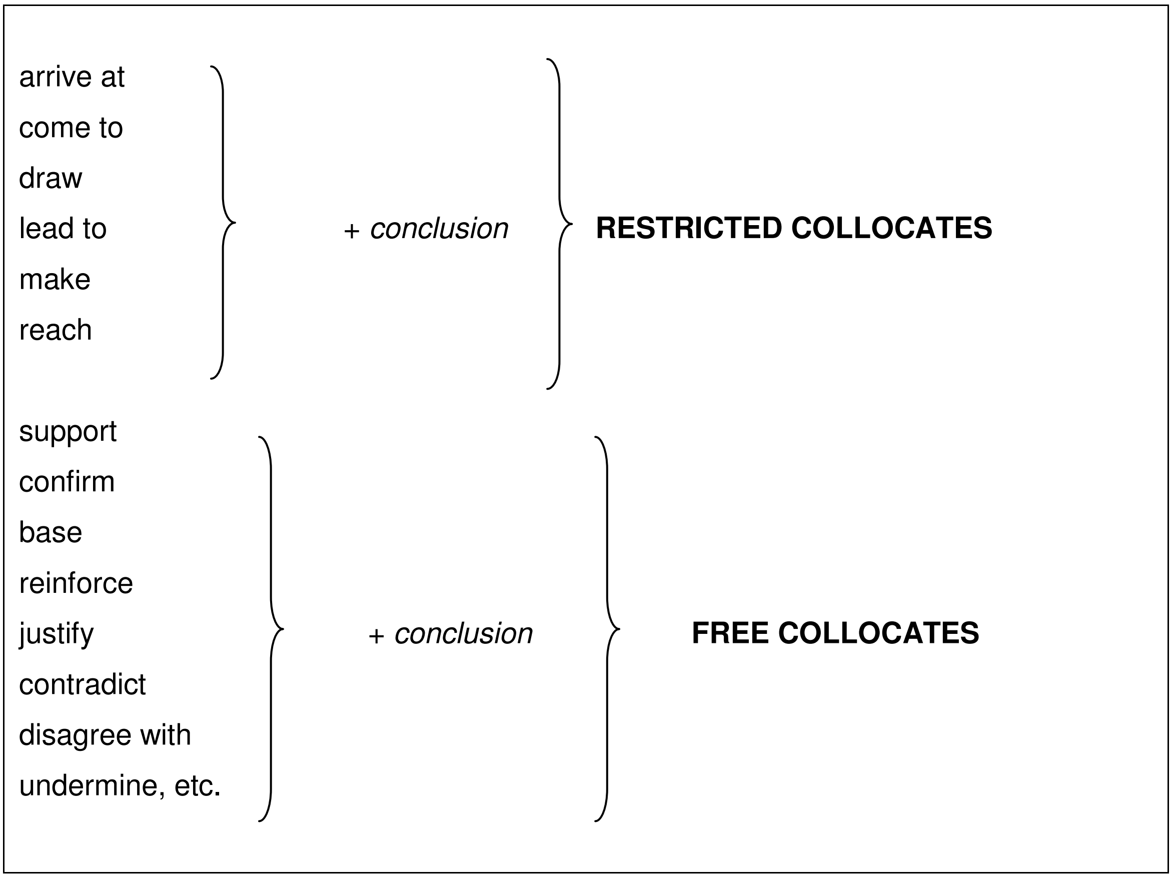

arrive at
come to
draw
lead to
make
reach
+ conclusion
RESTRICTED COLLOCATES
support
confirm
base
reinforce
justify
contradict
disagree with
undermine, etc.
+ conclusion
FREE COLLOCATES

Figure 12 ***Verb + conclusion*** collocates in the *HSC*

As can be seen in Figure 12, there are several verbs that collocate with *conclusion*, so it seemed necessary to group them according to the different patterning and meanings associated with each of them. When combining with verbs, this abstract noun may perform the syntactic function of Object or Oblique (introduced by dummy prepositions) and can follow different types of verbs. From a syntactic point of view, thus, these verbs share the same syntactic behaviour, whereas semantically-driven differences can be identified.

The frequent use of this abstract noun in combination with a group of **restricted collocates** is of particular interest. All these verbs are used in a specialised or figurative sense and have undergone, as will be further discussed in section 4.8, a certain degree of **delexicalisation** and **semantic bleaching**; that is, it is the noun itself that mostly provides the semantic content to the whole unit. On the contrary, *arrive at*, *come to*, *draw*, *lead to*, *make* and *reach* have lost their full semantic content of *reach a place*, *move*, *decide*, *take someone somewhere* and *arrive*, respectively, and form part of extended lexical units. As Tognini-Bonelli (2001:116) points out, these verbs simply carry the verbal function and at the same time allow the noun *conclusion* to remain as a noun, but they have lost their primary meaning and have adopted a more grammatical role.

Another significant feature of these restricted collocates lies in the fact that they are all equivalent to the corresponding verb of *conclusion*; that is, *conclude*, and, therefore, they are equally interchangeable. The following figures show the frequency and percentage of occurrences of both restricted and free collocates, based on the overall number of occurrences of **Verb +** ***conclusion*** (192) in the *HSC*:

Verbs that collocate with the abstract noun ***conclusion*** **in the** ***HSC***	**Number of occurrences (%)**
DRAW	26 (39,4%)
LEAD TO	17 (25,7%)
REACH	13 (19,7%)
MAKE	6 (9,1%)
ARRIVE AT	2 (3,03%)
COME TO	2 (3, 03%)
Overall number of occurrences: 66 (100%)	

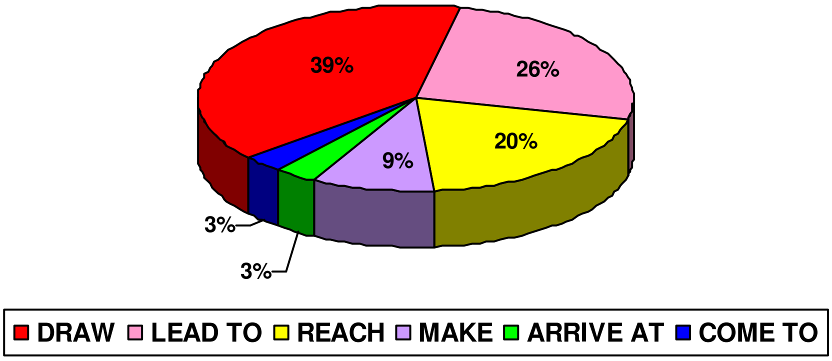


Figure 13 Percentages of occurrences of restricted collocates of the type **Verb +** ***conclusion***

Verbs that collocate with the abstract noun *conclusion* in the *HSC*	**Number of occurrences (%)**
SUPPORT	67 (53,2%)
CONFIRM	9 (7,1%)
BASE	6 (4,8%)
JUSTIFY	3 (2,4%)
REINFORCE	3 (2,4%)
CONTRADICT	3 (2,4%)
LIMIT	2 (1,6%)
PRODUCE	2 (1,6%)
EXPLAIN	2 (1,6%)
DISAGREE WITH	2 (1,6%)
CRITICIZE	2 (1,6%)
ACCEPT	2 (1,6%)
Others: accompany, add weight to, affect, allow, alter, challenge, dismiss, doubt, emphasize, express, flaw, group, highlight, influence, modify, publish, question, refute, strengthen, suggest, test, undermine, verify	1 (0,8% each = 18,4% in total)
Overall number of occurrences: 126 (100%)	

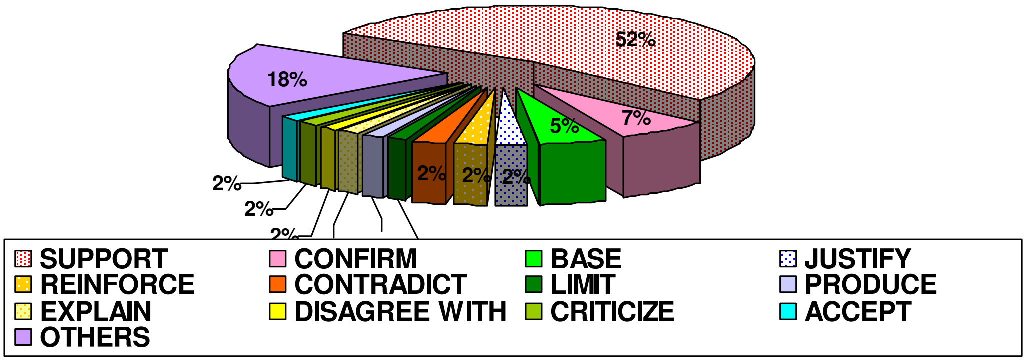


Figure 14 Percentages of occurrences of free collocates of the type **Verb + *conclusion***

In the following pages, **restricted collocations** will be examined. According to Firth's statement (1957:179): "*you shall know a word by the company it keeps*", there is no doubt that these restricted collocates give the observer a great deal of information. Each verb has its own specific particularities, but an undeniable generalisation can be made about the pattern they share.

With a total of 26 occurrences (8 in the active voice and 18 in the passive), *draw* is the most frequently used delexicalised verb in combination with *conclusion*. Consider the following examples extracted from the *HSC*:

```
WordSmith Tools -- 07/05/2008 16:22:09

1       rials are therefore   needed before any conclusion can be drawn about the
2       ntal transfer are necessary before any  conclusions can be drawn about the
3       nt quality to draw definitive           conclusions          Limited evidence
4       tal. There were too few twins to   draw conclusions. Several variables were not
5        in Drosophila.  Where possible we draw conclusions that have broad implication
6       lity should be considered when  drawing conclusions from observations in chimer
7       the student has little insight into how conclusions are  drawn. Here, the redee
8       rrors (discussed below). Two important  conclusions were drawn from this study.
9       lami.(24) Overall, the most interesting conclusion to  be drawn from these obse
10      ion mutants were generated in total. No conclusion  drawn in this manuscript wa
11      f distinct genes, we can draw   several conclusions about Nipped functions and
12      d, GAL4±Bcd or Bcd±VP16.        Several conclusions can be drawn from the gene
13      n-Alder and Bennett (Æ81) drew similar  conclusions for animals generally. Esti
14      culate, quite reasonably, that similar  conclusions might be drawn about many d
15      s no effect on body weight.  The simple conclusion drawn from these genetic obs
16       h45 mutation).   Before drawing strong conclusions from the above DNA sequenc
17       across-generation analyses support the conclusions drawn from the   within-gen
18      etween vital and nonvital genes, so the conclusions drawn here should apply to
19       by N17  Rac, and therefore support the conclusion drawn from the use of  the V
20        uteri of mice mated to t/1 males. The conclusion drawn is that  TRD cannot be
21      velopment to term was much better. The  conclusion drawn from this elegant stud
22      t  there were no interferences with the conclusions drawn  because of alkanes f
23      sm driving the outbreaks. In  turn, the conclusions that one can draw  from thi
24      er  field conditions. Nonetheless, the  conclusions that can be drawn from  eac
25      bias.              Discussion       The conclusions that can be drawn from our
26      but the data do not permit  unequivocal conclusions to be drawn. Using CTC
```

Figure 15 Overall number of occurrences of ***draw* + *conclusion*** in the *HSC*

This periphrastic use reveals a syntactic preference for passive constructions (18 out of 26 occurrences), which is typical of academic writing. In these cases, the source of the *drawn conclusion* is expressed by means of a prepositional phrase, introduced by the dummy preposition "*from*", whose dependency (a noun phrase) refers to scientific processes and mechanisms: *observations*, *studies*, *experiments*, *procedures*, etc.

The examples given so far demonstrate that concepts such as **delexicalisation** and **semantic bleaching** have become central in the identification of word meaning. A semantic shift from the core meaning of *draw* as "make, create something", in its literal sense, towards the meaning of "*draw a conclusion* = *conclude*" has undoubtedly taken place. What also seems worth noting is that this delexicalised use of *draw* also entails a change in the semantic roles of the participants involved in "drawing a conclusion". Consider the following example from the corpus: "*where possible we draw conclusions that have broad implications*". The subject "*we*" can no longer be interpreted as the agent (active instigator of an action) but as the experiencer who concludes something after discussing it carefully.

What has been posed so far about ***draw* + *conclusion*** can also be applied to *lead to* (17 occurrences), *reach* (13 occurrences), *make* (6 occurrences), *arrive at* (2 occurrences) and *come to* (2 occurrences). Unlike *reach*, though, the other delexicalised verbs tend to appear in passive constructions. It seems worth mentioning, however, that the periphrasis "*reach a conclusion*" hardly subcategorises for an animate volitional Subject but for noun phrases associated with demonstrable, evident data. The same could be said about "*arrive at / make a conclusion*", as the following examples extracted from the *HSC* show:

1. Studies of ftsZ transcription in E. Coli have reached conflicting conclusions.
2. Other cells with multiple spindles have reached conclusions that agree with those drawn from studies...
3. Previously published studies could not have reached this conclusion.
4. The analysis of this pathway allowed several important conclusions to be made.
5. Three recent studies arrived independently at the conclusion that of those genes in the yeast genome for which...

As regards *arrive at* (2 occurrences) and *come to* (2 occurrences), it should be stressed that these collocates are rare in the *HSC*. In contrast, a random search for *conclusion* in *The British National Corpus* demonstrates that they are frequent collocations. From a corpus-based approach these two verbs should have been identified as prominent collocates of this abstract noun, but evidence from the health science corpus would not have supported such statement. A corpus-driven approach, on the contrary, takes the evidence as the starting point. This method, as Francis (1993:155) suggests, allows the investigator "*considerable freedom of movement*". Some expressions may be frequent in some contexts but almost non-existent in others, as Figures 16 and 17 report:

```
1       sublethal  infections and came to four conclusions. (1) Individuals  survi
2       345go. Cooper and Penny(8) come to similar conclusions for  modern bird gr
3       23 nt studies arrived independently at the conclusion  that of those genes
4       501endocytic pathway. We   arrived at this conclusion because endocytosed
```

Figure 16 Overall number of occurrences of *arrive at* / *come to* + *conclusion* in the *HSC*

Results of your search

Your query was

conclusion

JY1 **1885** It was a foregone conclusion, I soon realised, that after university he'd start up his own firm with some money his father left him.

HNM **60** The assumptions made earlier in this chapter may be used to derive what has become known as the `separation theorem'; which underpins the foregoing conclusion.

ACH **887** I have for myself come to the conclusion that owing to the conditions which exist in the world today, having regard to the economic environment, having regard to the situation of our country, if we go on pattering along as we are we shall have grave unemployment with us to the end of time, and I have come to the conclusion that the only way of fighting this subject is by protecting the home markets.

ASF **767** St Augustine came to the conclusion that we can measure time only if the mind has the power of holding within itself the impression made by things as they pass by even after they are gone.

CB9 **974** In 1984, radio astronomers from Columbia University identified electro-magnetic vortices in our Milky Way and arrived at a similar conclusion to the theories of Alfven and Perratt.

EVX **139** We have come to this conclusion quickly.

FBM **445** This was the closest Neil had ever come to Wembley and on the way home that night he came to the conclusion that this was the closest he would ever come in his entire life.

FBT **134** I reach that conclusion, I hope, having properly considered the pros and cons of the situation; the fact that he has this particular advantage as a police officer in marshalling evidence; considering the issues.

H81 **1252** I think I should add very shortly that having considered the many authorities cited, even if I had come to a different conclusion on the issue about consideration, I would have come to the same decision adverse to the owners on the question whether the payments were made voluntarily in the sense of being made to close the transaction.

Figure 17 Random selection of the search word *conclusion* from *The British National Corpus*

On the other side of the coin, a wide range of free verbs have been found in colligation with the abstract noun *conclusion*. Contrary to what was observed with restricted collocates, verbs such as *support*, *confirm*, *base*, *reinforce*, *limit*, *contradict*, etc. do not equal the meaning of the verb *conclude*. Thus, the phenomenon of delexicalisation is not present here since each of these free collocates add their own particular meaning to the object *conclusion*. Take the following examples as an illustration[66]:

6. *These across-generation analyses* ***support*** (≠ conclude) *the conclusions drawn from the within-generation models*. ~ **These across-generation analyses* ~~***conclude***~~ *the conclusions drawn from the within-generation models*.
7. *These across-generation analyses support the conclusions* ***drawn*** (= conclude) *from the within-generation models* ~ **These across-generation analyses support what the within-generation models* ***concluded****.*

The verb *support* deserves special attention as it stands out as the most frequent free verb collocate (a total of 67 occurrences out of 126) in combination with *conclusion*. Although there is a wide range of verbs that collocate with this abstract noun, some of them appear just once in the corpus (see Figure 14 for exact percentages and number of occurrences). The verb *support* is much more frequently used (71,6%) in the active voice and the subjects it subcategorises for are hardly ever performed, but for one exception[67], by animate volitional entities. On the contrary, they are realised by NPs that fall into the same semantic field (i.e. *observations, results, data*, etc.) See the examples below:

8. *The data presented in this paper support the following conclusions.*
9. *These observations strongly support our conclusion.*
10. *These findings were used to support the conclusion.*
11. *Considerations of ultimate and proximate function support the conclusion that...*
12. *Taken together, our results support the conclusion.*
13. *The present experiment seems to support this conclusion.*

---

[66] The asterisked examples have not been taken from the *HSC*.

[67] "*We support the conclusion that relatively small but vital changes in experimental protocols play a key role in underpinning the recent successes.*" (*HSC* example)

After having described the behaviour of the **Verb + Noun [*conclusion*]** pattern in the *HSC*, it can be concluded that when such a pattern does not equal the meaning of its cognate verb *conclude*, the use of a periphrastic construction is undeniable. At the other end of the continuum, it is an evident fact that the occurrences of *conclude* (418) outnumber the occurrences of restricted collocates (66), but as syntax and lexis are co-selected, "*we cannot look at either of them in isolation*" (Francis 1995). If *reach*, *make*, draw, *lead to*, *arrive at* and *come to* equal the verb *conclude*, what is the point in using periphrastic constructions? Is there any difference between analytical and synthetic structures? One specific finding which clearly emerges from the corpus data is that there seems to be a preference for the periphrastic use (i.e. delexicalised verb + *conclusion*) in passive structures, whereas in the active voice there is a strong tendency for "*conclude*" to be used.

c) *conclusion* + preposition

The abstract noun *conclusion* is commonly followed by prepositional phrases dependent of nouns. See the Table below:

conclusion OF~	analysis, investigation, results, studies || Hatini et al., Peschken and Sawchyn, Templeton and Read
conclusion ABOUT~	efflux, toxicities || levels, programs, structures || risks
conclusion ON~	this || the scale's invulnerability
conclusion FROM~	this ||analysis, data, observations, studies
conclusion FOR~	findings || (bird) groups

Table 7 Frequent NPs as Prepositional complements of the ***conclusion* + preposition** collocational pattern

A significant property that can be applied to the prepositions *about* (notably dependent of the noun *conclusion* as it states the nature / typology of what is being concluded) and *of* is what Huddleston and Pullum (2005:136) refer to as **grammaticised uses** of prepositions. The role of such prepositions here is not to serve to indicate spatial and / or temporal relationships of nominals to each other, typical of prepositions, but to mark certain grammatical functions. Consider the following *HSC* instances:

14. *Ignoring chemical persistence may lead to erroneous conclusions about the risk that a chemical poses to an ecosystem.*
15. *The biological conclusions about the program are not justified.*
16. *We also highlight an important conclusion of these studies.*
17. *Willis and Wiese [1997] further criticize the conclusions of Templeton and Read [1983, 1984]*

The preposition *about* in the Prepositional Phrases in examples (14) and (15) marks the complement of a noun (*conclusion*). This noun allows a Prepositional Phrase introduced by a particular preposition, which is specified by the noun itself. Thus, this preposition determines the kind of *conclusion* the writer is referring to. This function can also be performed by a "*that*-clause", which usually (71 occurrences) follows the abstract noun under study. See the examples:

18. *This lends a degree of robustness to our conclusion that universal screening would be cost effective.*
19. *Thus, these results support the conclusions that Africans are the most divergent population.*

The underlined sequence in example (17) is a NP within which *of* marks the NP (*Templeton and Read*) that equals the Subject of the corresponding clause ("*What Templeton and Read [1983, 1984] concluded was further criticized by Willis and Wiese [1997]*"). According to Huddleston and Pullum (2005:136), "*what makes 'grammaticised' an appropriate term for such prepositions is that where they are placed in sentences depends not on what they mean but entirely on rules of the grammar.*" This is especially noticeable in the case of the preposition *of*, which has no meaning of its own and there is no possibility of replacing it by any other preposition. This statement cannot be applied to *about* as it can also be replaced by the preposition *on* (2 occurrences), which is much less frequently used, though.

d) preposition + *conclusion*

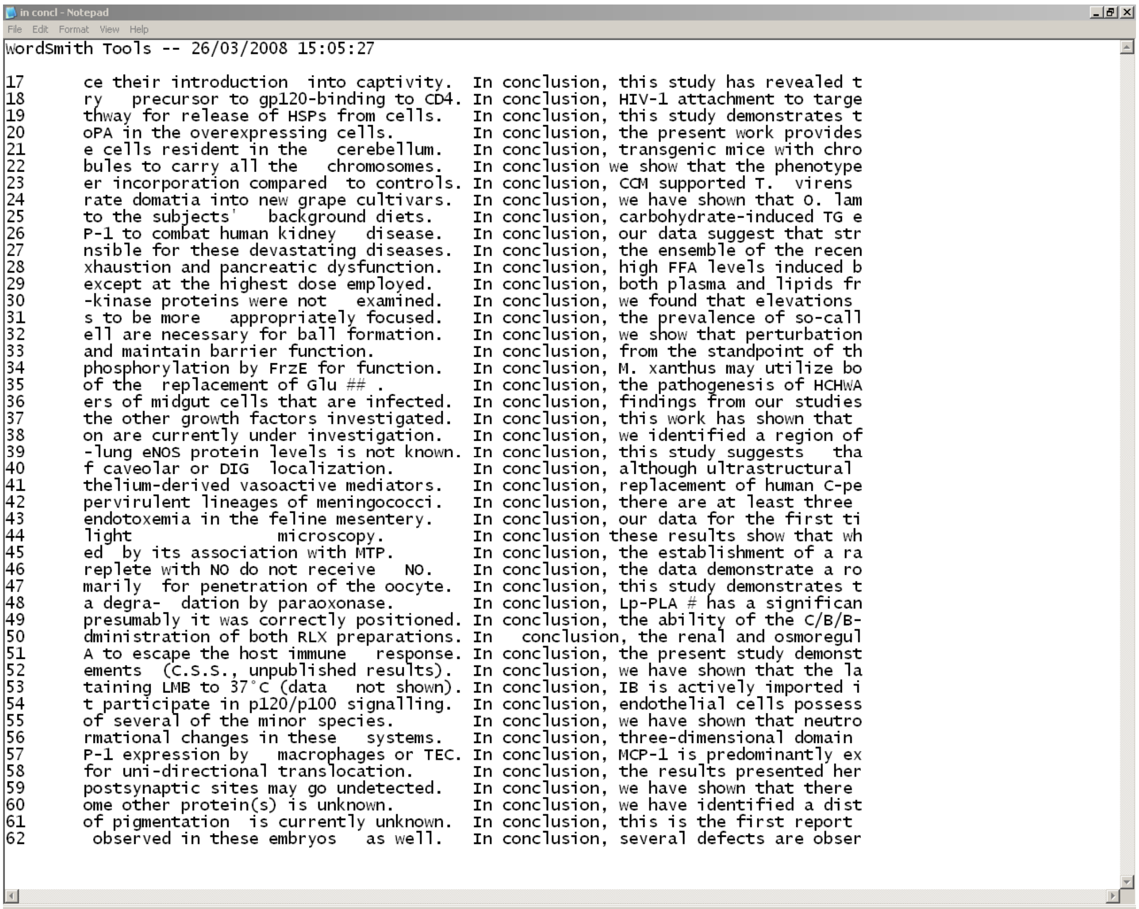
The pattern *in conclusion* is particularly salient in the *HSC* (65 occurrences) due to its frequent use as a discourse marker. It is obvious throughout the corpus that this construction constitutes an extended unit of meaning and must thus be understood as a whole, rather than the juxtaposition of the individual meanings of *in* + *conclusion*. Likewise, by a further process of grammaticalisation, such a construction in this particular linguistic context of scientific discourse has come "*to serve (a) grammatical function and, once grammaticalized, continues to develop new grammatical functions*" (Hopper & Traugott, 1993:18).

Understood as a lexical bundle[68], this holistic unit appears to have undergone a process of semantic shift, being equivalent to "sum up, finish, end up with something". It definitely serves the discourse function of introducing a summary and anticipating that the discourse is about to come to an end. It is also particularly notable the fact that it is separated from the rest of the text by means of a comma, which contributes even more to underline its entity as a boundary marker; more specifically, as a summative conjunct.

As observed with other kinds of collocates, the noun phrases associated with the abstract noun *conclusion* are characteristic of the described register; that is, they all belong to the field of results, findings, studies, reports, analyses and measurable data. Consequently, it comes as no surprise the fact that the construction *in conclusion* is commonly used introducing summaries and overt evaluations related to that semantic field, as shown below:

---

[68] Cf. Biber et al. 1999 and Cortes 2002.



```
WordSmith Tools -- 26/03/2008 15:05:27

17      ce their introduction  into captivity.  In conclusion, this study has revealed t
18      ry   precursor to gp120-binding to CD4. In conclusion, HIV-1 attachment to targe
19      thway for release of HSPs from cells.   In conclusion, this study demonstrates t
20      oPA in the overexpressing cells.        In conclusion, the present work provides
21      e cells resident in the   cerebellum.   In conclusion, transgenic mice with chro
22      bules to carry all the   chromosomes.   In conclusion we show that the phenotype
23      er incorporation compared  to controls. In conclusion, CCM supported T.  virens
24      rate domatia into new grape cultivars.  In conclusion, we have shown that O. lam
25      to the subjects'   background diets.    In conclusion, carbohydrate-induced TG e
26      P-1 to combat human kidney   disease.   In conclusion, our data suggest that str
27      nsible for these devastating diseases.  In conclusion, the ensemble of the recen
28      xhaustion and pancreatic dysfunction.   In conclusion, high FFA levels induced b
29      except at the highest dose employed.    In conclusion, both plasma and lipids fr
30      -kinase proteins were not   examined.   In conclusion, we found that elevations
31      s to be more   appropriately focused.   In conclusion, the prevalence of so-call
32      ell are necessary for ball formation.   In conclusion, we show that perturbation
33      and maintain barrier function.          In conclusion, from the standpoint of th
34      phosphorylation by FrzE for function.   In conclusion, M. xanthus may utilize bo
35      of the  replacement of Glu ## .         In conclusion, the pathogenesis of HCHWA
36      ers of midgut cells that are infected.  In conclusion, findings from our studies
37      the other growth factors investigated.  In conclusion, this work has shown that
38      on are currently under investigation.   In conclusion, we identified a region of
39      -lung eNOS protein levels is not known. In conclusion, this study suggests   tha
40      f caveolar or DIG  localization.        In conclusion, although ultrastructural
41      thelium-derived vasoactive mediators.   In conclusion, replacement of human C-pe
42      pervirulent lineages of meningococci.   In conclusion, there are at least three
43      endotoxemia in the feline mesentery.    In conclusion, our data for the first ti
44      light               microscopy.         In conclusion these results show that wh
45      ed  by its association with MTP.        In conclusion, the establishment of a ra
46      replete with NO do not receive   NO.    In conclusion, the data demonstrate a ro
47      marily  for penetration of the oocyte.  In conclusion, this study demonstrates t
48      a degra-  dation by paraoxonase.        In conclusion, Lp-PLA # has a significan
49      presumably it was correctly positioned. In conclusion, the ability of the C/B/B-
50      dministration of both RLX preparations. In   conclusion, the renal and osmoregul
51      A to escape the host immune   response. In conclusion, the present study demonst
52      ements  (C.S.S., unpublished results).  In conclusion, we have shown that the la
53      taining LMB to 37°C (data   not shown). In conclusion, IB is actively imported i
54      t participate in p120/p100 signalling.  In conclusion, endothelial cells possess
55      of several of the minor species.        In conclusion, we have shown that neutro
56      rmational changes in these   systems.   In conclusion, three-dimensional domain
57      P-1 expression by   macrophages or TEC. In conclusion, MCP-1 is predominantly ex
58      for uni-directional translocation.      In conclusion, the results presented her
59      postsynaptic sites may go undetected.   In conclusion, we have shown that there
60      ome other protein(s) is unknown.        In conclusion, we have identified a dist
61      of pigmentation  is currently unknown.  In conclusion, this is the first report
62       observed in these embryos   as well.   In conclusion, several defects are obser
```

Figure 18 Randomly selected occurrences of the construction ***in conclusion*** in the *HSC*

e) adjective + *conclusion*

The pattern **Adjective +** ***conclusion*** is exemplified by utterances such as "*Inappropriate handling of missing response can produce misleading conclusions*". Some adjectives like *misleading* associated with this pattern are concerned with **opinions**; in other words, they describe the view that the writer has as regards a given *conclusion*. These **evaluative descriptors** (in order of frequency) are the following: *important*, *conflicting*, *misleading*, *strong*(*er*), *believable*, *erroneous*, *interesting*, *logical*, *obvious*, *popular*, *proper*, *realistic*, *significant*, *surprising*, *tentative*, *unequivocal*, *unexpected*, *untenable* and *straightforward*.

All these adjectives describe the position of the writer towards the discussed *conclusion*. However, it must be highlighted that this list is far from exhaustive. The fact that only **opinion** descriptors collocate with the noun *conclusion* in the native corpus does not prevent other types of descriptors from combining with it.

Thus, it would not be surprising to consider that other adjectives such as the antonyms of the list above may also be associated with this pattern. In this respect, Hunston and Francis (2000:102) point out that "*the lack of corpus evidence does not indicate that the missing occurrences are 'incorrect English*". Similarly, Moon (2007b) further argues:

> *"However large and representative a corpus, and however sophisticated its tools, some items, or some forms of item, may simply not be found. Yet it is not always clear how to interpret negative evidence. While in some cases it may be reasonable to assert that items do not exist in current English, it is usually preferable to moderate the assertion with hedges such as 'has not been found' or 'is unlikely to occur in'. In general, corpus data shows up tendencies, and powerful observations can be made on the basis of those tendencies; but to recast tendencies as rules would be inappropriate."* (Moon 2007b:1049)

Although most adjectives can appear both in attributive and predicative positions, the *HSC* corpus only shows **evaluative** adjectives being used attributively and predicatively. It is true, however, that the list of predicative adjectives describing the noun *conclusion* is fairly limited: *consistent*, *paradoxical*, *controversial*, *uncertain*, *tentative*, *robust*, *surprising*, *premature* and *unwarranted*. The examples below show that they are all subcategorised by intensive verbs (i.e. *be*, *seem* and *remain*) and that they relate to a predicand (*conclusion*/*s*) which is Subject of a complex-intransitive construction:

The last conclusion	*is*	*consistent with that observation.*
The latter conclusion	*is*	*paradoxical.*
The conclusions	*remain*	*controversial and highly uncertain.*
This conclusion	*is*	*unwarranted.*
This conclusion	*is*	*not at all surprising.*
This conclusion	*is*	*extremely tentative.*
This conclusion	*seems*	*premature in view of continuing uncertainty*
SUBJECT	**PREDICATOR**	**PREDICATIVE COMPLEMENT**

Table 8 Predicative adjectives in combination with the abstract noun *conclusion*

There are also some other attributive adjectives preceding *conclusion*. They differ from the ones already discussed in the sense that they are **relational/classificational/restrictive**[69] and more objective than the former. Here are some examples, with bracketed numbers indicating the overall occurrences in the *HSC*: *similar* (9), *firm* (4), *general* (3), *same* (3), *two* (3), *several* (2), *definite* (2), *current*, *different*, *final*, *last*, *latter*, *main*, *major*, *second*, *long-standing* (1 occurrence each).

Finally, a third type of attributive adjectives that appear in combination with *conclusion* refers to **topical/defining** adjectives, such as the following: *biological* (7), *molecular* (4) and *qualitative* (2). These adjectives delimit the kind of *conclusion* that is being described; in other words, they show the subject area or a relationship with the noun described.

The examination of the patterning of the node word *conclusion* in the *HSC* has allowed to identify the main traits of its phraseology, namely, being used as an introductory item, in combination with restricted (delexicalised) and free verbs, co-occurring with several prepositions giving rise in some cases to the creation of discourse markers (i.e. *in conclusion*) and collocating with both attributive and predicative adjectives. Table 9 below summarises the main colligational and collocational patterns of this abstract noun.

---

[69] Following Biber et al.'s classification of adjectives (1999).

COLLIGATION	COLLOCATION	SEMANTIC PROSODY	DISCOURSE FUNCTION
NPs (optional, either preceding or following the node word and in coordination with it)	Summary, future directions, prospects, unresolved issues, speculation.	Introduces a new section, separated from the main body of the discourse by means of a comma.	INTRODUCTORY WORD
Verbs (active / passive voice)	**RESTRICTED COLLOCATES** 1. DRAW 2. LEAD TO 3. REACH 4. MAKE 5. ARRIVE AT 6. COME TO	States a conclusion	PERIPHRASTIC USE DELEXICALISED USE (the Verb becomes a more grammatical verb)
	FREE COLLOCATES Support, confirm, base, justify, reinforce, contradict, produce, limit, explain, disagree with, criticize, accept…	Predicates something about a conclusion	PERIPHRASTIC USE
Prepositions (following the node word)	• ~ ON • ~ ABOUT • ~ FROM • ~ OF • ~ FOR	States the nature / type of what is being concluded.	Grammaticised use of prepositions; it marks a grammatical function: the complement of a noun. Grammaticised use; it marks the NP that corresponds to the subject of the corresponding clause.
Preposition (preceding the node word)	• IN ~	Introduces a summary	Conjunct, discourse marker, lexical bundle.
Adjectives 1. Attributive OPINION adj. EXPLANATORY adj. DEFINING adj. 2. Predicative OPINION adj.	important, conflicting, misleading, strong… firm, similar, general, long standing… biological, molecular, qualitative… paradoxical, controversial, tentative, premature, surprising	Describes the position of the writer towards the discussed conclusion. Explains some characteristics of the discussed conclusion. Defines the type of conclusion. Classifies / describes the position of the writer towards the discussed conclusion.	RELATION

Table 9 Main colligational and collocational patterns of the abstract noun *conclusion* found in the *HSC*

4.4 The patterns of the noun *agreement*

A close look at the grammatical and lexical patterning around the abstract noun *agreement* reveals that its recurrent phraseologies are associated with particular syntactic and semantic contexts. This noun shows particularly striking features; not only due to the fact that it forms part of many delexicalised structures but also because it seems to have substituted its corresponding verb *agree* in many fixed expressions, as will be discussed in the following pages.

a) verb + *agreement*

The linguistic analysis of *agreement* in the *HSC* shows that it often collocates with a verb. Grammatically, these verbs can be divided into two main categories[70]: **restricted** (*reach*, *find*, *lead to*, *make* and *be*) on the one hand, and **free** (*require*, *evaluate*, *test*, *establish*, *express*, *lack*, *expect*), on the other.

The behaviour of this **Verb + Noun** pattern is strongly reminiscent of the "verb + *conclusion*" combination (cf. section 4.3), in the sense that these restricted collocates are often used as delexical verbs. Verbs such as *reach*, *find*, *lead to* and *make* that collocate with *agreement* as their Object or Oblique[71] simply indicate that someone performs an action; that is, someone agrees on/with something/someone. What seems really important is that the literal meaning of these verbs has weakened and has adopted the role of a grammatical prop for the noun (Howarth 1996:113). Therefore, it can be concluded that a certain degree of **semantic bleaching**[72] has taken place and, as a result, the noun *agreement* has gained more semantic weight.

As will be discussed later on, this process of delexicalisation is especially relevant when dealing with abstract nouns. In the case of *agreement*, for instance, it is obvious that the frequent verbs it collocates with have gradually lost most of their original sense and become part of an extended unit of meaning; that is, the meaning of the whole expression is now distributed "*across a number of words*" (Sinclair 1987:110). These delexical uses are thus seen as figurative extensions from their literal meanings. Table 10 below charts a clear difference between the delexical uses of *reach*, *find*, *lead to* and *make* and its other literal and/or more conventional uses, which have been provided by the *Oxford Collocations Dictionary for Students of English* (2002).

---

[70] See classification in Howarth (1996).
[71] Notice that this syntactic function of Oblique is only applicable to the *to*-PP in the pattern *lead to an agreement*.
[72] Cf. Footnote 64

	DELEXICAL USES	**PRIMARY / CONVENTIONAL MEANINGS**
REACH	conclusion, agreement, decision	a peak, a stage, height, weight
FIND	conclusion, agreement	a way, species, food, hosts
LEAD TO	conclusion, agreement	a place, a cavity
MAKE	conclusion, agreement, decision, comparison	a hole, work, extracts, a study

Table 10 Delexical and primary uses of the verbs *reach*, *find*, *lead to* and *make*

However, if instead of looking at restricted collocates, one considers the list of free-choice verbs that combine with *agreement*, striking differences stand out. First and foremost, frequency of use must be taken into account because the general impression seems to be that the wider range of verbs reported, the less frequently they are used. The results are given in Figure 19:

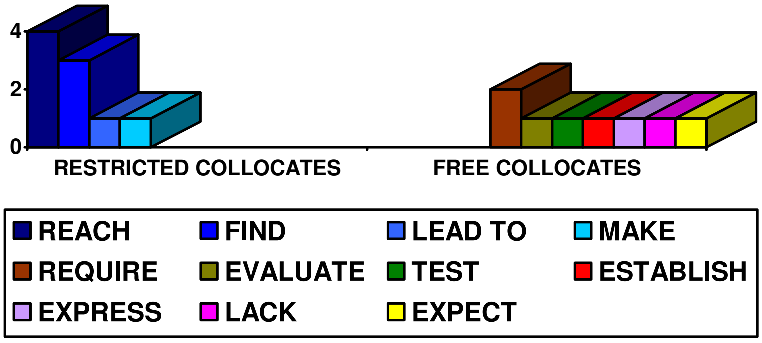


Figure 19 Number of occurrences of restricted and free collocates in combination with the noun *agreement*

Another remarkable feature lies in the fact that unlike restricted collocates, all instances of free collocates + *agreement* are used in the active voice. The most frequently used delexicalised verbs (*reach*, *find*[73]), on the contrary, are much more common in passive constructions, as the *HSC* examples show:

1. Can any *agreement* on principle be found?
2. *Agreement* on normality was found.
3. (...) in successive water samples until *agreement* was reached to + - 1 mmHg for samples with...
4. Despite almost universal use, *agreement* has not been reached on the need to administrate...

The collocation **BE (in) + *agreement***[74] has been purposely given a section on its own as it is the most frequently used "**V + agreement**" collocate and it shows certain particularities that should be further discussed. In a sense, this pattern should be regarded as a restricted collocate since the auxiliary verb "*be*" does contribute little to the meaning of the whole expression. However, it seems a bit risky to state that this periphrastic use always equals "*agree*", especially when being used in the past and when the Subject is animate (see examples 12, 14, 16). Semantically, "*be in agreement*" appears to entail some meaning extension from the action of "*agree*". The logical sequence would be as follows:

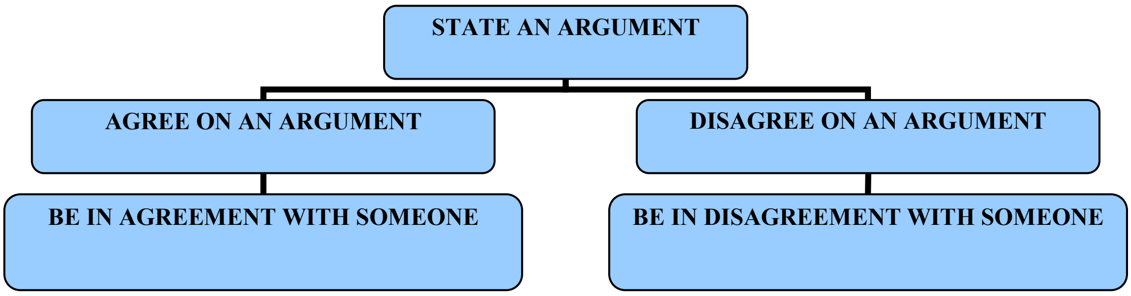


Figure 20 *agree* vs. *be in agreement*

---

[73] *reach* (3 occurrences in the passive as opposed to 1 in the active) and *find* (2 occurrences in the passive as opposed to 1 in the active).

[74] Here we will only refer to the periphrastic use of "*be in agreement*". "*Be (an) agreement*" will be analysed later on in combination with a fixed expression introduced by the existential "*there*".

Consider these pairs of examples[75]:

5. The results from the two methods were in agreement
6. The results from the two methods agreed. =

7. Our findings are in agreement with a large body of evidence.
8. Our findings agree with a large body of evidence. =

9. These observations are in agreement with the meiotic drive hypothesis. =
10. These observations agree with the meiotic drive hypothesis.

BUT

11. To my great pleasure she also agreed to collaborate in this venture.

≠

12. To my great pleasure she was also in agreement to collaborate in this venture

13. Men and women in the study sample agreed to be examined by a geriatrician.

≠

14. Men and women in the study sample were in agreement to be examined by a geriatrician.

15. All the participants agreed that the workshop approach was valuable.

≠

16. All the participants were in agreement that the workshop approach was valuable.

Taking into account that syntax and semantics are mutually dependant, there seems to be a semantic motivation in the syntactic patterns associated with the construction "*be in agreement*". The typology of noun phrases performing the syntactic function of Subject is of key importance in the analysis of such a periphrastic structure. A careful examination of the following examples is required:

---

[75] Odd examples (i.e. 5, 7, 9, 11, 13 and 15) have been taken from the *HSC*, whereas even examples (i.e. 6, 8, 10, 12, 14, 16) have been created *ad hoc* so as to illustrate whether the verb *agree* and its periphrastic counterpart, *be in agreement*, are semantically interchangeable or not.

AGREE

***agree* as "accept"**

17. It was assumed that all pregnant women offered the test would *agree* to be tested.
18. Professors *agreed* to conduct seminars.
19. If you were this patient would you *agree* to this standard treatment if it added 1 week to your life?
20. JPN *agreed* the funding for the project.

***agree* as "be of same opinion"**

21. In sum, we *agree* with Schullery and Whittlesey.
22. We *agree* with Dawkins [1980] that the presence of an abnormal behavior...
23. Most biologists *agree* with the latter authors.
24. They sometimes appear not to *agree* with each other.

***agree* as "tally, coincide, be equivalent"**

25. These findings agree with those from the Nuf2p-GFP analysis.
26. The results of both these studies *agree* with our forecasts.
27. These data *agree* with a direct binding solid-phase assay.
28. Our results *agree* with those obtained by Janknecht et al.
29. The models *agree* with the data qualitatively.

BE IN AGREEMENT

***be in agreement* as "be of same opinion"**

30. One researcher from the US and one from Poland *are in* broad *agreement*.
31. Foster et al 1988 and Kalderon and Rubin 1988 *are in agreement*.

***be in agreement* as "tally, coincide, be equivalent"**

32. These findings *are in* full *agreement* with our evidence.
33. These values *are in* good *agreement* with electro-physiological estimates.
34. These fluorescence data *are in* good *agreement* with observations on R. capsulate.

35. The results from the two methods *were in agreement*.
36. The MIC data *are* entirely *in agreement* with the topological model.
37. These observations *are in agreement* with earlier findings by our group.

The examples listed above report a clear preference for the periphrastic use (*be in agreement*) over *agree* when referring to coincident figures, data, values, etc. (70 occurrences as opposed to 37). Once again, it is worth mentioning that the NPs that perform the syntactic function of Subject of this construction are all related to the same semantic field of quantifiable, evident data.

Regarding the meaning of "be of same opinion", the corpus data show that this sense is less common. Despite this low frequency, the full verb "*agree*" is more likely to be used. Needless to say that this semantic use requires an animate volitional Subject, so the noun phrases found performing this function are realised by personal pronouns, mainly *we* and *they*, and/or proper names.

A third semantic sense traced is related to "accepting something". No instances of this meaning are expressed by means of the periphrastic construction *be in agreement*. Examples (17-20) **cannot** be paraphrased as follows:

38. ? It was assumed that all pregnant women offered the test would *be in agreement* to be tested.
39. ? Professors *were in agreement* to conduct seminars.
40. ? If you were this patient would you *be in agreement* to this standard treatment if it added 1 week to your life?
41. ? JPN *were in agreement* the funding for the project.

The fact that this use of *agree* in example (20) selects an Object as part of its subcategorisation makes the periphrastic structure in example (41) an unacceptable substitute in this case. The following Figure summarises the polysemy associated with these two expressions:

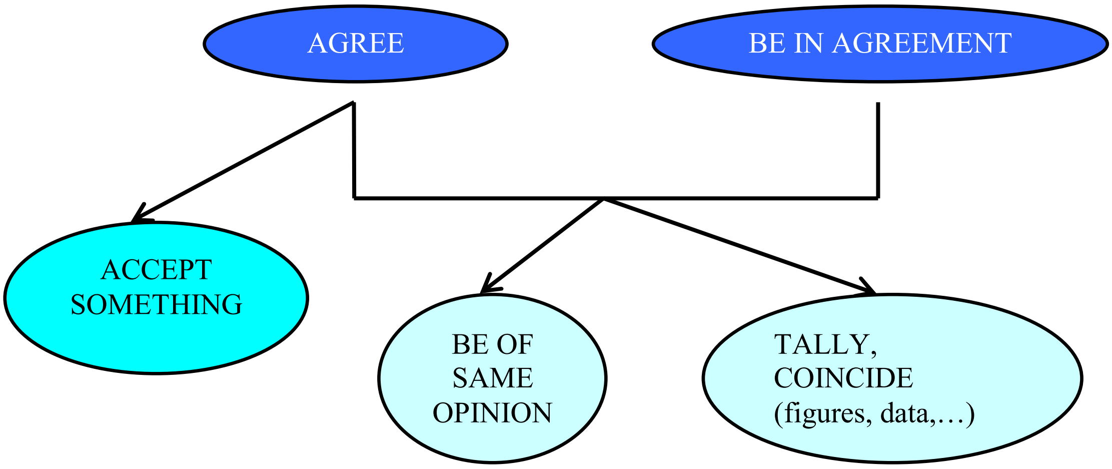


Figure 21 Semantic connotations of *agree* vs. *be in agreement*

Finally, it should be underlined that in comparison with other abstract nouns, *agreement* is characterised by its semantic load as a noun; in other words, it is more likely to appear in combination with other units (i.e. as a prepositional complement or a predicative complement, and as notional Subject) rather than in combination with either full or delexical verbs. In such cases, though, the noun contributes to a great extent to the meaning of the extended unit of meaning it belongs to.

b) adjective + *agreement*

Attention should now be addressed to the modifiers of the abstract noun *agreement* found in the *HSC*. There is a wide range of adjectives modifying *agreement* and most of them (96,7%) do appear in attributive position. In fact, there are only two instances in the corpus where adjectives modifying this noun are used predicatively:

42. Agreement regarding data abstraction was **good**.
43. However, agreement between symptoms and signs in people with varicose veins is so **poor**.

The lexico-grammatical patterning of *agreement* with reference to its modification is fairly straightforward. The adjectives associated with it on the concordance can be easily grouped into two main semantic domains:

descriptors and **classifiers**[76]. Descriptors found within this collocational span refer to either the extent of the *agreement* or the opinion the writer has with respect to a given *agreement* by denoting judgement and emphasis (**evaluative/emotive descriptors**). In contrast, classifying adjectives restrict the referent of a given noun (i.e. *agreement*). Adjectives found under the latter category can be grouped into subclasses, including **relational**/**classificational**/**restrictive** and **topical** adjectives.

Interestingly, descriptive adjectives stand out as the most frequent modifiers of *agreement*. The adjective *good* (20 occurrences) has gradually depleted its original qualitative meaning and embraces a new quantitative sense, in the same line as *broad*, *strong*, *general* or its negative counterpart, *poor*. See some *HSC* examples:

44. The results were in *good* agreement.
45. This finding is in *general* agreement with results reported previously.
46. They are in *broad* agreement.
47. This possibility is in *strong* agreement with the proposal that...
48. The poor predictive value of clinical signs and *poor* agreement between observers are major limitations.

As can be seen in Figure 22, the most frequent adjectives (*good*, *general*, *strong*, *close*) belong to the most objective category (**descriptors of extent**), whereas opinion adjectives (**evaluative**) are not that common. A possible reason to account for this particularity may lie in the fact that nouns being agreed or not are all related to demonstrable data, such as, findings, results, observations, figures and the like. Thus, more neutral adjectives are more likely to collocate with the abstract noun under study.

---

[76] Cf. Footnote 69

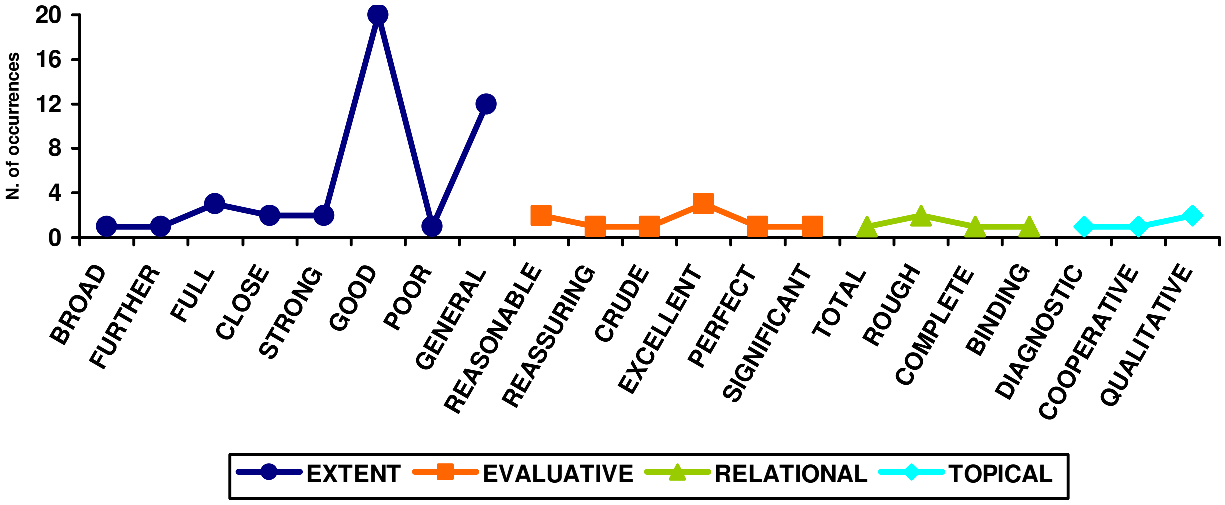


Figure 22 Frequency of occurrence of **adjective + *agreement*** collocations in the *HSC*

c) ***agreement*** + preposition

According to Biber (1999:634), "*prepositional phrases are by far the most common type of postmodifier in all registers*" although they are much more frequent in academic writing. Postmodifying prepositional phrases as dependencies of the abstract noun *agreement* can be introduced by different prepositions. Thus, it seems essential to examine in pretty much detail the choice of preposition in each case.

Prepositional phrases beginning with *among* (2 occurrences), *between* (18 occurrences) and *on* (6 occurrences) are less common than *with* (106 occurrences), but all of them are used to represent different semantic relations:

AGREEMENT ON → introducing a relationship of *respect*; that is, the topic of *agreement* (*on the matter of*)

HSC examples:

- a general *agreement on* the need to (...)
- general agreement on the format and criteria
- *agreement on* normality

AGREEMENT AMONG / BETWEEN → establishing common *agreement* shared by two or more people and/or figures, observations, results, etc.

HSC examples:

- *agreement between* zookeepers
- perfect *agreement between* the two methods
- general *agreement among* evolutionary biologists

AGREEMENT WITH → indicating the relationship *respect* (*with respect to*, *as regards*, *with reference to*)

HSC examples:

- in *agreement with* previously reported accounts
- in *agreement with* our hypothesis

(...) *agreement with* the results obtained

Table 11 Semantic relations expressed by the different prepositions postmodifying *agreement*

As stated earlier, the use of a *with*-PP (106 occurrences) postmodifying *agreement* outnumbers by far the occurrences of *on*-PPs (6 occurrences). Both postmodifiers represent a similar meaning, but the predominance of the former is due to the fact that the expression "*in agreement with*" is a highly frequent, recurrent sequence in scientific register. Also, the Prepositional complements of *with*-PPs may refer to "something or someone", whereas the Prepositional complements of *on*-PPs only refer to the agreed content ("something").

Regarding the combination "*agreement among / between*", it should be noted that it tends to be preceded by a fairly recurrent expression, "*the level of*", which must be interpreted as a lexical bundle; in other words, a sequence of items that occur together as a set. As examples (49-52) show, this 5-word bundle is open to the addition of an adjective (i.e. *high*, *substantial*) and the use of the (in)definite article is optional.

49. *The level of agreement between* the presence of symptoms and trunk varices is probably too low to be clinically useful.
50. We found significant *levels of agreement between* keepers.
51. There was a *high level of agreement between* the two methods.
52. Reasonably *high levels of agreement between* zookeepers can be accounted for.

Biber (1999:634) points out that in many cases, postmodification by prepositional phrases can be rephrased by means of a relative clause. The abstract noun *agreement* is hardly postmodified by a relative "*that*-clause" (only 7 occurrences) in the *HSC*, but it is true that this structure could replace an *on*-PP without changing its meaning. Compare the examples[77] below:

53. They lead to a general *agreement on* the need to re-establish populations.
54. They lead to a general *agreement that* populations need to be re-established.
55. Resolution of this problem requires *agreement on* the primitiveness of indirect development in various phyla.
56. Resolution of this problem requires *agreement that* indirect development is primitive.

d) preposition + *agreement*

This pattern is realised by the structure *in agreement with something*, which should be analysed as a lexical bundle that has undergone a remarkable process of delexicalisation. Biber (1999:999) states that most lexical bundles do not represent complete structural units. Instead, they span two structural units. This is also the case of *in agreement with*, which consists of a prepositional phrase (*in agreement*) + the beginning of another *with*-prepositional phrase. He also notes that most lexical bundles end in a function word; in this case, the preposition *with*.

---

[77] Odd examples (i.e. 53 and 55) have been taken from the *HSC*, whereas even examples (i.e. 54 and 56) have been created *ad hoc* so as to illustrate the fact that the *on*-PP postmodifying the noun *agreement* can be substituted by a *that*-clause.

Once again, certain degree of delexicalisation takes place as the noun *agreement* has lost its original semantic load and now it forms part of an extended unit of meaning. What is more interesting, though, is that this delexicalisation process has given rise to the lexicalisation of a new grammatical structure. Notice that the abstract noun *agreement* has lost its grammatical category as a noun and has become part of a prepositional phrase semantically equivalent to "*equally*, *likewise*, *in the same way*", corroborating what has already been stated and/or as a further demonstration of what has been found. The following examples show this correspondence:

57. *In agreement with* these results (=similarly to these results), earlier studies of lac permease have suggested that 31-residue insertions assembly competence.
58. *In agreement with* this concept (=following this concept), it was recently found that the same mechanism works as well with other sites.
59. *In agreement with* this observation (=following this observation), the average number of cells per colony on day 10 of culture was 1,400.

Another feature that reinforces this grammaticalisation is the fact that this construction is separated from the rest of the text by means of a comma and it is not premodified by any other element either. In this collocation the noun *agreement* forms part of a multiword sequence which shows not only a high degree of semantic unity but a process of **grammaticalisation** as well. This mechanism involves a transformation process in which lexical items become grammatical forms. Indeed, *agreement* is not used as a noun anymore but as an element of a complex phrase that functions semantically and syntactically as a whole grammatical unit.

e) The lexical bundle *there is* + (adjective) *agreement*

Last but not least, there is another lexicalised expression that deserves some attention as a lexical bundle characteristic of scientific discourse. Taking into account that "*only a few lexical bundles begin clauses in academic prose*" (Biber 1999:1023), the fact that 47% of the occurrences of "*there is* (adjective) + *agreement*" begin clauses in the *HSC* is significant.

This four-word bundle consists of the existential *there* as a Subject, the copula *be* (inflected for Present, Past and / or Present Participle), an optional premodifier (*general*, *good*, or the comparative *more*) and the abstract noun *agreement* as the notional Subject. This type of phrase, referred to as "stance bundle" by Cortes (2004:409), introduces certain degree of impersonality to what is being reported in the research articles under study. Once again a preference for periphrastic structures defines this genre.

4.5 The patterns of the noun *comparison*

As already pointed out, the key notion that lies behind the concept of lexical grammar is that certain patterns occur with restricted lexis (Hunston & Francis, 2000:96). When identifying the most typical patterns abstract nouns appear in, one easily realises that some words tend to appear in combination with others and this may enable the user of a language to make generalisations about the likelihood of a particular word to occur with a particular pattern in a given context.

However, Hunston and Francis (2000:96) further argue that "*it does seem to be the case that speakers of a language sometimes use a word with a pattern it does not typically have.*" Therefore, it should be speculated that when a given pattern consists of words associated with a particular meaning, speakers, by a process of analogy, tend to use similar words with that same meaning, so "*what words belong to a list is in a state of flux*" (ibid.) Several examples of this stated analogy can be detected in the patterns found with the abstract noun *comparison* in the *HSC*.

The first striking trait of this abstract noun refers to its frequency. Both the noun and its corresponding cognate verb, *compare*, are frequently used in the corpus (1196 and 2296 occurrences, respectively). While it is recognised that frequency counts are not the only reliable source of information with regard to pattern description, it is nonetheless clear that frequency is an important factor to validate generalisations made from corpus data. The next section will explore the patterns associated with the abstract noun *comparison* so as to present a comprehensive description of its main senses and uses.

a) verb + *comparison*

Once again, the verbs listed as having this pattern fall into two main categories: **restricted collocates*** that equal the meaning of *compare* (*make* / *perform* / *draw* / *establish* / *do* / *undertake* / *accomplish* / *conduct* / *carry out a comparison*) and **free collocates**. The latter constitutes a long list of verbs but, as stated with the free collocates of the noun *agreement*, their frequency is rather low in comparison with restricted collocates.

The most common verb in combination with *comparison* is *make* (28 occurrences). There are other verbs (i.e. *perform*, *use*, *accomplish*) that show a much smaller number of occurrences, but might have been used by a process of analogy, since they are synonyms of *make* and share the same core meaning of “create, produce”.

Following the idea that patterns and meanings are closely related, it is worth mentioning that not only do these verbs share a similar meaning when combining with *comparison* but also share a similar syntax. With the only exception of *draw*, the other verbs including *make* show a syntactic preference for passive structures, as data in Figure 23 demonstrate:

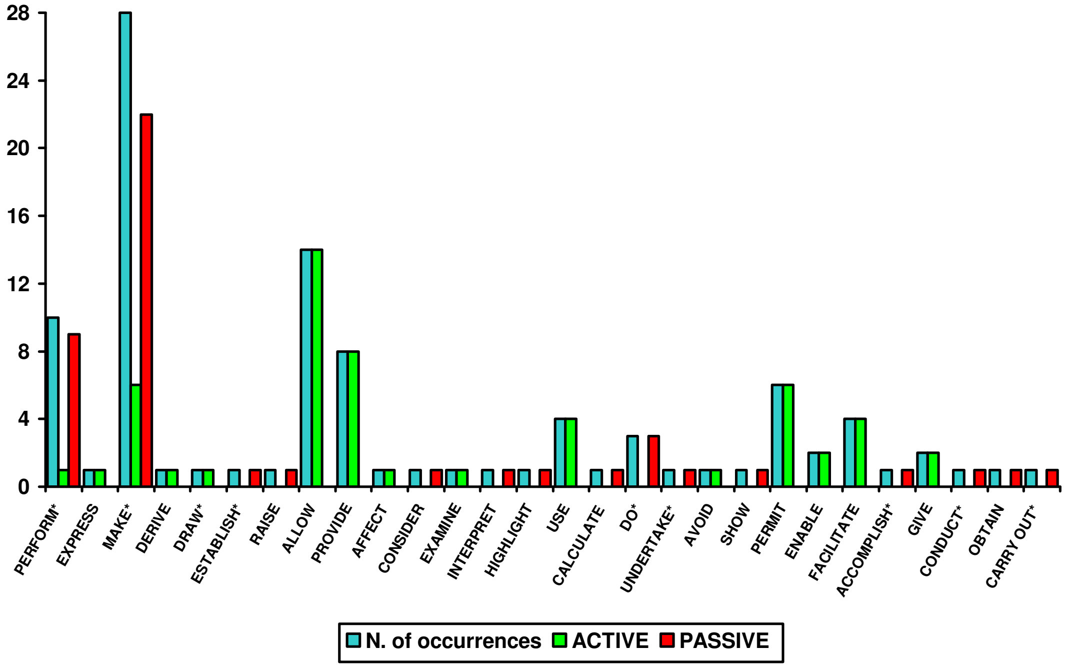


Figure 23 Collocational patterns of the noun *comparison*

Delexicalisation seems to be a commonplace when verbs that occur with abstract nouns equal the meaning of the noun's corresponding verb. *Comparison* is not an exception to this rule. These restricted verbs have lost their primary meanings and the abstract noun provides the semantic content to the whole unit.

Free collocates, on the contrary, outnumber restricted ones despite being much less frequently used. It would even be possible, by analogy, to enlarge that list with some other verbs such as *invite*, *bear*, *stand*, *analyse*, *value*, etc. Such limited occurrences (most of them are realised only once in the corpus) are not sufficient to conclude that they constitute representative collocations. On the other side of the coin, however, some other verbs similar in meaning to these free collocates could have also been used without altering the meaning of the pattern. These free verbs can be grouped semantically as follows:

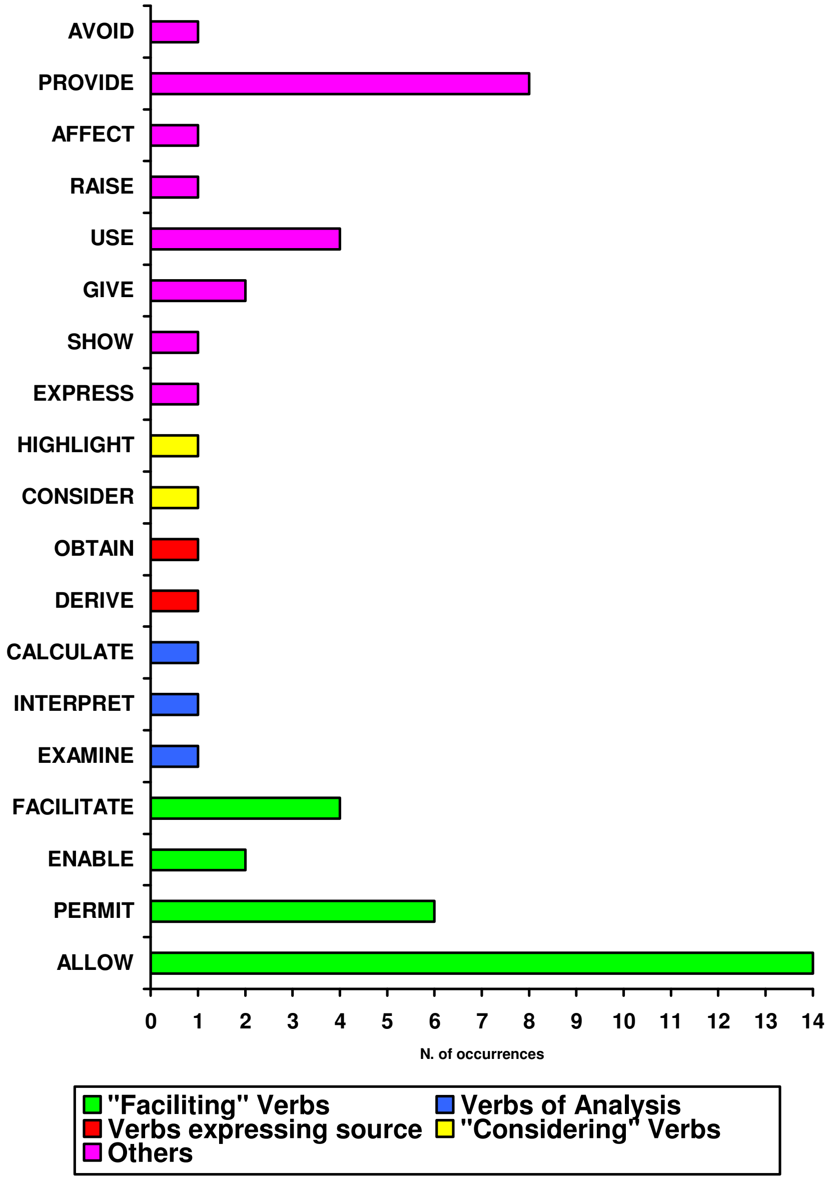


Figure 24 Semantic fields of free collocates of the type **V + *comparison***

In regard to the syntactic environment these collocates appear in, Figure 23 shows that, unlike restricted collocates, these verbs are mainly used in active constructions. Particularly noticeable as well as frequent enough is the group that embraces “facilitating” verbs, such as *allow*, *permit*, *enable*, *facilitate*. Here are some illustrating examples:

1. Our two estimates are based on the same two calibration dates to *allow direct comparisons* to previous results.
2. This has *allowed inter-specific and inter-gender comparisons*.
3. To *allow comparison* of the relative merits of alternative options we (...)

4. This *enabled a more direct comparison* with previous analyses that used such methods.
5. Use of bin assignments *enables comparison* among different maize mapping populations.
6. We chose these particular values because they *facilitate comparison* with Table 1.
7. The present study was conducted to *permit comparison* of serum 25 (OH)D concentrations.

To sum up, it must be indicated that not all delexical verbs that form part of restricted collocates show the same degree of delexicalisation. Comparing the examples (8-11) below, one realises that the verbs *undertake* and *carry out* still keep some semantic content in comparison with *make* and *draw*. Nevertheless, the syntactic environment of all these periphrastic structures has proven to be identical and all of them could also be substituted by the full verb *compare*.

8. The following series of *comparisons were undertaken* for each parasitoid species.
9. Pairwise *comparisons were made* using a paired t test.
10. In order to *draw comparisons* between the active site of calpain and those in other cysteine proteases...
11. Our *comparisons were carried out* on a uniform genetic background.

b) adjective + *comparison*

There are two important points to be made about this pattern: the noun *comparison* may be followed by a prepositional phrase beginning with *of*, and both descriptors and classifiers can be found modifying the abstract noun *comparison*.

DESCRIPTORS	CLASSIFIERS
valid (5): EVALUATIVE	direct (27): RELATIONAL
reliable (3): EVALUATIVE	detailed (4): RELATIONAL
sensitive (1): EVALUATIVE	exhaustive (1): RELATIONAL
relevant (1) : EVALUATIVE	statistical (15): TOPICAL
significant (1): EVALUATIVE	visual (2): TOPICAL
realistic (1): EVALUATIVE	
careful (3): DESCRIPTIVE	
recent (1): TIME	

Table 12 Classification of modifiers of the abstract noun *comparison* (number of occurrences in the *HSC*)

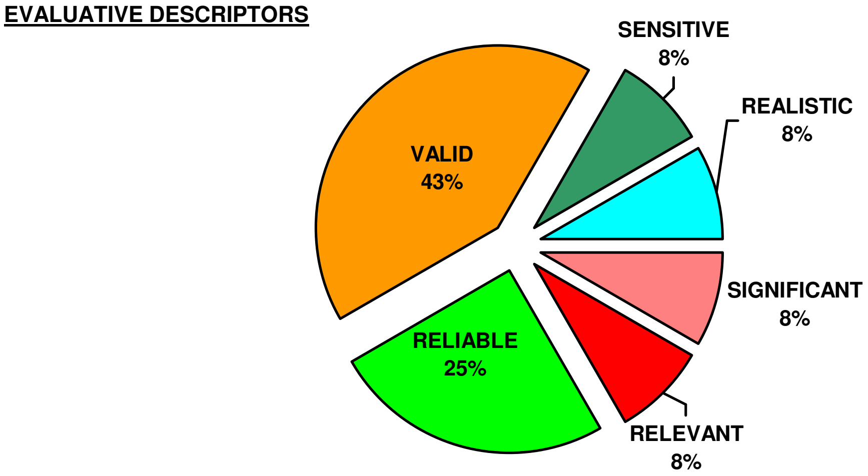


Figure 25 Distribution of evaluative premodifiers of the abstract noun *comparison*

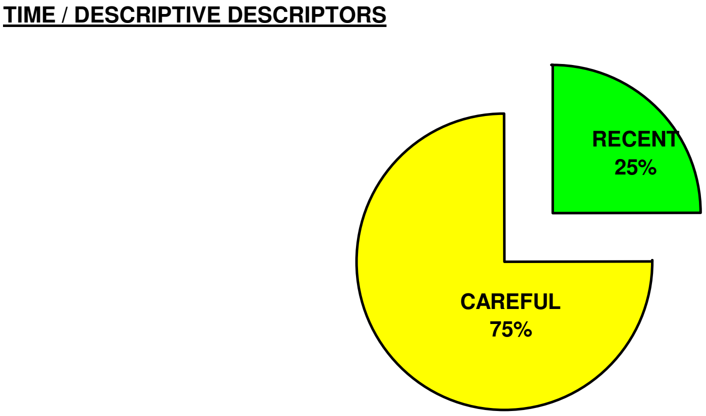


Figure 26 Distribution of time/descriptive premodifiers of the abstract noun *comparison*

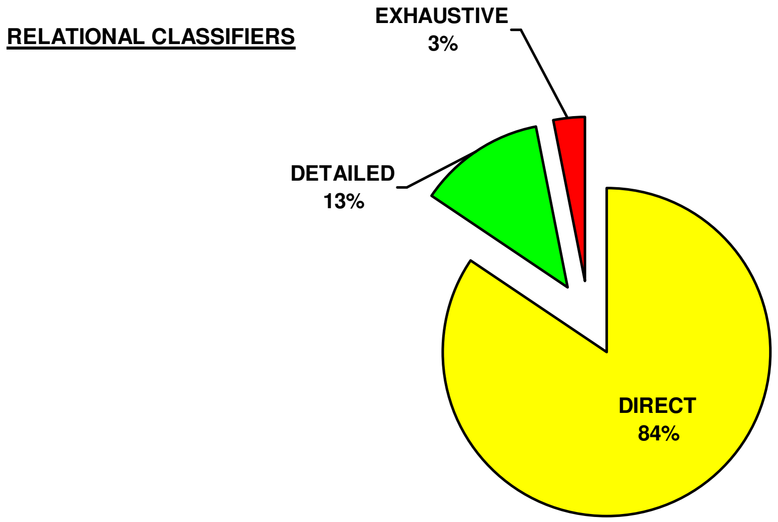


Figure 27 Distribution of relational premodifiers of the abstract noun *comparison*

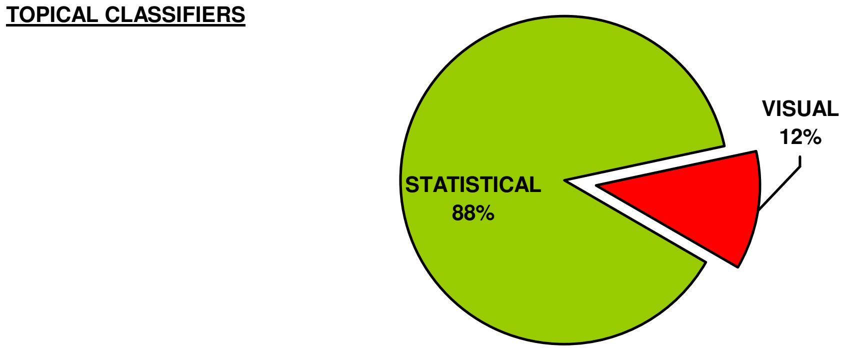


Figure 28 Distribution of topical premodifiers of the abstract noun *comparison*

As Figures 25-28 reveal, classifiers are much more common than descriptors. Most classifiers delimit the referent of *comparison*, whereas descriptors, such as *relevant*, *reliable* and *valid*, provide more subjective information concerning the suitability of the comparison made. This fact goes in line with Biber's et al. findings (1999:514), as they state that academic prose is characterised by its high use of classifiers, mainly relational and topical adjectives. Among the most frequent premodifiers of the abstract noun *comparison*, the relational adjective *direct* and the topical classifier *statistical* stand out as the most frequently used. In addition to showing a heavy reliance on relational adjectives, another notable feature of this pattern is that these adjectives appear in an attributive position, preceding, thus, the head noun *comparison*.

As will be discussed with much more detail when describing the pattern ***comparison* + preposition**, noun phrases in academic writing tend to have some kind of modifier. Consider the following *HSC* instances:

12. This theoretical argument is backed up by a *careful comparison* of the spectra of spontaneous (...)
13. In our recent study a *direct comparison* between both procedures was performed.
14. Eggplant in open plots did not produce any fruit, and thus no *statistical comparison* is needed.

The above examples are just a random selection of entries from the *HSC*, but the conclusion to be drawn from the examination of all occurrences (1196) is that the noun *comparison* always appears modified by premodifiers, usually adjective phrases, and/or postmodifiers (prepositional phrases). It is also frequent to find both kinds of modifiers in the same example. This finding should be taken very much into account because once again syntax is in the service of meaning. Needless to say that utterances such as "*several comparisons were made*" and "*this system can provide reliable comparisons of the relative amounts of RNA present at a given time*" differ to a great extent in the amount of information provided by modifiers.

In the former example, the lack of a longer context does not allow to infer what kinds of *comparisons* are being dealt with. Premodifiers, then, without the support of a postmodifier are seen as less explicit in helping identify the semantic relationship that exists between the modifier and the noun. The latter example, on the contrary, is much more specific as the abstract noun, *comparison*, is premodified by a semantically-driven adjective (*reliable*) and postmodified by an *of*-PP giving information about the compared data. Consequently, semantic motivations are more easily found in this type of pattern.

c) ***comparison*** + preposition

Most instances of the abstract noun *comparison* are postmodified by means of a prepositional phrase, which suggests that this combination of words occurs frequently. When investigating the different patterns of a word with the assistance of a concordancing program, one may be interested in sorting them into alphabetical order to the right or to the left. In the case of nouns and, especially, when the focus of attention lies in their complementation patterns, it seems advisable to sort them to the right. The following concordance lines taken from *WordSmith Tools*[78] have been sorted to the right and show some randomly selected postmodifiers of *comparison*.

```
WordSmith Tools -- 07/05/2008 17:35:28

298     ogous markers also allowed for a direct comparison of   the rate of meiotic rec
299         1998  in mosquitoes.   In the recent comparison of Arabidopsis and B. nigra,
300     643-651.[Medline]     Nuovo, G.J. 1991. Comparison of bouin solution and buffer
301      is               just out of view.     Comparison of the isomorphous crystal s
302     hard et al., 1997;   Lin et al., 1997). Comparison of the crystal structures of
303     ed with grey shading. (B)               Comparison of putative CKI motifs withi
304     fected cells are compared in Figure 3B. Comparison of the steady-state level
305     tney U test was applied for   pair-wise comparison of the data.   Intracellular
306     P65   and GRASP55-GFP to allow a direct comparison of the distributions of thes
307     ity observed for homologous proteins.   Comparison of the HMG-D-DNA structure w
308     te complex in the E.coli MalP enzyme.   Comparison of the binary complexes with
309     l basis of HMG domain specificity   The comparison of the bound HMG-D protein w
310     degradation. (A) Sequence               comparison of the first 26 residues of
311     proteases (Kamphuis et al., 1984)]. (B) Comparison of the D-IV-D-VI heterodimer
312     terminal extension (Nagy et al., 1998). Comparison of the   template-sized and
313     rows indicate a hypothetical route.     Comparison of histone H3/HMG-14 kinase
314      375±386 (Printed in Great Britain) 375 Comparison of the fibrin- binding activ
315     ange in a1.                   Fig. 9.   Comparison of the exposed surface of th
316     horImager.        Statistics        For comparison of differences between more
317     d reversed by unlabelled 14.4  kDa FBP. Comparison of the fibrin-binding affini
318     or `T',               respectively. (E) Comparison of the core metal-binding fo
319     ns of these two   proteins (Figure 4E). Comparison of the two images shows that
320     act of SPW3   (Fig. 6B). Interestingly, comparison of the total exoprotein prof
321     is. To permit an accurate and sensitive comparison of the effect of the   wild-
322     NA, thus   allowing direct and reliable comparison of the activities of the nat
323     gm for how the LTTR family operate.   A comparison of residues conserved, in te
324     FP fluorescence visualized (Figure 4D). Comparison of the   individual staining
325                                   Fig. 1.   Comparison of the GRASP55 and GRASP65 s
326     strates (Suzuki and Sorimachi, 1998).   Comparison of the active site and subst
327     (results not shown).        For further comparison of binding of HNF-3a and HNF
328     ournal Vol. 18, pp. 1598-1608, 1999   A comparison of in vivo and in vitro DNA-
329     x H12   that protrude from the monomer. Comparison of amino acid sequences indi
330                        Table I              Comparison of mRNA half-lives (min) bet
331     ls to the majority of genes. Based on a comparison of in vitro and in vivo DNA-
```

Figure 29 Concordance lines to the node word *comparison*

It has been previously mentioned that there are 1196 entries of *comparison* in the *HSC*, out of which 756 occurrences are explicitly postmodified by means of a prepositional phrase introduced by a wide range of prepositions, namely, *of*, *between*, *to* and *with*. Such prepositions are not content words on their own;

---

[78] Cf. Footnote 28 (p.48)

they are just function words that must necessarily be semantically-motivated. Figure 30 below gives account of the frequency of occurrence of each preposition in this pattern.

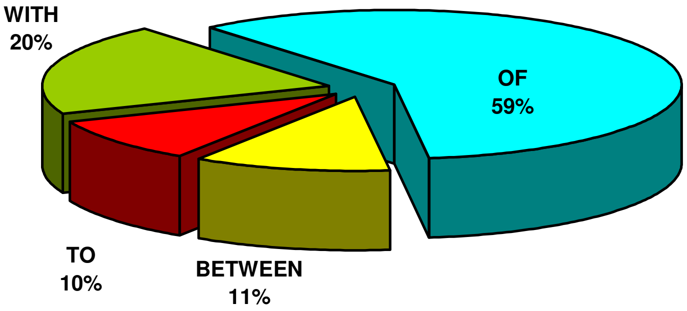


Figure 30 Prepositions following the noun *comparison*

What seems of vital importance is to analyse the noun phrases performing the syntactic function of Prepositional complement. As a first step, a representative sample of occurrences of each preposition following *comparison* will now be examined.

comparison + OF

N Concordance

1. Figure 1. *Comparison of* population growth for simulated population of pea
2. Table 1. *Comparison of* ELP scores for trainees attending the JWPT training
3. Berk, and B.P. Alter (1989) *Comparison of* erythroid progenitor cell growth in
4. et al., 1997; Lin et al., 1997). *Comparison of* the crystal structures of the
5. Figure 2. Comparison of relative frequency of total T cells and the CD4 and
6. transfer of endogenous Fig. 3. *Comparison of* decalcification procedures on
7. visualized (Figure 4D). *Comparison of* the individual staining patterns
8. both "subspecies." A *comparison of* frequencies of behaviors between the
9. a factor of three or more, a *comparison of* the results of the different methods
10. a *comparison of* treatment means (Table 2) by LSD test (P,0.05) revealed…

comparison* + *BETWEEN* / *comparison* + *AMONG

N Concordance

1. Values are mean ± SEM. *P < 0.05, *comparison between* dietary treatments.
2. a new window] Table 4. *Comparison between* children involved in road traffic
3. treated with caterpillar regurgitate. A *comparison between* the odors showed
4. *comparison between* patients with the acquired immunodeficiency syndrome
5. and fusions, direct *comparison between* rice and maize is facilitated by
6. Use of bin assignments enables *comparison among* different maize mapping

comparison* + *WITH

N Concordance

1. on estimates of nutrient composition and *comparison with* NRC guidelines for
2. cortisol levels from eland might serve as the best *comparison with* bongo.
3. lymphoma in the higher exposure categories. *Comparison with* other studies
4. *Comparison with* the thrombin receptor. J. Biol. Chem. 271:14910-14915.
5. This enabled a more direct *comparison with* previous analyses that used
6. To facilitate *comparison with* other studies, we used a Gaussian function for

comparison* + *TO

N Concordance

1. same two calibration dates to allow direct comparisons to previous results.
2. were determined by comparisons to a methylene blue-stained, 0.24- to
3. in origin, at least on the basis of comparisons to the proteobacterial genomes

Table 13 Random selection of right-sorted concordance lines of the pattern ***comparison*** + **preposition**

According to the above examples, the preposition *of* has revealed itself as the most common following the noun *comparison*. Without the intention of minimizing its frequency rates, it should be noted though that its high frequency is partly due to the fact that the structure "*comparison of*" tends to form part of the description of tables and figures in research articles, so, in a sense, some of these occurrences have not been taken from the body of the articles, but from the captions of such figures[79], as the ones shown below:

---

[79] And /or from descriptions referring explicitly to them.

15. Table 1. **Comparison of** cowpea and soybean ferritin subunits for mature proteins shows (...)
16. Figure 2. **Comparison of** relative frequency of total T cells.
17. Experiment 3: **Comparison of** Grasses to Nongrass plants.
18. Table 4. **Comparison of** effect on mortality of different drugs.

Noun phrases that perform the function of Prepositional complements can be semantically interpreted as noun phrases that form part of the same semantic field (i.e. *genes*, *populations*, *frequencies*, *procedures*, *structures*, *images*, *responses*, *results*, *patterns*, *treatments*, *measures*, *data*, *values*, *stages*, *doses*, *studies* and other quantifiable data.)

Similarly, the prepositions *between* and *among*, the latter being much less frequently used, introduce Prepositional complements that postmodify the noun *comparison* as well. So as to illustrate that the prepositions are just function words that carry little, if any, semantic load, the following examples show there is no semantic difference between these five prepositions as they select identical noun phrases.

19. *Comparisons among* these genes have begun to reveal some common features.
20. (...) comparative genomics, for most people, immediately brings to mind the study of the *comparison of* human disease genes.
21. It is difficult to make meaningful antibody titer *comparisons among* studies without standardization of the assay.
22. Our *comparison of* studies that manipulated vertebrate herbivores contradict these earlier, rather anecdotal reviews.
23. To facilitate *comparison with* other studies, we used a Gaussian function.
24. *Comparisons among* species at several institutions can illustrate differences.
25. Only one *comparison of each species* was conducted with N viridula.
26. *Comparison with* closely related species without the duplications
27. Although limited to *comparisons between* the three most acceptable plant species, did not indicate (...)
28. Initial observations and *comparison to* preclinical models
29. *Comparison of* the migration models reveals (...)

Even though it has been proved that *between*/*among* on the one hand, and *of*, on the other, can be equally interchangeable, there seems to be a preference for the use of *among* when the Prepositional complement refers to members of a group, species, companies, animals, etc.

d) preposition + *comparison*

Following Sinclair's idiom principle in that writers can use "*a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments*" (Sinclair, 1991:110), evidence for formulaic expressions of the type **preposition + *comparison* + preposition** is also noticeable from the analysis of the scientific texts under discussion.

In their study of **lexical bundles**, Biber et al. (1999) define them as "*recurrent expressions, regardless of their idiomacity and regardless of their structural status*" (1999:990); that is, they refer to simple strings of words that usually collocate in natural language production. In general, occurrences of such word strings are frequent in the *HSC*. Writers of scientific articles make use of lexical bundles of the form *as / in / by / for comparison* (*of / with / to*) when establishing a comparison. The bundles identified within this pattern present a straightforward grammatical grouping. In fact, Table 14 shows they are all parts of prepositional phrases (e.g. *in comparison with*, *for comparison of*).

	N. of occurrences
IN COMPARISON(,)	18
in comparison with	46
in a comparison with	1
in comparison to	30
in comparison of	1
in a comparison of	3
FOR COMPARISON(,)	21
for comparison of	10
for comparison with	7
BY COMPARISON(,)	19
by comparison with	19
by comparison to	14
as a comparison	5

Table 14 List of 2, 3 and 4-word lexical bundles with the structure **preposition + noun phrase fragment (*comparison*)**

As can be seen in the Table above, the sequences *in comparison with/to*, *for comparison* and *by comparison with* are the most frequently used in the corpus. The most common prepositions introducing these structures are *in*, *for* and *by*, which are intended to indicate different functions. Table 15 below summarises the discourse functions performed by these lexical bundles.

GRAMMATICAL GROUP	FUNCTION	EXAMPLES
PREPOSITIONAL PHRASES	INDICATE **COMPARISON**	in comparison, in comparison with/to, as a comparison ✓ *The plants with adults were in poor condition in comparison with the plants with no adults.* ✓ *No change in the pattern was observed in comparison with control cells.* ✓ *Thus, in comparison to non-invasive opaque forms, transparent bacteria present more cell wall choline.* ✓ *Although it requires updating, it was useful as a comparison for diets offered.*
	INDICATE THE **PURPOSE** OF AN ACTION	for comparison, for comparison of, for comparison with. ✓ *We use IAD as the baseline measure for comparison with the other two measures.* ✓ *They were combined to provide a powerful discovery tool for comparison of functions vs. phenotype.*
	INDICATE THE **MEANS** BY WHICH AN ACTION IS PERFORMED	by comparison, by comparison with, by comparison to ✓ *This achieved an "effective spatis temporal increase" in food distribution by comparison to baseline.* ✓ *The deduced aminoacid sequence was verified by comparison with sequences obtained from tryptic peptides.* ✓ *The defects were shown to be specific to mutant embryos by comparison to wild-type.*

Table 15 Discourse functions of the pattern **preposition +** ***comparison***

In the case of the preposition *in*, for instance, sequences simply introducing a *comparison* can be found. It is also worth highlighting that, as discussed in the description of the pattern ***comparison*** **+ preposition**, the choice of the preposition *with* / *to* postmodifying the unit *in comparison* does not entail any semantic change. Compare the following examples:

30. Worms at greater than 10ºC in the laboratory results in high mortality rates *in comparison to* worms maintained at 10ºC or below.
31. The overhead costs are also small *in comparison with* those of more developed regions.

The other prepositional phrases present varied meanings such as, ***FOR***-Prepositional phrases, which indicate the purpose of an action (e.g. "***For comparison of*** *differences between more that two means, data were subjected to analysis of variance*"), ***BY***-Prepositional phrases that express the means by which an action is performed (e.g. "*Efficiency of bacterial lysis was measured* ***by comparison of*** *the bacterial concentration before and after lysis*"), and ***AS***-Prepositional phrases, hardly used, which express merely a comparison (e.g. "***As a comparison****, we also did the calculations for a 10-kDsignal*"). See Figures 31-34 below:

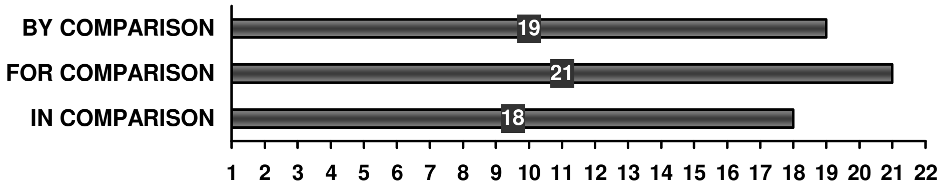


Figure 31 Overall number of occurrences of the pattern **preposition** + ***comparison*** (lexical bundles)

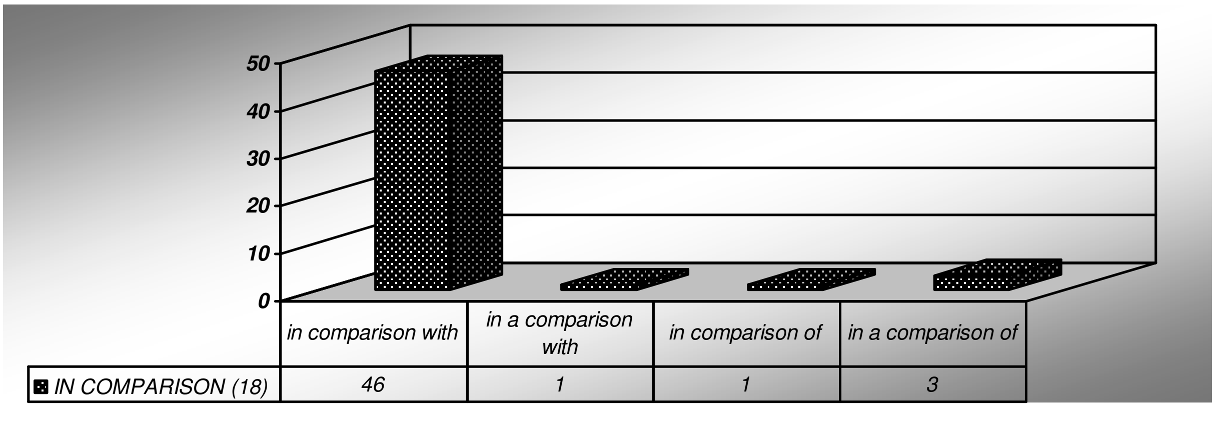


Figure 32 Realisations of the lexical bundle *in comparison*

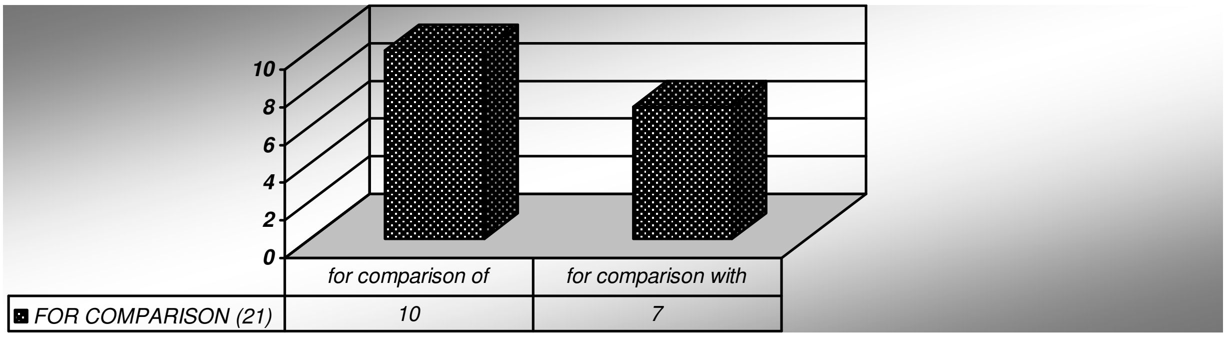


Figure 33 Realisations of the lexical bundle *for comparison*

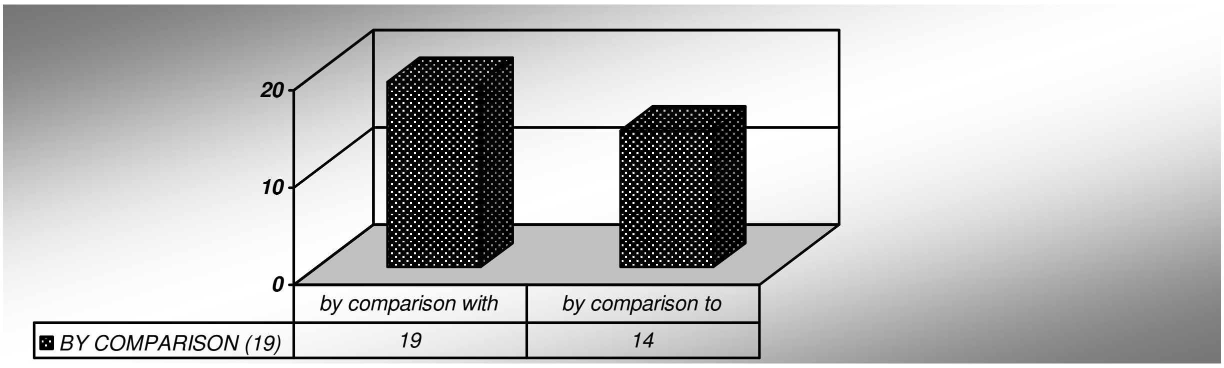


Figure 34 Realisations of the lexical bundle *by comparison*

Among all these lexical bundles, the most interesting ones seem to be *in comparison*, *for comparison* and *by comparison* when not being postmodified by any prepositional phrase. They can be interpreted as slightly different from postmodified structures since they appear to be more grammaticalised. These language chunks must be seen as **whole units of meaning**. The abstract noun *comparison* has lost part of its function as a noun (decategorialization) and has adopted along with the preposition it collocates with a more grammatical role as a **conjunct** that is intended to link two utterances or parts of it in the discourse.

When writing scientific papers, scientists are prone to make high use of these structures in order to organise their discourse, so it seems essential to study them more thoroughly. According to Hopper & Traugott (1993:106), "*the more frequently a form occurs in texts, the more grammatical it is assumed to be*". Consequently, the frequency of appearance of these three structures also contributes to reinforce the argument that they constitute extended units of meaning that function as linking adverbials. It should also be noted that these sequences do appear separated from the rest of the text by means of a comma, underlining even more their category as a multiword unit.

Finally, the position of these units in the discourse must be taken into account as well since their distribution may entail certain semantic connotations. Evidence from the *HSC* reveals a common preference for *in comparison* and *by comparison* to be used in initial position, which can thus be considered the unmarked (expected) distribution for these two linking adverbials. On the contrary, *for comparison* shows an almost identical number of occurrences both in initial (21 entries) and final position (23 entries). Whilst its initial placement highlights its role as a linking adverbial and marks its connection to the previous discourse, final placement makes this role almost unnoticed.

	Number of occurrences		**Context**
IN COMPARISON	Initial	16	followed by a comma
	Medial	2	
	Final	-	

HSC examples:

1. The remaining ovules (22%) were devoid of a normal embryo sac, and the gametophytic nuclei had apparently degenerated. ***In comparison***, ~5% of the ovules in a wild-type plant contain degenerated embryo sacs (MOORE et al.)

2. Sibling control study: The 15 Camelford participants who had eligible siblings were of similar age (41.0 (3.3) years) to the whole group of 55 (41.8 (2.1) years) and to their sibling pairs (42.7 (3.1) years, mean difference -1.7, P=0.36) and of similar pFSIQ as assessed by the reading test (114.7 (2.1)) to the group of 55 (114.4 (1.1)).Their siblings' results, ***in comparison***, are shown in figure 2 together with the results of the pFSIQ, Bexley Maudsley screening tests (standardised scores), and visual evoked potential flash-pattern differences.

Table 16 Context distribution of the *in comparison* bundle in the *HSC*

	Number of occurrences		**Context**
BY COMPARISON	Initial	19	followed by a comma
	Medial	1	following a semicolon
	Final	-	

HSC examples:

1. The lack of specific red fluorescence is shown for the same field in (D). ***By comparison***, in (E) and (F), cells were treated in a similar way except that they were permeabilized before treatment with rabbit anti-GFP.
2. In these control insects, lacZ signaling was detected at 14 and 16 hpi, but not at 12 hpi; ***by comparison***, no signals were detected at 4 hpi in the M2R-treated larvae inoculated with a similar dose, but at 6 hpi, lacZ was evident.

Table 17 Context distribution of the *by comparison* bundle in the *HSC*

	Number of occurrences		**Context**
FOR COMPARISON	Initial	21	followed by a comma
	Medial	4	between commas; immediately following the exclusive coordinator "*or*"
	Final	23	

HSC examples:

1. Two of the motifs of the insulin response sequence (IRS, [10]) are shown by open boxes. ***For comparison***, the sites of the mutations introduced into the vector CRE2mut are also shown, with the mutated bases underlined.
2. To test whether dAPC2 directly interacts with Arm, we used the yeast two-hybrid system (Figure 2 E), examining whether dAPC2's 15 and 20 amino acid repeats interact with the full set of Arm repeats of Arm (R1–13), or with the centralmost Arm repeats (R3–8; the binding site for Drosophila E-cadherin and dTCF). ***For comparison***, we tested the 15 and 20 amino acid repeats of dAPC (Hayashi et al. 1997).

3. In these samples, the total amount of insect cell lysate was kept constant to ensure that changes in apparent molecular mass arose from interactions among CBF3 proteins and not as a consequence of binding to contaminating insect proteins. (...) Elution profile of p64 activity from a column to which a mixture of extracts from cells coexpressing p23Skp1 and p58 and cells expressing p64 (solid line) or, ***for comparison***, extract containing only p64 (dotted line), was applied. (c) Elution profile of p58 activity from a column to which extract containing p23Skp1, p58, and p110 (solid line) or, ***for comparison***, extract containing p23Skp1 and p58 only (dotted line), was applied.

4. The residues of the S.cerevisiae 2 tail shown in bold are those that interact with the surface of the a1 hd. The corresponding residues in the other two tails are also shown in bold ***for comparison***.

5. However, no data as to the effects of brachial arterial infusion of adenosine on forearm NAR are available ***for comparison***.

6. Few other studies have examined the socioeconomic gradient in mortality in diabetic people,6 7 8 and none has included a non-diabetic population ***for comparison***.

Table 18 Context distribution of the *for comparison* bundle in the *HSC*

4.6 The patterns of the noun *contribution*

Unlike *conclusion*, *agreement* and *decision*, the noun *contribution* shows fewer entries (249), whereas its corresponding verb, *contribute (to)* is more overtly used in the *HSC*. A careful examination of the morphology of the noun *contribution* indicates that this is a countable noun used to refer to an abstraction. In Biber's et al. (1999) view, "*countability is not a simple reflection of things observed in the external world (...) with reference to discrete concrete objects, but also to abstractions which do not so obviously or naturally come as distinct entities*." (1999:242)

Taking into account that an obvious feature of countable nouns is their variation in number, the abstract noun *contribution* must be regarded as a fully countable entity. Grammatically, the noun *contribution* is characterised by number variation (i.e. it inflects for the plural) and its co-occurrence with determiners, mainly central determiners: definite and indefinite articles, demonstrative and possessive determiners. However, it should be mentioned that this noun shows a preference for its base form (166 occurrences) over its inflected counterpart, *contributions* (83 occurrences). The next sections will focus on the different patterns this noun collocates with.

a) verb + *contribution (to)*

Amongst the various verbs combining with this abstract noun, there is a limited group of verbs that stand out as being semantically equivalent to *contribute*. This group of restricted collocates consists of the following verbs: *make*, *provide* and *produce*. Consider some examples[80]:

1a. Bruce Weir has made many important *contributions to* population genetic inference theory.
1b. Bruce Weir contributed to population genetic inference theory.

2a. This book should make a significant *contribution to* the reemergence of the field.
2b. This book should contribute significantly to the reemergence of the field.

3a. Embryos were organized like cellular jigsaw puzzles, each cell of which was prespecified to produce its own precisely delimited *contribution to* the mosaic that was the developing organism.
3b. Embryos were organized like cellular jigsaw puzzles, each cell of which was prespecified to contribute to the mosaic that was the developing organism.

4a. It provides only marginal *contributions to* binding as judged by inhibition studies.
4b. It contributes marginally to binding as judged by inhibition studies.

---

[80] Examples 1a, 2a, 3a, 4a and 5a have been taken from the *HSC*, whereas their counterpart ("b") examples have been produced *ad hoc* so as to compare the periphrastic structures (i.e. **V +** ***contribution***) against the lexical verb *contribute*.

5a. These multivitamins did not appear to provide significant *contributions to* the parameters stated.

5b. These multivitamins did not appear to contribute significantly to the parameters stated.

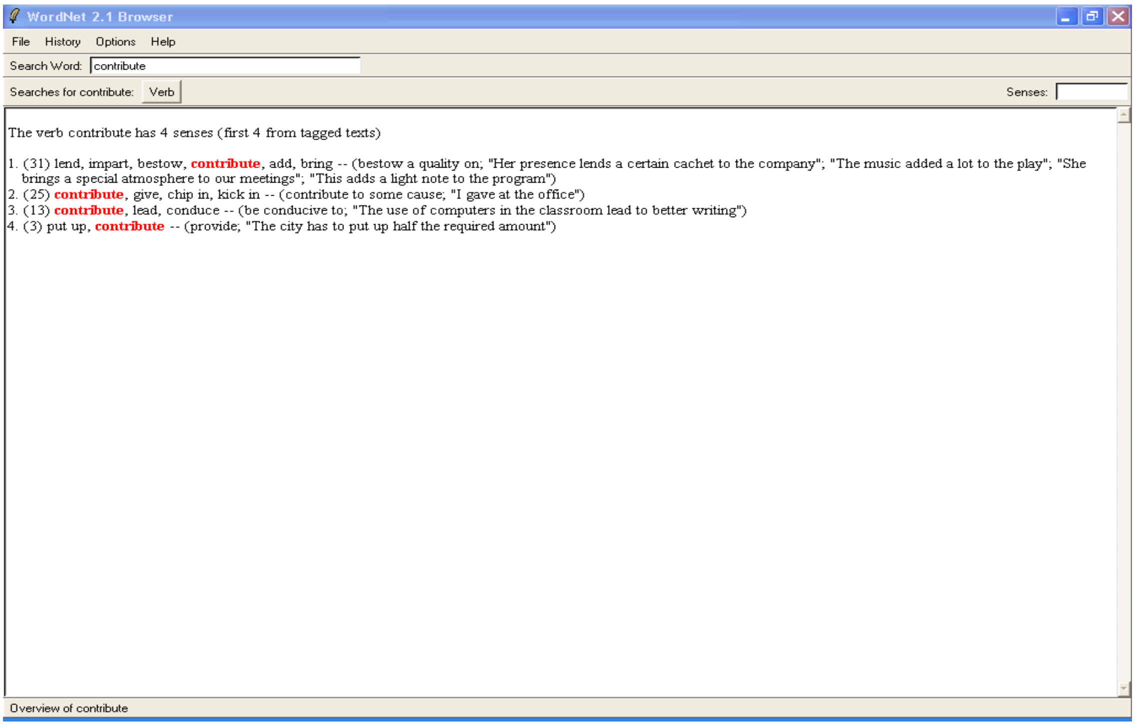
As can be seen in the examples above, the periphrastic structures of the type **"V+ *contribution*"** equal the verb *contribute*, given the fact that they convey the same meaning. In this respect, it is particularly relevant the fact that *make*, *provide* and *produce* have undergone a process of **delexicalisation**, by which they have gradually lost their primary sense of "making, creating something" in favour of the meaning provided by the abstract noun they collocate with.

From a semantic point of view, there arise two key questions which this analysis attempts to answer: what do the periphrastic structures *make a contribution to*, *provide a contribution to*, *produce a contribution to* mean? Are they as polysemic as their equivalent, *contribute to*? The answers to these two questions are of central importance as one of the main fundamentals lying behind **pattern grammar** is that the meaning of one of the items in a collocation is tied to its co-occurrence with the other item. Thus, for instance, the meaning of *make* in "*make a contribution to*" is semantically constrained by the collocation it occurs with.

Likewise, the full verb *contribute to* is highly polysemic in general English. *WordNet*[81] identifies four different senses of such a verb: 1) bestow a quality on; 2) provide money, time, knowledge, assistance, etc. along with others to a common supply, fund, etc.; 3) be conducive to and 4) contribute to some cause –for instance, furnish works for publication (cf. Figure 35 below):

---

[81] http://wordnet.princeton.edu/perl/webwn



Search Word: contribute

Searches for contribute: Verb

The verb contribute has 4 senses (first 4 from tagged texts)

1. (31) lend, impart, bestow, **contribute**, add, bring -- (bestow a quality on; "Her presence lends a certain cachet to the company"; "The music added a lot to the play"; "She brings a special atmosphere to our meetings"; "This adds a light note to the program")
2. (25) **contribute**, give, chip in, kick in -- (contribute to some cause; "I gave at the office")
3. (13) **contribute**, lead, conduce -- (be conducive to; "The use of computers in the classroom lead to better writing")
4. (3) put up, **contribute** -- (provide; "The city has to put up half the required amount")

Overview of contribute

Figure 35 Overview of the node word *contribute* extracted from *WordNet 2.1*

On the contrary, the occurrences of both *contribute to* and "***V + contribution to***" found in the *HSC* show a narrower field of use. All the examples refer to a more figurative sense of "*contribute* ~ *contribution*" rather than being associated with money, time and the like. A possible paraphrase of these units in the *HSC* could be "play a significant part / help cause something".

As mentioned in the characterisation of the patterns of *conclusion*, *agreement* and *comparison*, with the above sense of "play a significant role", the restricted collocates of the type ***V + contribution to*** and its related cognate verb *contribute to* are semantically interchangeable. This fact goes reasonably well with the phenomenon of delexicalisation observed in the verbs *make*, *provide* and *produce*, since the meaning of "playing a significant part" is mainly conveyed by the abstract noun *contribution* in these periphrastic structures and by the verb *contribute to* in synthetic uses.

Moving now on to syntax, there are a few points that should be considered as well. It has been long acknowledged that in scientific writing, writers often make use of the passive voice in order to avoid the recurrent repetition of personal references (i.e., *I / my / me*; *we / our / us*) and to make the text look more impersonal, neutral and objective (Swales 1990; Biber et al. 1998/1999; Hyland 2008). This syntactic feature (**form**) contributes to a great extent to placing the emphasis of a given message on processes and experimental procedures (**meaning**), which is a widely used device in academic writing.

The use of *make a contribution*, however, differs from what has been stated above. With a total of 25 occurrences in the corpus, there has only been found one instance of this sequence in the passive:

> “First, mediastinal tissue analysis of AIDS patients from autopsy revealed that five of seven (71%) patients had either no thymus or no areas of thymopoiesis, demonstrating that **no contribution** to the peripheral T-cell pool **was being made by the thymus** (...)”

Such a preference for active constructions seems to be semantically motivated. The sense of this pattern, which could be paraphrased as “X plays an important part in Y” requires an explicit specification of the Agent (i.e., the maker / causer of the action described); it can, by no means, demote the Subject because it represents an important focus of attention. Below are some examples that illustrate the gist of this argument:

6. **This amplification makes** too small **a contribution to the total amount of lac DNA** to be detected by (...)
7. The challenge for the future will be to determine not simply that such altered **cell biology** could have an effect but that **their effects** are large enough to **make a significant contribution to age-related changes**.
8. **Additional transcripts of abundance class D make** the largest and decisive **contribution to the colon-cancer phenotype**.

Last but not least, notice that in the only example of *make a contribution* in the passive found in the corpus (see Figure 36 below), the prepositional phrase (AGENT “*by*”-PP) of passive sentences has not been omitted (“[...] *by the thymus*”) Again this underlines the fact that this typically optional element in passive sentences is considered to be relevant in this case.

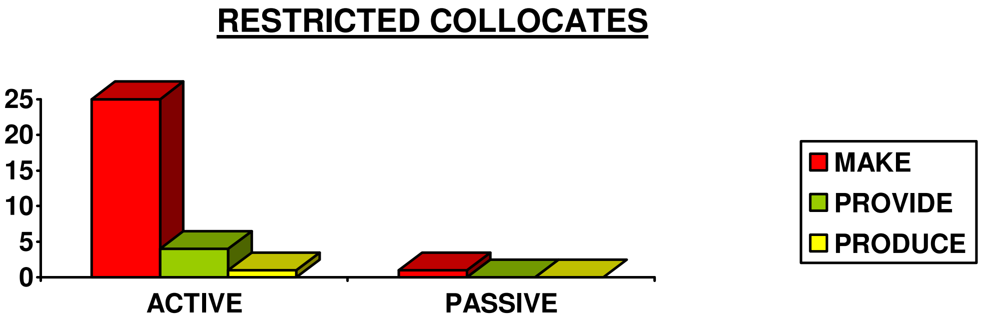


Figure 36 Active and passive constructions with the pattern **restricted verb + *contribution***

In sharp contrast, there is a wide group of verbs that combine with *contribution* but do not equal semantically the verb *contribute*. Most of these free collocates only appear once in the *HSC* (see Figure 37 and Table 19 below for frequency rates), but they are used to convey a really wide range of meanings.

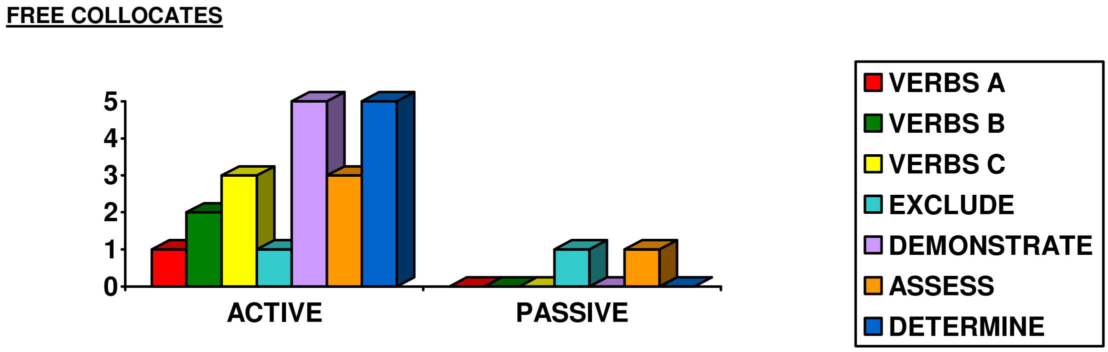


Figure 37 Active and passive constructions with the pattern **free verb + *contribution***

VERBS A (1 occurrence)	VERBS B (2 occurrences)	VERBS C (3 occurrences)
CONSIDER	SUGGEST	EXAMINE
ELIMINATE	INDICATE	EVALUATE
SHED LIGHT ON	MODEL	INVESTIGATE
SUSTAIN	ESTIMATE	
REFLECT	DISSECT (OUT)	
SCORE	ENABLE	
QUANTIFY	LIMIT	
APPRAISE	TEST	
OVERLOOK		
DISCERN		
DISCUSS		
ELUCIDATE		
DEFINE		
REVEAL		
ACKNOWLEDGE		
DISENTANGLE		
MINIMIZE		

Table 19 Frequency rates of the pattern **free V + *contribution***

Especially noticeable is the use of active structures (i.e. only the verbs *exclude* and *assess* are used once in the passive voice) as well as a common preference for verbs connected with scientific procedures: *assess*, *demonstrate*, *determine*, *examine*, *evaluate*, *investigate*, etc.

For the present study, these free collocates have been grouped into different semantic fields, consisting of similar verbs that convey similar meanings and, consequently, can be encompassed under the same category. From the very beginning, all these semantic fields seem to follow a logical line of thought that can be summed up as follows:

1) X **makes** a contribution to Y.
2) Z **examines** that contribution of / to something.
3) Z **evaluates** that contribution of / to something.

4) Z **excludes** that contribution of / to something

5) Z **appraises** that contribution of / to something.

Such a sequence of actions can be expressed by a variety of verbs and described as part of a logical sequence of events, as illustrated in Figure 38:

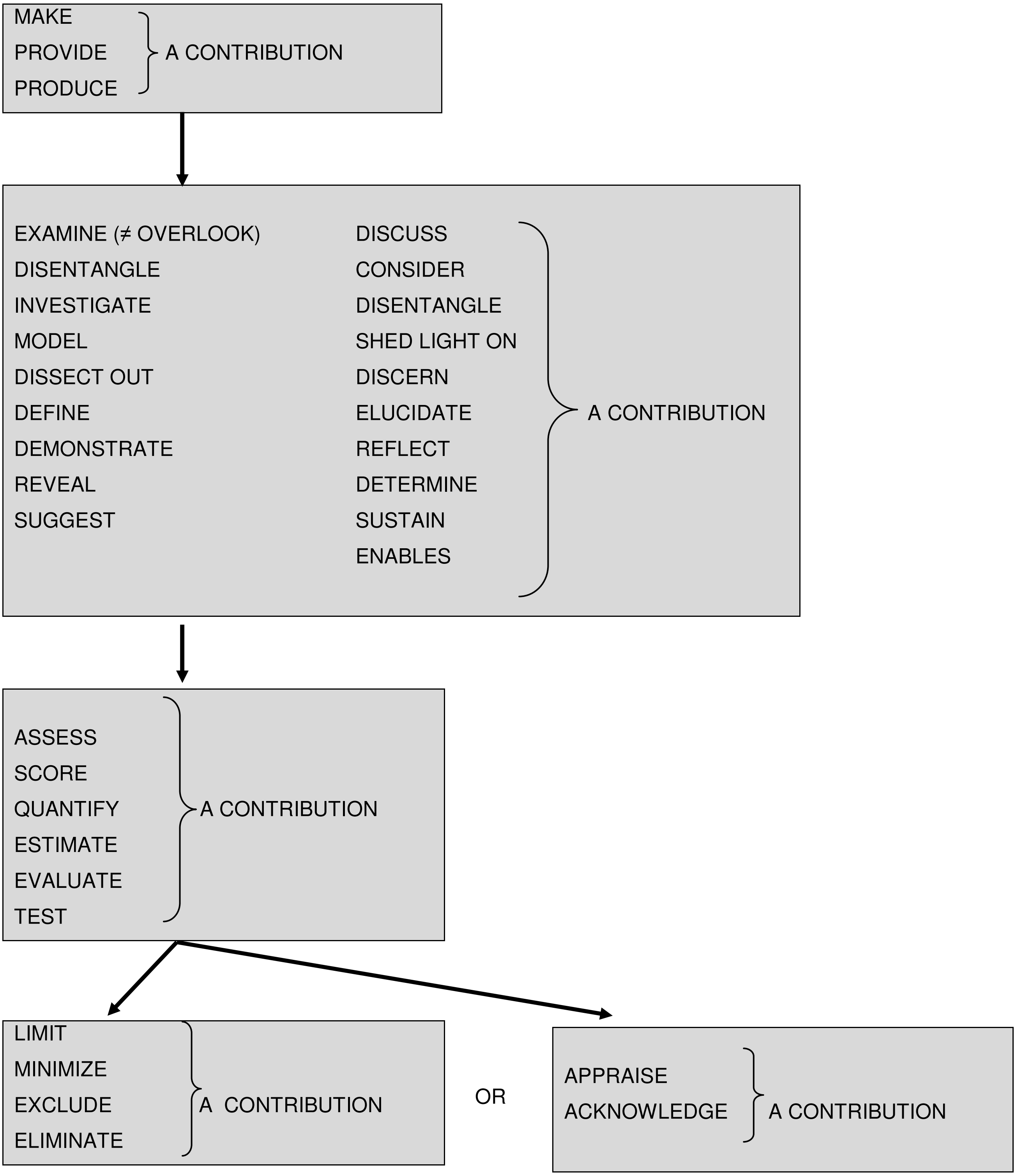

MAKE
PROVIDE
PRODUCE
A CONTRIBUTION
EXAMINE (≠ OVERLOOK)
DISENTANGLE
INVESTIGATE
MODEL
DISSECT OUT
DEFINE
DEMONSTRATE
REVEAL
SUGGEST
DISCUSS
CONSIDER
DISENTANGLE
SHED LIGHT ON
DISCERN
ELUCIDATE
REFLECT
DETERMINE
SUSTAIN
ENABLES
A CONTRIBUTION
ASSESS
SCORE
QUANTIFY
ESTIMATE
EVALUATE
TEST
A CONTRIBUTION
LIMIT
MINIMIZE
EXCLUDE
ELIMINATE
A CONTRIBUTION
OR
APPRAISE
ACKNOWLEDGE
A CONTRIBUTION

Figure 38 Restricted and free collocates of the pattern **V+ *contribution*** grouped into semantic fields and described as part of a logical sequence of events

b) adjective + *contribution*

There is a wide range of attributive adjectives occurring to the left of the abstract noun *contribution*. Although both **descriptors** and **classifiers** can be found as modifiers of *contribution*, the **descriptors** are outnumbered by the **relational** and **topical classifiers** (see Table 20 for frequency counts of modifiers of this noun). As for the former, there are adjectives covering the semantic domains of **size**, **quantity** and **extent** (*small*, *minor*, *minimal*, *heavy*, *lower*, *massive*), **time** (*new*, *early*) and **evaluation** (*outstanding*, *biased*, *decisive*, *favorable*, *functional*, *sympathetic*).

ADJECTIVES + *contribution*					
DESCRIPTORS				**CLASSIFIERS**	
EVALUATIVE	**SIZE**	**EXTENT**	**TIME**	**RELATIONAL**	**TOPICAL**
SIGNIFICANT (10)	MAJOR (4)	MINIMAL	NEW	RELATIVE (23)	CALORIC (2)
IMPORTANT (7)	MINOR (3)	HEAVY	EARLY	FUNCTIONAL (4)	TECHNICAL
SUBSTANTIAL (3)	SMALL (2)	LOWER		MATERNAL (4)	BIOLOGICAL
OUTSTANDING	MASSIVE			PARTICULAR (3)	PHYSIOLOGIC
BIASED				FRACTIONAL (2)	PLACODAL
FAVORABLE				INDIVIDUAL (2)	CELLULAR
SYMPATHETIC				SPECIFIC (2)	NUTRIENT
				UNEQUAL (2)	GENETIC
				UNIQUE	
				OVERALL	
				ADDITIONAL	
				DELIMITED	
				DETECTABLE	
				DIFFERENTIAL	
				EXCLUSIVE	
				INDEPENDENT	
				INTRINSIC	
				QUANTITATIVE	
				SECONDARY	
				SIMILAR	
				SEPARATE	

Table 20 Frequency rates of **descriptors** and **classifiers + *contribution***

Although descriptive adjectives do not usually collocate with the noun *contribution*, it should be highlighted that the list of descriptors seems to be limited to the constraints of the genre in question. Scientific writing appears to rely more on **relational/classifying** (*relative*, *functional*, *maternal*); **affiliative** (*African*) and **topical adjectives** (*technical*, *physiologic*, *caloric*, *cellular*, *nutrient*, *genetic*).

This phenomenon is what could be referred to as “stylistic preference”; the most common adjectives in this type of genre are **classifiers** because academic writing is concerned with delimiting, defining, classifying and focusing on demonstrable data rather than on making judgements or personal evaluations, which tend to be more common of fiction and literary writing.

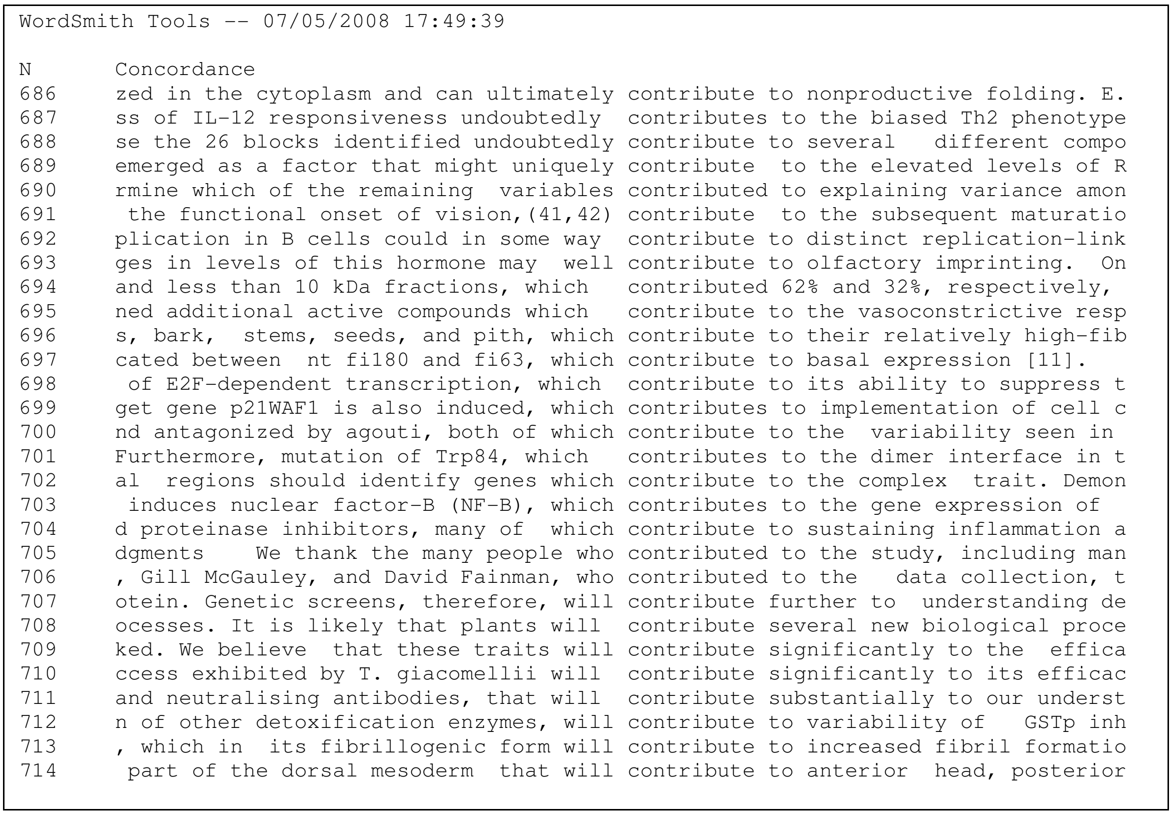
It is wellknown that many adjectives by a process of derivational affixation become adverbs by suffixing -*ly* to the base form of an adjective. This universal truth in grammar plays an important part when analysing the most frequent adjectives in combination with *contribution*: *significant*, *substantial*, *important*. Examining the environment of the verb *contribute* in the *HSC* shows that it usually collocates with adverbs derived from the most common adjectives combining with *contribution*: *significantly*, *substantially*, *importantly*.

The following concordance lines extracted from *WordSmith Tools* show enlightening examples for *significant contribution* and *contribute significantly* in context:

```
WordSmith Tools -- 07/05/2008 17:42:26

133      et al., Æ89). To evaluate the relative contribution  of embryonic gene transcr
134     T. Menes, and E. Hanski (1997) Relative contributions of        hyaluronic aci
135     es and an  assessment of their relative contributions. The success of  these te
136     “selective adhesionö  makes a secondary contribution: “In  consequence of direc
137     lexly interrelated that their separate  contributions cannot be determined with
138     1984] clearly demonstrate a significant contribution of ancestral inbreeding to
139     oa. Given the acknowledged significant  contribution of ''male factor'' inferti
140     e. This book should make a significant  contribution to the reemergence of the
141      70/85 kDa S6 kinases make  significant contributions to events resulting in ce
142     cells but could make a more significant contribution  to assays of synaptic exo
143     ment for proteinuria made a significant contribution to the model,   but barely
144     are large  enough to make a significant contribution to age-related changes.  A
145     e. This book should make a significant  contribution to the reemergence of the
146      did not appear to provide  significant contributions to the parameters tested,
147     athione is likely to make a significant contribution to the  mechanisms by whic
148        1992; Braun et al., 1995). A similar contribution to the stabilization of th
149     crossover loop makes a relatively small contribution to the   energetics of sub
150     rons on the O3 atom also to make some   contribution to the stabilization of th
151      impossible to dissect out the specific contributions of individual  species to
152     c a1   proteins was due to non-specific contributions of the linker or of the t
153     rs at  SCI, with a possible substantial contribution from the provisioned diet.
154     ipogenesis did not make   a substantial contribution of fatty acids (<5%) to fa
155     s thus unlikely to make a   substantial contribution to rejection of cytosine.
156     ng of dense SS RBC (13).   Despite such contributions, trans-species studies ar
157     romoting Th2 cell growth and survival.  Contributions from non-CD41 cells  In a
158     gh ratios would indicate  a sympathetic contribution. A similar approach of  an
159     hem. This is SeaWorld-Florida technical contribution  No. 9606-F.
160     izes and few resources often  limit the contribution an individual institution
161     in the embryo  [14]. Alternatively, the contribution by individual Bcd sites  m
162     s consumed normally. Furthermore, the   contribution from de novo lipogenesis t
163     used in these experiments to limit the  contribution from native LPP. The resul
```

Figure 39 Concordance lines of the node word *contribution*



```
WordSmith Tools -- 07/05/2008 17:49:39

N       Concordance
686     zed in the cytoplasm and can ultimately contribute to nonproductive folding. E.
687     ss of IL-12 responsiveness undoubtedly  contributes to the biased Th2 phenotype
688     se the 26 blocks identified undoubtedly contribute to several   different compo
689     emerged as a factor that might uniquely contribute  to the elevated levels of R
690     rmine which of the remaining  variables contributed to explaining variance amon
691      the functional onset of vision,(41,42) contribute  to the subsequent maturatio
692     plication in B cells could in some way  contribute to distinct replication-link
693     ges in levels of this hormone may  well contribute to olfactory imprinting.  On
694     and less than 10 kDa fractions, which   contributed 62% and 32%, respectively,
695     ned additional active compounds which   contribute to the vasoconstrictive resp
696     s, bark,  stems, seeds, and pith, which contribute to their relatively high-fib
697     cated between  nt fi180 and fi63, which contribute to basal expression [11].
698      of E2F-dependent transcription, which  contribute to its ability to suppress t
699     get gene p21WAF1 is also induced, which contributes to implementation of cell c
700     nd antagonized by agouti, both of which contribute to the  variability seen in
701     Furthermore, mutation of Trp84, which   contributes to the dimer interface in t
702     al  regions should identify genes which contribute to the complex  trait. Demon
703      induces nuclear factor-B (NF-B), which contributes to the gene expression of
704     d proteinase inhibitors, many of  which contribute to sustaining inflammation a
705     dgments    We thank the many people who contributed to the study, including man
706     , Gill McGauley, and David Fainman, who contributed to the   data collection, t
707     otein. Genetic screens, therefore, will contribute further to  understanding de
708     ocesses. It is likely that plants will  contribute several new biological proce
709     ked. We believe  that these traits will contribute significantly to the  effica
710     ccess exhibited by T. giacomellii will  contribute significantly to its efficac
711     and neutralising antibodies, that will  contribute substantially to our underst
712     n of other detoxification enzymes, will contribute to variability of   GSTp inh
713     , which in  its fibrillogenic form will contribute to increased fibril formatio
714      part of the dorsal mesoderm  that will contribute to anterior  head, posterior
```

Figure 40 Concordance lines of the node word *contribute*

Compare the *HSC* examples below:

9. Food transfers seem to be an **important contribution** to the total food obtained by young.
10. All these mechanisms **contribute importantly** to the shaping of animal embryos.
11. It is thus unlikely to make a **substantial contribution** to rejection of cytosine.
12. That will **contribute substantially** to our understanding of the overall biological significance.
13. Abnormalities in potassium channel function are also unlikely to **substantially contribute** to conduction disturbances.

With regard to the semantic classification of these adverbs, three of them refer to adverbs of manner which give account of how an action (*contribute*) is performed. As can be seen from the examples above this general sense may present specific connotations depending on the context they appear in. For instance, the adverb *substantially* in example (12) refers to the overall importance of a given *contribution*, whereas in example (13) the same adverb is used in its more specific contextualised statistical / quantitative sense.

c) noun + *contribution*

Not only do attributive adjectives premodify the abstract noun *contribution*. Premodification by means of another noun also occurs in the *HSC* corpus. However, it should be pointed out that nouns as premodifiers of *contribution* are not as common as adjectives. Among the most frequent N+N sequences, the following chunks could be identified: *call contributions*, *cell contributions*, *gene contributions*, *group contributions*, *mutation contributions*.

All these sequences contain content words and express a **source relationship** between the modifying noun (on the left) and the head noun *contribution* (on the right). In other words, the modifying noun expresses the source of the head: *contributions* are provided by cells, genes, a group, etc., as is clearly observable in the examples below:

Source relationship	**N+N sequence**	**Meaning / paraphrase**
	call contributions →	contributions provided by the calls
HSC example: "Because our subjects differentially contributed to the number of calls recorded in each session, it is possible that our analyses were biased by calls from particularly vocal contributors. However, the extremely high percentages of correctly classified syllables suggests that unequal **call contributions** played little role in our discriminant function classification results."		
	cell contributions →	contributions from cells
HSC example: "Glucose phosphate isomerase (GPI) analysis (Papaioannou and Johnson, 1993) was performed on spleen samples from the two mice with ALL to estimate the contribution originating from the CCB-derived ES cells, injected into the C57Bl/6 blastocysts. Both mice had significant CCB ES **cell contribution** in the spleen (data not shown) suggesting that the lymphoblastic tumours are of ES cell origin."		
	gene contributions →	contributions from genes

HSC example:

"Unlike these HOXA and HOXD genes, the HOXC **gene contribution** is not restricted so closely to the 58 end of the cluster."

group contributions → contributions of / from a group

HSC example:

"Prior probabilities options were utilized to weight unequal **group contributions**, and syllables were analyzed and classified separately."

Table 21 Noun + *contribution* collocation pattern

d) *contribution* + preposition

This pattern has proven to be frequent in the *HSC*. It is particularly noticeable the fact that 83% of the occurrences of *contribution* are postmodified by a prepositional phrase that delimits the meaning of the abstract noun. *Contribution* is postmodified by the prepositions *of* (126 occurrences), *to* (64 occurrences) and *from* (17 occurrences).

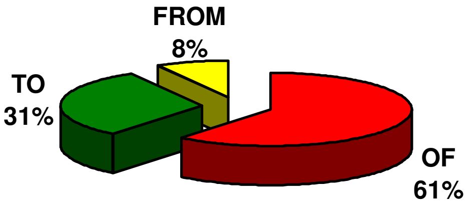


Figure 41 Prepositions following the noun *contribution*

It is important to mention that they are not simply function words (dummy prepositions) but, on the contrary, they are subcategorised by the abstract noun *contribution* in order to convey different meanings. In this context, each of them is used to refer to the source / origin / contributor (*of* / *from*) or the goal / purpose (*to*) of a *contribution*.

What seems to be rather remarkable is the fact that the prepositions *of* and *from* are equally interchangeable; that is, both of them refer to the origin / source of the *contribution*. There is no difference in between them and, in fact, they collocate with the same noun phrases. Consider some illustrating examples:

14. (...) with a significant *contribution from* surrounding cells.
15. (...) they did enable the extent of the *contribution of* the donor cell(s) within different tissues.
16. (...) that enabled the relative *contributions from* maternal and embryonic gene expression.
17. (...) it provides an opportunity to test the *contributions of* specific genes to neurobiological processes.
18. JGW wrote the paper with core *contributions from* all authors.
19. Barrow especifically discussed the *contributions of* women to the field of ornithology.

Despite being semantically equivalent, there are two worth mentioning aspects. Firstly, frequency rates show a preference for the use of the preposition *of* (88%, 126 out of 143 occurrences) over *from* (12%, 17 out of 143 occurrences). Secondly, there seems to be a recurrent tendency to use *of* in correlation with the preposition *to*. In other words, the combination "contribution *from* X *to* Y" does appear (e.g. "*In the present study, a* ***contribution from*** *unlabeled hepatic lipid stores* ***to*** *TG synthesis may be less likely, because the subjects had been fasted for 24 hours by the end of the infusion test, which should substantially reduce hepatic lipid stores*") but less frequently than the sequence "*contribution of* X *to* Y". Here are some examples:

20. ***Contributions of*** blastocyst micromanipulation ***to*** the study of mammalian development (…)
21. ***Contribution of*** prostaglandin EP2 receptors ***to*** renal microvascular reactivity (…)
22. We wished to determine more systematically the intrinsic ***contributions of*** individual Bcd-binding sites ***to*** transcriptional activation by Bcd.
23. Our ability to assess the relative ***contributions of*** genic and chromosomal factors ***to*** the genetic barrier (…)

Regarding the noun phrases that play the role of Prepositional complement of these sequences, they correspond to both animate [+/- volitional] (e.g., *authors*, *cells*, *parasitoids*) and inanimate entities (e.g., *genes*, *processes*, *enzymes*, *metabolism*). They all refer to scientific processes, substances and mechanisms which are characteristic of this specific genre. The following Table shows frequent collocates found in combination with the pattern “***contribution* + *of* / *to* / *from***” (with the number of occurrences in brackets[82]).

contribution* + *of **(126 occ)**		***contribution* + *to*** **(64 occ.)**		***contribution* + *from*** **(17 occ.)**
genes (6)	studies	gene activation	differences	**cells** (2)
site (5)	age	(4)	**mutation**	stores
factors (5)	lineages	phenotype (4)	the total	**mechanisms**
cells (4)	reactions	**cell/gene pool**	catalysis	parasitoids
receptors (4)	routes	(2)	properties	cell pool
proper names (3)	**binding**	project (2)	anaphase	authors
proteins (2)	activities	the host (2)	**binding**	biscuits
effects (2)	transcription	interface (2)	positioning	helicases
thymus (2)	**mechanisms**	**process** (2)	synthesis	nitrosyl
mutation (2)	metabolisms	stabilization (2)	specificity	(heme)hemoglobin
diet (2)	infertility	**binding affinity**	understanding	**diet**
fatty acids (2)	species	total amount of	genetic	de novo lipogenesis
fat accumulation	**process**	values	isolation	neighbours
box	interactions	mosaic	development	
membrane	experience	work	cancer	
substances	pathway	assembly	progression	
gluconeogenesis	**model**	theory	blood	
enzymes	rearrangement	**metabolism**	pressure	
component	errors	problem	degenerations	
plasma	agents	strength	**mechanisms**	
males & females	source	regulation	parameters	
	variables	**proteins**	events	
	individuals		assays	
			model	
			rejection of	

Table 22 Noun Phrases as Prepositional complements of the sequence *contribution of / to / from*

---

[82] Notice that **coloured** noun phrases appear as Prepositional complements of more than one of these prepositions.

4.7 The patterns of the noun *decision*

Despite being the least frequently used noun in comparison with the other four abstract nouns analysed so far, its environment in the *HSC* shows some special characteristics that justify its essential presence in this list. It is especially relevant the fact that the idea of “form an opinion after considering some facts” is more commonly conveyed by means of a verb in combination with *decision* than by the lexical verb *decide*, which shows a remarkably low frequency (61 entries, comprising all its formal realisations). Such a preference for periphrastic structures with an abstract noun will be discussed in much more detail in the following pages.

a) verb + *decision*

A total of 39 occurrences give account of this periphrastic use, out of which 17 entries (43,6%) show restricted collocates whereas 22 (56,4%) consist of free collocates. On the one hand, restricted collocates deserve special attention as the verbs that collocate with the abstract noun *decision* have undergone certain degree of **delexicalisation**. *Make*, *take* and *reach* in combination with *decision* are characterised by a loss of semantic content. They fulfil a verbal function but carry no meaning apart from this to contribute to the periphrasis in which they occur. The semantic content of the whole multiword unit, on the contrary, is being provided by the abstract noun.

Of great importance is the close relationship between the **delexicalisation** of these verbs and their syntactic behaviour. Data reveal that these periphrastic structures tend to be expressed in passive constructions (see Figure 42). The case of “*make a decision*” is worth mentioning as such a preference for passive structures is most outstanding.

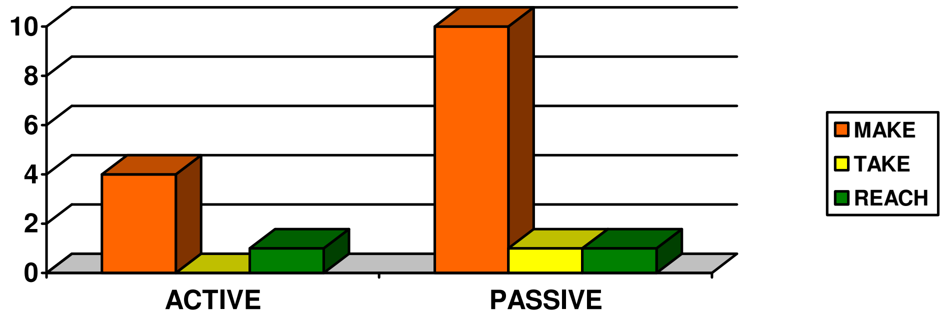


Figure 42 Syntactic structures related to restricted collocates of the type **V+ *decision***

The focus of attention in these structures is not the agent but the *decision* itself being *made*. Concerning the "decision maker", it should be remarked that it tends to refer to the research team conducting the experiments described in the articles. Thus, they are not often made syntactically explicit as they are obvious from the context. See some examples expressed in the passive form:

1. (...) being present well before a **decision** about gender **is made**.
2. (...) other issues mentioned in this guide should also be considered before any **decision is made** about which clinic to attend.
3. (...) they are often derived from populations dissimilar to the individuals for which **a decision must be made**.
4. Also **the decision** to differentiate per se **may be taken** independently of the decision as to which lineage pathway to adopt.
5. Management and research **decisions are usually reached** by consensus.

As for examples in the active voice (a total of 5), it must be stressed that they are either associated with passive constructions somehow (see examples 6 and 7 below) or, alternatively, the "decision maker" is felt to be necessary as it is rather specific (examples 8-10):

6. I did not notice anything extremely controversial, but it still disturbs me because in conservation biology research results **may be used to make far-reaching decisions**.
7. If **it was not accessed in making the decision** to import the bruchids, then it highlights an important deficiency in biocontrol practice.

8. All patients were treated with chemotherapy so why would only **a quarter of them** **make the same decision** again?
9. The choice of experimental system is inevitably a compromise, and the different **Parasite Genome Initiatives** **have reached different decisions** based on the characteristics of the individual organisms.
10. Although such brute force comparative approaches cannot replace creativity, insight, and good experimental technique, they can allow clear analysis of past progress and enable **the new group** to **make well-informed decisions** regarding the direction of future research.

On the other hand, free collocates consist of fully lexical verbs in combination with the abstract noun *decision*. With no exception, all these structures, realised by 14 different verbs, namely: *govern*, *support*, *influence*, *contemplate*, *control*, *allow*, *force*, *reconsider*, *base on*, *formulate*, *communicate*, *guide*, *face*, *think about*, are used in the active voice. The noun phrases performing the syntactic function of Subject have proven to be determinant in the description of these free collocates. Consider the following Figures:

GROUP 1

mechanisms
analysis
tools
methods
scientific evidence
strategies
pathways
modules
genes / genotypes

+ (govern, support, influence, control, allow, force, base something on, formulate, communicate) **+** a *decision*

SUBJECT + **PREDICATOR** + **OBJECT (*decision*)**

Figure 43 Syntactic and semantic environment of **free verb collocates + *decision*** found in the *HSC* (Group 1)

GROUP 2

proper names
the research group } **+** (contemplate, reconsider, face, think about) **+** a *decision*
patients

SUBJECT + PREDICATOR + OBJECT (*decision*)

Figure 44 Syntactic and semantic environment of **free verb collocates + *decision*** found in the *HSC* (Group 2)

Verbs in Group 1 (Figure 43) subcategorise for causers, tools and means as Subjects, whereas verbs in Group 2 (Figure 44) require agents. Both groups of verbs select the abovementioned semantic roles and give them the syntactic position of Subject. This fact explains why these free collocates are not used in passive structures. Compare the following examples:

11. This does not prove that **genes of that class will influence the decision** (*HSC* instance)
12. This does not prove that **the decision will be influenced by genes of that class** (*ad hoc* instance)

The passive structure in example (12) shifts the focus of attention from the *genes* (**causer**) to the *decision* itself. An active structure is felt to be more appropriate in order to focus on the noun phrases that give account of the very specific causers / tools / experiencers that instigate the action described by each verb. Once again, it can be concluded that **syntax is driven by lexis**.

b) *decision* + noun

Unlike the rest of the abstract nouns dealt with so far, this noun in combination with other nouns forms part of an endocentric (i.e. syntactically and semantically-headed) compound. A close look at these recurrent strings of words both from a lexicological and morphological perspective must be taken.

Compounds understood as "*stems consisting of more than one root*" (Jackson & Zé, 2000:79) constitute a productive word formation process in English.

Although the orthographic treatment of compounds is not systematic at all throughout written articles, the examples consisting of $\mathbf{N_1}$ **(*decision*) +** $\mathbf{N_2}$ found in the corpus are written as two separate words with or without a hyphen between the two bases. According to Jackson and Zé (2000:80), compounds may be distinguished from phrases on syntactic and semantic grounds since they do have their own specific syntactic properties with regard to word order, modification and inflection.

For the position of each base in a compound, it must be underlined that they show an unexpected (**marked**) order. Taking the example of *decision making process*, possibly paraphrased as "a process that makes / contributes to make decisions", it can be observed that the Object of the verb *make*, *decision*, precedes the deverbal base. On the contrary, Objects follow verbs in sentence structure (i.e. *to make a decision*).

In addition to word order, the close relationship between the two bases of a compound should not be overlooked either; there is no element interrupting the strong connection between the two nouns. Compare the following examples:

13. Although not providing precise **decision rules**, the principles of risk assessment are really helpful. (*HSC* example)
14. Although not providing precise **decision of the rules**, the principles of risk assessment are really helpful. (*ad hoc* example)

The inclusion of a preposition and a determiner between *decision* and *rules* makes the unity of the two bases disappear and, as a consequence, such an structure cannot be regarded as a compound any more. On the same note, bearing in mind that compounds must be interpreted as a whole, their bases cannot be modified separately. Take the following example: "*As expected for molecules that act at* ***a key decision point****, the activities of cyclinD- and cyclinE-dependent kinases are tightly controlled and subject to multiple tiers of regulation*". Notice that the compound *decision point* is premodified by the adjective "*key*", which affects the whole unit. However, premodifiers of base$_2$ (*point*) would not be acceptable:

15. As expected for molecules that act at a ***decision key point**, the activities of cyclinD- and cyclinE-dependent kinases are tightly controlled and subject to multiple tiers of regulation.

16. As expected for molecules that act at a ***key decision important point**, the activities of cyclinD- and cyclinE-dependent kinases are tightly controlled and subject to multiple tiers of regulation.

Likewise, compounds can only be inflected in terms of their grammatical category; that is, their bases cannot show inflection separately. To give some examples, consider the following[83]:

decision makers	*decisions maker/s
decision making	*decisions making
decision model/s, point/s, tree/s, rule/s	* decisions model/s

Table 23 Inflected forms of ***decision* + noun** compounds

In fact, the inflected forms for the plural (i.e., *decision models*, *decision points*, *decision trees*, *decision rules*) refer to the two bases, which means the tree diagram that describes such structures should give account of this point:

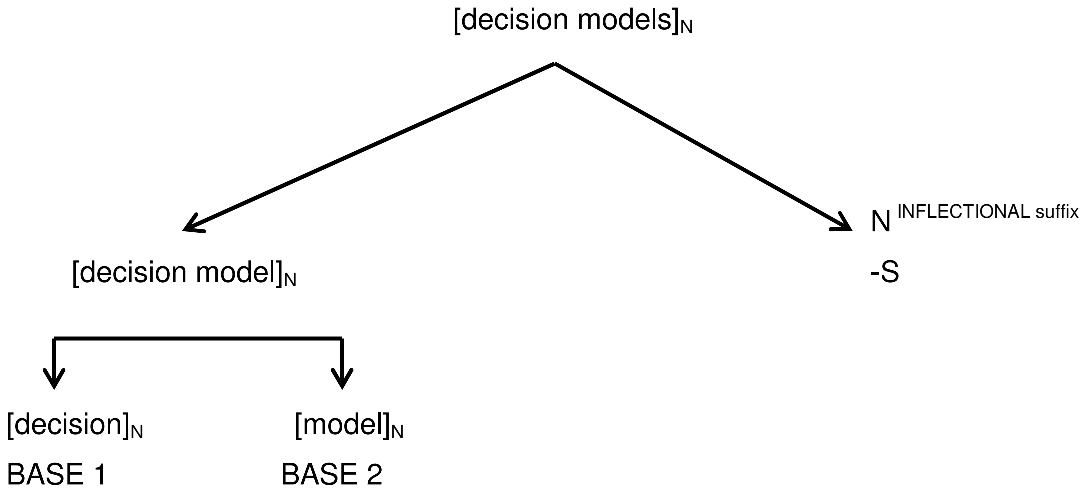


Figure 45 Tree diagram for the compound *decision models*

---

[83] The asterisked examples on the right column are considered to be ungrammatical utterances.

Semantically, the compounds found within this pattern have acquired more specialised meanings. Their meaning is not a simple juxtaposition of the meaning of their constituents. Below are some examples of the endocentric verbal compound *decision making* in the different contexts it appears:

17. There were problems in the initial **decision making process**.
18. They propose a plan for conservation, reintroduction, and eventual restoration of the black-footed ferret that they claim will integrate biological and technical knowledge with the organizational structure of the **decision-making body**, the social behavior of the participants, local interests and human needs, and the legal parameters of reintroduction.
19. Our findings would ultimately provide a baseline **decision-making strategy** regarding the use of fungal pathogens.
20. Several organisations are responsible for **decision making** in transfusion safety.
21. Previous transfusion information was not always used in **decision making**.

As can be seen in the examples above, *decision making* (see its tree diagram in Figure 46 below) can be used as a noun (examples 20 and 21) or as a noun modifying another noun (i.e., *process*, *body*, *strategy*) in examples (17-19).

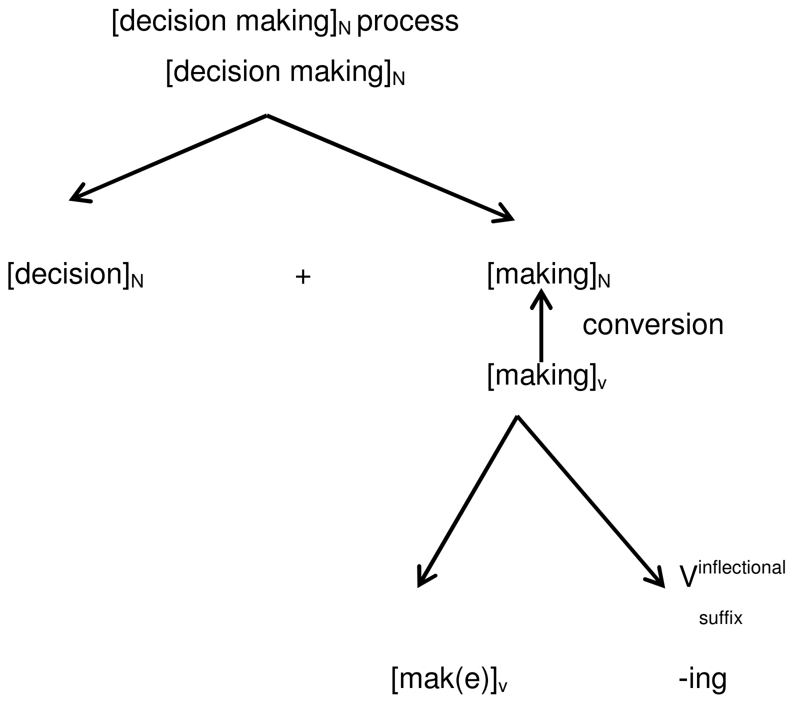


Figure 46 Tree diagram for the compound *decision making*

Another worth mentioning compound in the corpus is *decision maker*, which consists of a complex head *maker* and a non-head constituent realised by the abstract noun *decision*:

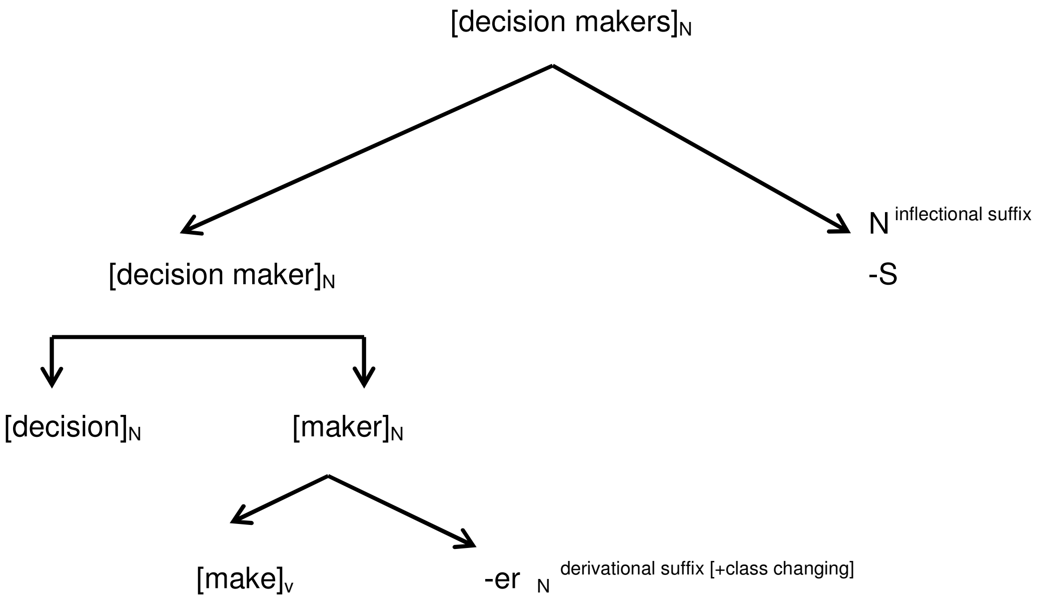


Figure 47 Tree diagram for the compound *decision makers*

At a first glance, this noun (i.e. *decision maker*) should be described as an endocentric verbal compound; that is, semantically equivalent to "someone who makes a decision". However, a more careful look at its environment reveals that this compound embraces a more opaque meaning. "Someone" in this context refers to an actual patient who has to make a decision about whether they wish to undergo an experiment / a treatment, etc. Consider some illustrative examples:

22. Estimates of lifetime costs (discounted at 6% per year) and life years (discounted at 2% per year) of HIV infected women and their children according to whether the woman's infection was known about during pregnancy. For **a decision maker** willing to pay up to £10 000 to gain one life year, the net benefit of diagnosis would be £49 090 ((6.392×£10 000)£14 833)), and it would be cost effective to devote this sum to detect one additional infected woman in an antenatal testing programme.

23. However, **a decision maker** assuming the most pessimistic net benefit scenario, but willing to pay up to £15 000 per life year gained would spend the

> same sum to identify an additional maternal infection as **a decision maker** assuming base case net benefit and willing to pay up to £10 000 per life year gained.

In the light of the above examples, it seems clear that this compound has drifted away from its original compositional meaning (in the sense that its interpretation cannot be worked from the sum of the literal meanings of its constituents) and has adopted a more opaque meaning. This gradual lack of compositionality makes the whole unit fall under the category of semantically-headless (i.e. exocentric) compounds.

c) *decision* + preposition

Contrary to what was expected, the abstract noun *decision* is hardly ever postmodified by prepositional phrases regarding the **content** of the *decision made*. But for 6 occurrences of prepositions following *decision*, the other entries for this noun show no postmodification by means of a prepositional phrase.

Two prepositions (i.e. *about* and *on*) appear within this pattern in combination with *decision/s*. Both of them (the preposition *on* to a much lesser extent, though) select noun phrases intended to specify what sort of *decision* is being made by an agent; they introduce a relationship of "respect", similar to *regarding*, *with regard to*, etc. Although both prepositions share the same meaning, the corpus data reveal a preference for *decision about* over *decision* on. When comparing the noun against its cognate lexical verb, figures reverse: the collocation *decide on* is more frequent than *decide about* (see Table 24 below).

Similarly, the concordance lines for *decide* provided by *WordSmith Tools* show that the same prepositional phrases, introduced by the same prepositions and providing the same semantic contents, may form part of the complementation of the lexical verb *decide*.

DECISION ABOUT	***DECISION ON***
N. of occurrences: 5	N. of occurrences: 1
The **decision about** whether to vaccinate against typhoid fever (...)	However, as noted by Norton [1995], an ethic that bases all moral **decisions on** a single criterion such as Regan's subject-of-a-life is "likely to conclude that sacrifice of individuals for species survival is always wrong because the individuals cannot fulfill the key requirement of voluntary acceptance of risk" [p. 113].
(...) being present well before a **decision about** gender is made.	
Decisions about whether child abuse has occurred must then be made after consideration of all these investigation together with the multiagency child protection team.	
For this reason, **careful decisions about** oxygen therapy and referral are required to avoid overburdening the referral system and depleting scarce oxygen supplies.	
DECIDE ABOUT	***DECIDE ON***
N. of occurrences: 1	N. of occurrences: 4
Conclusion: In 1997, between 80 000 and 100 000 patients in the United States will have to **decide about** having chemotherapy for the treatment of advanced non-small cell lung cancer.	Once one has **decided on** the appropriate permutational space for the null hypothesis given the experimental design, the second step is to implement a computational procedure to explore that state space.
	Prior to the development of promising formulations of artificial diet, we **decided on** three standards that are absolutely essential for successful commercial-based mass rearing: production of continuous generations on diet (with no recourse to real prey); continuous increase in colony size; and ability to kill and feed on natural prey (Cohen, 1992).

Table 24 *HSC* examples of the pattern ***decision* + preposition** vs. ***decide* + preposition**

d) ***decision*** **+** ***to*-infinitive**

With reference to the kind of *decision* being *made*, the most common pattern shows the abstract noun *decision* followed by a "*to*-infinitive" clause. As observed previously, the abstract noun *decision* is unlikely to be found postmodified by a prepositional phrase. On the contrary, postmodification by means of a "*to*-infinitive" clause is much more common.

24. There are ethical and pragmatic issues associated with **a decision to modify** a screening programme which are beyond the scope of this study.
25. This updated result is much less affected by **the decision to stop** treatment early.

These noun clauses present the complete content of the head noun *decision*, which refers to an action. This goes reasonably well with the fact that the typical head nouns selecting "*to*-infinitive" clauses tend to be nominalised equivalents of verbs controlling "*to*-complement" clauses. Consider the following examples:

26. It will also be feasible to test whether the patients' **decision to live or die** can be controlled by the balance of active MAPKs in the cell, as has been suggested in mammals.
27. It will also be feasible to test whether it can be controlled **if parents decide to live or die**.
28. The **decision to have another child after late loss in pregnancy** is a personal one.
29. Parents **decide whether to have another child after late loss in pregnancy** (or not).

Another remarkable syntactic aspect of this pattern is that the "*to*-infinitive" may be preceded by *whether*; thus, giving rise to a subordinate interrogative clause that includes an infinitive clause. Such an alternative interrogative clause is likely to occur following a noun as *decision* since this abstract noun usually entails a choice (see example 30). Sometimes the correlative "(...) *or*" is not syntactically present as can be easily understood from the context (e.g. "*They might attempt to influence his decision whether to help his mother*").

30. On the other hand, ripe oocytes are heavily outnumbered by lymphocytes, and a cell population that is a minority within the body could still exercise an effective veto [Haig 1992] on **the decision whether to ovulate**.

From a semantic point of view, the type of infinitives that form part of these "*to*-infinitive clauses" are similar both as complements of the abstract noun *decision* and as complements of its corresponding verb, *decide*. Of the 123 entries for *decision*, 19 are of this type ("*decision to*", 15; "*decision whether to*", 4). Similarly, of the 61 entries for *decide*, 31 follow the pattern "**V + *to*-clause**" ("*decide to*", 27; "*decide whether to*", 4). Compare some examples:

DECISION **+ TO-infinitive**

decision to

~use
~import
~undergo
~differentiate
~modify
~request
~stop
~live
~form
~become
~terminate
~have a child
~be a parent
~help
~ovulate

DECIDE **+ TO-infinitive**

decide to

~change
~initiate
~restrict
~stop
~promote
~proceed
~release
~sequence
~focus
~explore
~examine
~use
~investigate
~analyse
~characterise
~terminate
~administer
~see
~isolate
~help

Figure 48 Examples of *to*-infinitive clauses following *decision* and its corresponding verb ***decide***

e) adjective + *decision*

Regarding the premodifiers of the abstract noun *decision*, there are two main categories: **adjectives** and **nouns**. The latter (e.g. *cell fate decisions*, *management decision*, *research decisions*) are not very common. Instead, there is a wide range of adjectives, both in attributive and predicative positions[84], modifying the noun *decision*, for which semantic groupings can be established.

---

[84] Notice that only two predicative adjectives have been found (*difficult* and *non-medical*), which brings us to conclude that *decision* is much more commonly modified by attributive adjectives.

It seems useful to distinguish two main semantic groups of adjectives: **descriptors** and **classifiers**. As for descriptors, the list is significantly low (see Table 25 below for examples) in comparison with classifying / relational adjectives. This type of finding points to the fact that, as already discussed, the defining traits of scientific discourse make it prone to refer to precise processes and mechanisms, rather than expressing the attitudes of the writer with respect to the discussed issues.

DESCRIPTORS		CLASSIFIERS	
SIZE, EXTENT, TIME	**EVALUATIVE**	**RELATIONAL**	**TOPICAL**
Describing / denoting observable data with regard to time, extent, frequency and age	Expressing the attitude of the writer towards a *decision*; making judgements	Delimiting the referent of the *decision*	Referring to the subject area and/or showing a relationship between the adjective and the noun
recent early poor careful precise complex	(im)moral	independent maximum initial*(3) different important*(2) key operational far-reaching	clinical*(2) programming patterning spatial developmental behavioral food-sharing

Table 25 Semantic groupings of the attributive adjectives that collocate with the abstract noun *decision*[85]

4.8 Discussion of overall results

The main goal of the whole chapter was to provide a detailed account of the colligational and collocational patterns of abstract nouns in the *HSC*, so as to contribute to the characterisation of such phraseological units in medical science writing. After having investigated the various patterns the five abstract nouns in question (i.e., *conclusion*, *agreement*, *comparison*, *contribution* and *decision*) collocate with, several conclusions can be drawn from these findings.

---

[85] *The asterisked adjectives indicate that they have been used more than twice within this pattern.

Four processes need to be highlighted from the overall results: a) various degrees of delexicalisation in restricted verb collocates, b) gradual grammaticalisation involving semantic bleaching in prepositional bundles, c) preference for the use of periphrastic structures (i.e. verb + abstract noun) in passive forms and d) a general tendency for adjective collocates to appear in attributive positions. Each of these processes deserves further discussion.

Firstly, the five abstract nouns explored appear in restricted combinations of the type **V+ Noun**; in particular, the patterning associated with the lexical verbs *make*, *reach*, *draw* and *take*. Of the several senses related to these polysemic verbs, it has been proved that their core original meanings are not the most frequently used. On the contrary, the most prominent use of the above verbs points to the general meaning of “doing”; that is, the lexical verb has become a grammatical form with very little semantic content on its own. Thus, it can be stated that these verbs have undergone a process of what scholars (Sinclair 1992; Partington 1993; Howarth 1996, Hunston & Francis 2000, among others) refer to **delexicalisation** and/or **semantic depletion** and, as a consequence, their original meaning is no longer the most prominent.

Howarth (1996:94) further claims that such delexical senses are being identified as semantically depleted not only on the basis of the semantic nature of the verb itself but also on the kind of noun in combination with it and the correspondence of the collocation to a morphologically related verb of equivalent meaning. These last two tests for semantic depletion will now be further discussed.

Following Howarth’s (1996) theory, the nature of the noun may contribute to a certain extent to the fact that the **V + Noun** pattern be regarded as a restricted collocation. In his point, abstract nouns, such as the ones examined in this study (i.e. *conclusion*, *agreement*, *comparison*, *contribution* and *decision*), are more likely to be combined with semantically empty verbs; in other words, abstract nouns can activate the semantic weakening of the verb they collocate with: “*the less concrete the noun is, the less possible it is for the primary literal sense of the verb to be activated*” (Howarth 1996:97). In *make chemical* *products* the noun (i.e. [*chemical*] *products*) is concrete and the verb (i.e. *make*) keeps its literal meaning of "creating, producing", whereas in *make a conclusion* the noun (i.e. *conclusion*) is abstract and the sense of the verb has faded away and, consequently, does not refer to the notion of "producing" anymore.

It seems thus that one of the main tests so as to label a given verb sense as delexical is the possible transformation from a **V + Noun** construction (i.e. *make a contribution*) to a lexical verb (i.e. *contribute*). In *make a contribution*, the verb could be replaced by *provide* and/or *produce* and such a transformation would keep working perfectly. However, if the noun (i.e. *contribution*) were substituted in the periphrastic structure, the sense of the whole multiword unit would be significantly altered. As Howarth (1996) states, "*if the delexical verb is regarded as contributing little or no independent meaning to the whole collocation (having the role of a grammatical prop for the noun), then one is justified in considering that no substitute for a noun can exist for the verb in that sense*." (Howarth 1996:112-113).

Concerning the equivalence of the periphrastic structure to a cognate verb, it is worth noting that this is an outstanding feature of the abstract nouns selected for the present analysis: *make a conclusion* ~ *conclude*; *reach an agreement* ~ *agree*; *make a comparison* ~ *compare*; *make a contribution* ~ *contribute*; *make a decision* ~ *decide*. In this respect, Howarth asserts that "*collocations involving these verbs demonstrate a strong tendency to collocate with nouns that have cognate lexical verbs*" (Howarth 1996:98). But for the case of *comparison* and *contribution*, the other three nouns outnumber their corresponding cognate verbs as far as frequency is concerned. Figure 49 below displays their exact number of occurrences in the *HSC*.

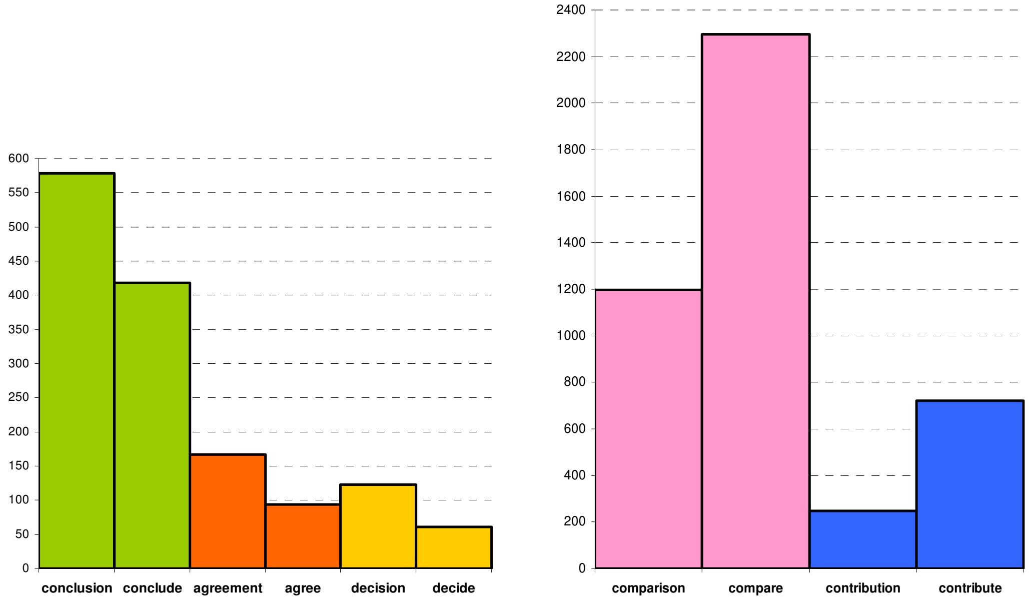


Figure 49 Frequency of abstract nouns in comparison with their corresponding verbs

It has been proved that periphrastic structures consisting of **restricted verb collocates + Noun** are semantically equivalent to the abstract noun's cognate lexical verb. The main question is then raised as to what syntactic features distinguish the periphrastic uses (i.e. **delexical verb + Noun**) from the full lexical synthetic constructions. Taking into account that "*syntax is driven by lexis*" (Francis 1993:142), it has been observed that despite being semantically interchangeable, there is a preference for the use of the periphrastic structure (i.e. delexical verb + *conclusion*/ *agreement* / *comparison* / *contribution* / *decision*) in passive constructions[86]. On the other side of the coin, their equivalent cognate lexical verbs are more commonly used in the active voice.

Free collocates have also been analysed. On average, the abstract nouns under study collocate with a wide range of free combination lexical verbs. Unlike restricted collocates, this pattern tends to be used in active constructions and its absolute frequency is much lower. Interestingly, it seems that the wider range of free collocates an abstract noun combines with, the less frequently such verbs are used. Figures have revealed that, but for few exceptions (e.g. *support a conclusion*, 67 occurrences), the absolute frequency of free combination verbs is hardly significant as in most cases it consists of no more than two or three instances with a given noun (e.g. *criticize a conclusion* [2 occurrences]; *establish an agreement* [1 occurrence]; *express a comparison* [1 occurrence]; *examine a contribution* [3 occurrences]; *formulate a decision* [2 occurrences]).

---

[86] The only exception to this preference can be found in *make a contribution*, which tends to be used in the active voice (cf. section 4.6).

The examination of abstract noun patterns in the *HSC* has also revealed a gradual process of grammaticalisation and semantic bleaching[87] in multiword units like "*in agreement with*" and "*in / by / for comparison*", where the lexical items *agreement* and *comparison* in combination with certain prepositions have been transformed into grammatical forms and serve to link parts of a discourse; that is, they are used as connective operators. These two examples will now be considered in turn.

As stated earlier, the abstract noun *agreement* has substituted its corresponding cognate verb *agree* in many fixed expressions. In fact, in correlation with "*in ... with*" it must be regarded as a highly frequent lexical bundle which shows a high degree of unity. The abstract noun *agreement* in such a context forms part of a complex string of words that functions both semantically and syntactically as a discourse marker.

A similar process of decategorialisation (i.e. the noun loses part of its original function as a noun) has been found in other word strings, such as *for comparison*, *in comparison* and *by comparison*. In these contexts, the literal meaning of the abstract noun *comparison* is not applicable and, through a process of semantic loss, it is seen to acquire a new grammatical role (Gabelentz 1891; Sweetser 1988; Aitchison & Lewis 2003) as a conjunct.

A further feature that is striking regarding the collocational patterns of abstract nouns in the *HSC* is that there is a general tendency for adjective collocates to be used as internal pre-head modifiers to the following abstract noun; that is, there seems to be a preference for the use of adjectives in attributive position. This trait has proven to be true for the five adjectives under study, since the number of occurrences of adjectives functioning as predicative complements in clause structure is remarkably low in all cases. Taking, for example, the case of the abstract noun *agreement*, corpus data reveal that there are just two instances of predicative adjectives modifying this noun.

---

[87] *Verbleichen* and *verblassen* , respectively, in Gabelentz's words (1891:242)

In addition, the typology of adjectives found in combination with abstract nouns in the *HSC* deserves special attention. As pointed out before (section 4.3), the kind of adjectives used to denote properties of abstract nouns fall into two main semantic groups: **descriptors** and **classifiers**. The properties conveyed by the former may relate to opinions (**evaluative** adjectives) and **size**, **quantity**, **extent** or **time** descriptors, whereas the latter refer to **relational** or **topical** adjectives.

With the only exception of *agreement* and *conclusion*, the other three nouns (i.e. *comparison*, *contribution* and *decision*) show a strong preference for **relational** (e.g. *firm conclusion*, *direct comparison*, *relative contribution*, *initial decision*) and **topical** (e.g. *biological conclusion*, *statistical comparison*, *caloric contribution*, *clinical decision*) classifying adjectives. These findings corroborate Biber's et al. (1999) assertion of the high use of classifiers in academic prose. Due to the fact that scientific writing is more concerned with defining and describing demonstrable data from an objective point of view, both topical and relational adjectives showing the subject area or a relation with the noun described are likely to be more frequently used.

In contrast, the nouns *agreement* and *conclusion* usually collocate with **quantity**, **extent**, **time** and **evaluative** descriptors (e.g. *controversial conclusion*, *general agreement*) in the *HSC*. The case of *conclusion* is particularly worth noting. Contrary to what might be expected in academic prose, this noun is more commonly modified by evaluative descriptors (e.g. *misleading conclusion*, *controversial conclusion*). Such unexpectedness lies in the fact that evaluative adjectives describe the position of the writer towards the discussed *conclusion*. According to Biber et al. (1999), this function of making personal judgements is considered to be more appropriate of fiction and literary writing. Thus, this seems to be a distinctive feature of the collocational patterning of the abstract noun *conclusion* in the *HSC*.

Having identified and analysed the most salient patterns of the five abstract nouns under study, the chief emphasis of the following chapter will be placed on providing a detailed account of the informants' command of the phraseological patterning of abstract nouns in medical English. So as to examine their proficiency in the use of the collocations associated with abstract nouns, the data obtained from the worksheet of exercises as well as the information from the Survey Content-Based Questions will be studied.

On the one hand, the former will enable to identify the collocations that appear to be particularly difficult for informants. The analysis of how the collocations produced by these non-native speakers deviate from native-like collocation will reveal which collocations are specifically liable to confusion and which elements in the phraseological units considered turn out to be problematic for informants. On the other hand, the latter will inform about the subjects' perception with regard to the problem areas they encounter when writing their medical articles in English as well as their writing up process.

Once these problem areas have been identified, particularly useful resources and tools will be presented (cf. chapter 6). As will be further discussed in the following chapter, I firmly believe that the idiosyncrasies of noun patterns in scientific English justify the need to provide non-native speakers with suitable tools to help them become proficient in the use of the correct collocations of words when writing a paper in English.

CHAPTER 5: INFORMANTS' DATA ANALYSIS

In the previous chapter, a detailed account of the colligational and collocational patterning of abstract nouns in the *HSC* corpus was set out. This chapter will introduce the second part[88] of the survey conducted among a community of twenty-four Spanish doctors. The first and more substantial part of this chapter (section 5.1) comprises a set of exercises, aimed to test the informants' familiarity with the usage of nouns described in the previous chapter. The analysis of informants' difficulties when dealing with the abstract nouns in question will show how valuable the corpus-based research conducted in this study is. This is followed by a summary of the overall results (cf. section 5.2) concerning the informants' command of the collocational and colligational patterning of abstract nouns in medical English, which will, on the one hand, stress the importance of collocations in EFL writing and, on the other, validate the relevance of the corpus data analysis undertaken in chapter 4.

The last section of this chapter (cf. section 5.3) presents a further part of the abovementioned survey (i.e. Survey Content-Based Questions), which offers a more self-evaluative view of the participants' language skills as well as their needs. This part was included to offer informants a chance to state their own views and opinions on their writing process in English and also the resources they would like to have at their disposal so as to facilitate their written production in English.

It is worth pointing out that neither the list of deviant structures identified in the informants' responses to the worksheet of exercises (cf. section 5.1) nor the problem areas highlighted by informants themselves (cf. section 5.3) can of course be considered exhaustive since informants' publications in English have not been explored, but the results obtained can still be seen as an indication of what type of language help is relevant to provide to Spanish medical doctors who aim to publish internationally.

---

[88] The first part (i.e. Survey Profile Questions) provides information about the informants' profile, which is described in chapter 3.

5.1 Results obtained from the worksheet of exercises

As stated in chapter 3, with the aim of identifying the sort of difficulties researchers usually encounter when using abstract nouns in their academic papers, four exercises were handed out to the community of twenty-four informants involved in this study. Appendix 5 summarises the description and the aims pursued in each of the activities included in the worksheet distributed among participants.

This section presents the major results of the analyses of the abovementioned exercises (5.1.1 to 5.1.4), which are in turn compared against the results extracted from the *HSC* corpus (sections 4.3 to 4.7). The typology of exercises designed for this research ranges from “filling in the gaps” and “matching” exercises to “sentence translation from informants’ L1 (Spanish) into English”. Therefore, the sort of information provided by each of them is also varied. While in Exercises 2 and 3 participants were exposed to instances of real language in context, the “matching” exercises (i.e. Exercises 1 and 4) asked them to produce collocations in isolation[89]. Consequently, the retrieval of data from the former provided, as will be seen, much more significant information than the latter.

5.1.1 Fill in the Gaps Exercise

Exercise 2. Predict the words which are missing. Complete the sentences below with a suitable word.

1. The research unit, within four days, ____________ the conclusion that the government was not giving enough.
2. Comparisons between mice have been ____________.

(…)[90]

The main aim of Exercise 2 was to identify the major types of difficulties informants had regarding the lexico-grammatical patterning of the five abstract nouns considered in the present research. As described in section 3.5, participants were requested to predict the missing words in a list of fifteen sentences, randomly extracted from the *HSC* corpus (see Appendix 4). Each example contained either one or two (i.e. sentences 4, 9, 10 and 12) blank spaces that informants had to complete with a suitable lexical item. No additional guidance about the word class expected was supplied.

---

[89] Cf. Appendix 4.

[90] The whole exercise is displayed in Appendix 4.

Unlike Exercise 1 and Exercise 4, the data displayed in this exercise provided informants with contextualised real examples; that is, they were not asked to associate lexical items in isolation but within a given context. The fact that the examples selected were real instances typical of the medical register was also intended to offer informants a more recognisable linguistic context. In addition, as will be discussed in the pages to follow, it should be borne in mind that some sentences allowed for more than one possible answer.

The findings obtained from this exercise give account of the informants' collocational competence with respect to three main combinatorial patterns; namely: restricted verb collocates (i.e. sentences 1, 2, 5, 6, 7, 8, 9 and 15), free verb collocates (i.e. sentences 3, 11, 13 and 14) and lexical bundles (i.e. sentences 4, 10 and 12).

a) Group 1: restricted verb collocates

Evidence indicates that informants seemed to be aware of the light verbs most commonly used in combination with the abstract nouns examined (e.g. *reach/ draw/ lead to a conclusion*, *make a comparison*, *make a contribution*, *reach an agreement*). However, the use of verbs such as *give* (e.g. *give a conclusion*), *do* (e.g. *do a contribution*, *do a comparison*), *be* (e.g. *be a contribution*) and *achieve* (e.g. *achieve an agreement*, *achieve a conclusion*) where a different or more appropriate light verb was required reveals that these verb-noun combinations pose certain problems for non-native speakers.

Furthermore, the fact that a wide repertoire of free collocates was used when a restricted verb collocate was expected (e.g. *report/ give/ adopt/ publish a conclusion*; *give/ bring/ report/ generate/ mean a contribution*) reinforces my assumption that restricted verb-noun collocations are liable to confusion. The sentences in this group will now be analysed in turn.

Sentence 1: The research unit, within four days, *reached* the conclusion that the government was not giving enough.

Data show (cf. Figure 50) that informants were able to identify the most common restricted verb collocates associated with the noun *conclusion* (i.e. *reach*, *draw*, *make*, *lead to* and *arrive at*). The light verbs *draw* and *reach* (produced by 6 informants each) stand out as the most commonly used. However, as data from the *HSC* revealed (cf. section 4.3), the combination *draw + conclusion* tends to be more frequently used in passive constructions whereas *reach + conclusion* is much more common in the active form. Informants did not show a stronger preference for the latter in sentence 1, though.

	n	%
DRAW	6	25,0
REACH	6	25,0
ARRIVE(D) AT	2	8,3
MADE	1	4,2
LEAD TO	1	4,2
WILL GIVE/GIVES	2	8,3
WILL REPORT	1	4,2
WILL ADOPT	1	4,2
WILL PUBLISH	1	4,2
GOT	1	4,2
OBTAINED	1	4,2
IS ANNOUNCING	1	4,2

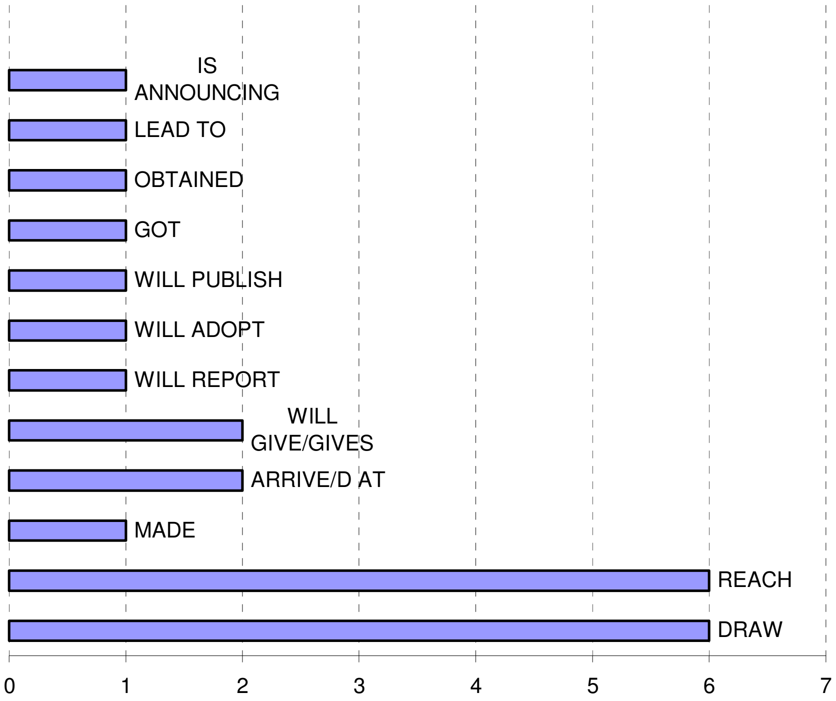


Figure 50 Collocations provided by informants in Sentence 1 (Exercise 2)

Despite the high percentage of the restricted verb collocates provided (66,7%), there is still a remarkable 29,2% of informants who produced deviant collocations with free collocates as the following: *give/ report/ adopt/ get/ obtain/ announce a conclusion*. But for *publish a conclusion*, none of the other free collocates selected was found in the *HSC*.

Sentence 2: Comparisons between mice have been made.

Figure 51 below shows that the word combinations *make a comparison* (8 informants) and *do a comparison* (7 informants) stand out as the most commonly produced collocations in sentence 2. On the contrary, data from the *HSC* corpus show a stronger preference for the use of *make* in combination with the abstract noun *comparison*, while the collocation *do + comparison* hardly ever occurred in the native corpus[91].

	n	%
MADE	8	33,3
DONE	7	29,2
SIGNIFICANT	2	8,3
WRONG	1	4,2
PERFORMED	1	4,2
CLEAR	1	4,2
CONTRADICTORY	1	4,2
FAR-REACHING	1	4,2
DRAWN	1	4,2
CORRECT	1	4,2

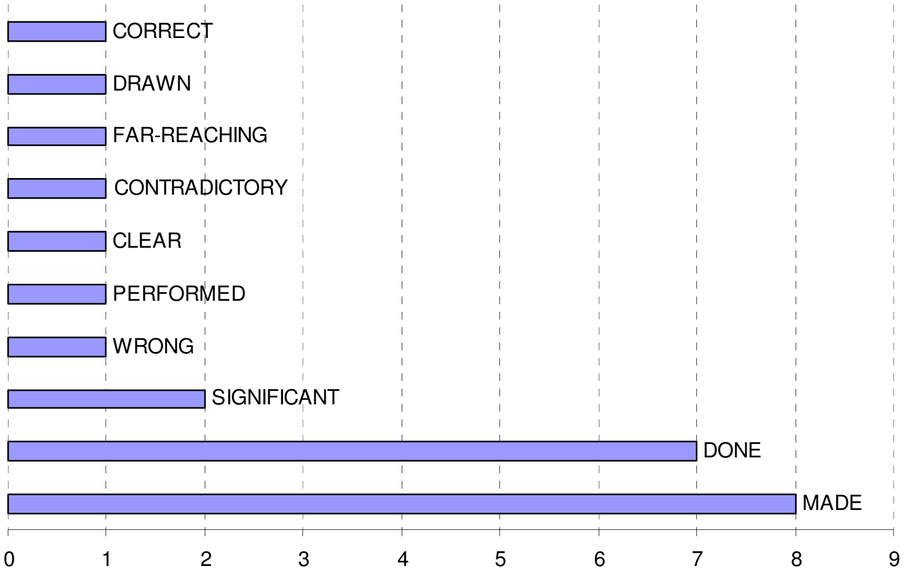


Figure 51 Collocations provided by informants in Sentence 2 (Exercise 2)

Another interesting feature in this sentence lies in the fact that 25% of informants produced an adjective rather than a passive form of a light verb. With the only exception of the adjective *significant*, the other five adjectives provided were not found as collocates of the noun *comparison* in the *HSC*. In fact, there were very few *HSC* examples of predicative complements (i.e. *comparison is unique/ valid*; *comparisons are difficult*) modifying the noun *comparison*. 6 informants (25%), on the contrary, associated evaluative predicative adjectives (e.g. *correct*, *contradictory* and *wrong*) as common collocates of the noun *comparison*. This type of deviation therefore indicates that the use of delexicalised verbs (i.e. *make*, *do*) in combination with abstract nouns is a problem area for non-native speakers.

---

[91] There were 28 occurrences of the combination *make + comparison* as opposed to 3 instances of *do + comparison* (cf. section 4.5).

Sentence 5: Further trials are needed before any *conclusion* can be drawn about the protocol's efficacy.

Unlike the rest of the sentences in this group of verb-noun collocates, where a verb was expected, in sentence 5 informants were asked to elicit the abstract noun (i.e. *conclusion*) subcategorised by the restricted verb *draw*, which was the most frequently verb collocate of the noun *conclusion* in the *HSC* corpus[92].

	n	%
CONCLUSION	19	79,2
CONCLUSIONS	2	8,3
AGREEMENT	1	4,2
INFORM	1	4,2
NO REPLY	1	4,2

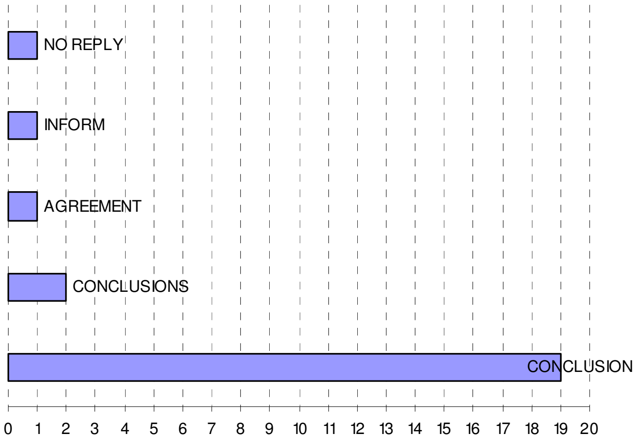


Figure 52 Collocations provided by informants in Sentence 5 (Exercise 2)

As can be inferred from Figure 52, a total of 21 informants (87,5%) identified the abstract noun *conclusion* as the best collocate for the restricted verb *draw*. What seems worth mentioning in this case is the fact that 19 informants chose the form *conclusion* instead of its inflected counterpart *conclusions*, which was only chosen by 2 informants. Both forms are equally acceptable in this context but it should be borne in mind that the analysis of the noun *conclusion* in the *HSC* showed a strong preference for the use of the inflected form (145 occurrences) as opposed to its base form (39 occurrences).

The other two[93] examples provided correspond to two inappropriate uses of the abstract noun *agreement* (i.e. "*before any agreement can be drawn*"), on the one hand, and the non-existent noun **inform* (i.e. "*before any *inform can be drawn*"), on the other.

---

[92] 39,4% of the total number of occurrences of the pattern restricted verb + *conclusion* occurred with the verb *draw*. See further details in section 4.3.

[93] There was one informant who did not answer this item.

Sentence 6: Clinical examination and biochemical screening *make* an important contribution to the detection of abnormalities.

The analysis of the abstract noun *contribution* in the *HSC* (cf. section 4.6) revealed that the most common restricted collocate of that noun was *make* (25 occurrences) and, to a much lesser extent, *provide* (4 occurrences) and *produce* (1 occurrence). With respect to informants in this study, 29,2% (7 informants) of those inquired associated *make* with the noun *contribution* whereas a higher percentage (37,5%; 9 informants) chose the verb *be* (in its various inflected forms), as Figure 53 shows. This collocation should be more carefully looked at. Although the string *be a contribution* did appear in the native corpus, its frequency of occurrence was particularly low. The only three examples found across the 167 hits of the node word *contribution* in the *HSC* are the following:

(1) This chapter is a most outstanding contribution to our understanding of deer evolution and behavioural ecology.

(2) Food transfers seem to be an important contribution to the total food obtained by young beyond weaning age.

(3) […] this computationally inefficient procedure is a unique contribution of Willis and Wiese.

One possible reason for the informants' high use of *be* in this context could be that this verb may be used with a wide range of nouns and may have different translation equivalents in Spanish. Nesselhauf (2005) in her study of collocations in learner corpora pointed to the same reasons with regard to collocations and free combinations with the lexical verb *have*.

	n	%
IS(2)ARE(6)HAVE BEEN(1)	9	37,5
MADE/MAKE	7	29,2
GET	1	4,2
CONSTITUTES	1	4,2
MAY HAVE/PROVIDE	1	4,2
SUPPORT	1	4,2
HACEN[94]	1	4,2
PLAY	1	4,2
BECOME	1	4,2
GIVE	1	4,2

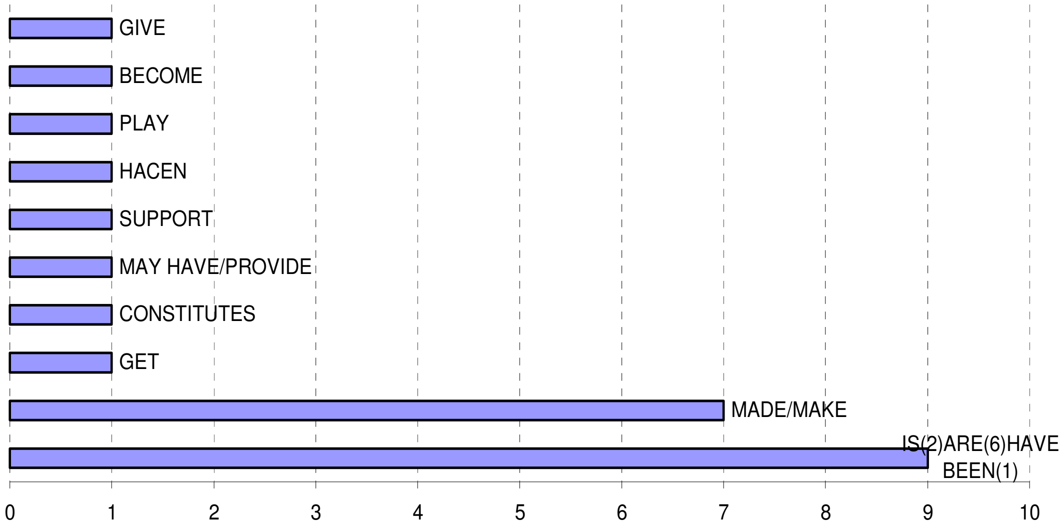


Figure 53 Collocations provided by informants in Sentence 6 (Exercise 2)

Deviations concerning the use of inappropriate free combinations are also present in examples such as *support/ play/ become/ give/ get a contribution*. Some other examples like "?*constitute a contribution*" seem to be influenced by the informants' L1 since the word string "*make an important contribution to*" could be translated into Spanish as "*constituye/ supone una contribución importante a*".

Sentence 7: They *reached* the final agreement that clinical examinations for defects in hips, vision and hearing, and other congenital abnormalities is less well founded on scientific evidence.

In line with the findings described in section 4.4, 13 informants (54,2%) associated the restricted verb *reach* with the abstract noun *agreement*. However, the individual verbs listed in Figure 54 reveal that six other verbs (i.e.

---

[94] This informant provided the Spanish word form *hacen* rather than the English *make*.

achieve, conclude, support, be, have got and **porpouse*[95]) were produced inappropriately by at least one participant in the study. A total of 8 informants (33,3%) provided inappropriate verb collocates in sentence 7, which contributes to reinforce the idea that the collocations of the type verb-abstract noun are remarkably difficult for the informants.

	n	%
REACHED	13	54,2
ACHIEVED	2	8,3
CONCLUDED	2	8,3
NO REPLY	2	8,3
SUPPORTED	1	4,2
LEAD TO	1	4,2
ARE	1	4,2
HAVE GOT	1	4,2
PORPOUSE	1	4,2

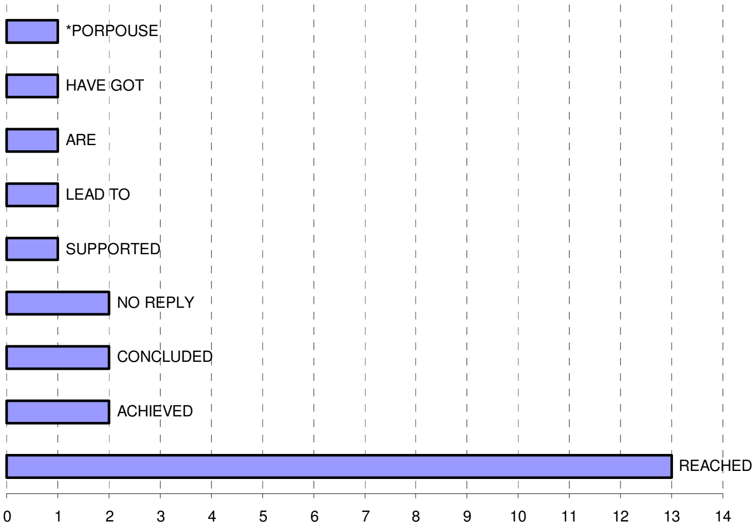


Figure 54 Collocations provided by informants in Sentence 7 (Exercise 2)

Sentence 8: Bruce Weir has *made* many important contributions to population genetic inference theory.

As already noted in sentence 6, the restricted verb *make* proved to be the most common verb in combination with the noun *contribution* in the *HSC* corpus. Whereas in sentence 6 the verb *make* was only produced by 29,2% of those inquired, in sentence 8 the use of *make* is significantly higher (58,3%). See Figure 55 below:

	n	%
MADE	14	58,3
DONE	4	16,7
PROVIDED	1	4,2
CONTRIBUTED	1	4,2
GIVEN	1	4,2
BROUGHT	1	4,2
REPORTED	1	4,2
NO REPLY	1	4,2

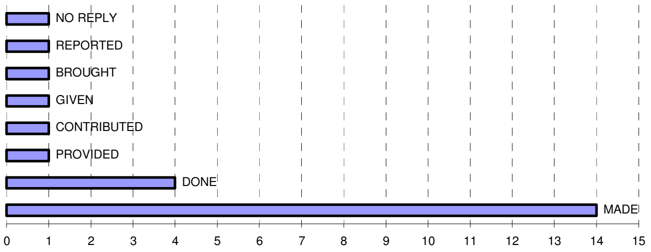


Figure 55 Collocations provided by informants in Sentence 8 (Exercise 2)

---

[95] This non-existent form may be referred to the verb *suggest* ("*proponer*" in Spanish).

Concerning the other verb collocates listed in the Figure above, the most outstanding deviation corresponds to the word string *do a contribution*, produced by 16,7% of the informants. This inappropriate choice of the verb *do* reflects the widespread confusion (Kaszubski 2000; Nesselhauf 2005) of certain pairs of high frequency verbs (i.e. *do* for *make*) used as light verbs in verb-noun collocations. Although this confusion is relatively low (i.e. 4 occurrences of *do* as opposed to 14 of *make*), the choice of light verbs in combination with abstract nouns appears to be problematic for the medical community under study. Further indications of this assertion can also be found in the fact that another 16,7% of the informants resorted to the use of free collocates such as ?*give a contribution*, ?*contribute a contribution*, ?*bring a contribution* and ?*report a contribution*, which were not found in the native production.

Sentence 9: These results *led to* some important conclusions.

As indicated in Figure 56 below, the collocations most frequently produced in this sentence were the following: *lead to a conclusion*, *draw a conclusion* and *reach a conclusion*. These three word combinations were in turn the most frequent ones in the *HSC*. However, *draw a conclusion* should not be regarded as an acceptable answer in sentence 9 since informants were required to provide two word forms rather than one.

	n	%
LEAD TO	7	29,2
DRAW	5	20,8
HAVE REACH/ED	2	8,3
NO REPLY	2	8,3
CAN SUPPORT	1	4,2
HAVE ACHIEVED	2	8,3
DO ALLOW	1	4,2
HAVE SUPPOSED	1	4,2
CAN PROVIDE	1	4,2
HAVE BROUGHT	1	4,2
MADE UP	1	4,2

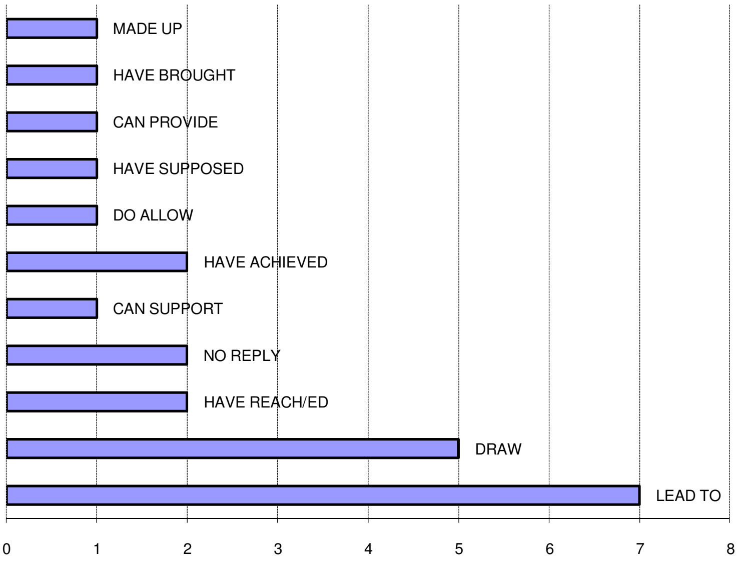


Figure 56 Collocations provided by informants in Sentence 9 (Exercise 2)

The above requirement made the other 10 informants who completed the sentence use a variety of strategies: two informants provided a modal verb in combination with a free verb collocate (e.g. *can support*, *can provide*), six informants made use of the auxiliary "*have* perfective" collocating with either the restricted verb collocate *reach* (e.g. *have reached*) or some unacceptable free verb collocates (e.g. *have achieved, have supposed, have brought*), one informant used the emphatic *do* with the free collocate *allow*, which was not found in the *HSC* either, and, finally, another informant chose an inappropriate phrasal verb (e.g. "?*these results made up some important conclusions*").

Sentence 15: These multivitamins do not appear to *provide* significant contributions to the parameters tested.

The results found in sentence 15 (cf. Figure 57) are very similar to what was discussed for sentence 8 above. In this case 45,8% of the informants wrote *make* as an appropriate restricted verb collocate of the noun *contribution*. In the *HSC* example, the chosen verb was *provide*, which, as noted in section 4.6, was less frequently used. In line with this trend, there was no instance of *provide* in the repertoire of verbs produced by informants.

	n	%
MAKE	11	45,8
HAVE (A)	3	12,5
ADD	2	8,3
NO REPLY	1	4,2
GENERATE	1	4,2
MODIFICAR	1	4,2
PROVIDE	1	4,2
GET	1	4,2
MEAN	1	4,2
GIVE	1	4,2
CONTRIBUTE	1	4,2

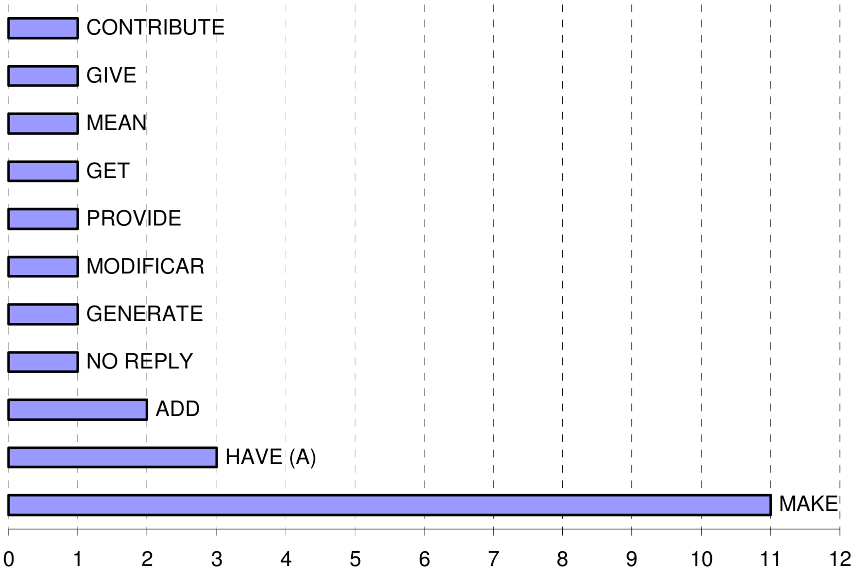


Figure 57 Collocations provided by informants in Sentence 15 (Exercise 2)

With regard to the other verbs used in this sentence, a further type of inappropriate collocations is worth discussing. It consists of free collocates that were not found in combination with the noun *contribution* in the *HSC*. Examples of this category are: *add / generate / *modify*[96] */ get / mean / give / ?contribute contributions*. The second type of questionable collocations, produced by three informants (12,5%), corresponds to the word string *have a contribution*. As mentioned earlier, the light verb *have* tends to be used in combination with a wide range of verbs in English (e.g. *have access to*, *have an influence*, have *a need*, *have significance*, *have relevance*). However, the collocation *have a contribution* was not found in the native corpus.

b) Group 2: free verb collocates

The second combinatorial pattern to be examined in this exercise is the informants' competence in using free verb collocates with abstract nouns. So as to explore whether this was a particularly difficult area for non-native speakers, four examples were selected (i.e. sentences 3, 11, 13 and 14) from the *HSC*, which will be now analysed in detail.

Sentence 3: Such a conclusion is *supported* by our observations.

37,5% of the informants referred in their production to either *support a conclusion* or *confirm a conclusion*, which had also revealed as the most usual free combinations in the *HSC* (cf. section 4.3). Five informants produced an inappropriate word class, an adjective, giving rise to unexpected instances (e.g. "*such a conclusion is contradictory* [3 informants] / *valid* [1 informant] / *wrong*[97] [1 informant] *by our observations*"). The reasons for that inappropriacy lie in the fact that the "*by*-Prepositional Phrase" requires a passive participle rather than an adjective in the example provided.

Among the other free verb collocates provided by informants, it should be stressed that they did not appear as collocates in the native corpus. Thus, examples such as *take a conclusion*, *prove a conclusion*, *share a conclusion*, *give a conclusion* and *obtain a conclusion* were not found in the *HSC*. The word string **enforce a conclusion*, though inappropriate in its morphological form, should not be grouped along with the already cited deviant collocations since it seems it refers to *reinforce a conclusion*, which was used three times in the native data.

---

[96] The actual example provided was in Spanish: "**modificar" contributions*.

[97] But for "*such a conclusion is wrong*", which was found once in the *BoE* (*Bank of English*), the other two adjectives provided by informants (i.e. *contradictory* and *valid*) were not found in the reference corpora.

	n	%
SUPPORTED	6	25,0
CONFIRMED	3	12,5
CONTRADICTORY	3	12,5
TAKEN	2	8,3
MADE	2	8,3
ENFORCED	1	4,2
DRAWN	1	4,2
PROVEN	1	4,2
VALID	1	4,2
WRONG	1	4,2
SHARED	1	4,2
GIVEN	1	4,2
OBTAINED	1	4,2

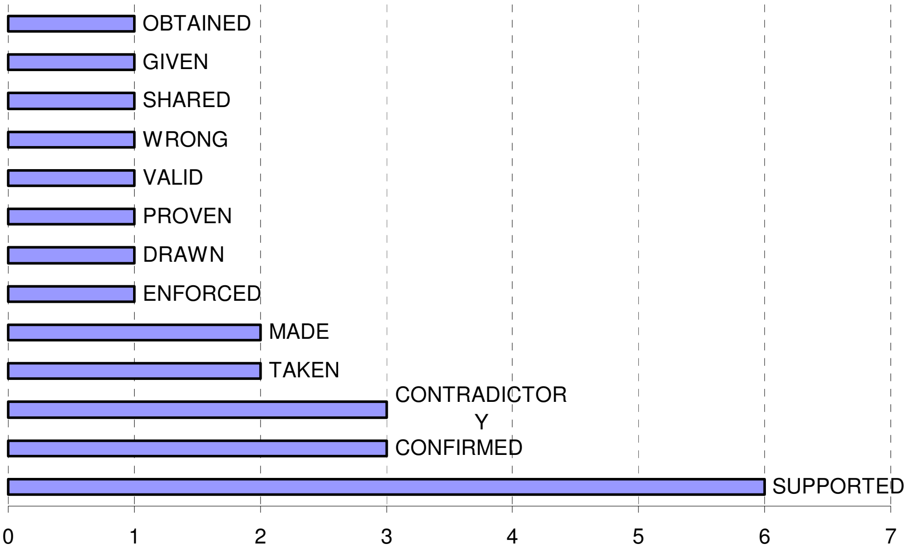


Figure 58 Collocations provided by informants in Sentence 3 (Exercise 2)

Another remarkable feature is the use of restricted verb collocates in this context. Data from Figure 58 above reveal that two informants wrote *made* and *drawn* in their answers to this sentence. These two inappropriate uses of support verbs indicate that informants find it difficult to discriminate between restricted and free collocates in combination with the abstract noun *conclusion*.

Sentence 11: Despite these conservative features, our analysis broadly *supports* the recent decision by the Department of Health that HIV testing should be offered universally.

	n	%
SUPPORT/S	15	62,5
LED/LEAD TO	2	8,3
NO REPLY	2	8,3
AGREEMENT	1	4,2
ACCEPTED	1	4,2
CONCLUDE	1	4,2
CONFIRM	1	4,2
CORROBORATE	1	4,2

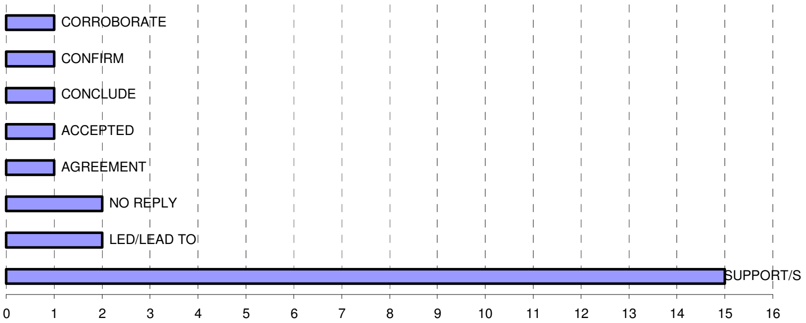


Figure 59 Collocations provided by informants in Sentence 11 (Exercise 2)

As shown in Figure 59, most informants (15; 62,5%) produced the verb *support* as the most common free verb collocate of the noun *decision*. With the only exception of two informants who provided an inappropriate restricted verb collocate (e.g. "*our analysis broadly ?led to the recent decision by the Department of Health that HIV testing should*") and another informant who appeared to have confused the verb *agree* with the abstract noun *agreement* and thus produced an ungrammatical sentence (i.e. "*our analysis broadly *agreement the recent decision by the Department of Health that HIV testing should be offered universally*"), the rest of the informants' contributions correspond to free verb collocations:

(4) Our analysis broadly accepted the recent decision by the Department of Health that HIV testing should be offered universally.
(5) Our analysis broadly conclude(s) the recent decision by the Department of Health that HIV testing should be offered universally.
(6) Our analysis broadly confirm(s) the recent decision by the Department of Health that HIV testing should be offered universally.
(7) Our analysis broadly corroborate(s) the recent decision by the Department of Health that HIV testing should be offered universally.

None of the above collocations was observed in the examination of the *HSC*. In this respect, Nesselhauf (2005:210) claims that a possible reason for this type of inappropriate free combinations lies in the fact that in these cases both the verb and the noun are chosen independently of each other, which may bring about an inappropriate selection of individual elements.

Sentence 13: It was not possible to *determine* the relative contribution of this glucose derived directly from fructose.

37,5% of those inquired wrote either *assess* (20,8%; 5 informants) or *determine*[98] (16,7%; 4 informants) in their responses to sentence 13, which seems to go in line with the findings observed in section 4.6: the most frequent free collocates of the noun *contribution* in the native corpus were *determine* (5 occurrences) and *assess* (3 occurrences).

---

[98] Two of those informants, though, produced an ungrammatical form (i.e. **determinate*) to refer to the verb *determine*.

	n	%
ASSESS	5	20,8
DETERMINE/ *DETERMINATE	4	16,7
KNOW	2	8,3
MEASURE	2	8,3
SHOW	2	8,3
PROVE	2	8,3
DRAW	1	4,2
NO REPLY	1	4,2
ASSUME	1	4,2
CONCLUDE	1	4,2
REACH	1	4,2
PREDICT	1	4,2
CONFIRM	1	4,2

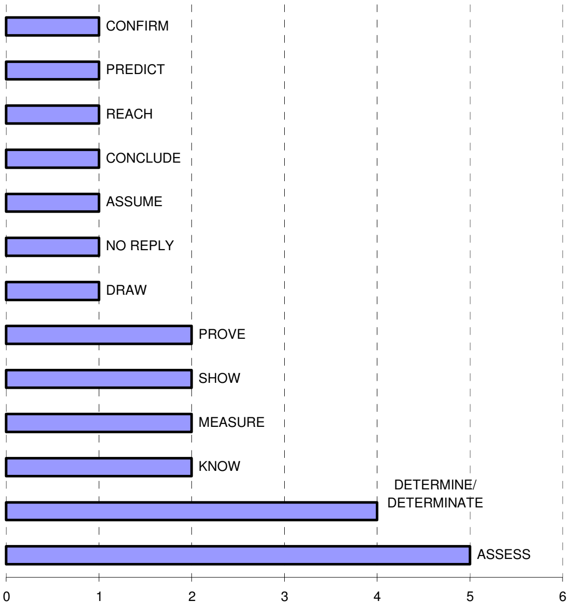


Figure 60 Collocations provided by informants in Sentence 13 (Exercise 2)

As noted before, however, there is a general tendency across the non-native production to use a wide range of free collocates in combination with the noun *contribution*. This was also observed in the native corpus but, interestingly, none of the free collocates provided by the informants in this sentence (cf. Figure 60 above) coincides with the ones found in the native corpus (cf. Table 19 in section 4.6). This mismatch stresses, as will be discussed later on, the convenience of providing non-native speakers with the collocational patterns that typically occur in medical discourse.

Sentence 14: They will certainly *support* the decision the group has made.

As evidenced by the informants' responses to the previous example (i.e. sentence 13), some participants appeared to be aware of the most frequent free verb collocates that occur with the abstract noun *decision*. A remarkable 41,7% wrote either *support* (5 informants) or *accept* (5 informants) in the space provided in sentence 14. Both *support a conclusion* and *accept a conclusion* were also frequently used combinations in the *HSC*. However, data in Figure 61 reveal that more than half of the informants (54,1%) either offered free verb-noun collocations that did not occur in the *HSC* data (41,6%; 10 informants) or produced restricted verb-noun collocations with the verb *make* (12,5%; 3

informants), which resulted in ungrammatical instances (e.g. "*they will certainly *make the decision the group has made*").

	n	%
SUPPORT	5	20,8
ACCEPT	5	20,8
ADOPT	3	12,5
MAKE	3	12,5
RESPECT	1	4,2
OBBEY	1	4,2
AGREEMENT	1	4,2
NO REPLY	1	4,2
ENCOURAGE	1	4,2
BELIEVE	1	4,2
AGREE	1	4,2
CONFIRMED	1	4,2

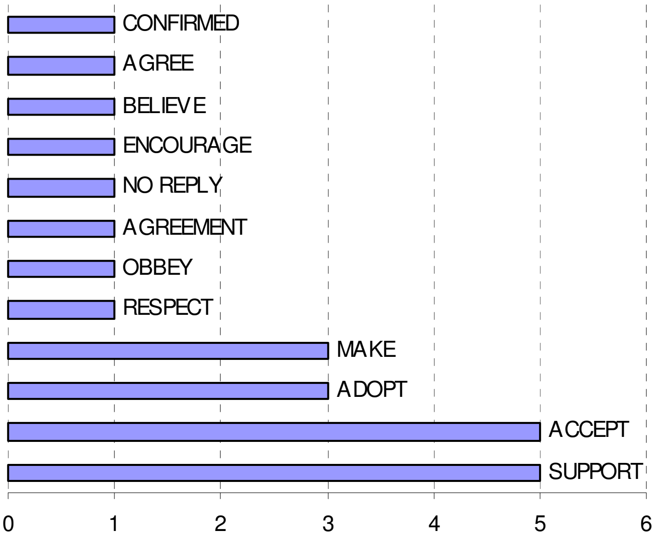


Figure 61 Collocations provided by informants in Sentence 14 (Exercise 2)

c) Group 3: lexical bundles

The third aspect to be explored in this exercise concerned participants' competence regarding the use of three main lexical bundles; namely, *in agreement with*, *there is* (adjective) *agreement that* and *in comparison with/to* which were very productive[99] in the native *HSC* corpus.

Sentence 4: The data presented here are not in agreement with the model of Studier and Bandyopadhyay.

A total of 9 deviant collocations were found in sentence 4. Evidence from the participants' responses reveals that only 11 informants (45,8%) produced the lexical bundle *in agreement with* correctly. The following Figure displays all the collocations provided:

---

[99] Cf. sections 4.4 for the analysis of the combinations *in agreement with* and *there is* (adjective) *agreement that* and section 4.5 for the description of *in comparison with/to*.

	n	%
IN…WITH	11	45,8
NO REPLY	4	16,7
TOTALLY…WITH	2	8,3
FULLY…WITH	2	8,3
REPORTED…WITH	1	4,2
INCLUDED IN THE…OF	1	4,2
VALID…IN	1	4,2
VALID…FOR	1	4,2
ENOUGH…WITH	1	4,2

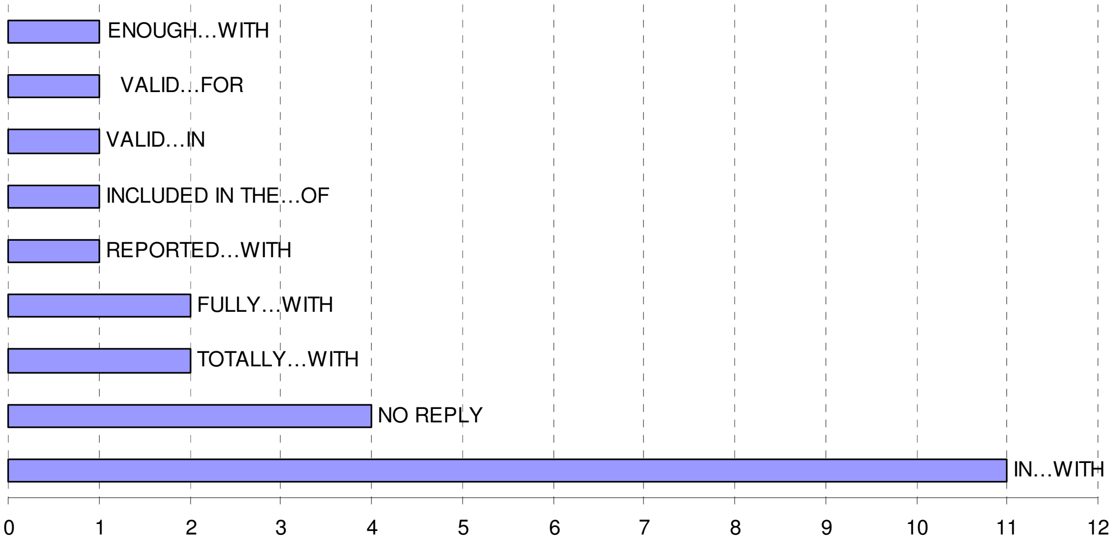


Figure 62 Collocations provided by informants in Sentence 4 (Exercise 2)

Among the inappropriate collocations, two different types stand out. The most salient one is the use of an adjective (i.e. *enough*), a passive participle (i.e. *reported*) or an adverb (i.e. *totally*, *fully*) for the preposition *in*. See the following examples:

(8) *The data presented here are not enough [1 informant] agreement with the model of Studier and Bandyopadhyay.

(9) *The data presented here are not reported [1 informant] agreement with the model of Studier and Bandyopadhyay.

(10) *The data presented here are not totally [2 informants] agreement with the model of Studier and Bandyopadhyay.

(11) *The data presented here are not fully [2 informants] agreement with the model of Studier and Bandyopadhyay.

The choice of adverbs in examples (10) and (11) above deserves special attention. Despite the fact that both of them result in deviant combinations, these two adverbs are very common collocates of the verb *agree* and, similarly, the adjectives *full* and *total* proved to be frequent premodifiers of the abstract noun *agreement* in the *HSC* (cf. section 4.4). This seems to indicate that informants were familiar with some collocates of the abstract noun *agreement* as well as with its cognate lexical verb *agree*, but did not control the lexical bundle *in agreement with*, which, as discussed in section 4.4, shows a high degree of semantic unity and works as a whole grammatical unit.

The other type of deviant collocations in this sentence corresponds to those combinations produced by three informants where the two elements associated with the noun *agreement* were unacceptable:

(12) *The data presented here are not included in the agreement of [1 informant] the model of Studier and Bandyopadhyay.
(13) *The data presented here are not valid agreement in [2 informants] the model of Studier and Bandyopadhyay.
(14) *The data presented here are not valid agreement for [2 informants] the model of Studier and Bandyopadhyay.

The three examples above suggest that these informants were not familiar with the structure of the lexical bundle required in this sentence.

Sentence 12: *There* is general agreement *that* the primitive host cell was anaerobic.

The *HSC* revealed that 47% of the occurrences of the four-word bundle "*there is* (adjective) *agreement that*" were used to introduce clauses (cf. section 4.4). Thus, I was especially interested in examining how competent informants were in producing this lexicalised expression.

	n	%
THERE…THAT	12	50,0
NO REPLY	5	20,8
IT…THAT	2	8,3
THERE…IN	1	4,2
THERE…ABOUT	1	4,2
FINALLY…THAT	1	4,2
THE RESULT…WITH	1	4,2
THIS…WITH	1	4,2

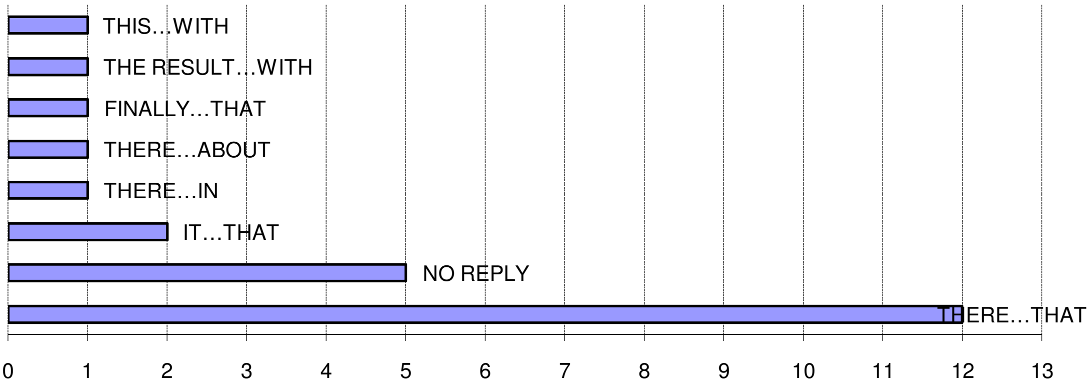


Figure 63 Collocations provided by informants in Sentence 12 (Exercise 2)

Percentages in Figure 63 evidence that half of the participants in the study produced the lexical bundle *there is (general) agreement that* accurately, while the other half had problems with at least one of its constituents. The existential *there* introducing this word string was not provided by 5 informants, among whom 4 either used different types of subjects, such as the impersonal pronoun *it* (2 informants), a noun phrase (e.g. *the result*; 1 informant) or the demonstrative pronoun *this*, whereas the fifth informant provided no subject (e.g. "**finally is general agreement that*"). In both cases, informants' examples resulted in unacceptable word combinations.

On the other hand, there were two informants who had problems with the function word introducing the "*that*-clause" immediately following the abstract noun *agreement*. These two informants, as the examples below show, produced an inappropriate preposition (i.e. *in*, *about*) instead:

(15) There is general agreement *in [1 informant] the primitive host cell was anaerobic.

(16) There is general agreement *about [1 informant] the primitive host cell was anaerobic.

Another remarkable feature in this sentence is the high percentage (20,8%) of informants who supplied no answer. This fact reinforces the idea that this type of collocation was particularly problematic for the Spanish medical community examined in this investigation.

Sentence 10: To assess the cost effectiveness of universal antenatal HIV screening in comparison with selective screening in the UK.

This example has proven to be especially difficult for the participants in the study, as the data displayed in Figure 64 below indicate:

	n	%
IN…WITH	8	33,3
NO REPLY	7	29,2
IN…TO	4	16,7
WE'LL MAKE A…WITH	1	4,2
REQUIRE…WITH	1	4,2
AS…TO	1	4,2
WE MADE A…BETWEEN	1	4,2
STATISTICAL…LEADED TO	1	4,2

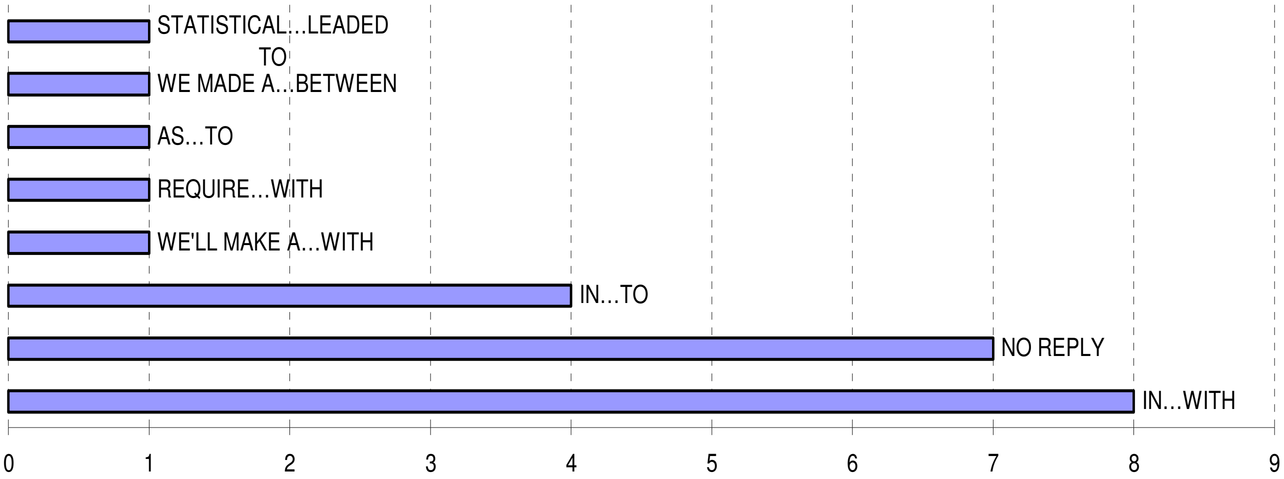


Figure 64 Collocations provided by informants in Sentence 10 (Exercise 2)

Altogether, 12 informants (50%) wrote the appropriate collocations *in comparison with* (8 informants) or *in comparison to* (4 informants). These figures are also comparable to the results obtained in the *HSC*. As was discussed in section 4.5, the occurrences of *in comparison with* outnumbered the instances of *in comparison to* by 46 to 30.

Nevertheless, 5 participants (20,8%) produced other word combinations consisting of inappropriate uses of clauses (cf. examples (17) and (18)), verb phrases (cf. example 19), prepositions (cf. example (20)) and adjectives (cf. example (21)):

(17) ?To assess the cost effectiveness of universal antenatal HIV screening we'll make a comparison with [1 informant] selective screening in the UK.

(18 ?To assess the cost effectiveness of universal antenatal HIV screening we made a comparison between [1 informant] selective screening in the UK.

(19) To assess the cost effectiveness of universal antenatal HIV screening *require comparison with [1 informant] selective screening in the UK.

(20) To assess the cost effectiveness of universal antenatal HIV screening *as comparison to [1 informant] selective screening in the UK.

(21) To assess the cost effectiveness of universal antenatal HIV screening statistical comparison *leaded to [1 informant] screening in the UK.

Examples (17) and (18) were quite elaborated and included restricted verb collocates which typically occur with the noun *comparison* (i.e. *make a comparison*). However, one of the objects of *comparison* was missing in both of them (i.e. "*we'll make a comparison* [of]… *with*…", "*we made a comparison between selective screening in the UK* [and]…" In Nesselhauf's view,

> "a verb-noun collocation is not only understood as comprising a verb and a noun but also the central determiners, noun complementation structures, and so forth that are present. All elements in an expression such as *come to the conclusion* that would therefore be considered to belong to the collocation, and deviations in any of these elements are considered to be deviations in the collocation." (Nesselhauf 2005:71)

Accordingly, word combinations such as the above (i.e. examples (17) and (18)) must also be regarded as unacceptable. Likewise, examples (19), (20) and (21) are inappropriate in spite of the fact that they include some lexical items such as the adjective *statistical* and the preposition *as* which, as noted in section 4.5, are usually associated with the noun *comparison* in other word combinations (i.e. *statistical comparison*, *as a comparison*).

One last outstanding feature in this sentence is the high percentage of informants (29,2%; 7 informants) who did not provide any answer. As mentioned in the analysis of sentence 12, these figures seem to prove unmistakably the informants' lack of knowledge about the appropriate use of the lexical bundles explored in this exercise.

All in all, the analysis of informants' production in Exercise 2 has yielded very interesting findings regarding their command of restricted verb-noun collocates and free combinations of the type **V + abstract noun** on the one hand, and their knowledge of the use of some lexical bundles (e.g. *in agreement with*, *in comparison with/to*, *there is* (adjective) *agreement that*), on the other.

Although most informants seemed to be familiar with frequent restricted verb-noun collocates (e.g. *reach / draw a conclusion*, *make a comparison*, *make a contribution*, *reach an agreement*), a high percentage of free combinations was used where a restricted collocation would apply more often in the native production. Within the group of restricted verb-noun collocations, a great number of verb deviations were also observed in the use of high-frequency support verbs in collocations, such as *reach* vs. *achieve an agreement*, *make* vs. *do a contribution*, *be* vs. *have a contribution*. This finding highlights the fact that these pairs of light verbs hinder the participants' written production in English.

Concerning the overall use of free combinations, it should be noted that while informants were able to produce the most common free collocates found in the *HSC*, they also proved to have difficulties in discriminating between appropriate and unacceptable word combinations since some of their productions were deviant collocations not traced in the native corpus (e.g. **take a conclusion*, **adopt a decision*). Exercise 2 has also highlighted participants' lack of awareness of the correct use and structure of some prefabricated expressions (i.e. lexical bundles), which typically occur in medical writing.

In the analysis of Exercise 3 that follows, the mechanisms used by informants when asked to translate some sentences into English will be explored. The examination of their translations will inform about participants' knowledge of the multiword units under discussion.

5.1.2 Translation Exercise

Exercise 3. Provide a suitable translation into English for the following sentences.

1. Se produjo un acuerdo casi unánime entre ambos métodos.

____

____

2. No se realizaron comparaciones para determinar diferencias de género ni de edad.

____

____

(…)[100]

Exercise 3 aimed at analysing informants' written production with regard to the collocational patterning of the selected abstract nouns. Participants were required to provide suitable translations into English for a total of eight sentences, each of which was related to some of the patterns discussed in chapter 4 (i.e. *there* + *be* + (adjective) *agreement*, *in conclusion*, *contribute to* vs. *contribution*, restricted verb collocate + (adjective) + *comparison*, free verb collocate + (adjective) + *conclusion* and *decision making*).

Unlike the other three exercises in the worksheet, Exercise 3 supplied not only information about participants' command of the combinatorial patterning of abstract nouns in medical English, but also their knowledge of other related morphological and syntactic aspects such as word order, the use of passive constructions and the different word forms of lexical units.

Bearing in mind that the main focus of this thesis is on the informants' production with reference to the behaviour of abstract nouns, this linguistic analysis will not cover other deviations observed in the collected data, such as the misuse of the definite article (e.g. *the researchers* vs. *researchers*), misspelling of certain lexical items[101] (e.g. *organisations* ~ **organitations*,

---

[100] Cf. Appendix 4

[101] The examples displayed in the analysis of Exercise 3 are the actual examples provided by informants, which means such deviations (e.g. spelling mistakes, inappropriate word forms) have been preserved.

scientists ~ **scientificts*), confusion with modal verbs (e.g. *must* vs. *have to*), problems with verb tenses (e.g. present perfect vs. past simple) as well as some difficulties with dummy prepositions (e.g. *responsible* **in* for *responsible for*). The individual translations for each sentence are shown in Appendix 6.

Sentences were divided into five groups according to the kind of pattern they were examples of; namely, a) lexical bundles (sentences 1 and 3), b) the pattern restricted verb + *comparison* (sentences 2 and 6), c) the use of the verb *contribute* as opposed to its cognate noun *contribution* (sentence 4), d) restricted and free verb collocates associated with the noun *conclusion* (sentences 5 and 7) and e) the compound *decision making*. The results obtained in each group will be discussed in the pages to follow.

a) Informants' command of the lexical bundles *there + be +* (adjective) *agreement* and *in conclusion*

So as to ascertain whether the production of the lexical bundles *there* + *be* + (adjective) *agreement* and *in conclusion* was particularly difficult for informants, they were asked to translate into English the Spanish expressions "*se produjo un acuerdo casi unánime"* (sentence 1) and "*en resumen"* (sentence 3). Here, the translation equivalents provided by participants will be shown so as to illustrate some relevant discussion points. As already stated, the translations of the full sentences produced by each informant are included in Appendix 6.

Sentence 1: *Se produjo un acuerdo casi unánime entre ambos métodos*.
HSC example: There was almost perfect agreement between the two methods.

The analysis of informants' instances exemplifies well that the lexical bundle *there* + *be* + (adjective) *agreement*, as already noted in Exercise 2, was not easily identified. Only 5 informants produced the string *there was agreement* as the equivalent for the Spanish construction "*se produjo un acuerdo*":

(22) There was almost total agreement […]
(23) There was an almost complete agreement […]
(24) There was a broad agreement […]

(25) There was an almost general agreement […]
(26) There was almost a unanimous agreement […]

A closer look at the examples above reveals that four informants included an indefinite article (i.e. *a*/*an*) between the copular verb and the notional subject that contained the abstract noun *agreement*. The *HSC* instance, on the contrary, does not contain any article. On the one hand, the fact that in the native corpus there is no determiner preceding the noun *agreement* should not be overlooked, since it underlines the idea that there is a strong connection among the elements involved in this multiword unit, which must be understood as a fixed expression. On the other hand, examples (23-26) appear to indicate that non-native speakers were not fully aware of the fixedness of the structure and, thus, might have used an indefinite article so as to translate literally "*un acuerdo*" into *an agreement*.

The other 16 informants[102], who did not use the existential *there* as an equivalent subject for the Spanish impersonal construction, produced deviant collocations. It seems clear from the data that although participants appeared to be aware of the fact that the subject in the English version had to be realized syntactically by means of a phrase, most of them failed to find an appropriate equivalent of the Spanish construction with the reflexive pronoun "*se*".

The most common deviation found in the data corresponds to the use of a passive construction instead of the expected impersonal structure with the existential *there*: a total of 11 informants converted the Spanish impersonal "*se*" into a passive sentence in English, in which the gap of the subject was filled by either the abstract noun *agreement* (8 instances; cf. examples 27-33) or by the impersonal pronoun *it* (3 instances, cf. examples 34-36):

(27) A nearly unanimous agreement […] was reached [2 informants].
(28) An almost unanimous agreement was achieved.
(29) A near unanimous agreement was obtained.
(30) An almost general agreement […] was taken.

---

[102] Three informants provided no translation in Sentence 1.

(31) An almost general agreement […] was reached.
(32) An agreement was done […]
(33) An almost unanimous agreement was reached.
(34) It was reached an agreement.
(35) It was produced an agreement.
(36) It was ??? a ???[103] agreement between both methods.

All the above examples reflect repeated efforts on behalf of the informants to cope with the problem of translating the reflexive pronoun "*se*" into English. While in instances (27-33) the notional subject *agreement*, along with its various premodifiers (e.g. *unanimous*, *general*) and postmodifiers (e.g. *between both methods*), occupies the initial position and, thus, constitutes a deviant example in which the end-weight principle is not met, examples (34-36) are clear instances of literal translations of "*se produjo*" into "*it* + *be* + passive participle", ranging from "*it was reached an agreement*", where an appropriate restricted verb collocate is used, to "*it was produced an agreement*", where the literal translation is most evident.

The extended use of the passive across participants' translations in this sentence has given rise to their production of several restricted and free verb collocates in the form of passive participles. With the only exception of *reach* (i.e. *an agreement was reached*), which was used by 5 informants, the other verbs produced (i.e. *take*, *produce*, *achieve*, *obtain* and *do*) constitute inappropriate verb collocates of the abstract noun *agreement*. None of them was found in the *HSC* verb + *agreement* occurrences analysed in section 4.4. This is a further indication of informants' lacking awareness of the collocational patterning of this abstract noun.

High degrees of L1 influence can also be observed in the selection of the adjective *unanimous* as a premodifier of the noun *agreement*. As can be seen in examples (37-41) below, 8 informants chose "*unanimous agreement*" as the translation equivalent of the Spanish "*acuerdo unánime"*. Although the descriptor *unanimous* can be found as a common collocate of other nouns such as *decision*, *verdict* and *support*, it hardly co-occurs with the noun *agreement*. In fact, it was only traced in the *BoE* (63 occurrrences), since no instance of the string "*unanimous agreement*" could be found either in the *HSC* or in the *BNC*. Informants' consistent production of this word combination seems to be thus influenced by their L1:

---

[103] This informant was unable to produce neither a suitable passive participle nor an equivalent of the Spanish adjective *unánime* and left both spaces blank.

(37) *a nearly unanimous agreement
(38) *an almost unanimous agreement [3 informants]
(39) *a near unanimous agreement
(40) *an unanimous agreement [2 informants]
(41) *a unanimous agreement

For the 14 cases where an attributive adjective was provided, only 6 instances contained premodifiers which were also found in the *HSC*: *general agreement* (3 informants), *total agreement* (1 informant), *broad agreement* (1 informant) and *complete agreement* (1 informant). Besides, data reveal that very few informants (8) provided an equivalent of the adverb *casi* premodifying the adjective *unánime*: 6 participants chose the appropriate equivalent *almost* whereas the other two produced *near*, which might have been influenced by their likely confusion between the adverbs *near* and *nearly*.

Finally, it seems worth mentioning that four informants produced free translations that do not resemble the Spanish structure:

(42) Agreement between both methods was almost unanimous[104].
(43) Both methods showed almost the same result.
(44) Two methods led to almost the same results.
(45) The two methods reached an (?) in practice unanimous agreement.

Except for examples (42) and (45), which show structures which were not found in any of the reference corpora (i.e. **in practice*, **the agreement was almost*

---

[104] The *BoE* includes only one occurrence of the predicative adjective *unanimous* being referred to the noun *agreement*: "*the Saudi minister said, adding that the agreement was unanimous*" [guard/uk].)

unanimous[105]), the other two examples (43 and 44) constitute appropriate expressions from a collocational point of view. What could be argued, however, is whether the free translations provided are faithful to the original sentence and to what extent they convey the same meaning as the *HSC* instance. The inclusion of the noun phrase *results* incorporates a new information topic not mentioned in the original sentence and, as a consequence, suggests that informants were interpreting, rather than merely translating the Spanish construction. Although these two examples should then be labelled as inappropriate equivalents, they undoubtedly show that informants tried to avoid the difficulty of translating the impersonal construction with the reflexive pronoun *se* into English.

A number of other deviations in lexical bundles could also be traced in sentence 3, where informants' translation equivalents of the Spanish expression *en resumen* were examined.

Sentence 3: *En resumen, los resultados de nuestros estudios sugieren que la vaginitis bacteriana no afecta en absoluto a la concepción*

HSC example: In conclusion, findings from our studies suggest that bacterial vaginosis has no influence on conception.

All informants considered the Spanish construction *en resumen* as a clear distinct unit understood as a whole and separated from the rest of the text by means of a comma. Unlike what was observed in the analysis of informants' translation of the lexical bundle "*there* + *be* + (adjective) *agreement*", the string *en resumen* seems to have been interpreted as an extended unit of meaning which serves the function of a discourse marker, i.e. introducing the summary of what has already been discussed, and, consequently, all translations, but for three ungrammatical cases: **as summary*, **in resume* and **summing*, referred to frequently used formulae when expressing summaries and overt evaluations in academic writing. Some of these expressions, however, were not found in the *HSC* corpus, as Table 26 shows:

---

[105] As the native corpora show, the adjective *unanimous* was not used predicatively in combination with the noun *agreement*, which indicates that the informant lacks knowledge of the collocational patterning of the noun under study. However, the structure subject + verb + complements (SVC) seems to be another strategy used to deal with the problem posed by the reflexive pronoun "*se*" in the Spanish sentence.

HSC		**Informants' data**
(97 occurrences)	In summary,	(11 occurrences)
(65 occurrences)	In conclusion,	(6 occurrences)
(9 occurrences)	To summarize,	(1 occurrence)
		*As summary, (1 occurrence)
		To sum up, (1 occurrence)
		*In resume, (1 occurrence)
		To end, (1 occurrence)
		*Summing, (1 occurrence)
		*Summarising, (1 occurrence)

Table 26 Lexical bundles used to introduce conclusions and summaries in the *HSC* and the informants' data

Figures above indicate that 18 out of 24 informants produced the most commonly used combinations in the native corpus to anticipate that the discourse was coming to an end; that is, *in summary* (11 informants), *in conclusion* (6 informants) and *to summarize* (1 informant). Some other translations provided should be either considered unlikely to be found in medical discourse (e.g. *to sum up*, *to end*) or clearly influenced by the informants' L1 (e.g. **summing*, **summarising*, **in resume* and *as summary*).

While the confusion between *in resume* and *in summary* must partly be due to the fact that informants might have problems in distinguishing between the English *resume* and the Spanish *resumen*, the case of *as summary* is a clear example of transfer of structures from *como* and *a modo de resumen*. Likewise, *summing* and *summarising*, each of which was produced by one informant, might have been influenced by the colloquial forms "*resumiendo, en resumidas cuentas*".

It can be concluded that although most informants were successful in producing appropriate equivalents for the lexical bundle *en resumen*, some other deviant translations give account of their transfer of structures from their L1.

b) Informants’ use of the pattern restricted verb + *comparison*

The analysis of sentences 2 and 6 (see below) was aimed at finding out whether the use of restricted verbs in combination with the abstract noun *comparison* was problematic for the group of informants under investigation. Before a more in-depth examination of the combinatorial patterns provided, it must be noted that the most striking feature revealed by the observation of both sentences refers to the fact that a high percentage of participants (30,4% in sentence 2 and 41,7% in sentence 6) produced inappropriate word forms of the lexical item *comparison*. This was not to be expected since that noun was common in scientific procedures and they had already been using it in the previous exercises. Thus, they were supposed to be familiar with its spelling. The following deviant forms, however, seem to support the assumption that when producing the word themselves some participants tended to use word forms that resembled the Spanish spelling:

Sentence 2	Sentence 6
*comparations (5 informants)	*comparation/s (7 informants)
*comparitions (1 informant)	*comparisions (2 informants)
*comparisions (1 informant)	*comparition (1 informant)

Table 27 Deviant formal realisations of the noun *comparison* provided by informants in Sentences 2 and 6

Sentence 2: *No se realizaron comparaciones para determinar diferencias de género ni de edad.*

HSC example: Comparisons were not performed to determine gender or age differences.

As evidenced by informants’ data, non-native speakers seemed to be aware of the most common restricted verb collocates associated with the abstract noun *comparison* in the *HSC*. The verb *make*, and to a lesser extent, *do* were produced by 21 informants. These are some examples:

(46) *Comparations to determine gender and age differences weren’t done.

(47) No comparisons were made in order to determine age and gender differences.

(48) Comparisons to determine sex or age differences were not made.

(49) No comparisons were done about any differences of sex.

Regarding deviant uses of verbs collocating with *comparison*, there were only two informants who produced inappropriate word combinations with the verbs *realise* and *be*:

(50) *Did not realise comparisons to decide differences in gender and age.

(51) There were no comparisons to determine possible gender and age differences.

In addition, a high percentage of informants (83,3%) expressed these verb-noun collocations by means of a passive construction, which was also typically observed in the native corpus (cf. section 4.5). One of these informants, however, had problems with the inflection of the "*be* passive" and produced a deviant instance:

(52) Comparisons to analyse differences of gender and age *didn't be made.

Among those who did not express the restricted verb-noun collocation in the passive form (12,5%; 3 participants), it should be highlighted that, as illustrated in example (51) above, one informant made use of the existential *there* following the copular verb *be*, while the other two participants failed to provide grammatical instances since their translations either lacked a syntactic subject (see example 51) or appeared to be word-for-word translations from their L1 (e.g. "*comparisons didn't make to determine some differences nor gender neither age*").

With reference to the second half of the sentence introducing the purpose of the *comparisons made*; that is, "*para determinar diferencias de género ni de edad*" / "*to determine gender or age differences*", two main points must be underlined. On the one hand, although most informants (21) proved to be familiar with the subordinator *to* (18) / *in order to* (3) introducing the purpose adjunct[106] that followed the noun *comparison* (examples (53-55) below), there were also two deviant instances where the purpose adjunct turned into a prepositional phrase introduced by the preposition *about* (see example (56)) or an "*–ing* clause" (see example (57)):

---

[106] Following Huddleston's & Pullum's (2005) terminology.

(53) Comparisons were not made to determine differences between sexes and ages.
(54) No comparisons to detect differences in gender *nor age were done.
(55) No comparisons were made in order to determine age and gender differences.
(56) No comparisons were done about any differences of sex.
(57) Comparisons analyzing gender and age differences were not made.

On the other hand, an area that was of particular difficulty concerns the free verbs provided in combination with the noun *differences*. Although these deviant uses go beyond the scope of the present study, it should be at least mentioned that most free verb-noun collocations with the noun *differences* expressed by informants (e.g. *assess differences*, *decide differences*, *reach differences*, *probe differences*) were not found in the *HSC* corpus.

Sentence 6: *Se hicieron las mismas comparaciones en los tratamientos de control.*
HSC example: The same comparisons were made in the control treatments.

As stated in the analysis of informants' translations in sentence 2, deviations in the restricted verb produced were hardly found. 16 informants produced the collocation *make comparisons*, 6 used the support verb *do* (i.e. *do comparisons*) and one informant provided the verb *perform* as a translation equivalent of the Spanish *hacer* (see examples (58-60) below). The only deviant instance refers to a free translation where some misunderstanding of the original is noticeable (example (61)):

(58) The same comparisons were made in the control treatments.
(59) Same comparisons were done in control treatment.
(60) Similar comparisons were performed between control treatments.
(61) In control groups the *comparations were the same.

The most striking case of deviation concerning the pattern "verb + *comparison*" in this sentence is the difficulty encountered by some informants (5) when translating the impersonal reflexive pronoun "*se*" into English. Unlike sentence 1, where the overall numbers suggested that this was a particularly difficult area for informants, the percentage of deviant translations of the Spanish *se* in Sentence 6 is much lower. While most participants made use of a passive construction, in which the noun *comparisons* was the subject, so as to express the impersonality conveyed in Spanish by means of an impersonal pronoun (examples (62-63)), five informants produced syntactically (i.e. problems with notional subject-verb agreement, e.g. "*it was made *comparations*") and/or morphologically deviant (e.g. **comparation*, **comparations*) translation equivalents:

(62) In control treatments the same comparisons were made.
(63) The same comparisons were done in the control treatments.
(64) *They made the same *comparation in the control treatment.
(65) *It was made the same comparisons in the control treatments.
(66) *It was made similar *comparations on control treatments.
(67) *They did the same comparisons in the control treatments.
(68) *They made the same *comparations in the control treatments.

In examples (64), (67) and (68) above an unspecified personal pronoun (i.e. *they*) was used as the subject of *comparisons*, whereas in examples (65) and (66) an impersonal *it* was produced so as to fill the same subject position. These two deviant constructions emphasise the fact that, as noted by Neff van Aertselaer (2008), the translation of Spanish *se* structures into English passive sentences poses certain problems for non-native speakers and, thus, some transfer of structures from their L1 becomes evident.

c) Informants' use of the verb *contribute* vs. the noun *contribution*

Since the restricted verb + noun collocations analysed in the present investigation corresponded to a morphologically related verb (i.e. *reach an agreement* ~ *agree*; *make a comparison* ~ *compare*; *draw a conclusion* ~ *conclude*; *make a contribution* ~ *contribution*; *make a decision* ~ *decide*), it seemed worth investigating whether the collocational patterning of those cognate verbs would also be a difficult area for non-native speakers. To this respect, sentence 4 attempted to inform about participants' command of the lexical verb *contribute* in combination with the adverb *significantly*, which had proven to be a common collocation in the *HSC*. Similarly, the word string *significant contribution* was also frequently found in the native corpus.

Sentence 4: *Las sustancias analizadas han contribuido de forma poco significativa a disminuir los síntomas de la enfermedad.*
HSC example: The analysed substances have not contributed significantly to lessen the disease symptoms.

While 17 informants produced the verb *contribute* as the translation equivalent of the Spanish *contribuir*, 5 participants made use of a periphrastic structure of the type "restricted verb + *contribution*" and the other two informants provided inappropriate translations: one, with the ungrammatical adjective **significative* (e.g. "*the contribution of analysed substances is marginally *significative*") and the other, with a free verb (i.e. *affect*) which changed the meaning of the original (e.g. "*the analyzed substances have not significantly affected the symptoms of the disease*").

On the one hand, in those five instances where a restricted verb-noun collocation was produced, *make* stands out as the most commonly used (4 occurrences), which goes in line with the observations made in the *HSC* corpus (cf. section 4.6), whereas the deviant verb collocate, *have*, was only produced once (cf. example (73)):

(69) […] analyzed substances have made a poor significant contribution to […]

(70) The substances *analized have made a little *significative contribution to diminish […]

(71) The analized substances have made a little significant contribution in decreasing the symptoms of the disease.

(72) The *analized *substans have made a contribution in a short *significative way to decrease the illness *symthoms.

(73) The analyzed substances *have a little contribution to diminish the symptoms of the disease.

On the other hand, the seventeen instances where the verb *contribute* was provided show that informants had no difficulty in identifying the English equivalent of the Spanish dummy preposition "*a*" since all their translations contained the preposition "*to*" following the lexical verb (i.e. "*contribuir a*" vs. "*contribute to*"). Consequently, data suggest that most informants seemed to be aware of both the usual support verb in combination with the abstract noun *contribution* as well as the dummy preposition following its cognate verb *contribute*.

What seems to be particularly difficult in this sentence, however, is the translation of the Spanish phrase "*de forma poco significativa*". Figures are revealing: only three informants produced either the appropriate adverb equivalent in English, i.e. *significantly*, in the structure "*contribute significantly to*" (2 informants) or the adjective *significant* premodifying the noun *contribution* in verb-noun constructions. The other twenty-one examples are clear attempts at providing literal translations of the Spanish structure and also give account of the fact that the most problematic element in the string "*de forma poco significativa*" appeared to be the Spanish adverb *poco*.

As a further indication of informants' inclination to word-for-word translations, it seems worth noting that only 5 of their productions were expressed in the negative form of the verb *contribute*, as was the case in the *HSC* example:

(74) The tested substances have not contributed to diminish […]
(75) The *analized substances haven't contributed in a significant way to decrease […]
(76) The tested substances have not contributed significantly to relief […]
(77) The substances studied don't contribute to decrease […]
(78) The analyzed substances have not significantly affected the symptoms of the disease.

This finding is especially relevant because the original sentence in Spanish was declarative and, in fact, it was the prepositional phrase "*de forma poco* *significativa*” what provided the sense of “lack of contribution of the analysed substances”.

On the contrary, 13 informants produced the verb *contribute* in the affirmative form and, as a consequence, found it very difficult to translate the word string “*de forma poco significativa*” in English. Rather than producing the verb in its negative form in combination with the adverb *significantly* (i.e. “*[…] have not contributed significantly to […]*”), informants seem to have referred back to their L1 by providing a literal translation of the original.

As can be seen in the examples below, such literal translations often consisted of deviant structures which made it evident not only informants’ unawareness of the morphological form of the adjective *significant*, as well as its adverb counterpart, *significantly*, (see examples (79-83)) but also the problems posed by the adverb *poco*:

(79) Analysed substances have contributed in a few *significative way to diminish disease symptoms.
(80) The *analized substances have contributed to minimize the illness symptoms in a little *significative way.
(81) Analysed substances have contributed not much *significant to decrease the symptoms.
(82) The *analized *substans have made a contribution in a short *significative way to decrease the illness symthoms.
(83) The substances analized have made a little *significative contribution to diminish the disease symptoms.
(84) Analyzed substances have made a *poor significant contribution to decrease disease symptoms.
(85) Substances analyzed have *poor significantly contributed to reduce disease symptoms.
(86) The *analized substances haven’t contributed in a significant way to decrease the symptoms of the disease.
(87) Analyzed substances have contributed a little to decrease the symptoms of the illness.

(88) Analysed substances have *contribute to decrease disease *synthons ?in a limited way.

(89) The analyzed ?compounds have ?poorly contributed to lessen the symptoms of the disease.

(90) The *analised substances had contributed in ?poorly significant form to decrease the illness symptoms.

(91) The substances analyzed have contributed only slightly to reduce the symptoms of the illness.

(92) The analyzed substances have contributed in a not very significant way to diminish the illness symptoms.

The translation equivalents of the Spanish adverb *poco* provided by informants in their literal translations deserve further attention. Data from both the *HSC* and the *BNC* had revealed that the adjectives *significant*, *substantial*, *small*, *large*[107] and *little*, as well as their corresponding adverbs, i.e. *significantly*, *substantially*, *largely* and *little*, were usual collocates of the noun *contribution* and its cognate verb *contribute*, respectively. No instance of *short*, *poor/poorly* or *a few* could be traced in the native corpora. However, examples (79), (84), (85), (89) and (90) above suggest that informants failed to distinguish between *short* / *small* / *little* and *a few* and might have perceived them all as equivalents of the Spanish node word *poco*. Likewise, the deviant use of *poor/poorly* might have also been regarded as a synonym of *little* (*poco*, in Spanish).

Looking at the grammatical use that emerges from the non-native translations of sentence 4, it seems plausible that informants are not aware of the different collocational patternings of English quantifiers, such as *a few*, *a little* and *not much*. In addition, most of their morphologically (e.g. **significative*) and semantically (e.g. **in a short significative way*, **in poorly significant form*) deviant instances were highly influenced by their L1 and, thus, revealed their lacking collocational competence. Further examples of this tendency can also be traced in the following analysis.

---

[107] Unlike the rest of the examples in this list, the adjective *large* and its morphologically related adverb, i.e. *largely*, could only be found in the *BNC* in combination with either the noun *contribution* or the verb *contribute*.

d) Informants' command of the patterns associated with the noun *conclusion*

In order to get an overview of the main difficulties informants encountered when dealing with the phraseological patterns associated with abstract nouns in medical English, Exercise 3 contained two examples of the abstract noun *conclusion* in combination with restricted and free verb collocates (sentences 5 and 7, respectively). Besides, the abstract noun in both sentences was premodified by an adjective, which provided very useful information regarding informants' knowledge of the most common adjective collocates of the noun *conclusion*.

Sentence 5: *Antes de que se pueda llegar a conclusiones fiables, los investigadores deben aprender mucho más acerca de la función genética.*

HSC example: Before firm conclusions can be made, researchers must learn much more about genetic function.

The number of restricted verb collocates produced as equivalents of the Spanish "*llegar a*" in combination with the noun *conclusion* proved to be higher than the support verbs found in the *HSC* corpus. See the following Table:

HSC	**Informants' data**
draw	reach (14 informants)
lead to	obtain (3 informants)
reach	make (2 informants)
make	lead to (1 informant)
arrive at	draw (1 informant)
come to	have (1 informant)
	get (1 informant)
	achieve (1 informant)

Table 28 Restricted verb collocates + *conclusion* in the *HSC* and the informants' data (Sentence 5)

As can be inferred from the data in Table 28, the most frequently used verb was *reach* (14 informants), which was also very common in the *HSC*, and to a much lesser extent, other verbs like *make* (1 informant), *lead to* (1 informant) and *draw* (1 informant) were also produced by informants. However, some deviations in the verb could also be traced in examples like the following:

(93) Before *to obtain reliable conclusions, the investigators must learn much more about genetic function.
(94) Before having better evidence, researchers have to learn much more about genetic functions.
(95) Before *to get reliable conclusions, the researchers must learn much more about the genetic function
(96) Before *to achieve reliable conclusions, researchers must learn more about genetic function.

In addition to the question of what kind of verbs were most commonly chosen by informants, it was also worth investigating whether they were aware of the type of constructions those verbs typically occurred in. Data from the *HSC* corpus (cf. section 4.3) had shown that whereas the verbs *draw* and *make* tended to occur in the passive, the verb *reach* when combining with the noun *conclusion* was frequently used in active constructions in which the phrase performing the role of subject was referred to demonstrable and evident data (e.g. *studies have reached conflicting conclusions*, *previously published studies could not have reached this conclusion*) rather than to animate volitional entities.

Bearing in mind that 14 out of 24 informants chose the collocation *reach a conclusion* in their translations of "*llegar a una conclusión*", the fact that the active voice was very frequently used resembles the *HSC* findings. However, a closer look at the examples reveals that some deviations also took place:

(97) Before we can reach firm conclusions, researchers must learn much more about genetic function.
(98) Before *it could reach reliable conclusions, the research team must improve *his knowledge about the genetic function.
(99) Before we can draw reliable conclusions, investigators must learn much more about genetic function.

(100) Before they obtain valid conclusions, the investigators must learn much more about genetic function.

Probably due to the fact that the Spanish sentence contained an impersonal *se* construction, informants found it difficult to provide a suitable subject in their English translations and, thus, certain pronouns like *we*, *it* and *they* (see examples (97-100)) were used to refer anaphorically to "*researchers, investigators* or the *research team*". As stated above, this type of personal subjects was hardly found in the native corpus since the use of passive constructions (e.g. "*conclusions were made/ drawn*") served to avoid expressing the animate agent of the action. Likewise, the active structures in which the verb *reach* typically occurred did not include personal pronouns occupying the subject position in the *HSC*. Thus, examples like (97), (99) and (100) above were not characteristic of the medical discourse.

Other informants (15) managed to avoid the problem caused by the Spanish reflexive pronoun *se* by producing a non-finite clause following the preposition *before*. These are some examples:

(101) Before reaching reliable conclusions investigators must learn more about the genetic function.
(102) Before *to obtain reliable conclusions, the investigators must learn much about genetic function.
(103) Before *reach a reliable conclusion, researchers must learn much more about genetic function.
(104) Before having better evidence, researchers have to learn more about genetic function.
(105) Before *to achieve reliable conclusions, researchers must learn more about genetic function.

The use of a non-finite clause following the preposition *before* illustrated in the above examples might have been perceived as a safe option since it allowed no explicit specification of the subject. However, some expressions like "*before* + *to obtain /reach/ to achieve*" made it evident informants' lacking awareness of the expected verb form following a preposition in English. A total of 8 participants produced deviant examples with both the "*to*-"and the bare infinitive, where a gerund would have been expected.

Regarding the use of the passive, it must be stressed that it was hardly used in this sentence. Only 4 informants provided a passive construction to translate the Spanish "*antes de que se pueda llegar a conclusiones fiables*". Interestingly, among these very few uses of the passive, the verb *reach* was again the most frequently selected. The following examples illustrate such a preference:

> (106) Before reliable conclusions can be reached, the researchers must learn more about the genetic function.
> (107) Before reliable conclusions could be reached, investigators should learn more about the genetic function.
> (108) The researchers have to learn much more about genetic function, before reliable conclusions can be reached.

This overt use of the verb *reach* to the detriment of other possible restricted verb collocates like *draw*, *make* and *lead to* appeared to be influenced by the fact that informants might perceive the literal meaning of *reach* as closer to the Spanish "*llegar a*" and, thus, the former would be seen as a more accurate translation equivalent. Again, this further indicates their preference for literal translations and underlines the difficulties that the use of verbs in figurative senses entails.

Another final issue to be looked at refers to the informants' translation equivalents of the adjective "*fiables*" modifying the noun *conclusion*. Out of 22 instances of evaluative adjectives (i.e. *reliable*, *firm*, *valid* and **verificable*[108]) preceding the node word, 19 included the adjective *reliable* whereas the other three adjectives were only produced once. These findings do not resemble what was observed in the native corpus. The *HSC* did not show any occurrence of the collocation *reliable conclusions* and, by comparison, only two instances of this word combination were traced in the *BNC*. On the contrary, the most likely equivalent of "*conclusiones fiables*", that is, "*firm conclusions*" was only expressed by one informant:

---

[108] Misspelling of the adjective *verifiable*.

> (109) Before we can reach firm conclusions, researchers must learn much more about genetic function.

The present data thus suggest that informants' collocational performance with respect to the most usual premodifiers of the noun *conclusion* show little correspondence with native-like written production.

Sentence 7: *Estas conclusiones biológicas han sido justificadas por múltiples tipos de análisis estadísticos.*
HSC example: These biological conclusions are justified by multiple types of statistical analyses.

The most common deviation observed in this sentence that sticks out as being especially difficult for informants is related to morphological issues. Interestingly, seven participants produced an inappropriate word form for the adjective *biological* (i.e. **biologyc*). Unlike the evaluative adjective *firm* discussed in sentence 5, the topical adjective *biological* in the current example was not expected to pose problems for the non-native speakers in this study as they were supposed to be familiar with such a classifying adjective, so widely spread as a general term among the medical community.

In the same line, morphological deviations were also traced in the most commonly used free verb collocate selected by sixteen informants; that is, the verb *justify*. Although participants did not seem to have problems in providing an appropriate equivalent of the Spanish word string "*justificar una conclusión*" (i.e. *justify a conclusion*), partly because both the English and the Spanish verbs are morphologically related, a total of five informants produced ungrammatical inflected forms of the passive participle of the verb *justify* (i.e. **justifiqued*, **justificated* and **justifyed*):

> (110) These conclusions have been *justificated by multiple types of statistical analysis.

(111) These biological conclusions have been *justifyed by a lot of kinds of statistical analysis.
(112) *This biological conclusions have been *justificated by several kinds of *stadistical analysis.
(113) *This biological conclusions have been *justifiqued for a lot of different *stadistics ways.
(114) These biological conclusions were *justifiqued by different statistical analysis.

Other free verb collocates expressed by informants were *support* (5 instances) and *verify* (1 instance):

(115) These biological conclusions have been supported by multiple types of *stadistical analysis.
(116) The *biologyc conclusions have been *verificated[109] by a lot of types of statistical analysis.

Comparing the constructions produced by informants against the original *HSC* example, it should also be pointed out that a high percentage (91,7%) of instances were of both the Spanish sentence and the *HSC* instance were expressed in the passive form. Only two informants provided free translations in the active voice by changing the focus of attention from the *conclusions* to their causer (i.e. *statistical analyses*):

(117) *To many statistical analysis led to these biological conclusions.
(118) Multiple statistical analyses support these biological conclusions.

All in all, as the examples above show, the free combination examined in this sentence (i.e. *justify a conclusion*) was by no means a particularly problematic area for non-native speakers since both the *HSC* example and the Spanish translation were very similar (i.e. the verb in both cases was morphologically related and both sentences were expressed in the passive). What seemed to be more complex, on the contrary, was the morphology associated not only with the passive participle of the lexical verb provided (i.e. *justified* and *verified*) but also with the adjective collocate (i.e. *biological*) of the noun *conclusion*.

---

[109] Notice that morphological problems with the word form *verified* were also present in this example.

e) Informants' knowledge of the compound *decision making*

As already discussed in chapter 4, the abstract noun *decision* tended to occur in combination with some other nouns (e.g. *decision rules*, *decision point*, *decision model*, *decision making*), giving rise to the creation of compounds. Therefore, it seemed worth investigating how informants would deal with the production of those structures. The compound noun selected for this analysis was *decision making*.

Sentence 8: *Varias organizaciones son las responsables de la toma de decisiones por lo que respecta a la seguridad de las transfusiones.*
HSC example: Several organizations are responsible for decision making in transfusion safety.

Given the fact that the translation equivalent of the endocentric compound *decision making* consisted of a periphrastic structure of the type "noun *de* noun combination" (i.e. "*toma de decisiones*"), rather than a compound lexeme made of two nouns, it was especially interesting to examine the expressions produced by informants.

Of the 24 instances, only 6 included the compound *decision making* as the translation equivalent of the Spanish "*toma de decisiones*", whereas sixteen[110] informants made use of other types of constructions, such as "*-ing*" clauses or simple nouns. These are some examples:

(119) Several *organitations are responsible *of making of decisions about security in transfusions.

(120) Several organizations are responsible for making decisions about transfusions security.

(121) Some *boureaus are *responsibles *of the decisions taken about transfusion safety.

---

[110] The translations provided by the other two participants were completely free and included some information (underlined in the following examples) which was not explicit in the original: "**transfussions* *follow strict protocols accordingly to* *some organizations*" and "*blood transfusion security* *is surveyed* *by some organizations*".

(122) Various organisations are responsible for taking the decision with respect to the safety of transfusion.

(123) Several organizations are responsible *in taking decisions about transfusion security.

(124) Some organizations are *responsable for making decisions regarding the security of transfusions.

The fact that most of the expressions created were periphrastic structures (i.e. *taking decisions about*, *making decisions with respect to*, *making decisions about*) instead of a compound (i.e. *decision making*), as it would have been more appropriate, suggests not only the informants' unawareness of the English compound, but also their general tendency to provide word-for-word translations. A clear example of this can be observed in example (119), where the word string "*making of decisions*" is in sharp correlation with the Spanish expression "*toma de decisiones*".

With respect to the support verbs being used in the informants' constructions, it should be highlighted that the restricted verb collocates *make* and *take* stand out as the most commonly used (5 and 6 occurrences, respectively), as was the case with the verb-*decision* collocations observed in the *HSC.* There were only two instances of deviant verb-noun collocations in informants' translations:

(125) *Very organisations are the *responsables *of the doing decisions about transfusions security.

(126) Regarding *to transfusions safety. Several organizations are responsible for *to adopt decisions.

Example (125) contains an inappropriate use of the support verb *do*, which was not traced in combination with *decision* in the *HSC*, whereas in example (126) a deviant use of the free verb collocate *adopt* can be observed.

From the analysis of informants' periphrastic constructions, another aspect that also emerges as being particularly susceptible to deviation is their translations of the lexical bundle "*por lo que respecta a*". As noted in section 4.7, the noun *decision* was hardly ever postmodified[111] by a prepositional phrase supplying information about the content of the *decision made*. However, in all informants' periphrastic structures (see examples (119-124)), the noun *decision* was postmodified.

Unlike the *HSC* where the only prepositions following the noun *decision* were *about* and *on*, the translation equivalents suggested by informants included a wider range of prepositions. Examples with the number of occurrences in square brackets are:

(127) Several organizations are responsible for making decisions about [9] safety of transfusions security.
(128) Several organisations are responsible for decision making with respect to [7] transfusion security.
(129) Several organizations are responsible for taking decisions regarding [2] blood transfusions safety.
(130) Several organizations are responsible of decision-making concerning [1] security of transfusions.

As can be seen in the examples above, a total of 10 informants made use of prepositions that had not been found in combination with the noun *decision* in the *HSC* (e.g. *with respect to*, *regarding* and *concerning*). However, they appear to refer literally to the Spanish "*por lo que respecta a* ". This ties in well with the earlier observation that informants showed a stronger preference for literal translations.

To sum up, the findings concerning informants' command of the compound *decision making* in sentence 8 suggest that most of their expressions are influenced by their L1 and thus constitute a further group of deviant collocations. In the following pages, participants' lexico-grammatical competence with regard to adjective-abstract noun and verb-abstract noun patterns will be discussed in the light of the results obtained in Exercises 1 and 4, respectively.

---

[111] In the *HSC* there were only 6 instances of this noun postmodified by a prepositional phrase [*about* / *on*]

5.1.3 Matching Exercise (adjective + abstract noun)

Exercise 1. Match each of the adjectives on the left with a suitable noun from the facing column. Try to use your intuition. Please notice that more than one combination is possible and some of the adjectives can be left out.[112]

logical, clear, firm, contradictory, valid, little reliable, short, late, statistical, far-reaching substantial, general, significant, large, small fashionable, broad, unforeseeable	+ conclusion/ agreement/ comparison/ contribution/decision

In Exercise 1 informants were asked to associate nineteen adjectives with the five abstract nouns analysed in this study (i.e. *agreement*, *comparison*, *conclusion*, *contribution* and *decision*). It was clearly stated in the instructions of the exercise that several combinations were possible; that is, each noun from the list provided could be matched with more than one adjective. Equally, they were told that some adjectives could be left out.

The main goal of this exercise was thus to analyse informants' intuition and performance with respect to certain adjectives collocating with a selection of abstract nouns, on the one hand, and comparing their associations against the actual collocations found in the *HSC*, on the other. Some conclusions concerning the participants' command of the adjective collocates in combination with abstract nouns will be drawn from such a comparison.

As pointed out in chapter 3, the selection of the nineteen adjectives discussed here (see Appendix 4) was determined by the fact that, with the only exception of four of them (i.e. *contradictory*, *fashionable*, *short* and *unforeseeable*), the rest had been found to be frequent collocates of at least one of the abstract nouns under study. The important point to make regarding the other four adjectives included is that none of them could be found in the corpora used for comparison (i.e. the *HSC* and the *BNC*). The focus of this exercise was then analysing how proficient informants were in dealing with these adjective collocates.

---

[112] See Appendix 4.

On the contrary, Exercise 1 was not expected to provide information about the participants' expertise in adjective order in English, which means they were not supposed to specify whether the adjectives provided tended to appear in either attributive or predicative positions. Given the layout of the exercise (cf. Appendix 4)[113], though, it is likely that they may have interpreted the clause pattern **adjective + noun** (e.g. "there was *general* agreement"), that is, the adjective appearing in an attributive position, rather than the predicative structure **Noun Phrase + copular Verb + Adjective Phrase** (e.g. "the contribution was *substantial*").

It had already been noted in chapter 4 that both **descriptors** and **classifiers** could be found adding some features to abstract nouns. However, the latter were more prone to collocate with the following nouns: *comparison*, *contribution* and *decision*. The abstract nouns *agreement* and *conclusion*, on the contrary, showed a stronger preference for evaluative, extent and time descriptors as premodifiers. As previously discussed in section 4.8, classifiers are seen to be more predictable in a medical register since academic writing in empirical research is chiefly focused on demonstrable data rather than on personal evaluations and opinions, expressed by means of qualitative descriptors, which are more typical of literary writing (Swales 1990; Halliday & Martin 1993; Biber et al. 1998/1999; Gledhill 2000; Hyland 2008).

As a consequence, doctors were expected to be more used to identifying relational and topical adjectives (**classifiers**) with abstract nouns (e.g. *statistical comparison*, *cellular contribution*, *clinical decision*) since this kind of collocates were highly frequent in medical writing. However, what would their performance be like when dealing with descriptors? The focus here lies on analysing how they would deal with evaluative and size adjectives (**descriptors**), less frequently found in the *HSC*, in combination with the abstract nouns under discussion. The following Table shows Biber's et al. (1999) classification of the adjectives selected for Exercise 1:

---

[113] Notice informants were required to produce adjective-noun collocations in isolation and not as part of a given context.

DESCRIPTORS				CLASSIFIERS	
EVALUATIVE	**SIZE**	**EXTENT**	**TIME**	**RELATIONAL**	**TOPICAL**
LOGICAL	LITTLE	GENERAL	LATE	FAR-REACHING	STATISTICAL
CLEAR	SHORT*	BROAD			
FIRM	LARGE				
CONTRADICTORY*	SMALL				
VALID					
RELIABLE					
SUBSTANTIAL					
SIGNIFICANT					
FASHIONABLE*					
UNFORESEABLE*					

Table 29 Semantic classification of the adjectives[114] selected in Exercise 1

Most adjectives from Exercise 1 were evaluative descriptors. The fact that their frequency of appearance in the *HSC* was fairly low in comparison with classifying adjectives made it very difficult to follow Halliday's (1961) & Sinclair's (1991) "statistical / textual approach"[115] to the measure of frequency as the sole indication of collocational significance. To put it another way, if judged by frequency of occurrence the adjective collocates used in this exercise would not be regarded as significant collocates of the *HSC*, since classifiers have proven to be much more frequent in combination with three (i.e. *comparison*, *contribution* and *decision*) of the five abstract nouns analysed here. The main goal of the present exercise, however, was not verifying that evaluative adjectives tended to co-occur with abstract nouns less often than classifiers in medical writing. On the contrary, the focus of attention was analysing whether the informants were able to identify the evaluative adjectives most frequently used in combination with the abstract nouns provided so as to show their knowledge of the limited number of opinion collocate adjectives found in the reference corpora.

Altogether, the results of the data analysis that follow are mainly quantitative; indicating in brackets the numerical findings of the frequency of occurrence of each **adjective + noun** collocation. The low percentage of occurrence of some evaluative adjectives with abstract nouns in the *HSC* (see Table 31) forced through some methodological decisions, as was already pointed out in chapter 3. Bearing in mind that very modest conclusions could be drawn from comparing the informants' collocates with the adjective collocates extracted from the *HSC*, it was decided to use the *British National Corpus* (*BNC*) as a third comparison measure. As a 100 million word collection of samples of written and spoken language from a wide range of sources, the *BNC* provided a higher context of co-occurrence and enabled to reach more significant conclusions as well.

---

[114] The asterisked adjectives were chosen at random and did not occur as collocates of any of the abstract nouns considered.

[115] Cf. section 2.3 for further discussion on this "statistical / textual approach" to the notion of *collocation*.

Prior to the comparison of results, several searches needed to be made. Firstly, it was necessary to establish what adjectives from the list collocated with each of the abstract nouns, both in the *BNC* and the *HSC*, so as to determine whether they were found to appear in combination and, if so, its frequency of occurrence. Secondly, the sample from the twenty-four informants was individually analysed. Tables 30, 31 and 32 below show a detailed account of the adjective collocates found in the *BNC*, the *HSC* and the informants' answers to Exercise 1, respectively.

AGREEMENT	COMPARISON	CONCLUSION	CONTRIBUTION	DECISION
GENERAL(268)	GENERAL(3)	GENERAL(52)	GENERAL(1)	GENERAL(1)
BROAD(49)	BROAD(2)	BROAD(2)		
LITTLE(20)	LITTLE(5)		LITTLE(8)	LITTLE(1)
CLEAR(11)		CLEAR(14)		CLEAR(20)
SUBSTANTIAL(7)			SUBSTANTIAL(54)	
FIRM(6)		FIRM(12)		FIRM(27)
SIGNIFICANT(4)	SIGNIFICANT(2)	SIGNIFICANT(2)	SIGNIFICANT(161)	SIGNIFICANT(4)
FAR-REACHING(2)				
VALID(1)	VALID(6)	VALID(2)	VALID(3)	VALID(1)
LARGE(1)			LARGE(12)	
	STATISTICAL(10)			STATISTICAL(1)
	LOGICAL(2)	LOGICAL(90)		LOGICAL(5)
	SMALL(1)		SMALL(15)	
		RELIABLE(2)		
			FASHIONABLE(1)	
				LATE(7)

Table 30 Adjective collocates in combination with *agreement* / *conclusion* / *comparison* / *contribution* / *decision* extracted from the *BNC*

AGREEMENT	COMPARISON	CONCLUSION	CONTRIBUTION	DECISION
GENERAL(12)		GENERAL(3)		
BROAD(1)	BROAD(1)			
LITTLE(1)			LITTLE(1)	
SIGNIFICANT(1)	SIGNIFICANT(1)		SIGNIFICANT(10)	
	RELIABLE(3)			
	VALID(5)			
	STATISTICAL(15)			
		FIRM(4)		FIRM(2)
		LOGICAL(1)		
			SUBSTANTIAL(3)	
			SMALL(2)	
				CLEAR(2)
				FAR-REACHING(1)

Table 31 Adjective collocates in combination with *agreement* / *conclusion* / *comparison* / *contribution* / *decision* found extracted from the *HSC*

AGREEMENT	COMPARISON	CONCLUSION	CONTRIBUTION	DECISION
GENERAL(12)	GENERAL(3)	GENERAL(9)	GENERAL(6)	GENERAL(3)
BROAD(8)	BROAD(1)	BROAD(5)	BROAD(7)	BROAD(3)
LITTLE(6)	LITTLE(2)	LITTLE(2)	LITTLE(7)	LATE(8)
VALID(6)	VALID(9)	VALID(13)	VALID(6)	VALID(6)
RELIABLE(6)	RELIABLE(2)	RELIABLE(10)	RELIABLE(3)	
FIRM(5)		FIRM(13)	FIRM(1)	FIRM(15)
SUBSTANTIAL(4)		SUBSTANTIAL(1)	SUBSTANTIAL(18)	
LATE(4)	LATE(1)	LATE(3)	LATE(3)	
UNFORESEEABLE(4)	UNFORESEEABLE(2)	UNFORESEEABLE(5)	UNFORESEEABLE(2)	UNFORESEEABLE(5)
CLEAR(3)	CLEAR(2)	CLEAR(12)	CLEAR(5)	CLEAR(5)
SIGNIFICANT(3)	SIGNIFICANT(4)	SIGNIFICANT(4)	SIGNIFICANT(13)	SIGNIFICANT(2)
FAR-REACHING(3)	FAR-REACHING(3)	FAR-REACHING(6)	FAR-REACHING(4)	FAR-REACHING(5)
LARGE(3)	LARGE(3)		LARGE(10)	
LOGICAL(3)	LOGICAL(7)	LOGICAL(16)	LOGICAL(2)	LOGICAL(12)
FASHIONABLE(1)	FASHIONABLE(2)			FASHIONABLE(3)
SHORT(1)	SHORT(3)	SHORT(2)	SHORT(6)	SHORT(1)
	SMALL(3)	SMALL(1)	SMALL(11)	
	STATISTICAL(21)	STATISTICAL(2)	STATISTICAL(1)	
	CONTRADICTORY(7)	CONTRADICTORY(12)	CONTRADICTORY(1)	CONTRADICTORY(7)

Table 32 Adjective collocates in combination with *agreement* / *conclusion* / *comparison* / *contribution* / *decision* found extracted from the informants' sample

The information displayed in the above Tables (30, 31 and 32) was so varied that comparisons among them seemed difficult to make. Therefore, a method of sampling on the basis of frequency of occurrence was adopted for practical reasons and an initial cut-off of the six highest frequency adjective collocates in the *BNC*, the *HSC* and the informants' sample was imposed. This gave rise to a more manageable number of collocates:

BNC	***HSC***	**Informants' responses to Ex.1**	
GENERAL(268)	GENERAL(12)	GENERAL(12)	**AGREEMENT**
BROAD(49)	BROAD(1)	BROAD(8)	
LITTLE(20)	LITTLE(1)	LITTLE(6)	
CLEAR(11)			
SUBSTANTIAL (7)			
FIRM (6)		FIRM (5)	
	SIGNIFICANT(1)		
		RELIABLE(6)	
		VALID(6)	
STATISTICAL(10)	STATISTICAL(15)	STATISTICAL(21)	**COMPARISON**
VALID(6)	VALID(5)	VALID(9)	
LITTLE(5)			
GENERAL(3)			
LOGICAL (2)		LOGICAL(7)	
SIGNIFICANT (2)	SIGNIFICANT(1)	SIGNIFICANT(4)	
	BROAD(1)		
	RELIABLE(3)		
		SHORT (3)	
		CONTRADICTORY(7)	
LOGICAL(90)	LOGICAL(1)	LOGICAL(16)	**CONCLUSION**
GENERAL(52)	GENERAL(3)		
CLEAR(14)		CLEAR(12)	
FIRM(12)	FIRM(4)	FIRM(13)	
VALID (2)		VALID(13)	
SIGNIFICANT (2)			
		CONTRADICTORY(12)	
		RELIABLE(10)	
SIGNIFICANT(161)	SIGNIFICANT(10)	SIGNIFICANT(13)	**CONTRIBUTION**
SUBSTANTIAL(54)	SUBSTANTIAL(10)	SUBSTANTIAL(18)	
SMALL(15)	SMALL(2)	SMALL(11)	
LARGE(12)		LARGE(10)	
LITTLE (8)	LITTLE(1)	LITTLE (7)	
VALID (3)		BROAD (7)	
FIRM(27)	FIRM(2)	FIRM(15)	**DECISION**
CLEAR(20)	CLEAR(2)		
LATE(7)		LATE(8)	
LOGICAL(5)		LOGICAL(12)	
SIGNIFICANT (4)			
VALID (1)		VALID (6)	
		CONTRADICTORY(7)	
	FAR-REACHING(1)	FAR-REACHING (5)	

Table 33 Highest frequency adjective collocates in the *BNC*, the *HSC* and the informants' sample

Table 33 above presents the six most frequent adjectives in combination with the abstract nouns under study in the *BNC*, the *HSC* and the informants' sample, with the totals for the number of occurrences added in brackets. Altogether, data reveal that similar adjectives are used in combination with the five abstract nouns across the three data sources. Overall, it appears that the latter identified *general*, *broad* and *little* as frequent collocates of *agreement*; *statistical* and *valid* as common collocates of *comparison*; *logical* in combination with *conclusion*; *significant*, *substantial*, *small* and *little* collocating with *contribution* and *firm* and *logical*, with *decision*. Thus, informants were in part successful in identifying those adjectives from the list that most frequently co-occur with the above mentioned abstract nouns in the native corpora.

In light of such results, it seems reasonable to state that these similarities between native speakers' performance (i.e. the *BNC* and the *HSC*) and non-native speakers' production should be interpreted as informants being aware of the typical adjective collocates in the medical genre. There are thus indications that in some collocations informants chose the adjectives most commonly found in the **adjective + noun pattern** (e.g. *general agreement*, *statistical comparison*, *valid comparison*, *logical conclusion*, *firm conclusion*, *significant contribution*, *firm decision*), so they do not seem to have difficulties in identifying those collocations that occur with greater frequency (cf. Table 33). However, such assertion should be taken very cautiously since a more careful look at the data displayed in Table 33 reveals that certain word combinations of the type adjective + *agreement* / *comparison* / *conclusion* / *contribution* / *decision* are likely to cause problems for the group of informants.

Although they have proven to successfully detect the most usual collocates, it is worth underlining that participants have also associated a wide range of adjectives with each of the nouns provided. In fact, there were very few adjectives which were not included[116] in their combinations. As Table 32 displays, 14 out of 19 adjectives were chosen by a minimum of three informants in combination with *agreement*, 15 adjectives were associated with *comparison*, *conclusion* and *contribution*, and 12 modifiers were used as collocates of the abstract noun *decision* by more than one participant.

---

[116] The following list shows those adjective-noun combinations that were not produced by any of the informants: *statistical/contradictory agreement*; *firm/substantial comparison*; *large/fashionable conclusion*; *fashionable contribution*; *substantial/small/large/little reliable/statistical decision.*

By comparing the above figures against data extracted from both the *BNC* (cf. Table 30) and the *HSC* (cf. Table 31), it can be seen that in both corpora the list of adjectives that collocate with the nouns provided is significantly smaller. As a consequence, close observations of the larger number of collocations produced by the non-native speakers, which often result in inappropriate word combinations, have revealed some inconsistencies in the informants' collocational competence, as far as the pattern **adjective + abstract noun** is concerned.

It must be remembered that Exercise 1 presented the various options in a decontextualized way; that is, informants were not required to establish either syntactic or semantic properties of the combinations found, which means they were not expected to be capable of setting the semantic differences, if any, between, for instance, *little* and *small* (*little contribution* vs. *small contribution*) or *large* and *broad* (*large agreement* vs. *broad agreement*). However, some of their choices provide information about their lacking knowledge of the above mentioned collocational patterns.

Focusing on those collocations not traced either in the *BNC* or in the *HSC* seemed to be more revealing than concentrating on their similarities. For instance, a striking feature regarding the use of **adjective + noun** collocations by non-native speakers is that some of the adjectives selected are very infrequent ones in the native corpora. The participants' sample demonstrates that the adjective *reliable* has been very frequently used in combination with *conclusion* (10)[117], *agreement* (6) and *contribution* (3). The same could be said of the time descriptor *late* collocating with *agreement* (4), *conclusion* (3) and *comparison* (1).

---

[117] The figures in brackets inform about the number of occurrences.

However, such collocations have not been found in either of the two native corpora used for comparison in the present study. In fact, the evaluative adjective *reliable* was only found 3 times in combination with *comparison* (*HSC*), whereas the descriptor *late* only occurred 7 times in combination with *decision* (*BNC*). Interestingly, neither *reliable comparison* nor *late decision* occurred in any of the twenty-four questionnaires. Needless to say that some of the collocations provided by informants (e.g. *reliable conclusion*, *late agreement*) could be labelled as acceptable in the sense that they are grammatically accurate and idiomatically correct, though, as both native corpora suggest, they constitute unusual combinations in English.

By contrast, there is additional evidence that informants tended to collocate these nouns with inappropriate adjectives as well. As already stated in chapter 3, four randomly selected adjectives were included in the study; namely, three opinion descriptors, *contradictory*, *fashionable* and *unforeseeable* and a size descriptor, *short*. None of these four adjectives could be identified as a collocate of any of the five abstract nouns considered in this investigation. On the one hand, it was expected that some informants would have difficulties in distinguishing between *little*, *small* and *short*, since they are similar in meaning and may have similar translations into Spanish (*poco* ~ *little*; *pequeño* ~ *little*, *small*, *short*). While the qualitative adjectives *little* and *small* occurred in the *HSC* in combination with *agreement* and *contribution* (e.g. *little agreement*, *little contribution*, *small contribution*), the descriptor *short* was not found collocating with any of the selected nouns.

The above mentioned initial expectation has proven to be true as the following data show a noticeable misunderstanding between *little* and *short* in the informants' production:

little agreement (6)[118] vs. **short agreement* (1)
little comparison (null) vs. **short comparison* (3)
little conclusion (2) vs. **short conclusion* (2)

---

[118] The figures in brackets indicate the number of informants who identified that collocation.

little contribution (null) vs. **short contribution* (6)
little decision (null) vs. *short decision* (1)

It was deemed then necessary to investigate the above deviation in the **adjective + noun pattern**, which was particularly frequent with the example *short contribution* (6 informants). As can be seen in the examples above, both *little* and *short* were found in combination with *agreement* and *conclusion* in the non-native speakers' production. The combinations **short agreement* and **short conclusion* must be regarded as deviated examples since they are not possible collocations in English. Similarly, the occurrences of **short comparison* (3) and **short contribution* (6) reveal that the distinction between the adjectives *little* and *short* poses some problems for the informants involved. This assumption is also supported by the fact that, as noted in chapter 3 (section 3.5), both adjectives are close in meaning and may be translated into participants' L1 (i.e. Spanish) by means of similar terms, such as *pequeño*, *poco* and *escaso*.

Inappropriate uses of the strings ***contradictory***, ***fashionable*** and/or ***unforeseeable*** **+ abstract noun** also occur in the informants' data. Such deviated combinations were not expected since the semantic properties of these three adjectives do not refer to demonstrable and quantifiable data or phenomena characteristic of the usual description of scientific processes and results. Therefore, it was unlikely that informants would produce combinations such as **unforeseeable conclusion*, **fashionable decision* and **contradictory comparison*, to name but a few. Noticeable is the wide range of deviant word combinations, such as **unforeseeable agreement*, **contradictory comparison*, and **fashionable decision*. Table 34 below charts the examples of **adjective + noun** collocates only found in the informants' questionnaires.

AGREEMENT	COMPARISON	CONCLUSION	CONTRIBUTION	DECISION
RELIABLE(6) *UNFORESEEABLE(4) LATE(4) LOGICAL(3) *SHORT(1) FASHIONABLE*(1)	*CONTRADICTORY(7) FAR-REACHING(3) LARGE(3) *SHORT(3) CLEAR(2) *UNFORESEEABLE(2) *FASHIONABLE(2) LATE(1)	*CONTRADICTORY(12) RELIABLE(10) FAR-REACHING(6) *UNFORESEEABLE(5) LATE(3) LITTLE(2) *SHORT(2) STATISTICAL(2) SUBSTANTIAL(1) SMALL(1)	*SHORT(6) CLEAR(5) FAR-REACHING(4) RELIABLE(3) LATE(3) LOGICAL(2) *UNFORESEEABLE(2) FIRM(1) *CONTRADICTORY(1) STATISTICAL(1)	*CONTRADICTORY(7) *UNFORSEEABLE(5) *FASHIONABLE(3) BROAD(3) SHORT*(1)

Table 34 Adjective collocates[119] only found in the informants' sample

Data from Exercise 1 demonstrate that informants on the whole were capable of identifying the highest frequency adjective collocates of the abstract nouns discussed in this research but, at the same time, they seemed to lack some collocational competence since they also produced some unexpected or even inappropriate word combinations. This fact serves to reinforce the idea that the informants' degree of unawareness of some typical adjectives in **adjective + noun** collocations may lead not only to the use of infrequent collocations but also to the production of inappropriate adjective collocates.

A second matching exercise was also included in the worksheet of exercises so as to test informants' performance with respect to the verb + abstract noun pattern. Like Exercise 1, this activity asked participants to combine lexical items, in this case a list of six light verbs with the nouns provided, which were not integrated as part of a text. Results obtained from their verb-noun collocations will be discussed below.

---

[119] The asterisked adjectives did not occur neither in the *BNC* nor in the *HSC* as collocates of any of the abstract nouns considered.

5.1.4 Matching Exercise (verb + abstract noun)

Exercise 4. *Make*, *draw*, *reach*, *lead to*, *take* and *do* are highly-frequent verbs in English. They tend to combine with a wide range of words to form fixed expressions. Combine the words below with the verbs provided in the chart.

E.g.: action, knowledge, a mistake (…) + make/draw/ reach/ lead to/ take/do (…)[120]

This exercise attempted to investigate with what nouns deviations in the verb in verb-noun collocations occurred. As already noted, informants were given a chart of six highly-frequent verbs in English (i.e. *make*, *draw*, *reach*, *lead to*, *take* and *do*) and were asked to associate them with a list of thirty-six nouns that had been found to collocate with at least one of the above verbs.

In order to analyse the vast amount of data (cf. Table 35) provided by informants, the repertoire of nouns was divided into three main groups:

a) Group A, consisting of the five abstract nouns under study (i.e. *conclusion*, *agreement*, *comparison*, *contribution* and *decision*).

b) Group B, involving a list of seventeen nouns from the *HSC* (i.e. *somebody's temperature*, *weight*, *a test*, *obesity*, *a tablet*, *research*, *action*, *damage*, *progress*, *attention*, *a suggestion*, *an effort*, *a photograph*, *changes*, *consensus*, *a mistake*, *significance*) which had been found in combination with at least one of the verbs in the exercise.

c) Group C, consisting of fourteen general nouns (i.e. *experience*, *a verdict*, *knowledge*, *a place*, *an exam*, *a guess*, *[financial] success*, *a profit*, *some writing*, *notes*, *a picture*, *skills* and *a goal*) which, despite being typical collocates of the list of verbs provided, had not been found in combination with any of those verbs in the *HSC* corpus.

The results obtained from the first two groups, Groups A and B, will be compared against the *HSC*, whereas informants' verb-noun collocations with the nouns from Group C will be compared against the *BNC*. Due to the fact that the nouns within this latter group were not typically distinctive of the medical discourse (and were not traced in the *HSC* in collocation with any of the verbs considered in this exercise either), it was necessary to use a general English corpus (i.e. the *BNC*) as a third comparison measure.

---

[120] Cf. Appendix 4.

All these comparisons are aimed at investigating the informants' collocational preferences concerning abstract nouns and other general nouns characteristic of the scientific discourse on the one hand, and general English nouns, on the other. The examination of such preferences will provide information with reference to the participants' production and command of the suggested verb-noun collocations.

	(n) MAKE	(%) MAKE	(n) DRAW	(%) DRAW	(n) REACH	(%) REACH	(n) LEAD TO	(%) LEAD TO	(n) TAKE	(%) TAKE	(n) DO	(%) DO
(financial) SUCCESS	3	12.5	0	0.0	6	25.0	12	50.0	0	0.0	0	0.0
A COMPARISON	17	70.8	1	4.2	0	0.0	1	4.2	0	0.0	4	16.7
A CONCLUSION	1	4.2	10	41.7	10	41.7	4	16.7	0	0.0	0	0.0
A CONTRIBUTION	15	62.5	0	0.0	0	0.0	3	12.5	1	4.2	3	12.5
A DECISION	10	41.7	1	4.2	6	25.0	3	12.5	7	29.2	0	0.0
A GOAL	2	8.3	1	4.2	14	58.3	3	12.5	0	0.0	1	4.2
A GUESS	7	29.2	3	12.5	2	8.3	0	0.0	6	25.0	1	4.2
A MISTAKE	19	79.2	0	0.0	0	0.0	4	16.7	1	4.2	2	8.3
A PHOTOGRAPH	1	4.2	0	0.0	0	0.0	0	0.0	23	95.8	0	0.0
A PICTURE	3	12.5	10	41.7	0	0.0	0	0.0	10	41.7	0	0.0
A PLACE	0	0.0	0	0.0	12	50.0	2	8.3	10	41.7	0	0.0
A PROFIT	12	50.0	1	4.2	1	4.2	3	12.5	1	4.2	0	0.0
A SPEECH	11	45.8	1	4.2	0	0.0	0	0.0	0	0.0	9	37.5
A SUGGESTION	17	70.8	0	0.0	0	0.0	1	4.2	0	0.0	4	16.7
A TABLET	1	4.2	2	8.3	0	0.0	1	4.2	17	70.8	1	4.2
A TEST	6	25.0	0	0.0	0	0.0	0	0.0	5	20.8	14	58.3
A VERDICT	2	8.3	4	16.7	11	45.8	3	12.5	2	8.3	0	0.0
ACTION	1	4.2	0	0.0	0	0.0	4	16.7	12	50.0	3	12.5
AN AGREEMENT	2	8.3	0	0.0	19	79.2	3	12.5	0	0.0	0	0.0
AN EFFORT	18	75.0	0	0.0	0	0.0	0	0.0	0	0.0	5	20.8
AN EXAM	4	16.7	0	0.0	0	0.0	0	0.0	6	25.0	14	58.3
ATTENTION	2	8.3	6	25.0	0	0.0	0	0.0	7	29.2	2	8.3
CHANGES	13	54.2	1	4.2	0	0.0	6	25.0	0	0.0	5	20.8
CONSENSUS	0	0.0	0	0.0	17	70.8	7	29.2	0	0.0	0	0.0
DAMAGE	7	29.2	1	4.2	0	0.0	7	29.2	0	0.0	13	54.2
EXPERIENCE	2	8.3	2	8.3	9	37.5	5	20.8	4	16.7	0	0.0
KNOWLEDGE	0	0.0	1	4.2	7	29.2	10	41.7	4	16.7	0	0.0
NOTES	0	0.0	3	12.5	0	0.0	0	0.0	21	87.5	0	0.0
OBESITY	0	0.0	0	0.0	2	8.3	18	75.0	1	4.2	0	0.0
PROGRESS	10	41.7	0	0.0	1	4.2	7	29.2	0	0.0	7	29.2
RESEARCH	3	12.5	1	4.2	0	0.0	2	8.3	1	4.2	13	54.2
SB'S TEMPERATURE	0	0.0	1	4.2	3	12.5	0	0.0	13	54.2	0	0.0
SIGNIFICANCE	1	4.2	1	4.2	10	41.7	4	16.7	2	8.3	1	4.2
SKILLS	0	0.0	3	12.5	8	33.3	2	8.3	1	4.2	2	8.3
SOME WRITING	0	0.0	3	12.5	0	0.0	0	0.0	2	8.3	16	66.7
WEIGHT	0	0.0	1	4.2	7	29.2	3	12.5	7	29.2	0	0.0

Table 35 Restricted verb-noun collocations (**Group A**: abstract nouns; **Group B**: nouns characteristic of the medical discourse; **Group C**: general English nouns not found in the *HSC*)

a) Group A: restricted verb + abstract noun collocations

Data from Exercise 4 evidence that informants seemed to be familiar with many of the restricted verb collocates of the abstract nouns examined in this research. As was the case in Exercise 2 (cf. section 5.1.1), participants in the study identified some appropriate verb-noun collocations that typically occur in scientific discourse (e.g. *draw a conclusion*, *reach a conclusion*, *lead to a conclusion*, *reach an agreement*, *make a comparison*, *do a comparison*). The following Table displays all the combinations of the type restricted verb-abstract noun provided by informants:

	(n) MAKE	(%) MAKE	(n) DRAW	(%) DRAW	(n) REACH	(%) REACH	(n) LEAD TO	(%) LEAD TO	(n) TAKE	(%) TAKE	(n) DO	(%) DO
A COMPARISON	17	**70.8**	1	4.2	0	0.0	1	4.2	0	0.0	4	16.7
A CONCLUSION	1	4.2	10	**41.7**	10	41.7	4	16.7	0	0.0	0	0.0
A CONTRIBUTION	15	62.5	0	0.0	0	0.0	3	12.5	1	4.2	3	12.5
A DECISION	10	41.7	1	4.2	6	25.0	3	12.5	7	29.2	0	0.0
AN AGREEMENT	2	8.3	0	0.0	19	**79.2**	3	12.5	0	0.0	0	0.0

Table 36 Group A: **restricted verb + abstract noun collocations**

Figures in the above Table reveal that informants' verb-noun collocations regarding the nouns *conclusion*, *agreement* and *comparison* are very similar to the findings observed in the *HSC* corpus (cf. sections 4.3, 4.4 and 4.5). Restricted word combinations such as *draw a conclusion*, *reach an agreement* and *make a comparison* appeared to be the most frequent collocations both in the *HSC* and the informants' data.

<table>
<tr><th><i>HSC</i></th><th>Informants' data</th></tr>
<tr><td colspan="2">draw a conclusion
lead to a conclusion</td></tr>
<tr><td>reach a conclusion</td><td></td></tr>
<tr><td colspan="2">make a conclusion</td></tr>
<tr><td colspan="2">reach an agreement
lead to an agreement
make an agreement</td></tr>
<tr><td colspan="2">make a comparison</td></tr>
<tr><td>draw a comparison</td><td>do a comparison</td></tr>
<tr><td>do a comparison</td><td>draw a comparison</td></tr>
</table>

Table 37 Most frequent collocates of **verb + *conclusion***, **verb + *agreement*** and **verb + *comparison*** in the *HSC* and the informants' data

However, several misuses of delexicalised verbs in combination with the nouns *contribution* and *decision* were also found in the data. Despite the fact that 62,5% of informants produced the delexical combination *make* + *contribution*, which was also the most common collocation of the type "support verb" + *contribution* in the *HSC*, there was also a 16,7% of participants who produced some other deviant collocations such as *lead to a contribution* (12,5%) and *take a contribution* (4,2%). Similarly, the abstract noun *decision* was associated with the most frequent delexicalised verbs (i.e. *make*, *take* and *reach*) in scientific discourse, but it was also used in combination with deviant verbs: *lead to a decision* (12,5%) and *draw a decision* (4,2%).

Thus, it can be stated that in line with the findings observed in Exercise 2, informants did not appear to have many problems with restricted verb collocates except for the production of some deviated verb-noun collocations with the nouns *contribution* and *decision*, in which clear cases of direct confusion between two delexical verbs (e.g. *make* vs. *do a contribution*, *make* vs. *lead to a decision*, *make* vs. *draw a decision*) were traced. These deviant collocations and some others that will be discussed in Groups B and C certify informants' uncertainty over appropriate collocability.

b) Group B: restricted verb + nouns characteristic of the medical discourse

As noted earlier, a total of seventeen nouns that had proven to be representative of medical discourse were included in Exercise 4. It was expected that informants would have very few difficulties in associating these nouns with appropriate verbs since they were common expressions in the medical literature they were used to reading. However, the analysis of their collocations, as will be seen, questions that initial expectation. See Table 38 below:

	(n) MAKE	(%) MAKE	(n) DRAW	(%) DRAW	(n) REACH	(%) REACH	(n) LEAD TO	(%) LEAD TO	(n) TAKE	(%) TAKE	(n) DO	(%) DO
A MISTAKE	19	79.2	0	0.0	0	0.0	4	16.7	1	4.2	2	8.3
A PHOTOGRAPH	1	4.2	0	0.0	0	0.0	0	0.0	23	95.8	0	0.0
A SUGGESTION	17	70.8	0	0.0	0	0.0	1	4.2	0	0.0	4	16.7
A TABLET	1	4.2	2	8.3	0	0.0	1	4.2	17	70.8	1	4.2
A TEST	6	25.0	0	0.0	0	0.0	0	0.0	5	20.8	14	58.3
ACTION	1	4.2	0	0.0	0	0.0	4	16.7	12	50.0	3	12.5
AN EFFORT	18	75.0	0	0.0	0	0.0	0	0.0	0	0.0	5	20.8
ATTENTION	2	8.3	6	25.0	0	0.0	0	0.0	7	29.2	2	8.3
CHANGES	13	54.2	1	4.2	0	0.0	6	25.0	0	0.0	5	20.8
CONSENSUS	0	0.0	0	0.0	17	70.8	7	29.2	0	0.0	0	0.0
DAMAGE	7	29.2	1	4.2	0	0.0	7	29.2	0	0.0	13	54.2
OBESITY	0	0.0	0	0.0	2	8.3	18	75.0	1	4.2	0	0.0
PROGRESS	10	41.7	0	0.0	1	4.2	7	29.2	0	0.0	7	29.2
RESEARCH	3	12.5	1	4.2	0	0.0	2	8.3	1	4.2	13	54.2
somebody’s TEMPERATURE	0	0.0	1	4.2	3	12.5	0	0.0	13	54.2	0	0.0
SIGNIFICANCE	1	4.2	1	4.2	10	41.7	4	16.7	2	8.3	1	4.2
WEIGHT	0	0.0	1	4.2	7	29.2	3	12.5	7	29.2	0	0.0

Table 38 Group B: **restricted verb + noun** characteristic of the medical discourse

The informants’ data contain 62 different collocations within Group B as opposed to 23 verb-noun collocations found in the *HSC*. The comparison with the *HSC* provides the following data:

<table>
<tr><th><i>HSC</i></th><th>Informants’ data</th></tr>
<tr><td colspan="2">take temperature</td></tr>
<tr><td></td><td>reach temperature</td></tr>
<tr><td></td><td>draw temperature</td></tr>
<tr><td></td><td>reach weight</td></tr>
<tr><td colspan="2">take weight</td></tr>
<tr><td></td><td>lead to weight</td></tr>
<tr><td></td><td>draw weight</td></tr>
<tr><td colspan="2">do a test</td></tr>
<tr><td colspan="2">make a test</td></tr>
<tr><td colspan="2">take a test</td></tr>
<tr><td colspan="2">lead to obesity</td></tr>
<tr><td></td><td>reach obesity</td></tr>
<tr><td></td><td>take obesity</td></tr>
</table>

take a tablet	
	draw a tablet
make a tablet	
	do a tablet
	lead to a tablet
do research	
	make research
	lead to research
	draw research
	take research
take action	
	lead to action
	do action
	make action
do damage	
	make damage
	lead to damage
	draw damage
make progress	
	lead to progress
	do progress
	reach progress
	take attention
draw attention	
	make attention
	do attention
make a suggestion	
	do a suggestion
	lead to a suggestion
make an effort	
	do an effort
take a photograph	
	make a photograph

<table>
<tr><td colspan="2">make changes</td></tr>
<tr><td colspan="2">lead to changes</td></tr>
<tr><td></td><td>do changes</td></tr>
<tr><td></td><td>draw changes</td></tr>
<tr><td colspan="2">reach consensus</td></tr>
<tr><td colspan="2">lead to consensus</td></tr>
<tr><td colspan="2">make a mistake</td></tr>
<tr><td colspan="2">lead to a mistake</td></tr>
<tr><td></td><td>do a mistake</td></tr>
<tr><td></td><td>take a mistake</td></tr>
<tr><td colspan="2">reach significance</td></tr>
<tr><td></td><td>lead to significance</td></tr>
<tr><td></td><td>take significance</td></tr>
<tr><td></td><td>draw significance</td></tr>
<tr><td></td><td>make significance</td></tr>
</table>

Table 39 Most frequent collocates of **verb + noun** characteristic of the medical discourse in the *HSC* and the informants' data.

The most striking feature in Table 39 is that informants provided a wide range of combinations in comparison with the *HSC* results. Percentages seem to indicate that more than a half of those inquired may be familiar with some verb-noun collocations such as *take somebody's temperature* (54,2%), *do a test* (58,3%), *lead to obesity* (75%), *take a tablet* (70,8%), *do research* (54,2%), *do damage* (54,2%), *make a suggestion* (70,8%), *make an effort* (75%), *take a photograp*h (95,8%), *reach consensus* (70,8%), *make a mistake* (79,2%) but at the same time they appeared to have problems with other restricted verbs, as the wide list of support verbs being used seems to prove.

With the only exception of the nouns *consensus* (e.g. *reach consensus*, lead *to consensus*) and *test* (e.g. *do a test*, *make a test* and *take a test*), the informants' use of the other fifteen nouns in Group B gave rise to deviant collocations like: *lead to weight* (12,5%), *reach obesity* (8,3%), *make research* (12,5%), *lead to action* (16,7%), *make damage* (29,2%), *do progress* (29,2%), *take attention* (29,2%), *do a suggestion* (16,7%), *do changes*

(20,8%), and *lead to significance* (16,7%). Such a long list exemplifies that verb deviance often occurred within this group. Although informants may be aware of the frequent occurrence of the six verbs provided in this exercise due to the fact that they are used to reading scientific articles in English, it is clear from the above deviant word-combinations that collocations of verbs with delexical senses cause some combinatorial problems to the analysed community.

As can be inferred from the informants' collocations in Table 39, an area that emerges as being particularly difficult for participants is the distinction between similar pairs of verbs. Clear cases of direct confusion between two delexical verbs can be seen in examples such as the following:

(131) *take weight* (54,2%) vs. **reach weight* (12,5%)
(132) *do research* (54,2%) vs. **make research* (12,5%)
(133) *do damage* (54,2%) vs. **make damage* (29,2%)
(134) *make progress* (41,7%) vs. **do progress* (29,2%)
(135) *make a suggestion* (70,8%) vs. **do a suggestion* (16,7%)
(136) *make an effort* (75%) vs. **do an effort* (20,8%)
(137) *make changes* (54,2%) vs. **do changes* (20,8%)
(138) *make a mistake* (79,2%) vs. **do a mistake* (8,3%)
(139) *reach significance* (41,7%) vs. **take significance* (8,3%)

In light of the above instances, the use of *do* and *make* appears as a noticeable problem in the present data. Although a minimum of 10 informants (41,7%) produced appropriate restricted verb collocates for the nouns *progress*, *suggestion*, *effort*, *changes* and *mistake*, the production of *do* and *make* was also confused by a 19,64% of participants.

Unlike other studies (Burgschmidt & Perkins 1985, Kaszubski 2000, Nesselhauf 2005), the present analysis seems to confirm that *make* and *do* are particularly liable to confusion. Nesselhauf (2005:77) argues that such a confusion "*while frequent in the beginning and intermediate stages of learning, decreases in the language of more advanced learners*". Evidence from this study, on the contrary, appears to indicate that the distinction between these two highly-frequent verbs remains a difficult area among informants, who had devoted between 3,5 and more than 30 years to the learning of English. (cf. section 3.4).

As already stated, the verb *do* can be regarded as the most problematic support verb for informants, in the sense that it is often produced when *make* would be more suitable. Out of the 39 confusions of light verbs in examples (131-139) above, *do* was produced inappropriately 23 times (vs. 10 inappropriate uses of *make*, 3 of *reach* and 2 of *take*), while the appropriate verb would have been *make*. Other deviant collocations involving verbs in delexical senses in combination with general English words were found in informants' responses to Exercise 4, and will be now discussed in the analysis of the nouns in Group C.

c) Group C: restricted verb + general English nouns

The third group of nouns included in Exercise 4 consists of fourteen randomly selected general English nouns (i.e. *experience*, *a verdict*, *knowledge*, *a place*, *an exam*, *a guess*, *a speech*, *[financial] success*, *a profit*, *some writing*, *notes*, *a picture*, *skills* and *a goal*), which were not characteristic of the medical discourse analysed in this research. This type of general nouns was included so as to find out whether informants would also have difficulties in associating them with restricted verb collocates or not. Table 40 below displays all the word combinations supplied by participants.

	(n) MAKE	(%) MAKE	(n) DRAW	(%) DRAW	(n) REACH	(%) REACH	(n) LEAD TO	(%) LEAD TO	(n) TAKE	(%) TAKE	(n) DO	(%) DO
(financial) SUCCESS	3	12.5	0	0.0	6	25.0	12	50.0	0	0.0	0	0.0
A GOAL	2	8.3	1	4.2	14	58.3	3	12.5	0	0.0	1	4.2
A GUESS	7	29.2	3	12.5	2	8.3	0	0.0	6	25.0	1	4.2
A PICTURE	3	12.5	10	41.7	0	0.0	0	0.0	10	41.7	0	0.0
A PLACE	0	0.0	0	0.0	12	50.0	2	8.3	10	41.7	0	0.0
A PROFIT	12	50.0	1	4.2	1	4.2	3	12.5	1	4.2	0	0.0
A SPEECH	11	45.8	1	4.2	0	0.0	0	0.0	0	0.0	9	37.5
A VERDICT	2	8.3	4	16.7	11	45.8	3	12.5	2	8.3	0	0.0
AN EXAM	4	16.7	0	0.0	0	0.0	0	0.0	6	25.0	14	58.3
EXPERIENCE	2	8.3	2	8.3	9	37.5	5	20.8	4	16.7	0	0.0
KNOWLEDGE	0	0.0	1	4.2	7	29.2	10	41.7	4	16.7	0	0.0
NOTES	0	0.0	3	12.5	0	0.0	0	0.0	21	87.5	0	0.0
SKILLS	0	0.0	3	12.5	8	33.3	2	8.3	1	4.2	2	8.3
SOME WRITING	0	0.0	3	12.5	0	0.0	0	0.0	2	8.3	16	66.7

Table 40 Group C: **restricted verb + general English nouns**

Given the fact that informants had claimed that they were used to reading scientific papers in English on a regular basis (cf. section 3.4), it could be expected that the number of deviations concerning the verb in Groups A and B would be smaller than the ones in Group C. However, data show that this is not quite the case. Looking carefully at the individual verb-noun collocations produced by informants and comparing them against the actual word combinations traced in the *BNC*[121] (cf. Table 41), 10 verbs were produced inappropriately by more than two informants (i.e. *reach experience*, *take experience*, *draw a verdict*, *lead to a verdict*, *take knowledge*, *draw a guess*, *lead to profit*, *draw some writing*, *draw notes*, *draw skills*).

---

[121] As already pointed out at the beginning of this section, the word combinations in Group C were compared against the collocations found in the *BNC* since the fourteen nouns in this group were not traced in the *HSC* in combination with any of the six verbs analysed in this exercise.

<table>
<tr><th><i>HSC</i></th><th>Informants' data</th></tr>
<tr><td></td><td>reach experience</td></tr>
<tr><td colspan="2">lead to experience</td></tr>
<tr><td></td><td>take experience</td></tr>
<tr><td></td><td>draw experience</td></tr>
<tr><td></td><td>make experience</td></tr>
<tr><td colspan="2">reach a verdict</td></tr>
<tr><td></td><td>draw a verdict</td></tr>
<tr><td></td><td>lead to a verdict</td></tr>
<tr><td></td><td>make a verdict</td></tr>
<tr><td></td><td>take a verdict</td></tr>
<tr><td colspan="2">lead to knowledge</td></tr>
<tr><td colspan="2">reach knowledge</td></tr>
<tr><td></td><td>take knowledge</td></tr>
<tr><td></td><td>draw knowledge</td></tr>
<tr><td colspan="2">take a place</td></tr>
<tr><td colspan="2">reach a place</td></tr>
<tr><td></td><td>lead to a place</td></tr>
<tr><td colspan="2">do an exam</td></tr>
<tr><td colspan="2">take an exam</td></tr>
<tr><td colspan="2">make an exam</td></tr>
<tr><td colspan="2">make a guess</td></tr>
<tr><td colspan="2">take a guess</td></tr>
<tr><td></td><td>draw a guess</td></tr>
<tr><td></td><td>reach a guess</td></tr>
<tr><td></td><td>do a guess</td></tr>
<tr><td colspan="2">make a speech</td></tr>
<tr><td colspan="2">do a speech</td></tr>
<tr><td></td><td>draw a speech</td></tr>
<tr><td colspan="2">lead to (financial success)</td></tr>
<tr><td colspan="2">reach (financial success)</td></tr>
<tr><td colspan="2">make (financial success)</td></tr>
</table>

<table>
<tr><td colspan="2">make a profit</td></tr>
<tr><td></td><td>lead to a profit</td></tr>
<tr><td></td><td>reach a profit</td></tr>
<tr><td></td><td>take a profit</td></tr>
<tr><td></td><td>draw a profit</td></tr>
<tr><td colspan="2">do some writing</td></tr>
<tr><td></td><td>draw some writing</td></tr>
<tr><td></td><td>take some writing</td></tr>
<tr><td colspan="2">take notes</td></tr>
<tr><td>make notes</td><td></td></tr>
<tr><td></td><td>draw notes</td></tr>
<tr><td colspan="2">make a picture</td></tr>
<tr><td>do a picture</td><td></td></tr>
<tr><td colspan="2">draw a picture</td></tr>
<tr><td colspan="2">take a picture</td></tr>
<tr><td colspan="2">reach skills</td></tr>
<tr><td></td><td>draw skills</td></tr>
<tr><td></td><td>lead to skills</td></tr>
<tr><td></td><td>do skills</td></tr>
<tr><td></td><td>take skills</td></tr>
<tr><td colspan="2">make a goal</td></tr>
<tr><td colspan="2">reach a goal</td></tr>
<tr><td colspan="2">lead to a goal</td></tr>
<tr><td></td><td>draw a goal</td></tr>
<tr><td></td><td>do a goal</td></tr>
</table>

Table 41 Most frequent collocates of **verb + general English nouns** in the *HSC* and the informants' data.

Similarly to what was observed in the verb-noun collocations of Group B, the number of deviations within this group is remarkably high. With the only exception of the nouns *picture* and *exam*, which were both found in appropriate collocations (e.g. *draw/ take/ make a picture* and *do/ take/ make an exam*), the other twelve were associated with at least one support verb which had not been found in the *BNC*.

According to the deviations displayed in Table 41, the nouns *skills*, *experience* and *verdict* stand out as the most problematic ones: four different inappropriate delexical verbs were associated with each of them (i.e. *draw skills* [12,5%], *lead to skills* [8,3%], *do skills* [8,3%], *take skills* [4,2%]; *reach experience* [37,5%], *take experience* [16,7%], *draw experience* [8,3%], *make experience* [8,3%], *draw a verdict* [16,7%], *lead to a verdict* [12,5%], *make a verdict* [8,3], *take a verdict* [8,3]). Such a wide range of verbs in collocations where only one verb from the six provided appeared to be appropriate in the *BNC* corpus (i.e. *lead to experience, reach skills* and *reach a verdict*) reinforces the idea that verbs in verb-noun collocations pose great difficulty to non-native speakers.

Another case worth mentioning as being particularly susceptible to deviation in the study is the verb *draw*. But for *draw a picture*, no other instance of *draw* was found in the *BNC* in combination with the other thirteen nouns analysed in this group. Interestingly, however, *draw* was produced by at least two informants in six deviant collocations, namely, *draw experience, draw a verdict, draw a guess, draw some writing, draw notes* and *draw skills*. As Table 41 above shows, a total of 22 occurrences of deviant collocations of the type "*draw* + noun" were provided. In all these cases *draw* was used inappropriately instead of *lead to*, in instances like *draw experience* or *draw knowledge*; *reach* (e.g. *draw a verdict, draw skills, draw a goal*); *make* (e.g. *draw a guess, draw a speech, draw a profit, draw notes*) or *do* (e.g. *draw some writing*).

A possible explanation for this consistent misuse of the verb *draw* is that informants might have found it difficult to distinguish between its literal sense (i.e. *draw a picture*) and its more figurative sense of "to obtain something from something else/somebody" (e.g. *draw a conclusion*) and, as a consequence, they produced the above unconventional collocations in which *draw* was used in a non-literal sense with a questionable noun (e.g. *draw a verdict, draw a goal, draw a profit*), giving rise to inappropriate word combinations.

Altogether, the six verbs analysed in Exercise 4 (i.e. *make, draw, reach, lead to, take* and *do*) were susceptible to deviation. Taking the thirty-six nouns into account, however, it is noticeable that *draw* and *lead to* stand out as the most problematic for informants. While *draw* was inappropriately produced in 18 different deviant constructions, *lead to* was inappropriately produced in 15 collocations. *Do* was combined with 13 inappropriate nouns, *take* with 12 and *make* and *reach* appeared in combination with 8 and 7, respectively, deviant nouns.

Regarding the distribution of the described verb-noun deviations across the three groups established in the exercise, that is, a) abstract nouns characteristic of medical discourse (Group A), b) general nouns often used in medical articles (Group B) and c) general English nouns (Group C), several points should be underlined. In general terms, the six verbs were particularly deviant with medical specific nouns and general nouns[122]. On the contrary, deviations of the type "verb + abstract noun" were less prevalent; probably due to the fact that on the one hand, the repertoire of abstract nouns was smaller (5) than the list of nouns in Groups B and C (17 and 14, respectively), and on the other, the abstract nouns examined had proven to co-occur with many of the verbs provided.

As Figure 68 (cf. Appendix 7) shows, inappropriate uses of *draw* were particularly frequent with general nouns (Group C) and nouns characteristic of the scientific discourse (Group B), i.e. *draw a verdict, draw a guess, draw temperature*. Deviant uses of *lead to* (cf. Figure 70), *make* (cf. Figure 67) and *do* (cf. Figure 72) were especially noticeable in combination with nouns from Group B (e.g. *lead to a tablet, lead to damages, lead to significance, do a tablet, do an effort, do a mistake*), whereas *reach* and *take* were produced inappropriately in combination with both general English nouns and nouns frequently found in the *HSC* (cf. Figures 69 and 71).

---

[122] Informants' verb-noun collocations in Exercise 4 can be seen in Appendix 7 (cf. Figures 67-72).

Altogether, it can be concluded that all the above verb deviations indicate the general difficulty of support verbs in restricted verb-noun collocations for non-native speakers (Howarth 1996/1998; Kaszubski 2000; Altenberg & Granger 2001; Nesselhauf 2005) and underline the fact that informants were not fully aware of the restrictions concerning the collocability of either scientifically-related (Groups A and B) or general nouns (Group C). A summary of the major results obtained from the informants' data analysis investigated in this chapter will be presented in the following section.

5.2 Summing up of findings

Since the intention of this chapter was to make a valid comparison between native speakers' performance and informants' collocational competence with reference to the selected five abstract nouns, all the observations made in the non-native speakers' data were checked against the *HSC* and the *BNC*. It was hoped that the examination of the informants' written production would reveal similarities as well as significant differences from native speakers' texts. As shown in the previous sections, although similarities could be traced in some collocational patterns successfully identified by participants (i.e. appropriate restricted verb-noun collocations and adjective + noun patterns), the collocational patterning of abstract nouns was found to be susceptible to deviation particularly often.

Major results of the analyses conducted in section 5.1 indicate that most deviations traced in the informants' production refer to the three most problematic areas which were also identified by participants in the Survey Content-Based Questions (cf. section 5.3). As will be seen in the last section of this chapter, when asked about the main difficulties they encountered while writing a paper in English, most of them pointed out three chief aspects; namely, a) the use of general words and expressions characteristic of the medical genre, b) the distinction between similar pairs of words and c) the correct application of grammatical rules when constructing sentences.

From the evidence of non-native speakers' collocational deviations, it is reasonable to suggest that they lack control of the phraseological behaviour of nouns in medical English. Deviations were found both in restricted verb-noun collocations as well as in free verb combinations. Whereas results show that they were able to produce appropriate restricted verb collocates in Exercises 2 (cf. section 5.1.1) and 4 (cf. section 5.1.4), such as *reach an agreement*, *make a comparison*, *draw a conclusion*, the fact that in Exercise 2 they also made use of a wide range of free verb collocates where a restricted verb would have been more suitable (e.g. **achieve an agreement*, *give/adopt/publish/report a conclusion*, **perform a comparison*) and that they failed to identify the appropriate delexicalised verbs in combination with the nouns *contribution* and *decision* in Exercise 4 (e.g. **lead to a contribution*, **draw a decision*) reveals that the distinction between both types of word combinations poses some problems for informants. This finding goes in line with Gouverneur's (2008) observation that delexicalised uses of these verbs "*represent a stumbling block to native-like proficiency.*" (Gouverneur 2008:223).

In addition, it was also observed that informants found it difficult to distinguish between similar pairs of verbs such as *do* vs. *make*, *reach* vs. *achieve* and *take* vs. *reach*. This deviation in the verb was especially noticeable in combination with abstract nouns (cf. section 5.1.4a) and nouns characteristic of medical English (cf. section 5.1.4c). Such a finding was unexpected since informants were supposed to be aware of the collocational patterning of medical every-day expressions like *take weight*, *do research*, *reach significance* and *do damage*. Part of the reason for this collocation difficulty might be that these support verbs can be used in combination with a wide range of nouns and may have similar translations into Spanish (i.e. *do* vs. **make* research ["*hacer, realizar una investigación*"], *take* vs. **reach somebody's temperature* ["*tomarle la temperatura a alguien*", "*alcanzar una temperatura*"], *take* vs. **reach weight* ["*alcanzar/ganar/coger peso*"], *reach* vs. **achieve an agreement* ["*alcanzar un acuerdo*"]. This semantic similarity therefore leads to confusion particularly often.

As to which of the six verbs in Exercise 4 (i.e. *make*, *draw*, *reach*, *lead to*, *take* and *do*) were more liable to deviation when combining with the list of nouns provided, it must be noted that deviations were found in all of them. However, the verbs *draw* and *lead to* clearly stand out as the most problematic ones, giving rise to inappropriate collocations such as **draw a decision*, **draw a tablet*, **lead to a contribution*, **lead to damage*, **lead to a decision*. Clear evidence is found here to support Howarth's opinion that "*they [learners] are perhaps not aware that figurative senses are more restricted in collocability than are literal senses and therefore require greater precision in their use*" (Howarth 1996:160). Thus, a possible explanation for the overuse of those two delexical verbs may lie in the fact that informants might have not been aware of the distinctive collocational patterns associated with the literal and the more figurative senses of both verbs.

Two further aspects that were investigated with regard to participants' command of the phraseological behaviour of abstract nouns in medical English were their knowledge of the collocational restrictions involved in fixed expressions like "*there* + *be* + (adjective) *agreement*" and *in conclusion* as well as the most common adjective collocates of the abstract nouns under study.

On the one hand, the failure to produce appropriate lexical bundles with the nouns *agreement* and *conclusion* emerged as a recurrent deviation in Exercises 2 and 3. The case of "*there* + *be* + (adjective) *agreement*" was often liable to confusion. A total of 19 informants were unable to produce the expression "*there was (unanimous) agreement*" appropriately and the deviant structures they provided, such as "*an agreement was done*" and "*it was produced an agreement*", clearly show a great deal of transfer from their mother tongue (i.e. "*se produjo un acuerdo*"). This strong influence from informants' L1 can be seen to be related with their writing process, as described in section 5.3. The fact that 20,8% of informants claimed that they first wrote a draft of a medical paper in Spanish and then translated it into English became evident in Exercise 3, where transfer from their L1 was most noticeable.

On the other hand, other individual collocations that were found to pose particular problems correspond to the adjectives selected as premodifiers of the abstract nouns examined. As a further indication of informants' lack of collocational competence concerning the use of abstract nouns, it should be underlined that some of the adjective-noun expressions provided consisted of questionable word combinations like ?*reliable agreement*, ?*reliable contribution*, ?*late agreement* and ?*late comparison*, which had not been traced in the native corpora.

Likewise, the use of semantically similar adjectives (i.e. *little*, *short* and *small*) was particularly difficult among participants. Examples such as **short agreement*, **short comparison* and **short contribution*, where the use of the adjective *little* would have been expected, give account of the fact that the semantic similarity between both adjectives as well as their similar translation equivalents into Spanish (i.e. *poco*, *escaso*, *pequeño*) may have contributed to the informants' misuse of these adjective collocates.

Another area which was also affected by informants' L1 interference was, unexpectedly, morphology. Morphological deviations in the word form of the nouns *comparison* (i.e. **comparation*, **comparition*, **comparision*) as well as the adjective *biological* (i.e. **biologyc*) are clear examples of informants' tendency to transfer morphological forms from their L1. Similarly, participants' attested preference for periphrastic structures like "?*responsible for the making of decisions*" over compound lexemes such as *decision making* serves to reinforce the assertion that non-native speakers show a strong inclination to literal translations.

A final issue worth mentioning refers to informants' uncertainty over the most common grammatical constructions typical of the word combinations examined in this study. As mentioned earlier, participants had already highlighted the use of appropriate grammatical rules as a problem area when writing their articles in English (cf. section 5.3). Deviations regarding the use of passive forms instead of the active voice (and vice versa), syntactic structures to express impersonality (e.g. "**it was produced an agreement*", "*before *they obtain valid conclusions*", "** it is general agreement that*") as well as the use of the appropriate non-finite forms following a preposition (e.g. "*before *to obtain/ before *reach + (*adjective*) conclusion*") also occurred in the informants' data.

All the above findings seem to indicate that the phraseological patterning of abstract nouns in medical English poses some problems for the discourse community under investigation. Although in Nesselhauf's view (2005:246) some other factors such as each informant's language aptitude and exposure to English, their motivation, their writing techniques and beliefs, among others, should also be taken into account when analysing non-native speakers' collocational performance, the present results, yet far from exhaustive, highlight informants' lack of knowledge over proper collocability and stress the need to provide them with useful resources and tools aimed specifically at improving their collocational competence.

Having investigated the informants' collocational performance in the worksheet of exercises, in the following section an attempt is made to briefly outline informants' views of their writing process in English so as to find out what linguistic areas are highlighted as the most problematic ones.

5.3 Informants' self-assessment (Survey Content-Based Questions)

As noted in chapter 3, the survey conducted among participants consisted of two clearly differentiated sorts of questions: a) the Survey Profile Questions already discussed in the characterisation of informants (cf. section 3.4) and b) the Survey Content-Based Questions which were more closely related to the participants' perception regarding their writing process in English, on the one hand, and the resources they highlighted as useful tools to help them improve their production of scientific papers in English, on the other.

Apart from assessing their linguistic competence when dealing with the collocational patterns selected for this study, it also seemed advisable to find out the steps they used to take when writing in English. Although informants' own views and perceptions of their written production were not the main aim of the present research, it appeared to be worth investigating how informants described their writing process as well as the areas they regarded as particularly difficult when producing their medical papers in English. Participants were also inquired about the resources they would like to have at their disposal so as to overcome those problem areas and facilitate their writing. In this section the results obtained from the informants' self-evaluative view of their linguistic skills and needs (i.e. Survey Content-Based Questions 7-9) are presented.

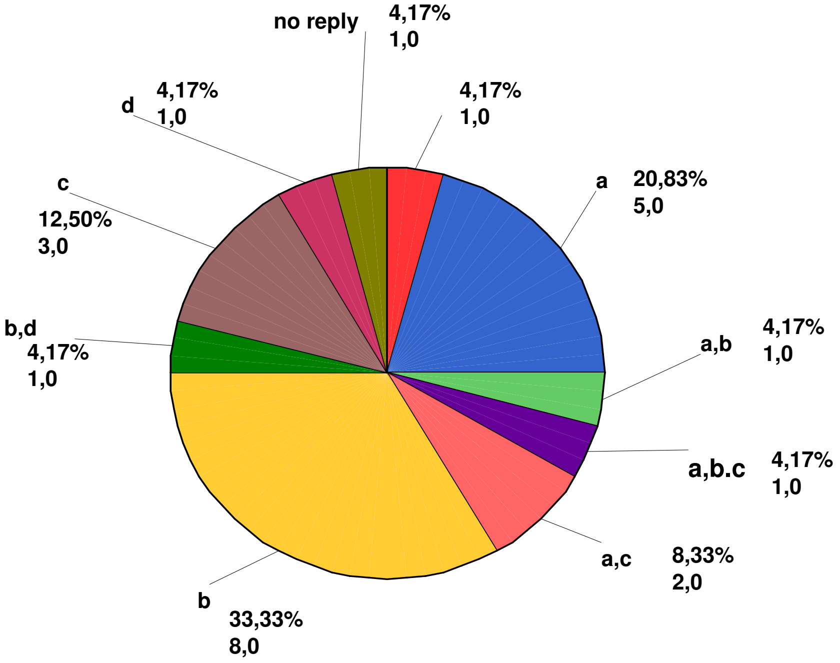
When being asked about the steps taken when planning and writing a scientific article in English (cf. Survey Content-Based Question 7), eight doctors (33,4%) pointed out that they first made a draft in English and, afterwards, they looked up unknown words or expressions in dictionaries and thesauruses, whereas some others (20,8%) wrote a first version in their mother tongue and then they themselves translated that into English. A very low percentage (12,5%), consisting of three informants, just resorted to asking for some expert advice before handing in their papers. Only one informant (**PH6**) stated that she used to write the text directly in English without making any further reference to her writing process.

Due to the fact that this was a closed-ended multiple-choice question[123], some participants chose more than one option so as to describe their writing process more thoroughly. As Figure 65 illustrates, two doctors (**PH3**, **PH10**) claimed that they first wrote their papers in Spanish, then translated them into English and, finally, asked for advice so as to polish their texts before sending them for publication. Informants **HN3** and **PH2** followed a similar writing process as both of them stated that they wrote a first version of their texts in their mother tongue and translated that version into English. Afterwards, they tried to improve the English draft by making use of dictionaries. Informant **PH2** included a final step by asking for expert advice before sending in his article to be considered for publication.

---

[123] In Dörnyei's terminology (Dörnyei 2007:106)

Finally, it should be mentioned that one informant (**ALB3**), who commented that she first made a draft of her papers in English, underlined the importance of consulting already published articles in search of useful expressions she could borrow for her own written productions. This assertion, however, should be taken cautiously as being familiar with certain useful recurrent expressions does not necessarily entail knowing how to use them correctly.



Survey Question 7: When writing a paper in English, what is your writing process?

a. first you write your contribution in your mother tongue and then translate it into English
b. you make a draft in English and then look up unknown words or expressions
c. you ask for some expert advice before handing in your paper
d. Other: (please specify) ______________________________

Figure 65 Informants' writing process

Informants were further inquired about the main linguistic areas they found more difficult when writing up their research papers in English[124]. In this question, in the form of a semantic differential scale, respondents were asked to indicate their answer by marking a continuum between the adjectives "easy" and "difficult" at the extremes. They were told to assign a number to a list of five aspects in order of difficulty (1-very easy; 2-easy; 3-not so easy; 4-difficult and 5-extremely difficult).

---

[124] cf. Survey Content-Based Question 8 in Appendix 8.

They were given the following aspects to consider: a) specific (medical) terminology in English, b) general words and expressions most commonly used in medical publications, c) distinction between similar pairs of words (e.g. *do* vs. *make*), d) distribution of paragraphs and e) grammatical rules when constructing sentences. There was also a sixth aspect, labelled as "Others (Please specify)", where participants were given the chance to mention some other difficulties, not included in the survey, they may encounter when writing in English[125].

Dealing with specific medical terminology was not regarded as difficult. In fact, most informants found it either "very easy" (41,7%) or "easy" (45,8%). On the contrary, the survey revealed that the use of general items and recurrent expressions typical of their specific medical community entailed more complexity for more than the half of the sample (58,3%).

Another area to be examined in detail was informants' command of similar pairs of words in terms of when (context) and how to use each of them. Most participants labelled that aspect as either a "not so easy" (41,7%) or "difficult" (45,8%) task.

The fourth aspect to be analysed was the informants' view of the distribution of paragraphs when editing and redrafting their texts. 54,2% of those interviewed considered it as a "not so easy" task, whilst a 29,2% thought it was "easy". Thus, it seems that the organisation of a text, on the whole, was not perceived as an extremely difficult aspect of writing.

The last feature mentioned in question 8 referred to the application of grammatical rules when constructing sentences. As was the case with participants' difficulty in discriminating similar pairs of words, nearly half of the sample considered that grammar was a problem area in the whole process of writing. It is also worth mentioning at this stage that unlike other aspects, the correct use of grammatical rules was labelled as “extremely difficult” by 12,5% of those interviewed. No higher percentage of the “extremely difficult” category could be traced in any of the previously described aspects.

---

[125] Informants' responses to this question can be seen in Appendix 8 (cf. Figures 73-77).

As already stated, informants were allowed to provide other aspects relevant to the writing process itself. Only four participants added their own ideas, which could be paraphrased as follows:

a) Introducing subordinate sentences (**ALB3**)
b) Using linking words to connect sentences (**PH5**)
c) Dealing with phrasal verbs and prepositions (**ALB2**, **PH12**)
d) Being able not to write in a Spanish-like way (**ALB2**)

All the above features were considered to be problem areas and, therefore, were labelled as “extremely difficult”.

Finally, informants were encouraged to think of any tool and/or resource they found to be useful in order to sort out the problems and difficulties they encountered when writing their medical papers in English (cf. Survey Content-Based Question 9). They were provided with some examples, such as a lexical database of general words and common expressions in academic writing and (specialised) dictionaries. A total of fourteen informants highlighted the importance of the abovementioned tools:

> *“I usually look for help in the on-line ‘Cambridge Dictionary for Advanced Learners’, where I can find really helpful examples.”* (**PH5**)
> *“A lexical database of words and common expressions is a good idea.”* (**HN3**)
> *“I think specialized dictionaries would be very useful.”* (**HN7**)
> *“I use a specialized dictionary.”* (**HN8**)

However, twelve doctors suggested some other tools that should be taken into consideration as well. All in all, their suggestions fall into three main categories: technical devices and on-line resources, human resources and the learner's attitude. The following Table shows their suggestions along with some of their quotations so as to illustrate their points:

Technical devices & on-line resources:	✓ Automatic on-line translators, collocation dictionaries (3 informants) *"I'd like an automatic on-line translator, but there are no good ones available."* (**ALB1**)
	✓ wordreference.com (1 informant)
Human resources:	✓ Training courses on writing techniques (2 informants) *"In general I use medical dictionaries in English, but some courses on style in medical English will help me a lot."* (**PH12**)
	✓ Building up lexical databases of frequent errors (1 informant) *"A lexical database of the most frequent mistakes non-native speakers make when writing in English will help me to avoid them."* (**ALB3**)
	✓ Linguists available to give support on writing (3 informants) *"I think it must be useful that a linguistic working in any University Department could help us not only to write but to resolve the most frequent questions."* (**PH4**)
Learner's attitude:	✓ Reading more often in English (1 informant) *"I try to learn paying attention when I read English journals."* (**ALB2**)
	✓ Writing more often in English (1 informant) *"My writing will improve if I wrote more often in English."* (**PH3**)

Table 42 Suggested tools and resources

As noted earlier, despite not being the main aim of the present investigation, informants' perceptions of their writing process in English serve to complement their characterisation. In light of their responses, participants in the sample (i.e. doctors from different Spanish hospitals who were working on different research areas; cf. section 3.4) seemed to face similar problems when writing, editing and rewriting a scientific text in English.

As already discussed in the description of the informants' profile and, particularly, in the data obtained in Survey Profile Question 6 (cf. section 3.4), respondents seemed reluctant to consider themselves as regular writers of scientific papers in English because the production of an article as a whole was seen as a team work in which each participant played a well-defined role: housemen were in charge of the field and laboratory work, whereas assistant doctors took decisions regarding which topic would be discussed in the research article and co-ordinated the whole process. With regard to the actual writing up of the article, they stated that the first draft was usually done in Spanish and then translated into English.

The above line of reasoning should not be overlooked when it comes to the descriptions informants made of the writing process they followed when producing an English article. Bearing in mind that writing the first draft in Spanish is usually the starting point (cf. section 3.4), it came as no surprise that a 20,8% of the informants claimed that they first wrote down their ideas in their mother tongue and then translated them into English. The obvious implications (i.e. literal translations, the confusion in the use of similar pairs of words, collocational problems) of such a methodological approach to the writing of an article in a foreign language were further discussed in the analyses of the worksheet of exercises (cf. section 5.1).

A noticeably high 33,4% claimed, though, that they wrote their first draft in English and then made use of dictionaries and thesauruses so as to improve their writing. They seemed to realise that it is doubtless essential to master the structures and the phraseology that resemble typical and consistent expressions in scientific English. Some other participants underlined the fact that their writing process was a combination of several of the options provided in question 7; that is, writing in Spanish and translating that text into English, re-writing the English draft and asking for some expert advice to sort out some editing problems and clarifying some doubts. Whatever their choice, they have acknowledged the relevance of the writing process in producing a good quality piece of writing.

It is also worth mentioning another strategy pointed out by informant **ALB3**. She stated the importance of reading already published articles since they were "*a very valuable source of good examples of recurrent expressions*".

Once more, informants seemed to perceive a symbiotic relationship between reading and writing.

Some of the most interesting pieces of information revealed by the survey are concerned with the areas non-native speakers of English highlighted as the most difficult when being faced with a white page or even with a Spanish text to be translated into English (cf. Survey Content-Based Question 8). One of the research questions in this study (cf. chapter 1) attempted to find out whether health scientists would have, as expected, more difficulties in dealing with general words than with specific terms as they were used to the specific terminology in medical English. In fact, in some cases they use English items rather than their Spanish equivalents to refer to some processes, illnesses and scientific mechanisms (i.e. *HIV* vs. *VIH*[126]; *DNA* vs. *ADN*[127]; *MHC* vs. *CMH*[128], among others). Although Tognini-Bonelli (2001) points out that the identification of specific terminology is extremely important in Language for Specific Purposes (LSP), it was expected that this sample of Spanish doctors, who were used to reading and writing their articles in English, would already be familiar with the specific terms characteristic of their own field of work.

Data drawn from the survey (Cf. Survey Content-Based Question 8) are fully congruous with that initial expectation since 58,3% of informants found dealing with general words either “not so easy” (45,8%) or “difficult” (12,5%). In line with this problem area, the distinction between similar pairs of words was also regarded as a central difficult aspect. A third somewhat problematic area highlighted in the survey was the use of accurate grammatical rules when writing in English.

Surprisingly, the distribution of paragraphs when building up the layout of a text was not seen as a main problem. The fact that texts tended to be exhaustively revised by correctors, translators or some other English experts seems to make researchers be less worried about linking devices between paragraphs than about the well-known phrasal verbs and prepositions, for instance. Yet it must be underlined that two of the four participants who contributed to the list of problem areas with their own aspects referred, essentially, to paragraph distribution as they claimed they had problems when introducing subordinate clauses as well as using cohesive devices to connect sentences within paragraphs. Last but not least, one informant stressed her constant fear of “thinking in Spanish” when writing in English. She pointed out that she found it very hard not to translate literally from her mother tongue.

---

[126] Human Immune Deficiency Virus (HIV) vs. *Virus de Inmunodeficiencia Humana* (*VIH*).
[127] Deoxyribonucleic Acid (DNA) vs. *Ácido desoxirribonucleico* (*ADN*).
[128] Major Histocompatibility Complex (MHC) vs. *Complejo Mayor de Histocompatibilidad* (*CMH*).

All these lexico-grammatical and stylistic worries expressed by participants raise the issue of the importance of lexical collocations in vocabulary learning and, by extension, in written production. These findings suggest that participants do not necessarily find discipline specific medical vocabulary particularly difficult. Rather, it is the selection and use of the general non-technical items along with their combinatorial potential what stands out, according to 58,3% of those interviewed, as one of the most difficult aspects of excelling in scientific writing.

Furthermore, when being inquired about what kind of tools and resources could help them improve the quality of their writing in English, more than half of the informants (58,3%) remarked the usefulness of checking a lexical database consisting of the most general items and expressions in medical English as well as writing with a specialised dictionary nearby.

Other worth notable suggestions put forward by informants pointed to the convenience of building up a lexical database of common errors typically made by non-native speakers. Some informants mentioned that there was a recurrent list of errors they constantly made which they found difficult to get rid of. Thus, they suggested they would benefit from checking errors made by their peers so as to avoid them.

Once again the need of training courses[129] on writing techniques emerged in the survey as a key factor in the informants' own progress on writing skills. A further piece of evidence in favour of the importance they granted to training is that it was also recommended the utility of getting linguistic assistance in their workplace on a regular basis.

Finally, two informants' perceptions drew the attention to the language learner attitude when writing. Fully aware of their leading role in their own writing process, they felt there was a cause and effect relationship in both reading and writing:

- the more attention one pays to what they read, the more significant learning will take place:
  "I try to learn paying attention when I read English journals." (**ALB2**)
- the more often one writes, the more their writing will improve:
  "My writing will improve if I wrote more often in English." (**PH3**)

The Survey Content-Based Questions have highlighted some of the informants' writing needs when producing their medical papers. The results obtained have brought to light not only informants' perception of their writing process but also their weaknesses as far as writing skills is concerned. In addition, several suggestions on possible effective tools which could have a positive effect on their written English have also been made. Such needs point to three areas in particular in the field of foreign language acquisition: the need for lexical devices, such as lexical databases, collocation dictionaries and automatic translators; training courses on writing skills, and an increase of exposure, on behalf of the learner, to reading and writing in English.

---

[129] Notice this was already mentioned in the Survey Profile Questions (cf. section 3.4, Question 4), where informants were asked whether they had taken any course related to their speciality in English.

CHAPTER 6: CONCLUSIONS, IMPLICATIONS AND FURTHER RESEARCH

This final chapter begins with a review of the research questions that motivated this dissertation (section 6.1) and a summary (section 6.2) of the major results of the analyses conducted in chapters 4 and 5. Then, some implications of these findings for the teaching and learning of collocations and phraseological units in medical English are examined (section 6.3). The last part of this chapter (section 6.4) points out new research directions in the study of the lexico-grammatical patterning of non-specialised vocabulary in science writing and presents a linguistic resource, *SciE-Lex*, which consists of a lexical database that takes account of the lexico-grammar of general terms in scientific English.

6.1 Referring back to the research questions underlying this thesis project

In the introductory chapter, three research questions (described in full in section 1.1) were formulated as the basis of the present investigation; the answer to each of which will now be outlined:

(i) Are Spanish doctors aware of the particularities of the medical genre?

This study has shown that Spanish doctors do not necessarily find discipline-specific terminological vocabulary difficult. Rather, it is non-technical words that pose more difficulties for them. Both the worksheet of exercises and the informants' view of their writing difficulties[130] have revealed that the idiosyncrasies and conventions of the medical register are problematic for non-native speakers. Despite being able to identify some domain-specific phraseological expressions, they lack knowledge of how to use them appropriately in their written production in English.

---

[130] Cf. Survey Content-Based Questions in section 5.3.

(ii) What are the main difficulties the phraseological patterns of abstract nouns in medical English pose for the non-native speakers' medical community?

The comparison between the native corpus (i.e. the *HSC*) and the non-native production has demonstrated that the lexico-grammatical patterns of the selected abstract nouns in medical discourse are liable to confusion. Evidence from informants' collocational deviations has underlined the following problem areas: (a) unfamiliarity with delexical uses of verbs in combination with abstract nouns (e.g. *do research*, *make a comparison*); (b) unawareness of common lexical bundles (e.g. *in agreement with*, *there is* [adjective] *agreement that*); (c) lack of knowledge of certain structures that typically occur in the passive (e.g. "*conclusions were drawn*") or rather in the active ("*their effects make a significant contribution to*[…]"); and (d) uncertainty about the different typology of the most frequent adjective collocates (e.g. *important conclusion*, *relative contribution*, *direct comparison*).

(iii) What are the prototypical usage patterns around the following abstract nouns: *agreement*, *comparison*, *conclusion*, *contribution* and *decision* in medical English?

The analysis of the collocational patterning of each of the above nouns has pointed to four linguistic processes, which are characteristic of their use in medical discourse: delexicalisation, grammaticalisation, passivisation of periphrastic structures of the type "verb + abstract noun" and an attested preference for the attributive use of adjectives. A summary of the main findings of this thesis project will follow up the above issues.

6.2 Summing up of main findings

The study of the empirical data gathered directly from corpus evidence has revealed reliable information on the usage of the following nouns: *conclusion*, *agreement*, *comparison*, *contribution* and *decision* in medical English. The native corpus investigation conducted in chapter 4 has provided a characterisation of the phraseological patterning of the selected abstract nouns in the register of the medical community under study. Such exploration of naturally-occurring instances as well as high-frequent abstract noun collocations has allowed the identification of four main linguistic processes which give account of the domain-specific features concerning the collocational patterns of the nouns in question:

> a) frequent use of delexicalised verbs in combination with abstract nouns in restricted verb-noun collocations (e.g. *draw a conclusion*, *reach an agreement*, *make a comparison*, *make a contribution*, *take a decision*)
>
> b) recurrent use of lexical bundles in which the abstract noun has acquired a more grammatical role and has become part of a distinctive connective operator (e.g. *in agreement with*, *in/by/for comparison*)
>
> c) high frequency of occurrence of periphrastic structures of the type "verb + abstract noun" in the passive voice as opposed to their equivalent cognate lexical verbs, which are more commonly used in active constructions (e.g. *several conclusions can be drawn* vs. *investigators concluded that both mechanisms are associated*)
>
> d) attested preference for the use of attributive adjectives premodifying abstract nouns as well as for descriptors to be associated with *agreement* and *conclusion*, and classifiers with *comparison*, *contribution* and *decision*.

The above processes have made it evident that the use of abstract nouns tends to be associated with specific grammatical patterns. From the point of view of medical discourse analysis and English for Specific Purposes (ESP), there is much to be said about the phraseological behaviour of abstract nouns in medical English.

Firstly, it has been shown that they are usually combined with semantically empty support verbs (e.g. *make*, *reach*, *take*, *draw*), whose original meaning has faded away. Through a process of delexicalisation, these verbs have gained a more grammatical role and simply perform a verbal function, whereas the abstract nouns they collocate with can be said to carry the semantic content of the whole multiword unit. This redistribution of meaning across lexical items underlines the close interrelationship between the verb and the abstract noun associated in collocation with it (e.g. *draw a conclusion*, *reach an agreement*, *make a comparison, make a contribution*, *take a decision*). In addition, such periphrastic uses have been found to correspond to a morphologically related verb of equivalent meaning (i.e. *conclude*, *agree*, *compare*, *contribute* and *decide*), which contributes to reinforce the delexicalised use of the restricted verb collocates in verb-noun combinations.

Despite being semantically equivalent, the native corpus revealed that the use of a periphrastic structure of the type “verb + abstract noun” was preferred to synthetic constructions with a cognate lexical verb (i.e. *conclude*, *agree* and *decide*) in three (i.e. *conclusion*, *agreement* and *decision*) out of the five abstract nouns analysed. Other noteworthy differences were for example observed between the use of restricted verb-noun collocations and the use of full lexical verbs (i.e. *draw/reach a conclusion*, *reach an agreement* and *make a decision* vs. *conclude*, *agree* and *decide*). While the former tend to appear in passive constructions, the latter are more frequently used in the active voice.

Secondly, the empirical observation of the lexico-grammatical patterning of the selected nouns in the *HSC* corpus has brought to light the observation that certain recurrent expressions like *in agreement with*, *in/by/for comparison* have undergone a process of grammaticalisation and semantic bleaching[131] and must therefore be analysed as multiword units in which the abstract noun cannot be interpreted in isolation, but as part of an extended unit of meaning that has acquired a new grammatical function. In the lexical bundles examined, the abstract nouns in combination with certain prepositions have taken on a more grammatical role as cohesive devices and thus serve to link different parts of the discourse.

---

[131] Cf. Footnote 64

Finally, the investigation of the most salient patterns of abstract nouns has also informed about the nature of the adjectives associated in combination with them. Data drawn from the *HSC* have addressed two main issues regarding adjective usage. The first issue relates to the fact that the adjective collocates of the five nouns analysed show a clear preference for attributive position. Very few instances of adjectives qualifying abstract nouns were used predicatively. The second general observation refers to the kind of adjectives being associated with each abstract noun. As pointed out in chapter 4, due to the fact that the publications examined in the *HSC* corpus are related to medical topics (in the form of observable, demonstrable data), classifiers were expected to outnumber descriptors as the most common premodifiers of abstract nouns in the native corpora.

Evidence from the *HSC* has revealed that while the nouns *comparison*, *contribution* and *decision* are usually premodified by classifiers, the nouns *agreement* and *conclusion* are, surprisingly, more frequently associated with extent, time and evaluation descriptors, which contradicts the observations from the scientific discourse where they are more rarely found (Biber et al. 1999).

The observed regularities have contributed to outline the phraseology around abstract nouns in medical English and also seem to have brought to the forefront the fact that corpus-based analyses are of paramount importance in the identification of the already discussed linguistic phenomena. As noted earlier, the compilation and subsequent examination of a corpus of medical journals has allowed the direct observation of naturally-occurring examples of abstract nouns in context, which otherwise would have been very difficult to analyse.

With the aim of determining whether informants in the study (i.e. a community of twenty-four Spanish practising doctors) were acquainted with the previously identified distinctive features and conventions with regard to the phraseological behaviour of abstract nouns in medical English, a survey was conducted among informants. Bearing in mind that they are intended to publish their current research in English, the analysis of their lexical expertise with respect to the phraseology of abstract nouns seemed relevant to this research. The results obtained from informants' performance in the worksheet of Exercises enclosed in the survey (cf. Appendix 4) have demonstrated that the already mentioned processes, involved in the collocational patterning of abstract nouns in the native corpus, pose problems for non-native speakers.

Although similarities with expert texts were also present in informants' identification of some usual verb-noun collocations (e.g. *reach an agreement, make a comparison, draw a conclusion*) and adjective-noun patterns (e.g. *general agreement, statistical comparison, logical conclusion, significant contribution, firm decision*), the study has shown that on the whole they lack knowledge of the collocational properties of the selected abstract nouns.

A considerable number of deviations found in the worksheet of exercises stress informants' uncertainty over proper collocability of the analysed lexical items. Firstly, several misuses of both restricted and free verb-abstract noun collocations (e.g. **achieve an agreement*, **perform a comparison*, **lead to a contribution*, **draw a decision*) as well as an apparent difficulty in distinguishing between similar pairs of words (e.g. *do* vs. *make*, *reach* vs. *achieve*) have shown that delexicalised uses of verbs are particularly liable to confusion. Whereas the *HSC* corpus had revealed that free verbs were hardly ever used in combination with the analysed nouns, informants did use a wide range of free verb-noun collocations in their responses (e.g. **perform a comparison*, **adopt a decision*, **generate a contribution*), where a restricted verb would have been expected.

In addition, some deviations in the verb appeared to occur in combination with abstract nouns (cf. section 5.1.4a) and other domain-specific nouns (cf. section 5.1.4c). Informants found it difficult to differentiate the preferred collocates of similar pairs of verbs, which led to inappropriate verb-noun combinations like the following: **make research*, **reach weight*, **achieve an agreement*. These findings seem to indicate that two are the main reasons for this confusion. On the one hand, the fact that these support verbs tend to collocate with a wide repertoire of nouns might make it difficult for non-native speakers to differentiate their usual collocates. On the other hand, the delexicalised senses of these verbs usually have similar translation equivalents in informants' L1, which may hinder even more their distinction.

A further finding related to informants' misuse of delexicalised verbs in association with abstract nouns suggests that the verbs *draw* and *lead to* were found to be remarkably problematic. Recurrent collocations like **draw a decision*, **lead to a contribution* and **lead to a decision* appear to verify that informants had difficulties in discriminating between literal and more figurative senses of both verbs.

As already stated, the influence of informants' L1 was found to be strong, as it led to different types of deviations, which were especially noticeable in the "Fill in the gaps exercise" (cf. section 5.1.1) as well as the "Translation exercise" (cf. section 5.1.2). These two exercises turned out to be particularly useful in the identification of problem areas since they provided participants with instances of real language in context rather than isolated examples of word combinations. Informants' responses to both exercises have supplied a good source of information regarding their familiarity with the structures under study.

Informants' failure to produce appropriate lexical bundles with the nouns *agreement* and *conclusion* emerged as a recurrent deviation which was clearly influenced by participants' L1. Data drawn from the *HSC* corpus had revealed that strings of words like "*there is* (adjective) *agreement that*" and "*in conclusion*" had undergone a process of grammaticalisation and, thus, should be regarded as compositional sequences of words in which the abstract noun, by losing its original semantic content, forms part of an extended unit of meaning which has acquired a more grammatical function.

On the contrary, results from non-native speakers' production seem to indicate that informants were not aware of the peculiarities of the abovementioned lexical bundles. A total of nineteen informants failed to produce the structure *there is agreement that* and used other deviant constructions like the use of the passive (e.g. "**an agreement was done*") and the anticipatory "*it*" (e.g. "**it was produced an agreement*"), as a means of disguising authorial interpretations. These deviations[132] underpin the argument that informants' lack of collocational competence made them resort to word-for-word translations. This distinct tendency may also be related to the fact that, as stated in the Survey Content-Based Questions (cf. section 5.3), a 20,8% of informants claimed to write the first draft of their articles in Spanish and then translated them into English.

The analysis of informants' material has evidenced that L1 influence was also present in some deviations regarding the use of passive forms. Whereas the *HSC* corpus had shown a preference for the use of verb-abstract noun collocations in passive constructions (e.g. *conclusions were drawn*, *an agreement was reached*, *a decision must be made*), some deviant uses of both the active and the passive voice occurred in informants' production. In particular, their difficulty in translating the Spanish impersonal construction with the reflexive pronoun *se* (e.g. "*no se realizaron comparaciones*", "*se hicieron las mismas comparaciones*") in Exercise 3 (cf. section 5.1.2) has given rise to inappropriate structures where transfer from L1 is most noticeable (e.g. "**did not realise comparisons*", "**they made the same comparisons*", "**it was made the same comparisons*").

Further instances of L1 interference were traced in participants' misuse of three semantically similar adjectives: *little*, *short* and *small*. A lack of knowledge of the adjective collocates of the selected abstract nouns was observed in the following deviations with the adjective *short*: **short agreement*, **short comparison* and **short contribution*, where *little* would have been the most appropriate adjective collocate. The semantic similarity among the three adjectives as well as their similar translation equivalents in Spanish may have contributed to informants' deviant collocations.

---

[132] Notice their resemblance to the Spanish sequence "*se produjo un acuerdo*" (cf. section 5.1.2).

In addition, the analysis conducted in chapter 5 revealed informants' unawareness of the most common adjective collocates of the nouns in question. This assertion seems to be supported by a number of questionable adjective-noun combinations (e.g. ?*reliable agreement*, ?*reliable contribution*, ?*late agreement*, ?*late comparison*), which neither the *HSC* corpus nor the *BNC* gave account of.

Other types of deviation that emerged as particularly frequent referred to morphological aspects. Surprisingly, participants proved to have problems with the morphological form of many general items such as *comparison* (**comparation*, **comparition*, **comparision*), *biological* (**biologyc*), *significant* (**significative*), *substance* (**substans*) and *analyze* (**analize*). These deviant forms were not expected since all these terms are often used in medical publications and, thus, informants were supposed to be familiar with them. However, such inappropriate forms provide further indication of informants' tendency to transfer structures (morphological forms, in this case) from their L1.

All in all, three general conclusions can be drawn from the analysis of the worksheet of exercises concerning non-native speakers' command over proper collocability of the phraseology around abstract nouns in medical discourse. Firstly, their responses to the exercises have shown that being familiar with certain collocations does not necessarily imply knowing how to use them correctly. Although informants were able to successfully identify acceptable adjective + noun collocations and restricted verb-noun combinations in Exercise 1 (cf. section 5.1.3) and Exercise 4 (cf. section 5.1.4), they also failed to produce those collocations appropriately when a context was provided (Exercise 2 and 4; cf. sections 5.1.1 and 5.1.2, respectively). The second observation made when comparing the mismatch between the *HSC* corpus and the informants' production points to the fact that the distinctive features of the lexico-grammatical patterning of the five abstract nouns pose many difficulties for non-native speakers.

A further assumption can be made from the survey data. The results obtained have demonstrated that the grammatical and collocational patterns of frequent general abstract nouns[133] are liable to confusion for informants. This goes in line with their perceptions with regard to their writing process in English (cf. section 5.3). When asked to reflect on their main problem areas when writing their medical articles, they claimed that the use of general words and common expressions as well as the appropriate use of grammatical rules when constructing sentences were difficult aspects of their actual writing up process.

These findings hold a number of implications for the teaching and learning of collocations and phraseological conventions in the disciplinary area under discussion, which will be considered in the following section.

6.3 Implications of the study

Several implications can be drawn from the analysis conducted in this study. The investigation of the defining traits of the phraseological behaviour of abstract nouns in a health science corpus (*HSC*) and the subsequent verification that non-native speakers were unaware of the lexico-grammatical patterning of such units brings to the forefront the relevance of corpus-based studies not only in the characterisation of language but also in the teaching and learning of phraseology in specialised registers.

Recent theory and corpus-based studies (Gledhill 2000, Tognini-Bonelli 2001, Tsui 2005, Nesselhauf 2005, among others) have provided substantiation of the convenience of using collocation evidence obtained from textual corpora in EFL and ESP settings so as to help non-native speakers focus on slices of real language as well as high-frequent combinations of words. As shown in this research, empirical corpus evidence has been essential so as to identify the main patterns of word co-occurrence associated with the selected abstract nouns.

---

[133] Although the present research was aimed at exploring the use of abstract nouns, this assertion has also proven to be true for other general vocabulary items characteristic of medical discourse, as Exercise 4 has shown (cf. section 5.1.4).

This analysis has also revealed informants' lacking familiarity with the most salient collocational patterns and lexical bundles that tend to occur in combination with the analysed nouns. To this respect, it is worth underlining that the use of a corpus-based approach has served a vital role in the identification of the phraseology of such nouns in medical English.

The observations made by rooting informants' lexical and collocational production suggest the need of large-scale studies based on corpus data that give account of the idiosyncrasies of medical discourse. In Gledhill's view, "*much of the language involved in (…)* [*scientific English*][134] *is idiomatic and highly stereotypical in nature*" (Gledhill 2000a:116), so in this perspective the examination of language in use is seen as extremely beneficial in providing information about how words are actually used and their natural context of occurrence; that is, showing words, not in isolation, but in terms of their semantic preferences (Sinclair 1991).

As demonstrated in the present dissertation, the control of general words and common expressions frequently found in medical English is regarded as essential in encoding messages to show near-native linguistic competence. This is especially relevant for the discourse community under discussion. Being part of the international medical research community, they are committed to ensuring rapid dissemination of their research findings. This inevitably means that they need to be aware of the conventions as well as the good academic style characteristic of medical writing, so that their research articles are accepted for publication in the prestige journals of their various specialised fields.

An issue recurring throughout the study of the errors produced by informants is that most instances that were particularly problematic consisted of frequent everyday expressions in medical written production. This finding goes in line with Gledhill's assumption that "*non-technical lexical items in science writing are involved in highly specific and consistent grammatical systems*" (Gledhill

---

[134] square brackets added.

2000b:214) and, thus, they often constitute sources of confusion. In this context, informants' failure to produce appropriate collocations highlights the need to improve non-native speakers' competence with regard to the collocational patterning distinctive of scientific discourse.

As already pointed out, researchers must adhere to the appropriate collocations and the conventionalized phraseological properties of the written community they belong to, so that their potential readership is not distracted by deviant expressions and can thus read fluently and focus on the content. The present research aimed at developing this generalisation by comparing the attested mismatch between collocation evidence obtained from the *HSC* corpus and non-native speakers' collocational performance in the worksheet of exercises, especially designed for this study.

Informants' deviations concerning the lexico-grammatical patterns associated with abstract nouns have also indicated that the analysis of collocations and phraseology in scientific English deserve further attention. According to Gledhill (2000b), the notion of **phraseology** is central in the characterisation of scientific discourse:

> "The notion of phraseology implies much more than inventories of idioms and systems of lexical patterns. Phraseology is a dimension of language use in which patterns of wording (lexico-grammatical patterns) encode semantic views of the word, and at a higher level idioms and lexical phrases have rhetorical and textual roles within a specific discourse." (Gledhill 2000b:202)

Within this frame, the study of the recurrent lexico-grammatical patterning of words, rather than isolated terms, becomes a substantial part of the defining characteristics of a given discourse community. As Barnbrook (2007) puts it:

> "Language does not seem to operate on the basis of syntactic slots available for filling with minimal structural

> restrictions, but instead largely on the basis of preconstructed phrases composed largely of delexicalised frequent words." (Barnbrook 2007:193)

Given this view, that a high proportion of the language produced by native speakers appears to consist of prefabricated recurrent chains of words in which at least one of the elements has been delexicalised, large-scale corpus studies showing the linguistic phenomena observed in chapter 4 (i.e. delexicalisation, grammaticalisation, distribution of meaning across different units) are seen as extremely useful for the description of the language involved in specialised registers.

To this respect, results from this study on the collocational and colligational patterns of the nouns *conclusion*, *agreement*, *comparison*, *contribution* and *decision* in the *HSC* corpus appear to suggest that native speakers make use of frequent prefabricated chunks (e.g. restricted verb-noun collocations in which the verb has lost its primary meaning, lexical bundles which have undergone a gradual process of grammaticalisation) of which non-native speakers do not seem to be aware.

What can be inferred from their uncertainty over proper collocability around abstract nouns is that their written production would fail to resemble native-like naturalness and fluidity. It has been claimed (Gledhill 2000a/b; Tognini-Bonelli 2001; Wray 2002) that a good command of the collocational patterns associated with lexical units leads to a more effective communication among the members of a given discourse community:

> "Collocations in science writing are undoubtedly selected as the best ways of expressing certain ideas, although this selection does not mean that these expressions are the best, or the only possible selections, the selection is largely a feature of convention and acceptability within the discourse community." (Gledhill 2000a:133)

Bearing in mind, thus, that native production is characterised by the use of conventionalized phraseological units which shape the discourse of a particular community, non-native speakers must be able to use not only grammatically correct structures and select appropriate lexical items but also, most importantly, choose the right combinations of words so as to show near-native competence.

Another inference about informants' inappropriate usage of the abstract nouns examined in the worksheet of exercises is that most deviations are related to the defining traits of the lexico-grammatical patterning of the nouns in question; namely, delexicalisation, grammaticalisation, use of periphrastic structures in passive constructions and the use of classifiers as opposed to descriptors as attributive adjective collocates. These attested deviations reinforce the assumption that the lexical patterning around abstract nouns poses apparent difficulties for non-native speakers.

A further implication of this study that has emerged from the identification of the above problem areas is that the teaching of phraseology should be central in EFL and ESP learning contexts so as to improve non-native speakers' collocational competence in English. As Gledhill (2000a) claims:

> "the direct correspondence between lexis and grammar is now so pervasive that it is difficult to conceive of a general characterization of science writing or the design of teaching materials for the benefit of science writers which can afford to ignore phraseology as a central level of analysis." (Gledhill 2000a:130-131)

Recent studies (Hunston 1995; Cowie 1998; Granger 1998; Howarth 1996; Hunston & Francis 2000; Gledhill 2000; Kaszubski 2000; Wray 2002; Nesselhauf 2005, among others) have stressed the important role played by prefabricated units in language learning. As the present analysis has shown, these units constitute a considerable proportion of native speakers' linguistic production, so it seems reasonable to conclude that they should be given central attention in EFL teaching methods and materials[135]. However, in Nesselhauf's (2005) view, further studies that take into account the difficulties non-native speakers have with collocational patterns are needed:

> "Although collocations have received increasing attention in recent years (Granger 1998c:159; Howarth 1998a:30), we are still far from the development of a coherent methodology and even further from a wide-spread and systematic treatment of collocations in language teaching materials and syllabi." (Nesselhauf 2005:3)

The results from this study (cf. chapter 5) indicate that informants appeared to be familiar with some restricted verb-noun collocations, which were frequently used in the *HSC* corpus, such as *draw a conclusion*, *make a comparison* and *reach an agreement*. Similarly, they also produced appropriate adjective-noun collocates like *general agreement*, *statistical comparison* and *logical conclusion*. It seems also worth noting that these acceptable word combinations were mainly found in the matching exercises (cf. sections 5.1.3 and 5.1.4). However, the analysis of the other two exercises (i.e. "Fill in the gaps exercise" [cf. section 5.1.1] and "Translation exercise" [cf. section 5.1.2]), where informants were expected to produce sequences of words in a given context, highlighted informants' inability to produce appropriate restricted verb-noun collocations as well as lexical bundles, which are central to the creation of scientific discourse.

The above finding reveals that mere exposure to or familiarity with certain multiword units does not necessarily entail that users will know how to use them correctly when writing their English texts. This assumption underlines even more the adequacy of teaching collocations explicitly (Kaszubski 2000; Nesselhauf 2005) in order to help non-native speakers improve their written production and feel more confident when expressing themselves in English. As discussed in section 5.3, data from the Survey Content-Based Questions informed about participants' view of their main problem areas when producing their medical articles in English. Their views pointed chiefly to three main aspects:

---

[135] For further discussion on teaching methodologies and materials that give account of phraseology, the interested reader can see, foe example, Conrad 1999/2000; Scott & Tribble 2006; Laso & Giménez 2007/2008 and Walker 2008.

a) the use of general words and common expressions in English
b) the distinction between similar pairs of words
c) the use of appropriate grammatical rules when building up a text

These three aspects along with the attested deviations discussed in sections 5.1 and 5.2 justify the usefulness of providing informants with resources and tools especially designed to improve their collocational competence in English. Although the main aim of the present research was, on the one hand, exploring the lexico-grammatical patterns of abstract nouns in medical English and, on the other, testing informants' knowledge of and ability to produce appropriate word combinations, it was also decided to inquire participants about the resources and tools that could facilitate their writing process in English.

As to the tools learners suggested, they could be grouped into three main categories which reflect informants' personal perception of what resources they felt that could have a positive effect on their writing:

a) **lexical devices and on-line resources**: lexical databases of common errors, dictionaries of collocations and automatic translators
b) **human resources**: training courses on writing techniques and linguistic support on behalf of language specialists.
c) **learner's attitude**: further exposure to reading and writing.

Clearly, along with the explicit teaching of collocations in EFL and ESP contexts, the development of a corpus-based linguistic reference tool that takes account of the lexico-grammar of general vocabulary in scientific English would extensively contribute to bridge the gap between real language in use and non-native speakers' uncertainty over proper collocability in medical discourse. In the following section new research directions raised by the results obtained from this study as well as some suggestions for the creation of such a tool will be put forward.

6.4 New research directions

As stated in the previous section, the findings drawn from the analysis of both the *HSC* and the informants' production with regard to the phraseological behaviour of abstract nouns in medical English have given rise to several implications for the teaching and learning of phraseology in specific genres. However, it is also worth underlying that the analysis carried out in this study has also some limitations which will constitute my future research directions.

The present research has left many dimensions of the lexico-grammatical patterning of medical discourse that remain unexplored and need to be investigated further. Firstly, the creation of useful analytic reference tools that improve Spanish scientists' usage of multiword units in medical English seems to be a fruitful area of research. It has been long acknowledged (Granger 1998; Altenberg & Granger 2001; Nesselhauf 2005, among others) that disambiguating word senses and how different lexical units are combined to form chunks of language may be tough for non-native speakers. Furthermore, some specialised dictionaries (i.e. the *Elsevier's Enciclopaedic Dictionary of Medicine*, the *Oxford Dictionary of Biochemistry and Molecular Biology*) lack contextual information, which often leads to unsatisfactory look-up experiences. Therefore, a reference tool which provides information about collocations in scientific English could be beneficial for the Spanish scientific community.

To this respect, a research group at the Department of English of the University of Barcelona[136] is currently developing a lexical database of general terms in scientific English, *SciE-Lex*, whose main aim is helping the Spanish scientific community to use the appropriate collocational patterns in their articles.

---

[136] See reference in Footnote 25 (cf. chapter 3).

With the aim of storing and assembling the information that is considered to be essential for the target user of such a tool, *SciE-Lex* includes the most usual collocations of English non-technical words used in scientific language. This lexical database displays information about the meanings and the grammatical and collocational patterns of general terms frequently produced in scientific discourse. In later stages, *SciE-Lex* will also incorporate lexical bundles, as these compositional recurrent sequences have also been found to be problematic among non-native speakers (cf. chapter 5).

Some other resources publicly available, such as already existing technical and scientific monolingual dictionaries[137], merely provide terminological and encyclopaedic information or, in the case of bilingual and multilingual dictionaries, they simply offer translation equivalents without further information about the context on which the meaning of a given lexical entry depends. Consequently, the development of lexical databases like *SciE-Lex* and specialised dictionaries that take into account the lexico-grammatical patterning of lexical units and acknowledge that meaning is highly dependent on the context of co-occurrence of the word (Barnbrook 2007:191) is considered to be highly useful to improve non-linguist user's performance.

In addition, the present study has also made it evident that further corpus-based studies on collocations of both native and non-native production in scientific English are needed so as to contribute to a thorough characterisation of the discourse community as well as pinpoint non-native speakers' difficulties and design efficient collocation teaching materials. For a better understanding of non-native speakers' main problems as far as their written production is concerned, the examination of a large text corpus consisting of their published articles in English would also be beneficial in order to evaluate their collocational performance against native corpus evidence in the light of greater amounts of data.

---

[137] Cf. References

There is also an urgent need to explore the materials used in EFL and ESP contexts for the teaching of collocations in scientific English. Informants to the sample stated that they used to attend both ESP and content-based training courses in English on a regular basis (cf. section 3.4). However, very little is known about the materials, in the form of coursebooks, writing manuals, workbooks and grammars, among others, used for the teaching of English collocations in health sciences. Future work should be applied to the analysis of these sources for further insights into the way collocations are taught. As collocational patterns are essential to the construction of knowledge in the disciplinary area, it is firmly believed that research on lexico-grammar will provide the framework for the creation of improved collocation teaching materials.

As can be inferred from all these considerations, there is clearly much work to be done in the area of phraseology and lexical patterning of general vocabulary in specialist registers. This thesis was meant to be an attempt at identifying the pervasive idiosyncrasies of collocations in medical English and the difficulties that the lexico-grammar of abstract nouns poses for non-natives speakers. It is hoped, however, that this research work will stimulate larger-scale follow-up corpus-based studies that contribute to a better understanding and teaching of the complexity of multiword units in scientific discourse.

REFERENCES

Aarts, J. & Meijs, W. (Eds.) 1984. *Corpus Linguistics*. Amsterdam: Rodopi.

Adams Smith, D. E. 1984. "Medical discourse: aspect of author's comment" *The ESP Journal* 3: 25-36.

Adams Smith, D. E. 1987. "Variation in field-related genres" *ELR Journal* 1: 10-32.

Aitchison, J. & Lewis, D. M. 2003. "Polysemy and bleaching" in Nerlich, B., Todd, Z., Herman, V., Clarke, D. (Eds.) *Polysemy: Flexible Patterns in the Mind*. Berlin: Mouton de Gruyter, 253-265.

Allerton, D. J. 1982. *Valency and the English Verb*. London/ New York: Academic Press Inc.

Altenberg, B. 1993. "Recurrent verb-complement constructions in the London-Lund Corpus" in Aarts, J., de Haan, P. & Oostdijk, N. (Eds.) *English Language Corpora: Design, Analysis and Exploitation. Papers from the 13th International Conference on English Language Research on Computerized Corpora, Nijmegen 1992*. Amsterdam: Rodopi, 227-245.

Altenberg, B. & Tapper, M. 1998. "The use of adverbial connectors in advanced Swedish learners' written English" in Granger, S. (Ed.) *Learner English on Computer*. London & New York: Addison Wesley Longman, 80-93.

Altenberg, B. & Granger, S. 2001. "The grammatical and lexical patterning of *MAKE* in native and non-native student writing" *Applied Linguistics* 22/2: 173-194.

Aston, G. 2001 "Learning with corpora: An overview" in Aston, G. (Ed.) *Learning with corpora*. Houston: Athlstan, TX, 7-45.

Bahns, J. & Eldaw, M. 1993. "Should we teach EFL students collocations?" *System* 21/1: 101-114.

Baker, M. 1993. "Corpus Linguistics and Translation Studies –Implications and Applications" in Baker, M., Francis, G. & Tognini-Bonelli, E. (Eds.) *Text and Technology. In Honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins, 233-250.

Baker, M., Francis, G. & Tognini-Bonelli, E. 1993. *Text and Technology. In Honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins.

Baker, M. 1996. “Corpus-Based Translation Studies: the Challenges that Lie Ahead” in Somers, H. (Ed.) *Terminology, LSP and Translation: Studies in Language Engineering*, *in Honour of Juan Sager*. Amsterdam and Philadelphia: John Benjamins, 175-186.

Banks, D. 1994. “Clause Organization in the Scientific Journal Article” *Alsed-Lsp Newsletter*. Vol 17/2: 4-16.

Barnbrook, G. & Sinclair, J. M. 1995. “Parsing COBUILD entries” in Sinclair, J. M., Hoelter, M. & Peters, C. (Eds.) *The Languages of Definition: The Formalization of Dictionary Definitions for Natural Language Processing* Luxembourg: Office for Official Publications of the European Community, 13-58.

Barnbrook, G. 1996. *Language and Computers*. *A Practical Introduction to the Computer Analysis of Language*. Edinburgh: Edinburgh University Press.

Barnbrook, G. 2007. “Sinclair on collocation” *International Journal of Corpus Linguistics* 12:2: 183-199.

Becker, J. D. 1975. “The phrasal lexicon” in Nash-Webber, B. L. & Schank, R. (Eds.) *Theoretical Issues in Natural Language Processing*. Cambridge, M.A.: Bolt Beranek & Newman, 60-63.

Bednarek, M. 2007. "Local Grammar and Register Variation: Explorations in Broadsheet and Tabloid Newspaper Discourse" *ELR Journal* 1.1, http://www.ejournals.org.uk/ELR

Bernardini, S. & Zanettin, F. (Eds.) 2000. *I Corpora della Didattica della Traduzione –Corpus Use and Learning to Translate*. Bologna: CLUEB.

Berry, R., Asker, B., Hyland, K. and Lam, M. (Eds). 1999. *Language Analysis, Description and Pedagogy*. Language Centre: The Hong Kong University of Science and Technology.

Bhatia, V. 1993. *Analysing Genre: Language in Use in Professional Settings.* London: Longman.

Biber, D., Conrad, S. & Reppen, R. 1998. *Corpus linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S & Finegan, E. 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.

Biber, D., Conrad, S. & Cortes, V. 2003. "Lexical bundles in speech and writing: An initial taxonomy" in Wilson, A., Rayson, P. & McEnery, T. (Eds.) *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*. Frankfurt/Main: Peter Lang.

Biber, D., Conrad, S. & Cortes, V. 2004. "If you look at…: Lexical bundles in university teaching and textbooks" *Applied Linguistics* 25/3: 371-405.

Biber, D. 2006. *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Bloomfield, L. 1933. *Language*. New York: Holt & Company.

Boas, F. 1940. *Race, Language and Culture*. New York: The Macmillan Company.

Bogaards, P. 1996. "Dictionaries for learners of English" *International Journal of Lexicography* 9/4: 277-320.

Bresnan, J. 1982. *The Mental Representation of Grammatical Relations*. Cambridge, Massachusetts: MIT Press.

Burgschmidt, E. & Perkins, C. 1985. EB *Fechlerkartei Englisch: Phraseologie: Kollocationen-Phraseme-Idiome*. Braunschweig: Private publisher.

Butler, C. 1985. *Computers in Linguistics*. Oxford: Basil Blackwell.

Carter, R. 1998 (2nd edition). *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.

Charles, M. 2006. "The construction of Stance in Reporting Clauses: a Cross-disciplinary Study of Theses" *Applied Linguistics* 27/3: 492-518.

Chief, L-C., Huang, C-R., Chen, K-J., Tsai, M-C. & Chang, L-L. 2000. "What can near synonyms tell us" *International Journal of Computational Linguistics and Chinese Language Processing* 5/1: 47-60.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Ma: MIT Press.

Clear, J. 1987. “Computing” in Sinclair, J. M. (Ed.) *Looking Up: An Account of the Collins COBUILD Project*. London: Collins ELT, 41-61.

Coates, J. 1983. *The Semantics of the Modal Auxiliaries*. London: Croom Helm.

Conrad, S. 1999. “The importance of corpus-based research for language teachers” *System* 27: 1-18.

Conrad, S. 2000. “Will corpus linguistics revolutionize grammar teaching in the 21st century?” *TESOL Quarterly* 34: 548-560.

Conrad, S. 2005. “Corpus Linguistics and L2 Teaching” in Hinkel, E. (Ed.) *Handbook of Research in Second Language Teaching and Learning*. New York: Lawrence Erlbaum Associates, 393-409.

Cortes, V. 2002a. “Lexical bundles in freshman compositions” in Reppen, R., Fitzmaurice, S. M. & Biber, D. (Eds.) *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins, 131-145.

Cortes, V. 2004. “Lexical bundles in published and student disciplinary writing: Examples from history and biology” *English for Specific Purposes* 23: 397-423.

Cowie, A. P. 1981. “The treatment of collocations and idioms in learners’ dictionaries” *Applied Lingusitics* 2/3 (thematic issue): 223-235.

Cowie, A. P. 1988. “Stable and Creative Aspects of Vocabulary use” in Carter, R. & McCarthy, M. J. (Eds.) *Vocabulary and Language Teaching*. London: Longman, 126-137.

Cowie, A. P. 1992. “Multiword Lexical Units and Communicative Language Teaching” in Arnaud, P. & Bejoint, H. (Eds.) Vocabulary and Applied Linguistics. Basingstoke: Macmillan, 1-12.

Cowie, A. P. 1994. “Phraseology” in Asher, R. E. & Simpson, J. M. Y. (Eds.) *The Encyclopedia of Language and Linguistics*, Vol. 6: 3168-3171. Oxford: Pergamon.

Cowie, A. P. (Ed.) 1998. *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press.

Cowie, A. P. 1998. “Introduction” in Cowie, A. P. (Ed.) *Phraseology: Theory, analysis and applications*, Oxford: Oxford University Press, 1-20.

Cowie, A. P. 1999. *English Dictionaries for Foreign Learners*. Oxford: Oxford University Press.

Croft, W. & Cruise, A. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

Danielsson, P. 2003. "Automatic extraction of meaningful units from corpora: A corpus-driven approach using the word stroke" *International Journal of Corpus Linguistics* 8(1): 109-127.

DeCarrico, J. S. & Nattinger, J. R. 1988. “Lexical phrases for the comprehension of academic lectures” *English for Specific Purposes* 7: 91-102.

Deignan, A. 2005. *Metaphor and Corpus Linguistics*. Amsterdam & Philadelphia: John Benjamins.

Dörnyei, Z. 2007. *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Ellis, N. C. 2007. “Learned attention in language acquisition: Blocking salience, and cue competition”. Paper presented at the *EuroCogSci07, the Second European Cognitive Science Conference*, May 23-27, Delphi, Greece.

Ellis, N. C. 2008. “Phraseology: The periphery and the heart of language” in Granger, S. & Meunier, F. (Eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins, 1-13.

Fillmore, C. J. 1979. “On fluency” in Fillmore, C. J., Kempler, D. & Wang, W. S-Y. (Eds.) *Individual Differences in Language Ability and Language Behaviour*. NY: Academic Press, 85-101.

Fillmore, C. J., Kay, P. & O’Connor, M. C. 1988. “Regularity and Idiomacy in Grammatical Constructions: the case of *let alone*” *Language* 64: 501-538.

Fillmore, C. J. 1989. “Grammatical Construction Theory and the Familiar Dichotomies” in Dietrich, R. & Graumann, C. F. (Eds.) *Language Processing in Social Context*. Amsterdam-New York-Oxford-Tokyo: North-Holland Linguistic Series, 17-38.

Fillmore, C. J. 1992. “Corpus linguistics” or “Computer-aided armchair linguistics” in Svartvik, J. (Ed.) 1992. *Directions in Corpus Linguistics*.

Proceedings of the Nobel Symposium 82: Stockholm 4-8 August 1991. Berlin and New York: Mouton de Gruyter, 35-60.

Firth, J. R. 1957 [1951]. *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.

Flowerdew, J. 2003. "Signalling nouns in discourse" *English for Specific Purposes* 22: 329-346.

Flowerdew, J. 2007. "The non-Anglophone scholar on the periphery of scholarly publication" in *AILA Review. Linguistic inequality in scientific community today* 20: 14-27.

Francis, G. 1993. "A Corpus-Driven Approach to Grammar. Principles, Methods and Examples" in Baker, M., Francis, G. & Tognini-Bonelli, E. (eds.) *Text and Technology. In Honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins, 137-156.

Francis, G. 1995. "Corpus-driven grammar and its relevance to the learning of English in a cross-cultural situation" in Pakir, A. (Ed.) *English in Education: Multicultural perspectives*. Singapore: Unipress.

Francis, G., Hunston, S. & Manning, E. 1996. *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.

Francis, G., Manning, E. & Hunston, S. 1997. *Verbs: Patterns and Practice*. London: HarperCollins.

Francis, G., Hunston, S. & Manning, E. 1998. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London: HarperCollins.

Gabelentz, G. 1891. *Die Sprachwissenschaft. Ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: Weigel.

Gabrys-Biskup, D. 1992. "L1 influence on learners' renderings of English collocations: a Polish/German empirical study" in Arnaud, P. & Béjoint, H. (Eds.) *Vocabulary and Applied Linguistics*. London: Macmillan, 85-93.

Gläser, R. 1988. "The Grading of Idiomacity as a Presupposition for a Taxonomy of Idioms" in Hüllen, W. & Schulze, R. (Eds.) *Understanding the Lexicon: Meaning, sense and world knowledge in lexical semantics*. Tübingen: Max Niemeyer, 264-279.

Gledhill, C. 1995. “Collocation and genre analysis. The discourse function of collocation in cancer research abstracts and articles” *Zeitschrift für Anglistik und Amerikanistik* 1: 1-26.

Gledhill, C. 1996. “Science as a collocation. Phraseology in cancer research articles” in Botley, S., Glass, J., McEnery, T. & Wilson, A. (Eds.) *Proceedings of teaching and language corpora 1996: UCREL technical papers* Vol. 9, The University of Lancaster, UK, 108-126.

Gledhill, C. 2000a. “The discourse function of collocation in research article introductions” *English for Specific Purposes* 19/2: 115-135.

Gledhill, C. 2000b. *Collocations in science writing*, Gunter Narr, Tübingen.

Gouverneur, C. 2008. “The phraseological patterns of high-frequency verbs in advanced English for general purposes: A corpus-driven approach to EFL textbook analysis” in Granger, S. & Meunier, F. (Eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins, 223-243.

Granger, S. 1998. “Prefabricated patterns in advanced EFL Writing: Collocations and formulae” in Cowie, A. P. (Ed.) *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press, 145-160.

Granger, S. 2003. “The International Corpus of Learner English: A new resource for foreign language learning and teaching and second language acquisition research” *TESOL Quarterly* 37: 538-546.

Granger S., Paquot M. & Rayson P. 2006. “Extraction of multi-word units from EFL and native English corpora. The phraseology of the verb *make*” in Häcki Buhofer, A. & Burger, H. (Eds.) *Phraseology in Motion I: Methoden und Kritik. Akten der Internationalen Tagung zur Phraseologie (Basel, 2004)*. Baltmannsweiler: Schneider Verlag Hohengehren, 57-68.

Granger, S. & Meunier, F. (Eds.) 2008. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins.

Greenbaum, S. 1970. *Verb-Intensifier Collocations in English. An Experimental Approach*. The Hague: Mouton.

Greenbaum, S. 1996. *The Oxford English Grammar*. Oxford: Oxford University Press.

Gries, S. T. & Stefanowitsch, A. (Eds.) 2006. *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*. Berlin-New York: Mouton de Gruyter.

Groom, N. 2005. “Pattern and meaning across genres and disciplines: An exploratory study” *Journal of English for Academic Purposes*. Vol. 4 Issue 3: 257-277.

Gross, M. 1993. “Local grammars and their representation by finite automata” in Hoey, M. (Ed.) *Data, Description, Discourse*. London: Harper Collins.

Hadley, G. 1997. *Sensing the Winds of Change: An introduction to Data-Driven Learning*. http://web.bham.ac.uk/johnstf/winds.htm.

Halliday, M. A. K. 1961. *Categories of the Theory of Grammar*. Department of English Language and General Linguistics Monographs. Edinburgh: Edinburgh University Press, 241-292.

Halliday, M. A. K. & Hasan, R. 1976. *Cohesion in English*. London: Longman.

Halliday, M. A. K. 1992. “Language as system and language as instance: the corpus as a theoretical construct” in Svartvik, J. (Ed.) *Directions in Corpus Linguistics.* Berlin: Mouton, 61-77.

Halliday, M. A. K. 1993. “Language in a changing world” Occasional paper no. 13 *Applied Linguistics Association of Australia*.

Halliday, M. A. K., & Martin, J. 1993*. Writing science: literacy and discourse power*. London: Falmer Press.

Halliday, M. A. K. 1998. “Things and Relations. Regrammaticising Experience as Technical Knowledge” in Martin, J. R. & Veel, R. (Eds.) *Reading Science: Critical and Functional Perspectives on Discourses of Science*. London: Routledge, 185-235.

Halliday, M. A. K., Teubert, W. Yallop, C., & Cěrmáková, A. 2004. *Lexicology and Corpus Linguistics*. London: Continuum.

Hasan, R. 1987/1996. “The grammarian’s dream: Lexis as most delicate grammar” in Halliday, M. A. K. & Fawcett (eds.) *New developments I n Systemic Linguistics. Vol. 1: Theory and Description*. London: Pinter (1987); and in Cloran, C., Butt, D. & Williams, G. (Eds.) *Ways of Saying: Ways of Meaning*. London: Cassell, 73-103.

Hausmann, F. J. 1989. “Le dictionnaire de collocations” in Hausmann, F. J., Wiegand, H. E. & Zgusta, L. (Eds.) *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Berlin: Mouton de Gruyter, 1010-1019.

Herbst, T. 1996. “What are collocations: sandy beaches or false teeth?” *English Studies* 77/4: 379-393.

Hoey, M. P. 1991. *Patterns of lexis in text*. Oxford: Oxford University Press.

Hoey, M. P. 2005. *Lexical priming: A new theory of words and language*. London: Routledge.

Hopper, P. J. & Traugott, E. C. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Hornby, A. S. 1954. *A Guide to Patterns and Usage in English*. London: Oxford University Press.

Howarth, P. A. 1996. *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Tübingen: Max Niemeyer Verlag.

Howarth, P. A. 1998a. “The phraseology of learners’ academic writing” in Cowie, A. P. (Ed.) *Phraseology: Theory, analysis and applications*. Oxford: Clarendon Press, 161-186.

Howarth, P. A. 1998b. “Phraseology and second language proficiency” *Applied Linguistics* 19/1: 24-44.

Huddleston, R. 1971. *The Sentence in Written English. A Syntactic Study Based on an Analysis of Scientific Texts*. Cambridge: Cambridge University Press.

Huddleston, R. & Pullum, G. K. 2005. *A Student’s Introduction to English Grammar*. Cambridge: Cambridge University Press.

Hunston, S. 1995. “Grammar in teacher education: The role of a corpus” *Language Awareness* 4/1: 15–31.

Hunston, S. & Francis, G. & Manning, E. 1997. “Grammar and vocabulary: showing the connections” *English Language Teaching Journal* 51/3: 208-216.

Hunston, S. & Francis, G. 1998. “Verbs observed: A Corpus-Driven Pedagogic Grammar” *Applied Linguistics* 19/1: 45-72.

Hunston, S. & Francis, G. 2000. *Pattern Grammar*. Amsterdam: John Benjamins.

Hunston, S & Sinclair, J. M. 2000. “A local grammar of evaluation” in Hunston, S. & Thompson, G. (Eds.) *Evaluation in Text: Authorial Stance and the Construction of Discourse*, Oxford: Oxford University Press, 75-101.

Hunston, S. 2002. “Pattern grammar, language teaching, and linguistic variation. Applications of a corpus-driven grammar” in Reppen, R., Fitzmaurice, S. M. & Biber, D. (Eds.) *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins, 167-183.

Hunston, S. 2007. “Semantic prosody revisited” *International Journal of Corpus Linguistics* 12/2: 249-268.

Hyland, K. 2008. “As can be seen: Lexical bundles and disciplinary variation” *English for Specific Purposes* 27: 4-21.

Jackson, H. & Zé Amvela, E. 2000. *Words, Meaning and Vocabulary. An introduction to modern English lexicology*. London: Cassell.

Johansson, S. & Hofland, K. 1994. “Towards an English-Norwegian Parallel Corpus” in Percy, C., Meyer, C. F. and Lancashire, I. (Eds.) *Synchronic Corpus Linguistics*. Amsterdam: Rodopi, 25-37.

Johns, T. 1991. “Should you be persuaded: Two examples of data-driven learning” in Johns, T. & King, P. (Eds.) *ELR Journal* Vol. 4. Birmingham: Birmingham University Press, 1-16.

Johns, T. 1994. “From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning” in Odlin, T. (Ed.) *Perspectives on Pedagogical Grammar*. New York: Cambridge University Press, 293-313.

Kaszubski, P. 1998. “Learner corpora: The cross-roads of linguistic norm” http://users.ox.co.ac.uk/~talc98/kaszubski.htm.

Kaszubski, P. 2000. *Selected Aspects of Lexicon, Phraseology and Style in the Writing of Polish Advanced Learners of English: A Contrastive, Corpus-Based Approach*. http://www.staff.amu.edu.pl/~przemka/rsearch.html#PhD

Kellerman, E. 1995. “Crosslinguistic influence: Transfer to nowhere?” *Annual Review of Applied Linguistics* 15: 125-150.

Kjellmer, G. 1984. “Some thoughts on collocational distinctiveness” in Aarts, J. & Meijs, W. (Eds.) *Corpus Linguistics*. Amsterdam: Rodopi.

Kjellmer, G. 1990. “Patterns of collocability” in Aarts, J. & Meijs, W. (Eds.) *Theory and practice in Corpus Linguistics*. Amsterdam: Rodopi.

Krishnamurthy, R. 2000. “Collocation: from *silly ass* to lexical sets” in Heffer, C., Sauntson, H. & Fox, G. (Eds.). *Words in Context: A Tribute to John Sinclair on his Retirement*. Birmingham: University of Birmingham, 31-47.

Langacker, R. W. 1991. *Foundations of cognitive grammar: Vol. 2: Descriptive Application*. Stanford, CA: Stanford University Press.

Langacker, R. W. 2000a. “A Dynamic Usage-Based Model” in Barlow, M. & Kemmer, S. (Eds.) Usage-based models of language. Stanford, CA: CSLI Publications, 1-63.

Langacker, R. W. 2000b. “Grammar and conceptualization” *Cognitive Linguistics Research* 14. Mouton De Gruyter.

Laso, N. J. & Verdaguer, I. 2005. “An invitation to explore conclusions and its restricted collocations” *Specific* 2: 47-54.

Laso, N. J. & Giménez, E. 2007. “Bridging the Gap between Corpus Research and Language Teaching” in Periñán, C. (Ed.) *Revisiting Language Learning Resources*. Newcastle: Cambridge Scholars Publishing, 49-64.

Laso, N. J. & Giménez, E. 2008. “Exploring the Impact of Concordancers in the Teaching of English Lexicology and Morphology” in Pérez, L., Pizarro, I. & González, E. (Eds.) *Estudios de Metodología de la Lengua Inglesa (IV)*. Valladolid: Universidad de Valladolid, Secretariado de Publicaciones e Intercambio Editorial, 281-290.

Lee, D. 2001. “Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle” *Language Learning & Technology* 5/3: 37–72.

Lee, D. & Swales, J. 2006. “A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora” *English for Specific Purposes* 25/1: 56-75.

Leech, G. 1991. “The state of art in corpus linguistics” in Aijmer, K. and Altenberg, B. (Eds.) *English Corpus Linguistics*. London: Longman, 8-29.

Leech, G. 1992. “Corpora and theories of linguistic performance” in Svartvik, Jan (Ed.) *Directions in Corpus Linguistics*. Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991 (= Trends in Linguistics. Studies and Monographs 65): Berlin/New York: Mouton de Gruyter, 105-122.

Lewis, M. 1993. *The Lexical Approach: The State of ELT and a Way Forward*. Hove: LTP.

Lewis, M. 1996. “Implications of a lexical view of language” in Willis, D. and Willis, J. (Eds.) *Challenge and Change in Language Teaching*. Oxford: Heinemann, 10-16.

Lewis, M. 2000. “Materials and resources for teaching collocation” in Lewis, M. (Ed.) *Teaching Collocation. Further Developments in the Lexical Approach*. Hove: Language Teaching Publications.

Luzón, M. J. 2000. “Collocational frameworks in medical research papers: A genre-based study” *English for Specific Purposes* 19: 63-86.

Malcolm, L. 1987. “What rules govern tense usage in scientific articles?” *English for Specific Purposes* 6: 31-44.

Mandelbaum, D. G. (Ed.) 1949. *Selected Writings of Edward Sapir in Language, Culture and Personality*. Berkeley: University of California Press.

Master, P. 1991. “Active verbs with inanimate subjects in scientific prose” *English for Specific Purposes* 10: 15-33.

May Fan, C. 1999. “An Investigation into the Pervasiveness of Delexical Chunks in Authentic Language Use and the Problems they present to L2 learners” in Berry et al. (Eds.), 162-175.

McEnery, T. & Wilson, A. 1996. Corpus Linguistics. Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R. & Tono, Y. 2006. *Corpus-Based Language Studies. An advanced resource book*. London & New York: Routledge.

Mel’čuk, I. A. 1998. “Collocations and lexical functions” in Cowie, A. P. (Ed.) *Phraseology: Theory, analysis and applications*, Oxford: Oxford University Press, 23-53.

Meunier, F. & Granger, S. 2008. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins.

Meyer, C. F. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Milton, J. & Freeman, R. 1996. “Lexical variation in the writing of Chinese learners of English” in Percy, C. E., Meyer, C. E. & Lancashire, I. (Eds.) *Synchronic Corpus Linguistics. Papers from the Sixteenth International Conference on English Language Research on Computerized Corpora* (ICAME, 16) Amsterdam & Atlanta: Rodopi, 121-131.

Moon, R. 1992. “Textual aspects of fixed expressions in learners’ dictionaries” in Arnaud & Bejoint (Eds.) *Vocabulary and Applied Linguistics*. Basingstoke: Macmillan, 13-27.

Moon, R. 1994. “The Analysis of Fixed Expressions in Text” in Coulthard, M. (Ed.) *Advances in Written Text Analysis*. London: Routledge, 117-135.

Moon, R. 1998. *Fixed Expressions & Idioms in English. A Corpus-Based Approach*. Oxford: Oxford University Press.

Moon, R. 2007a. “Phraseology in general monolingual dictionaries” in Burger, H., Dobrovol’skij, D., Kühn & Norrick, N. R. *An International Handbook of Contemporary Research*. Berlin/New York: Walter de Gruyter, 909-918.

Moon, R. 2007b. “Corpus linguistic approaches with English corpora” in Burger, H., Dobrovol’skij, D., Kühn & Norrick, N. R. *An International Handbook of Contemporary Research*. Berlin/New York: Walter de Gruyter, 1045-1059.

Moon, R. 2007c. “Sinclair, lexicography, and the Cobuild Project. The application of theory”. *International Journal of Corpus Linguistics* 12/2: 159-181.

Myers, G. 1989. “The Pragmatics of Politeness in Scientific Articles” in *Applied Linguistics* 10/1: 1-35.

Nattinger, J. R. & DeCarrico, J. S. 1989. “Lexical phrases, speech acts and teaching conversation” in Nation, P. & Carter, R. (Eds.) *AILA Review 6: Vocabulary Acquisition* Amsterdam: AILA, 118-139.

Nattinger, J. R. & DeCarrico, J. S. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Neff van Aertselaer, J. 2008. “Contrasting English-Spanish interpersonal discourse phrases” in Granger, S. & Meunier, F. (Eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins, 85-99.

Nesselhauf, N. & Tschichold, C. 2002. “Collocations in CALL: An investigation of vocabulary-building software for EFL.” *Computer Assisted Language Learning* 15/3: 251-279.

Nesselhauf, N. 2004. “What are collocations?” in Allerton, D. Nesselhauf, N. & Skandera, P. (Eds.) *Phraseological Units: Basic Concepts and Their Application*. Basel: Schwabe, 1-21.

Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.

Nevalainen, T. & Ramoulin-Brunberg, H. (Eds.) 1996. *Sociolinguistics and language history. Studies based on the corpus of early English correspondence.* Amsterdam & Atlanta, GA: Rodopi.

Newman, J. & Rice, S. 2006. “Transitivity schemas of English EAT and DRINK in the BNC” in Gries, S. & Stefanowitsch, A. (Eds.) *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*. Berlin-New York: Mouton de Gruyter, 225-260.

Noguchi, J. 2006. *The Science Review Article: An Opportune Genre in the Construction of Science*. Bern: Peter Lang.

Nuccorini, S. (Ed.) 2002. *Phrases and Phraseology – Data and Descriptions.* Bern: Peter Lang.

Oakes, M. 1998. *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Oakey, D. 2002a. “A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines” in Reppen, R., Fitzmaurice, S. M. and Biber, D. (Eds.) *Using corpora to Explore Linguistic Variation*. Philadelphia: John Benjamins, 111-129.

Oakey, D. 2002b. “Lexical Phrases for Teaching Academic Writing in English: Corpus Evidence” in Nuccorini, S. (Ed.) *Phrases and phraseology – data and descriptions*. Bern: Peter Lang, 85-105.

Palmer, H. E. 1933. *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.

Paquot, M. 2008. “Exemplification in learner writing: A cross-linguistic perspective” in Granger, S. & Meunier, F. (Eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins, 100-120.

Partington, A. 1993. “Corpus Evidence of Language Change. The Case of the Intensifier” in Baker, M., Francis, G. & Tognini-Bonelli, E. (Eds.) *Text and Technology. In Honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins, 177-192.

Partington, A. 1998. *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam: John Benjamins.

Partington, A. & Morley, J. 2004. “From frequency to ideology: investigating word and cluster / bundle frequency in political debate” in Lewandowska-Tomaszczyk, B. (Ed.) *Practical Applications in Language and Computers-PALC2003*. Frankfurt/Main: Peter Lang, 179-192.

Pawley, A. & Syder, F. H. 1983. “Two puzzles for linguistic theory: Nativelike selection and nativelike fluency” in Richards, J. C. & Schmidt, R. W. (Eds.) *Language and communication*. London: Longman, 191-225.

Pawley, A. 1985. “On speech formulas and linguistic competence.” *Lenguas Modernas* 12: 84-104.

Pecman, M. 2008. “Compilation, formalisation and presentation of bilingual phraseology” in Granger, S. & Meunier, F. (Eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins, 203-221.

Peters, A. M. 1983. *The Units of Language Acquisition*. Cambridge: Cambridge University Press.

Quirk, R., Greenbaum, S., Leech, G & Svartvik, J. 1972. *A Grammar of Contemporary English*. London: Longman.

Quirk, R., Greenbaum, S., Leech, G & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Renouf, A. 1987. “Corpus Development” in Sinclair, J. McH. (Ed.) *Looking Up: An Account of the COBUILD Project*. London: Collins, 1-40.

Renouf, A., & Sinclair, J. M. 1991. “Collocational frameworks in English” in Aijmer, K. & Altenberg, B. *English Corpus Linguistics*. London: Longman, 128-144.

Reppen, R., Fitzmaurice, S. M. & Biber, D. (Eds.) 2002. *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins.

Rissanen, M. 2000. “The World of English Historical Corpora: from Caedmon to the Computer Age” *Journal of English Linguistics* 28/1: 7-20.

Salager-Meyer, F. 1992. “A text-type and move analysis study of verb, tense and modality distribution in medical English abstracts” *English for Specific Purposes* 11: 93-113.

Salager-Meyer, F. 1994. “Hedges and textual communicative function in medical English written discourse” *English for Specific Purposes* 13: 149-170.

Scott, M. & Tribble, C. 2006. *Textual patterns*. Amsterdam: John Benjamins.

Schmied, J. & Schäffler, H. 1996. “Approaching Translationese Through Parallel and Translation Corpora” in Percy, C., Meyer, C. F. and Lancashire, I. (Eds.) *Synchronic Corpus Linguistics*. Amsterdam: Rodopi, 41-55.

Schmitt, N. (ed.) 2004. *Formulaic sequences*. Amsterdam: John Benjamins.

Scott, M. 1996. *Wordsmith Tools 4*. Oxford: Oxford University Press.

Simpson, R. C. 2004. “Stylistic Features of Academic Speech: The Role of Formulaic Expressions” in Connor, U. & Upton, T. A. (Eds.) *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, 37-64.

Sinclair, J. M. 1987a. “Collocation: a progress report” in Steele, R. & Threadgold, T. (Eds.) *Language Topics: Essays in Honour of Michael Halliday*. Amsterdam: John Benjamins, 319-332.

Sinclair, J. M. (Ed.) 1987b. *Looking Up: An Account of the Collins COBUILD Project*. London: Collins ELT.

Sinclair, J. M. 1990. (Ed-in-Chief). *Collins Cobuild English Grammar*. London: Collins.

Sinclair, J. M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. M. 1992a. "Trust the Text: the Implications are Daunting" in Davies, M. & Ravelli, L. (Eds.) *Advances in Systemic Linguistics. Recent Theory and Practice*. London and New York: Pinter Publishers, 5-19.

Sinclair, J. M. 1992b. "The Automatic Analysis of Corpora" in Svartvik, J. (Ed.) *Directions in Corpus Linguistics, Proceedings of Nobel Symposium 82*, Stockholm 4-8 August 1991, Berlin and New York: Mouton de Gruyter, 379-397.

Sinclair, J. M. 1997. "Corpus Linguistics at the Millenium" in Kohn, J., Ruschoff, B & Wolf, D. (Eds.) *New Horizons in CALL*. Szombathely: Bersenyi Daniel College.

Sinclair, J. M. 2004. *Trust the text: Language, corpus and discourse*, London: Routledge.

Singleton, D. 2000. *Language and the Lexicon. An Introduction*. London: Arnold.

Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. New York: Blackwell.

Swales, J. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Sweetser, E. 1988. "Grammaticalization and semantic bleaching" in Axmaker, S., Jaisser, A. & Singmaster, H. (Eds.) *Proceedings of the 14th annual meeting of the Berkeley Linguistics Society*. Berkeley, California: Berkeley Linguistics Society, 389-404.

Teubert, W. & Cěrmáková, A. 2004. "Directions in corpus linguistics" in Halliday, M. A. K., Teubert, W. Yallop, C., & Cěrmáková, A. (Eds.) *Lexicology and Corpus Linguistics*. London: Continuum, 113-165.

Thomas, S. & Hawes, T. P. 1994. "Reporting verbs in medical journal articles." *English for Specific Purposes* 13: 129-148.

Tognini-Bonelli, E. 2000. "Corpus Classroom Currency." *Darbai ir Dienos* 24: 205-243. (http://donelaitis.vdu.lt/publikacijos/bonelli.pdf)

Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.

Tsui, A. B. M. 2004. “What Teachers Have Always Wanted to Know – and How Corpora Can Help.” In John Sinclair (Ed.), *How to Use Corpora in Language* Teaching. Amsterdam: John Benjamins, 39–61.

Tsui, A.B.M. 2005. “ESL Teachers' Questions and Corpus Evidence.” *International Journal of Corpus Linguistics* 10/3: 335-356.

Van der Wouden, T. 1997. *Negative Contexts. Collocation. Polarity and Multiple Negation*. Routledge: London.

Varttala, T. 1999. “Remarks on the Communicative Function of Hedges in Popular Scientific English and Specialist Research Articles.” *English for Specific Purposes* 18/2: 177-200.

Verdaguer, I. 2003. “Collocations in Scientific Language: a contrastive study” in *Studies in Contrastive Linguistics. Proceedings of the 3rd International Contrastive Linguistics Conference*, 633-639.

Verdaguer, I. & Laso, N. J. 2006. “Delexicalisation in a Corpus of Scientific English” in Hornero, A., Luzón, M. J. & Murillo, S. (Eds.) *Corpus Linguistics*. Bern: Peter Lang, 417-428.

Walker, C. 2008. “Factors which influence the process of collocation”. In Boers, F. & Lindstromberg, S. *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*. Berlin/New York: Mouton de Gruyter.

West, G. K. 1980. “*That*-nominal constructions in traditional rhetorical divisions of scientific research papers.” *TESOL Quarterly* 14: 483-489.

Widdowson, H. G. 1979. *Explorations in Applied Linguistics*. Oxford: Oxford University Press.

Williams, I. A. 1996. “A Contextual Study of Lexical Verbs in Two Types of Medical Research Article” in *English for Specific Purposes*. 15/3: 175-197.

Williams, G. 1998. Collocational Networks: Interlocking Patterns of Lexis in a Corpus of Plant Biology Research Articles. *International Journal of Corpus Linguistics*, 3/1: 151–171.

Williams, G. 1999. “Looking in before looking out: Internal selection criteria in a corpus of plant biology” *Papers in Computational Lexicography. Proceedings of Complex ’99*. Hungary: Budapest, 195–204.

Williams, G. 2002. “In search of representativity in specialised corpora: Categorisation through collocation” *International Journal of Corpus Linguistics* 7/1: 43-64.

Williams, G. 2005. “Challenging the native-speaker norm: a corpus-driven analysis of scientific usage” in Barnbrook, G., Danielsson, P. & Mahlberg, M. (Eds.) *Meaningful Texts. The Extraction of Semantic Information from Monolingual and Multilingual Corpora*. London/ New York: Continuum, 115-127.

Willis, D. 1990. *The Lexical Syllabus. A New Approach to Language Teaching*. London & Glasgow: Collins.

Wray, A. 1999. “Formulaic language in learners and native speakers.” *Language Teaching* 32: 213-231.

Wray, A. & Perkins, M. 2000. “The functions of formulaic language.” *Language & Communication* 20: 1-28.

Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Zhang, X. 1993. *English collocations and their effect on the writing of native and non-native college freshmen*. Unpublished doctoral dissertation, Indiana University of Pennsylvania, Pennsylvania, USA.

Dictionaries

Macmillan English Dictionary for Advanced Learners (with CD), ed.-in-chief Michael Rundell. Oxford: Macmillan, 2007.

Longman Dictionary of Contemporary English, Fourth Edition, ed. Michael Rundell. Harlow: Longman, 2005.

Oxford Collocations Dictionary for Students of English. Oxford: Oxford University Press, 2002.

Oxford Dictionary of Biochemistry and Molecular Biology, eds. A. D. Smith et al. Oxford: Oxford University Press, 1997.

Cambridge International Dictionary of English, ed. Paul Procter. Cambridge: Cambridge University Press, 1995.

Medical Dictionary in Six Languages, compiled by Bert Spilker. NY: Raven Press, 1995.

Elsevier's Encyclopaedic Dictionary of Medicine. Part C. Biology, Genetics and Biochemistry in five languages, compiled by A. F. Dorian. Amsterdam/Oxford/N Y/Tokyo, 1989.

Collins Cobuild English Dictionary, ed. John Sinclair. London: HarperCollins, 1987.

Appendix 1

Proyecto de investigación: Tesis doctoral *"A Corpus-Based Study of the Phraseological Behaviour of Abstract Nouns in Medical English -A Needs Analysis of a Spanish Medical Community"*

Fichas de ejercicios (*Worksheets*)

Investigadora:
Natalia Judith Laso Martín
Departamento de Filología Inglesa y Alemana
Facultad de Filología, Universidad de Barcelona (UB)
Gran Via Corts Catalanes, 585 08007 Barcelona
Tel.: 93 4035674
E-mail: njlaso@ub.edu

Appendix 1

Full Name:
Nationality:
Age Group (please circle): 25-30 30-40 40-50 +50
University studies (Degree): ____________________________________
Graduated from the University of ________________________ in ______________.

1. How long have you been studying English?
   ____
2. Have you ever lived abroad –in an English speaking country-? If so, how long?
   ____
3. How long have you been working in this hospital? What is your *MIR* speciality (field / research work)?
   ____
   ____
4. Have you taken any course related to your speciality in English? When? What was the course about?
   ____
   ____
5. Do you ever read academic papers / scientific articles in English? If so, how often? Which books / journals, etc. do they come from?
   ____
   ____
   ____
6. Do you usually write scientific papers in English? Which journals do you usually publish in?
   ____
   ____
   ____
7. When writing a paper in English, what is your writing process:
   a) first you write your contribution in your mother tongue and then translate it into English.
   b) you make a draft in English and then look up unknown words or expressions.
   c) you ask for some expert advice before handing in your paper.
   d) Other (Please specify): ____________________________________
   ____

Appendix 1

8. What are your main problems when writing a paper in English? Rank the following aspects in order of difficulty:

EASY DIFFICULT

1 2 3 4 5

- Specific (medical) terminology in English
- General words / expressions most commonly used in this type of publications
- Distinction between similar pairs of words (i.e. *do* / *make*)
- Distribution of paragraphs
- Grammatical rules when constructing sentences
- Others (Please specify): ______________________________
  ____

9. Can you think of any tool / resource to help you improve your writing of scientific articles in English (e.g., a lexical database of the most frequent general words / expressions in academic writing, specialized dictionaries, etc.)?
   ____
   ____
   ____

10. How long did you spend completing this questionnaire?
    ____

THANK YOU very much for your HELP and TIME devoted to answering this questionnaire!

Appendix 2[*] Survey Question 10: Informants’ time estimates in completing the questionnaire

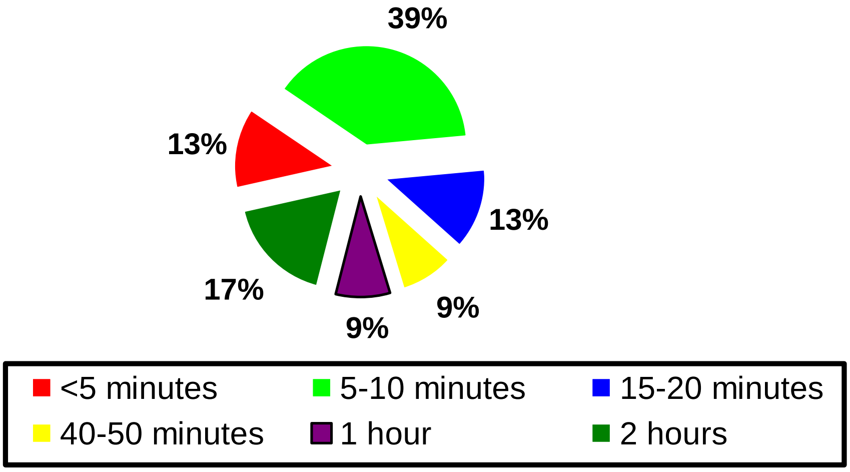


Figure 66 Survey Question 10: Informants’ time estimates in completing the questionnaire

---

[*] Notice only twenty-three informants answered this question. **HN8** provided no information about the time spent completing the questionnaires. Thus, hers was not taken into account in this Figure.

Appendix 3

FORMULARIO de CONSENTIMIENTO para UTILIZACIÓN de DATOS

Proyecto de investigación: Tesis doctoral ***"Delexicalisation and Semantic Bleaching of Noun Patterns in a Corpus of Scientific English"***[1]

Investigadora:
Natalia Judith Laso Martín
Departamento de Filología Inglesa y Alemana
Facultad de Filología, Universidad de Barcelona (UB)
Gran Via Corts Catalanes, 585 08007 Barcelona
Tel.: 93 4035674
E-mail: njlaso@ub.edu

Descripción del estudio:

Para la elaboración de la tesis doctoral arriba referenciada, la investigadora ha confeccionado unas fichas de ejercicios que serán distribuidas entre l@s participantes en este estudio. Los datos resultantes de dicho estudio serán analizados posteriormente y podrán ser objeto de publicaciones futuras.

Protección de datos:

La investigadora se compromete a preservar el anonimato de l@s participantes. Para ello, las fichas de ejercicios se analizarán sin incluir ningún dato personal que pudiese identificar a l@s sujetos. A cada participante se le asignará un código con el que se podrá hacer referencia a los datos obtenidos en sus cuestionarios. Una vez los datos hayan sido recogidos y se haya asignado un código a cada participante, la investigadora siempre utilizará dicho código para referirse a los datos concretos de cada sujeto.

Acceso a los resultados por parte de los sujetos:

Cualquier resultado y/o datos que se deriven de la participación de los sujetos en el presente estudio podrán ser facilitados a tod@s aquell@s que firmen el presente formulario de consentimiento.

Derechos de l@s participantes:

Se entiende que l@s sujetos pueden ejercer su derecho a negarse a contestar alguna pregunta del cuestionario si así lo estiman oportuno.

Nombre y firma del/ de la participante:

____

____

Nombre y firma de la investigadora:

Natalia Judith Laso Martín

____

---

[1] former title

WORKSHEETS

1. **Match each of the adjectives on the left with a suitable noun from the facing column. Try to use your intuition. Please notice that more than one combination is possible and some of the adjectives can be left out.**

logical clear firm contradictory valid little reliable short late statistical far-reaching substantial general significant large fashionable small broad unforeseeable	CONCLUSION
	AGREEMENT
	COMPARISON
	CONTRIBUTION
	DECISION

2. **Predict the words which are missing. Complete the sentences below with a suitable word.**

1. The research unit, within four days, ____________ the conclusion that the government was not giving enough.
2. Comparisons between mice have been ____________.
3. Such a conclusion is ____________ by our observations.
4. The data presented here are not ____________ agreement ____________ the model of Studier and Bandyopadhyay.
5. Further trials are needed before any ____________ can be drawn about the protocol's efficacy.
6. Clinical examination and biochemical screening ____________ an important contribution to the detection of abnormalities.

7. They ____________ the final agreement that clinical examination for defects in hips, vision and hearing, and other congenital abnormalities is less well founded on scientific evidence.
8. Bruce Weir has ____________ many important contributions to population genetic inference theory.
9. These results ________ ________ some important conclusions.
10. To assess the cost effectiveness of universal antenatal HIV screening ___________ comparison ____________ selective screening in the UK.
11. Despite these conservative features, our analysis broadly ____________ the recent decision by the Department of Health that HIV testing should be offered universally.
12. ____________ is general agreement ____________ the primitive host cell was anaerobic.
13. It was not possible to ____________ the relative contribution of this glucose derived directly from fructose.
14. They will certainly ____________ the decision the group has made.
15. These multivitamins do not appear to ____________ significant contributions to the parameters tested.

3. Provide a suitable translation into English for the following sentences.

1. Se produjo un acuerdo casi unánime entre ambos métodos.

2. No se realizaron comparaciones para determinar diferencias de género ni de edad.

3. En resumen, los resultados de nuestros estudios sugieren que la vaginitis bacteriana no afecta en absoluto a la concepción.

4. Las sustancias analizadas han contribuido de forma poco significativa a disminuir los síntomas de la enfermedad.

5. Antes de que se pueda llegar a conclusiones fiables, los investigadores deben aprender mucho más acerca de la función genética.

6. Se hicieron las mismas comparaciones en los tratamientos de control.

____

____

7. Estas conclusiones biológicas han sido justificadas por múltiples tipos de análisis estadísticos.

____

____

8. Varias organizaciones son las responsables de la toma de decisiones por lo que respecta a la seguridad de las transfusiones.

____

____

4. *Make*, *draw*, *reach*, *lead to*, *take* and *do* are highly-frequent verbs in English. They tend to combine with a wide range of words to form fixed expressions. Combine the words below with the verbs provided in the chart.

action	knowledge	a mistake
sb's temperature	a suggestion	a contribution
damage	a place	obesity
research	an effort	(financial) success
progress	a photograph	a profit
experience	an exam	some writing
a conclusion	an agreement	notes
a verdict	a guess	a picture
a decision	changes	a tablet
a comparison	a speech	skills
attention	a test	significance
weight	consensus	a goal

MAKE	DRAW	REACH	LEAD TO	TAKE	DO

Appendix 5 Typology of Exercises and Aims

TYPOLOGY OF EXERCISES	DESCRIPTION OF THE EXERCISE	AIMS
1. Matching Exercise (adjective + abstract noun)	Participants are asked to use their intuition and match several adjectives with the five abstract nouns under study (*conclusion, agreement, comparison, contribution, decision*). More than one combination is possible Some adjectives can be left out.	**Observe** informants' intuition as regards the clause pattern **adjective + noun** **Determine** whether they are able to identify frequent collocates **Verify** whether their performance corresponds to the frequent collocations found in the *HSC*.
2. Fill in the gaps Exercise	Informants are provided with fifteen sentences extracted from our *HSC*. These sentences must be completed with a suitable word, which depending on the context may be a preposition or a verb. However, participants are not restricted to use any of those. They are simply told to use "a suitable word". Even though this task could be interpreted as a very restrictive exercise, many combinations are possible.	**Explore** what collocates informants consider to be the most usual ones in combination with the abstract nouns under study. This will enable us the **analysis** of whether they are aware of the different collocational patterns those abstract nouns can be associated with in scientific English. The problem areas they may be confronted with will also be **identified**.

3. Translation Exercise (Spanish into English)	Informants are given eight sentences in Spanish and are asked to translate them into English. Such sentences contain the Spanish equivalents of the abstract nouns under study.	**Get to know** the strategies used by participants when facing how to express their ideas in English. **Find out** researchers' use of abstract noun patterns, verb tenses, active and passive voice and their use of prepositions. Thus, this exercise will provide highly valuable information concerning participants' command of English morphology (word forms), syntax (word order) and semantics (word selection).
4. Matching Exercise (verb + abstract noun)	Participants are asked to match certain verbs with some of their highly frequent collocates. Some nouns (*a verdict, a place, a picture, a photograph, effort, an exam, a speech, a mistake, some writing, notes, (financial) success*) introduced are not characteristic of the scientific register as such. However, they are considered to be common collocates of the general verbs provided.	**Determine** whether informants use their own intuitions when matching those parts of speech. One of the most remarkable features of this exercise is that there is not just one possible / correct answer in some cases, which will facilitate the **comparison** of the kind of collocates informants would use in their papers with the actual collocates found in the *HSC* corpus.

Appendix 6 Informants' translations to sentences in Exercise 3

1) Se produjo un acuerdo casi unánime entre ambos métodos **There was almost perfect agreement between the two methods**	
ALB1	There was almost total agreement between both methods.
ALB2	A near unanimous agreement between both methods was reached.
ALB3	There was an almost complete agreement between both methods.
PH1	An almost unanimous agreement between methods was achieved.
PH2	There was a broad agreement between both methods.
PH3	It was reached an agreement by nearly everyone between methods.
PH4	A near unanimous agreement between both methods was obtained.
PH5	There was an almost general agreement between both methods.
PH6	
PH7	There was almost a unanimous agreement between both methods.
PH8	
PH9	Agreement between both methods was almost unanimous.
PH10	Both methods showed almost the same result.
PH11	An almost unanimous agreement was reached between both methods.
PH12	Two methods led to almost the same results.
PH13	An almost unanimous agreement between both methods was reached.
HN1	An almost general agreement between both methods was taken.
HN2	It was produced an agreement between both methods.
HN3	An almost general agreement between both methods was reached.
HN4	An unanimous agreement took place between both groups.
HN5	It was ________a ________ agreement between both methods.
HN6	An agreement was done between these types of studies.
HN7	The two methods reached an in practice unanimous agreement.
HN8	

2) No se realizaron comparaciones para determinar diferencias de género ni de edad. **Comparisons were not made to determine gender or age differences.**	
ALB1	Comparations to determine gender and age differences weren’t done.
ALB2	Comparisons to analyse differences of gender and age didn’t be made.
ALB3	Comparisons analyzing gender and age differences were not made.
PH1	No comparisons to detect differences in gender nor age were done.
PH2	The comparisons to assess differences in gender or age were not done.
PH3	Comparitions were not made to get differents between gender neither age.
PH4	To probe differences between gender and age no comparisons were made.
PH5	There were no comparisons to determine possible gender and age differences.
PH6	No comparisons were made in order to determine differences about sex nor age.
PH7	No comparisons were made in order to determine age and gender differences.
PH8	
PH9	Comparisons were not made to determine differences between sexes and ages.
PH10	Comparations to assess differences in genre and age hadn’t been made.
PH11	Comparisons to determine any difference in gender or age were not made.
PH12	No comparations were made to make differences nor sex nor age.
PH13	Comparisons to determinate sex or age differences were not done.
HN1	Comparisons to determine sex or age differences were not made.
HN2	It wasn’t made comparations to determinate differences by genere or age.

HN3	No comparisons were made in order to determinate differences neither gender nor age.
HN4	Comparations to determine differences of genre or age were not done.
HN5	Comparisons didn’t make to determine some differencies nor gender neither age.
HN6	No comparisions were done about any differences of sex.
HN7	Comparisons were not made to determine sex or age differences.
HN8	Did not realise comparisons to decide differences in gender and age.

3) En resumen, los resultados de nuestros estudios sugieren que la vaginitis bacteriana no afecta en absoluto a la concepción. **In conclusion, findings from our studies suggest that bacterial vaginosis has no influence on conception.**	
ALB1	In summary, results of our investigations suggest that bacterian vaginytis doesn’t affect to conception.
ALB2	In resume, results from our studies suggest that vaginitis bacteriana doesn’t affect to conception at all.
ALB3	In summary, our studies’ results suggest that bacterial vaginitis does not affect at all to conception.
PH1	In summary, the results of our studies suggest that bacterial vaginitis has no influence on conception.
PH2	In summary, the results of our studies suggest that bacterian vaginitis does absolutely not affect conception.
PH3	To sum up, the results of our studies suggest that the bacterial vaginitis does not affect the pregnancy at all.
PH4	In summary, the results of our studies suggest that pregnancy was not affected by bacterian vaginitis.
PH5	Summarising, our results suggest that bacterial vaginitis does not impair conception/fertility.
PH6	In summary, the results of our studies suggest that bacterial vaginitis doesn’t affect to the __________.
PH7	In summary, the results of our studies suggest that bacterial vaginitis does not affect conception at all.
PH8	In conclusion, the results of our studies suggest that the bacterial vaginosis do not affect the conception.
PH9	In conclusion, the results of our study suggest that bacterial vaginitis does not affect conception in any way.
PH10	In summary, our results suggest that bacterial vaginitis doesn’t affect conception at all.
PH11	In conclusion our results suggest delivery is not affected by bacterian vaginitis.
PH12	As summary, the results of our research suggest that bacterial vaginities does not affect conception at all.
PH13	In summary, the results of our studies suggest that bacterial

	vaginitis doesn’t affect to the conception at all.
HN1	To summarize, the results of our studies suggest that bactetian vaginitis absolutely does not affect to conception.
HN2	To end, the results of our studies tell us that bacterian vaginitis don’t affect to the conception at all.
HN3	Summing, the results of our studies suggest the bacterial vaginitis don’t affect the conception at all.
HN4	In conclusion, our results suggest that bacterial vaginitis does not interfere with conception.
HN5	In conclusion, our studies results reachs that the bacterian vaginitys don’t affect to the concepcion.
HN6	In conclusion, the results of our studies suggest that bacterian vaginitis do not affect conception.
HN7	In summary, ours results suggest that the bacterial vaginitis don’t affect the conception at all.
HN8	In summary, the result of our studies suggest that bacterian vaginitis does not affect at all to conception.

4) Las sustancias analizadas han contribuido de forma poco significativa a disminuir los síntomas de la enfermedad. **The analysed substances have not contributed significantly to lessen the disease symptoms.**	
ALB1	Analysed substances have contributed in a few significative way to diminish disease symptoms.
ALB2	Analysed substances have contribute to decrease disease synthons in a limited way.
ALB3	The analyzed compounds have poorly contributed to lessen the symptoms of the disease.
PH1	Analyzed substances have made a poor significant contribution to decrease disease symptoms.
PH2	The analized substances have contributed poorly to decrease the symptoms of the disease.
PH3	The analized substances have contributed to minimize the illness symptoms in a little significative way.
PH4	The analyzed substances have a little contribution to diminish the symptoms of the disease.
PH5	The contribution of analysed substances on symptoms is marginally significative.
PH6	Analysed substances have contributed not much significant to decrease the symptoms.
PH7	The analyzed substances have contributed in a not very significant way to diminish the illness symptoms.
PH8	The analized drugs have contributed to reduce the symptoms of the illness.
PH9	The substances analyzed have contributed only slightly to reduce the symptoms of this illness.
PH10	Analyzed substances have contributed a little to decrease the symptoms of the illness.
PH11	Substances analyzed have poor significantly contributed to reduce disease symptoms.
PH12	The tested substances have not contributed to diminish the symptoms of the disease in a concrete way.
PH13	The analized substances haven’t contributed in a significant way to

	decrease the symptoms of the disease.
HN1	The substances analized have made a little significative contribution to diminish the disease symptoms.
HN2	The tested substances have contributed in a little significant way to reduce the illness symptoms.
HN3	The analized substances have made a little significant contribution in decreasing the symptoms of the disease.
HN4	The tested substances have not contributed significantly to relief the disease symptoms.
HN5	The analized substans have made a contribution in a short significative way to decrease the illness symthoms.
HN6	The substances studied don’t contribute to decrease the sintoms of the illness.
HN7	The analyzed substances have not significantly affected the symptoms of the disease.
HN8	The analised substances had contributed in poorly significant form to decrease the illness symptoms.

5) Antes de que se pueda llegar a conclusiones fiables, los investigadores deben aprender mucho más acerca de la función genética. **Before firm conclusions can be made, researchers must learn much more about genetic function.**	
ALB1	Researchers have to learn a lot about genetic function before to obtain conclusions.
ALB2	Before reaching reliable conclusions investigators must learn more about the genetic function.
ALB3	Before we can reach firm conclusions, researchers must learn much more about genetic function.
PH1	Before reaching reliable conclusions, investigators must learn much more on genetic function.
PH2	Before reaching reliable conclusions, investigators must learn much more on genetic function.
PH3	Before reliable conclusions can be made, scientificts must learn a lot about the genetic function.
PH4	Before reaching reliable conclusion investigators must learn to much about genetic function.
PH5	Investigators should improve their knowledge on genetic function in order to reach any reliable conclusion.
PH6	Before to obtain reliable conclusions, the investigators must learn much more about genetic function.
PH7	Before reaching reliable conclusions, the investigators must learn about the genetic function.
PH8	Before they obtain valid conclusions, the investigators must learn much about genetic function.
PH9	Before reliable conclusions can be reached, the researches must learn more about the genetic function.
PH10	Before reach a reliable conclusion, researchers must learn much more about genetic function.
PH11	Before we can draw reliable conclusions, investigators must learn much more about genetic function.
PH12	Before having better evidence, researchers have to learn much more about genetic functions.

PH13	Before reaching reliable conclusions, the researchers must learn much more about the genetic function.
HN1	Before reliable conclusions could be reached, investigators should learn much more about the genetic function.
HN2	Before it can lead to the reliable conclusions, researchers have to learn more about genetic function.
HN3	The researchers have to learn much more about genetic function, before reliable conclusions can be reached.
HN4	Before it could reach reliable conclusions, the research team must improve his knowledge about the genetic function.
HN5	Before to get reliable conclusions, the researchers must learn much more about the genetic function.
HN6	Before make some verificable conclusions, the investigators ned to learn a lot about the genetic function.
HN7	Before to reach reliable conclusions, the researchers have to learn more about genetic function.
HN8	Before to achieve reliable conclusions, researchers must learn more about genetic function.

6) Se hicieron las mismas comparaciones en los tratamientos de control. **The same comparisons were made in the control treatments.**	
ALB1	Same comparisons were done in control treatment.
ALB2	The same comparisons in control treatments were made.
ALB3	The same comparisons were made in the control treatments.
PH1	The same comparisons were made in the control treatments.
PH2	The same comparisons were done in the control treatments.
PH3	They made the same comparition in the control treatment.
PH4	In control treatments the same comparisons were made.
PH5	Similar comparisons were performed between control treatments.
PH6	Similar comparisons were made in the control treatments.
PH7	The same comparisons were made in the control treatments.
PH8	It was made the same comparisions in the control treatments.
PH9	The same comparisons were made with the control treatments.
PH10	The same comparations were made in control treatments.
PH11	In control treatments same comparisons were made.
PH12	In control groups the comparations were the same.
PH13	Similar comparations were done with the control treatments.
HN1	The same comparisons were made in control treatments.
HN2	It was made similar comparations on control treatments
HN3	The same comparations in the control treatment were done.
HN4	The same comparations were made in the control treatments.
HN5	They did the same comparisons in the control treatments.
HN6	The same comparisions were done in the control tratament.
HN7	The same comparisons were made in the control treatments.
HN8	They made the same comparations in the control treatments.

7) Estas conclusiones biológicas han sido justificadas por múltiples tipos de análisis estadísticos. **These biological conclusions are justified by multiple types of statistical analyses.**	
ALB1	These biologyc conclusions have been justified by multiple statistical analysis.
ALB2	This biological conclusions have been justified by numerous stadistical analysis.
ALB3	These biological conclusions have been justified by many types of statistical analyses.
PH1	These biological conclusions have been justified by multiple kinds of statistical analysis.
PH2	These biological conclusions have been supported by multiple types of stadistical analysis.
PH3	These biological conclusions have been justified by a lot of statistical tests.
PH4	To many statistical analysis lead to these biological conclusions.
PH5	Multiple statistical analyses support these biological conclusions.
PH6	These biological conclusions have been justified through multiples types of statistical analyses.
PH7	These conclusions have been justificated by multiple types of statistical analysis.
PH8	This biological conclusions have been supported by different statistical studies.
PH9	These biologyc conclusions have been justified by multiple types of statistical analysis.
PH10	These biological conclusions have been justifyed by a lot of kinds of statistical analysis.
PH11	These biological conclusions have been justified by different statistical analysis.
PH12	These biological conclusions have been justified by sever statistical analyses.
PH13	These biologyc conclusions have been justified by multiple kinds of statistical tests.

HN1	These biological conclusions have been justified by various types of statistical analysis.
HN2	This biological conclusions have been justificated by several kinds of stadistical analysis.
HN3	These biologyc conclusions have been supported by many kinds of statistical analysis.
HN4	This biologyc conclusions have been justified by many statistical analysis.
HN5	This biological conclusions have been justifiqued for a lot of different stadistics ways.
HN6	The biologyc conclusions have been verificated by a lot of types of statistical analysis.
HN7	These biological conclusions were justifiqued by different statistical analysis.
HN8	These biologyc conclusions had been supported by multiple pattern of statistic analysis.

8) Varias organizaciones son las responsables de la toma de decisiones por lo que respecta a la seguridad de las transfusiones.

Several organisations are responsible for decision making in transfusion safety.

ALB1	Several organitations are responsible of making of decisions about security in transfusions.
ALB2	Several organizations are responsible for making decisions about transfusions security.
ALB3	Several organizations are responsible for the decision-making with respect to transfusions security.
PH1	Several organizations are responsible of decision-making concerning security of transfusions.
PH2	Several organizations are responsible for decision making with respect to transfusion security.
PH3	Some boureaus are responsibles of the decisions taken about transfusion safety.
PH4	About transfusions security many organizations are responsible in taking decisions.
PH5	Several organisations are responsible in decision making with respect to transfusions safety.
PH6	Regarding to transfusions safety, several organizations are responsible for to adopt decisions.
PH7	Various organizations are the ones responsibles of the decision making with respect to the transfusion's security.
PH8	Several organizations are the responsible of the lack of decisions about the security of the transfusions.
PH9	Various organisations are responsible for taking the decision with respect to the safety of transfusion.
PH10	Several organizations are responsible for making decisions about safety of transfusions.
PH11	Several organizations are responsible for taking decisions regarding blood transfusions safety.
PH12	Transfussions follow strict protocols accordingly to some organizations.

PH13	Some organizations are responsable for making decisions regarding the security of transfusions.
HN1	Several organizations are responsible in taking decisions about transfusion security.
HN2	Very organisations are the responsables of the doing decisions about transfusions security.
HN3	Several organizations are responsible for making decision with respect to the safety of the transfusions.
HN4	Blood transfusion security is surveyed by some organizations.
HN5	Some organizations are responsibles in the decisions about transfusions security.
HN6	A couple of organizations are the responsable of the decisions about the transfusion security.
HN7	Several organizations are responsible in to take decisions about the safety of the transfusions.
HN8	Some organisations are responsible to decision-making with respect to transfusions security.

Appendix 7 Informants' verb-noun collocations in Exercise 4: Matching Exercise (verb + abstract noun)

MAKE[1]

(financial) SUCCESS	12,5%
A COMPARISON	70,8%
A CONCLUSION	4,2%
A CONTRIBUTION	62,5%
A DECISION	41,7%
A GOAL	8,3%
A GUESS	29,2%
A MISTAKE	79,2%
A PHOTOGRAPH	4,2%
A PICTURE	12,5%
A PLACE	0,0
A PROFIT	50,0%
A SPEECH	45,8%
A SUGGESTION	70,8%
A TABLET	4,2%
A TEST	25,0%
*A VERDICT	8,3%
*ACTION	4,2%

AN AGREEMENT	8,3%
AN EFFORT	75,0%
AN EXAM	16,7%
*ATTENTION	8,3%
CHANGES	54,2%
CONSENSUS	0,0
DAMAGE	29,2%
*EXPERIENCE	8,3%
KNOWLEDGE	0,0
NOTES	0,0
OBESITY	0,0
PROGRESS	41,7%
*RESEARCH	12,5%
somebody's TEMPERATURE	0,0
*SIGNIFICANCE	4,2%
SKILLS	0,0
SOME WRITING	0,0
WEIGHT	0,0

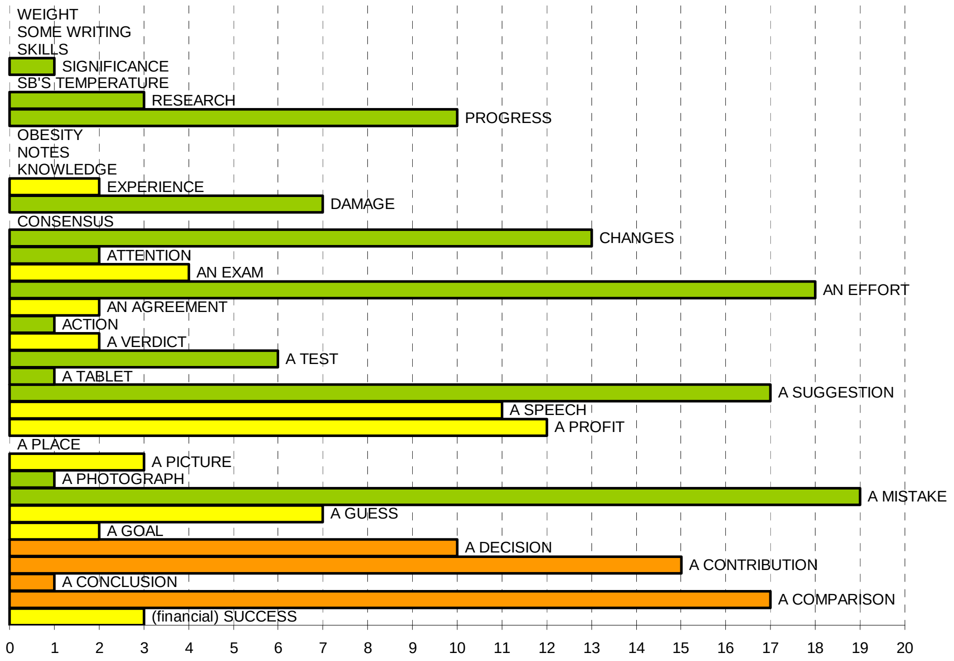


Figure 67 *make* + noun collocations provided by informants

---

[1] The asterisked combinations in Figures 67-72 correspond to deviant collocations produced by informants.

Appendix 7 Informants' verb-noun collocations in Exercise 4: Matching Exercise (verb + abstract noun)

DRAW

(financial) SUCCESS	0,0
A COMPARISON	4,2%
A CONCLUSION	41,7%
A CONTRIBUTION	0,0
*A DECISION	4,2%
*A GOAL	4,2%
*A GUESS	12,5%
A MISTAKE	0,0
A PHOTOGRAPH	0,0
A PICTURE	41,7%
A PLACE	0,0
*A PROFIT	4,2%
*A SPEECH	4,2%
A SUGGESTION	0,0
*A TABLET	8,3%
A TEST	0,0
*A VERDICT	16,7%
ACTION	0,0

AN AGREEMENT	0,0
AN EFFORT	0,0
AN EXAM	0,0
ATTENTION	25,0%
*CHANGES	4,2%
CONSENSUS	0,0
*DAMAGE	4,2%
*EXPERIENCE	8,3%
*KNOWLEDGE	4,2%
*NOTES	12,5%
OBESITY	0,0
PROGRESS	0,0
*RESEARCH	4,2%
*somebody's TEMPERATURE	4,2%
*SIGNIFICANCE	4,2%
*SKILLS	12,5%
*SOME WRITING	12,5%
*WEIGHT	4,2%

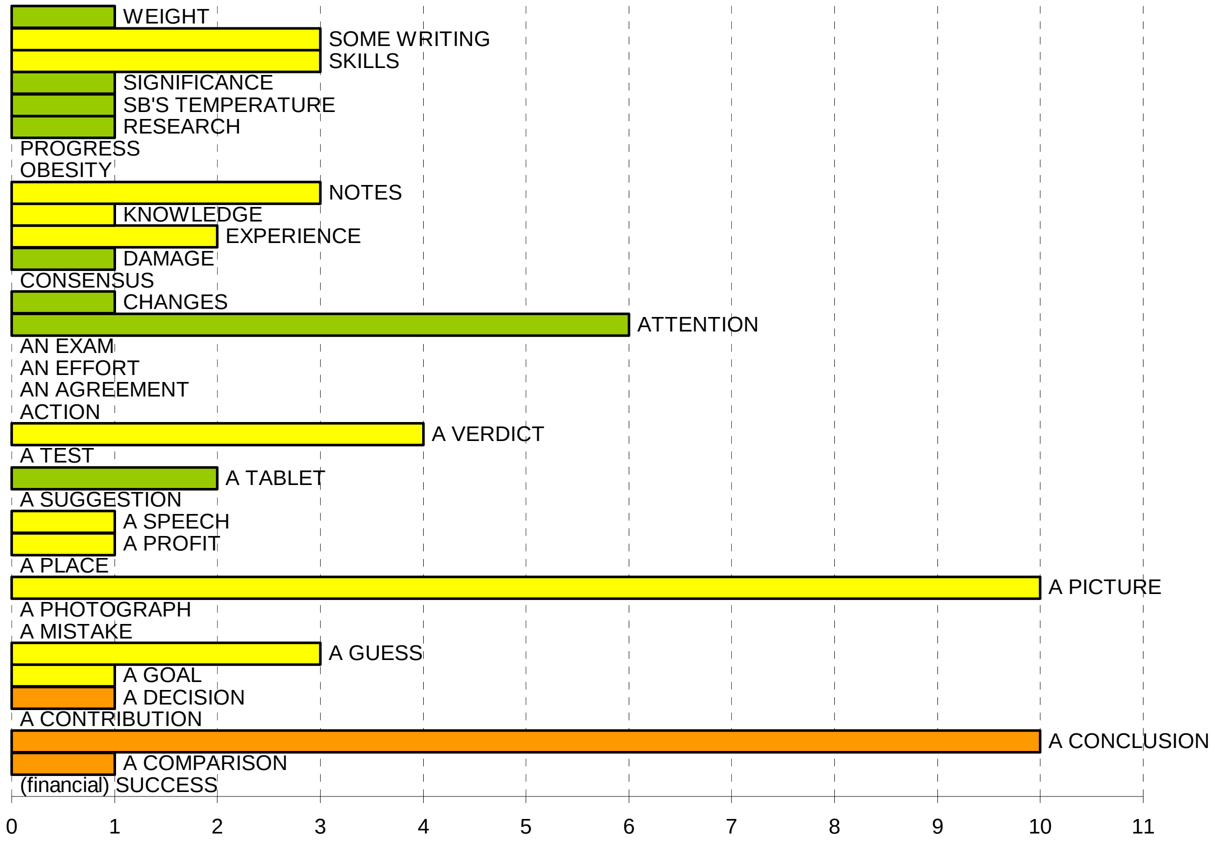


Figure 68 *draw* + noun collocations provided by informants

Appendix 7 Informants' verb-noun collocations in Exercise 4: Matching Exercise (verb + abstract noun)

REACH

(financial) SUCCESS	25,0%
A COMPARISON	0,0
A CONCLUSION	41,7%
A CONTRIBUTION	0,0
A DECISION	25,0%
A GOAL	58,3%
*A GUESS	8,3%
A MISTAKE	0,0
A PHOTOGRAPH	0,0
A PICTURE	0,0
*A PLACE	50,0%
*A PROFIT	4,2%
A SPEECH	0,0
A SUGGESTION	0,0
A TABLET	0,0
A TEST	0,0
A VERDICT	45,8%
ACTION	0,0

AN AGREEMENT	79,2%
AN EFFORT	0,0
AN EXAM	0,0
ATTENTION	0,0
CHANGES	0,0
CONSENSUS	70,8%
DAMAGE	0,0
*EXPERIENCE	37,5%
KNOWLEDGE	29,2%
NOTES	0,0
*OBESITY	8,3%
*PROGRESS	4,2%
RESEARCH	0,0
*somebody's TEMPERATURE	12,5%
SIGNIFICANCE	41,7%
SKILLS	33,3%
SOME WRITING	0,0
*WEIGHT	29,2%

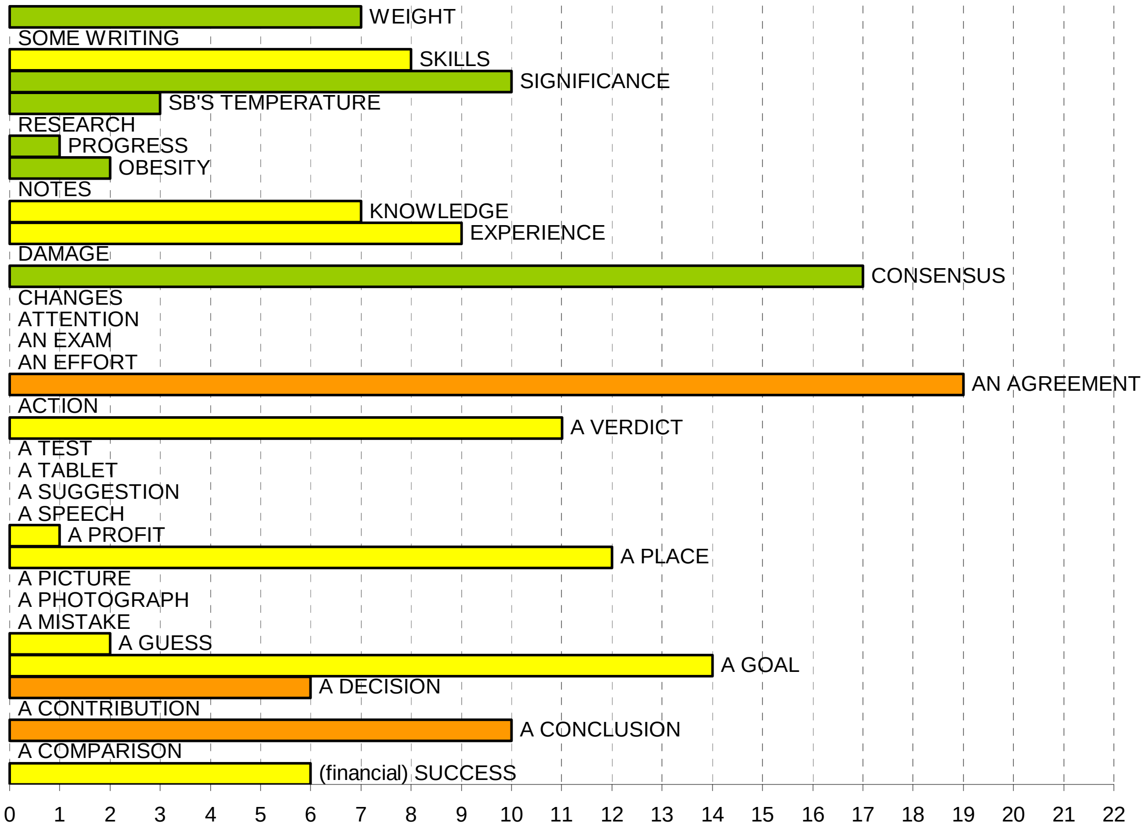


Figure 69 *reach* + noun collocations provided by informants

Appendix 7 Informants' verb-noun collocations in Exercise 4: Matching Exercise (verb + abstract noun)

LEAD TO

(financial) SUCCESS	50,0%
*A COMPARISON	4,2%
A CONCLUSION	16,7%
*A CONTRIBUTION	12,5%
*A DECISION	12,5%
A GOAL	12,5%
A GUESS	0,0
A MISTAKE	16,7%
A PHOTOGRAPH	0,0
A PICTURE	0,0
*A PLACE	8,3%
*A PROFIT	12,5%
A SPEECH	0,0
*A SUGGESTION	4,2%
*A TABLET	4,2%
A TEST	0,0
*A VERDICT	12,5%
*ACTION	16,7%

AN AGREEMENT	12,5%
AN EFFORT	0,0
AN EXAM	0,0
ATTENTION	0,0
CHANGES	25,0%
CONSENSUS	29,2%
*DAMAGE	29,2%
EXPERIENCE	20,8%
KNOWLEDGE	41,7%
NOTES	0,0
OBESITY	75,0%
*PROGRESS	29,2%
*RESEARCH	8,3%
somebody's TEMPERATURE	0,0
*SIGNIFICANCE	16,7%
*SKILLS	8,3%
SOME WRITING	0,0
*WEIGHT	12,5%

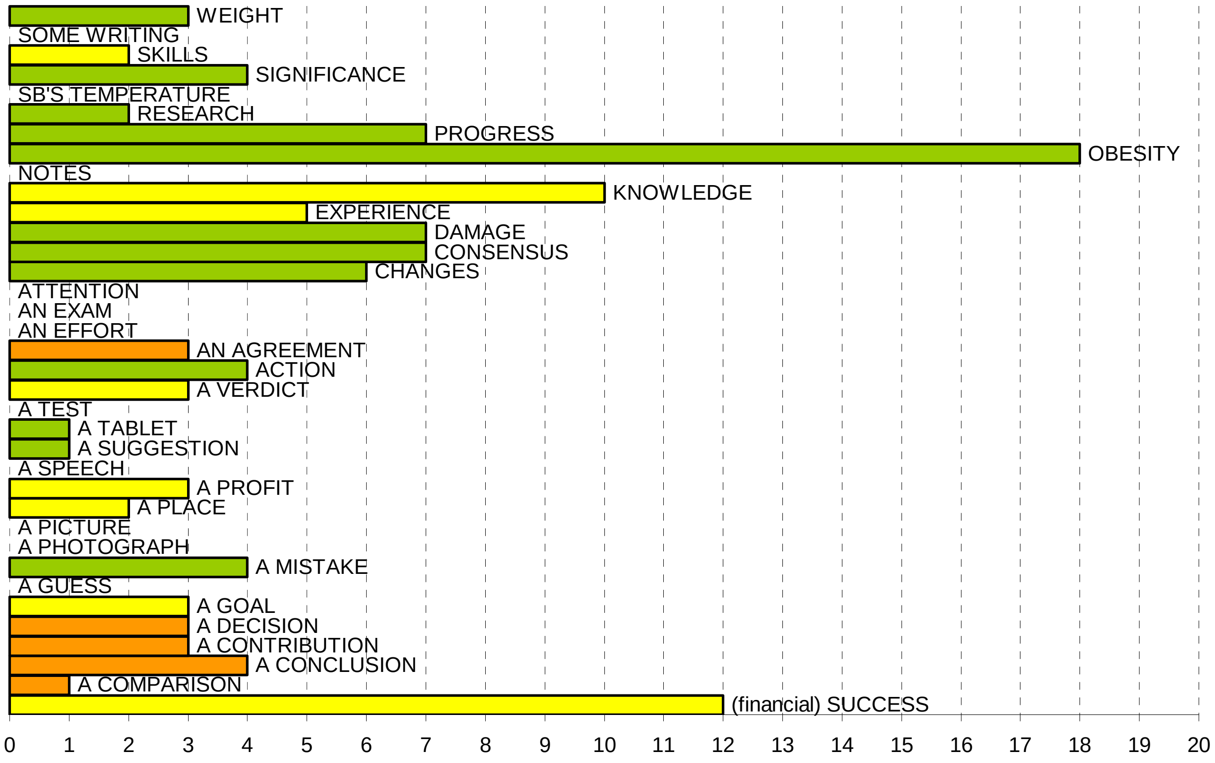


Figure 70 *lead to* + noun collocations provided by informants

Appendix 7 Informants' verb-noun collocations in Exercise 4: Matching Exercise (verb + abstract noun)

TAKE

(financial) SUCCESS	0,0
A COMPARISON	0,0
A CONCLUSION	0,0
*A CONTRIBUTION	4,2%
A DECISION	29,2%
A GOAL	0,0
A GUESS	25,0%
*A MISTAKE	4,2%
A PHOTOGRAPH	95,8%
A PICTURE	41,7%
A PLACE	41,7%
*A PROFIT	4,2%
A SPEECH	0,0
A SUGGESTION	0,0
A TABLET	70,8%
A TEST	20,8%
A VERDICT	8,3%
ACTION	50,0%

AN AGREEMENT	0,0
AN EFFORT	0,0
AN EXAM	25,0%
*ATTENTION	29,2%
CHANGES	0,0
CONSENSUS	0,0
DAMAGE	0,0
*EXPERIENCE	16,7%
*KNOWLEDGE	16,7%
NOTES	87,5%
*OBESITY	4,2%
PROGRESS	0,0
*RESEARCH	4,2%
somebody's TEMPERATURE	54,2%
*SIGNIFICANCE	8,3%
*SKILLS	4,2%
*SOME WRITING	8,3%
WEIGHT	29,2%

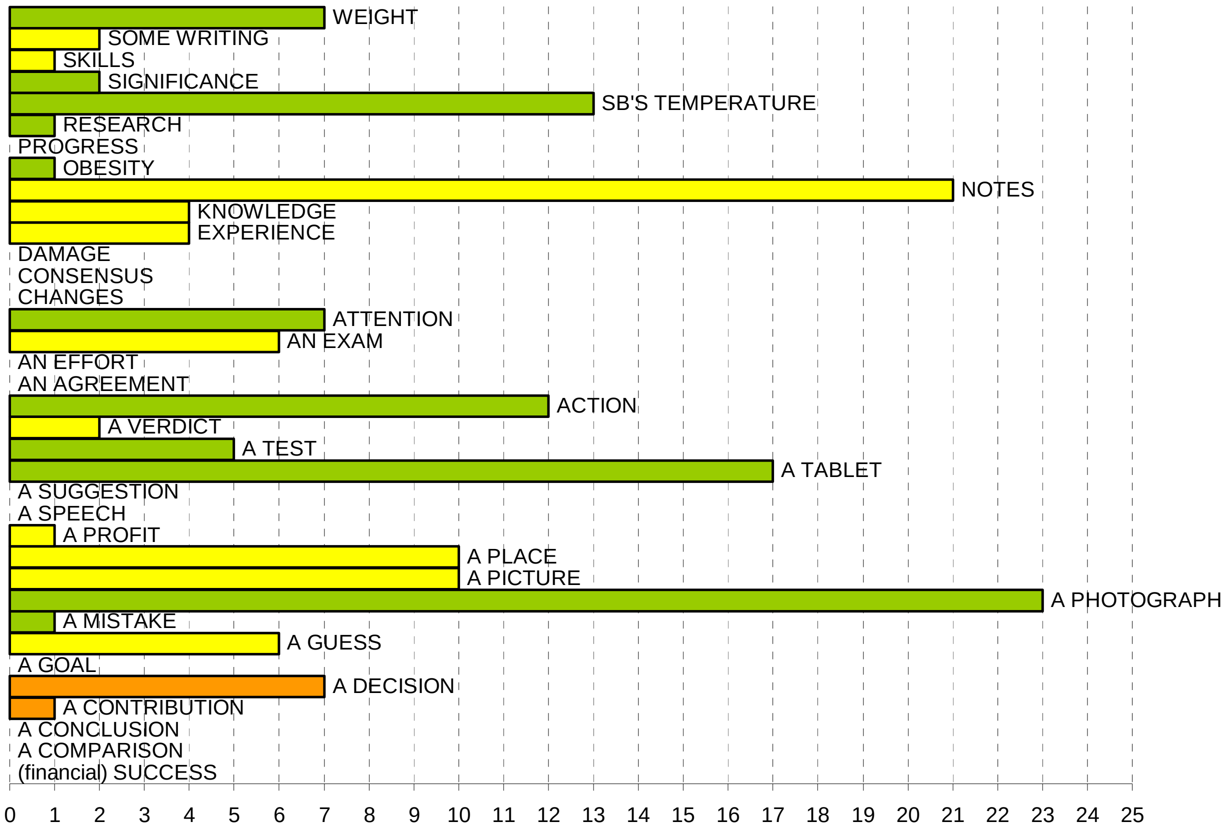


Figure 71 *take* + noun collocations provided by informants

Appendix 7 Informants' verb-noun collocations in Exercise 4: Matching Exercise (verb + abstract noun)

DO

(financial) SUCCESS	0,0
A COMPARISON	16,7%
A CONCLUSION	0,0
A CONTRIBUTION	12,5%
A DECISION	0,0
*A GOAL	4,2%
*A GUESS	4,2%
*A MISTAKE	8,3%
A PHOTOGRAPH	0,0
A PICTURE	0,0
A PLACE	0,0
A PROFIT	0,0
A SPEECH	37,5%
*A SUGGESTION	16,7%
*A TABLET	4,2%
A TEST	58,3%
A VERDICT	0,0
*ACTION	12,5%

AN AGREEMENT	0,0
*AN EFFORT	20,8%
AN EXAM	58,3%
*ATTENTION	8,3%
*CHANGES	20,8%
CONSENSUS	0,0
DAMAGE	54,2%
EXPERIENCE	0,0
KNOWLEDGE	0,0
NOTES	0,0
OBESITY	0,0
*PROGRESS	29,2%
RESEARCH	54,2%
somebody's TEMPERATURE	0,0
*SIGNIFICANCE	4,2%
*SKILLS	8,3%
SOME WRITING	66,7%
WEIGHT	0,0

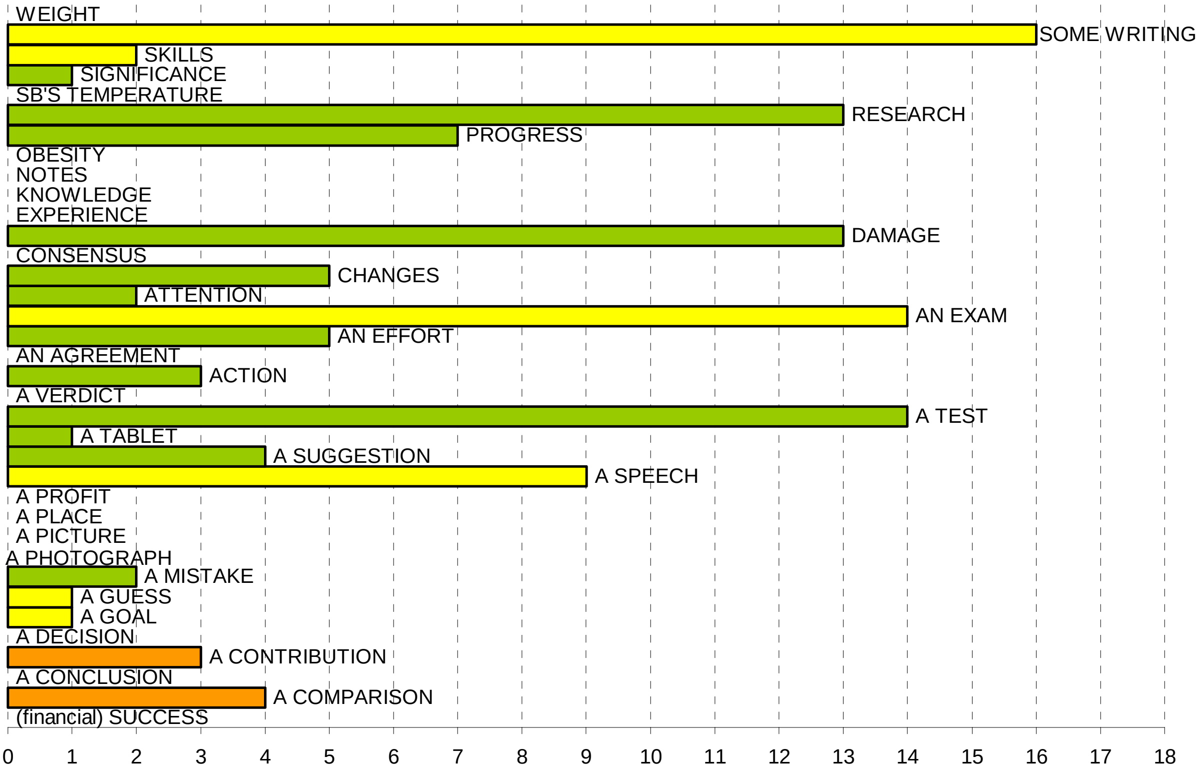


Figure 72 *do* + noun collocations provided by informants

Appendix 8 Survey Content-Based Question 8: Informants' self-evaluation of their main problem areas when writing in English

SPECIFIC TERMS

		Frequency	Percent
Valid	very easy	10	41,7
	easy	11	45,8
	not so easy	2	8,3
	difficult	1	4,2
	Total	24	100,0

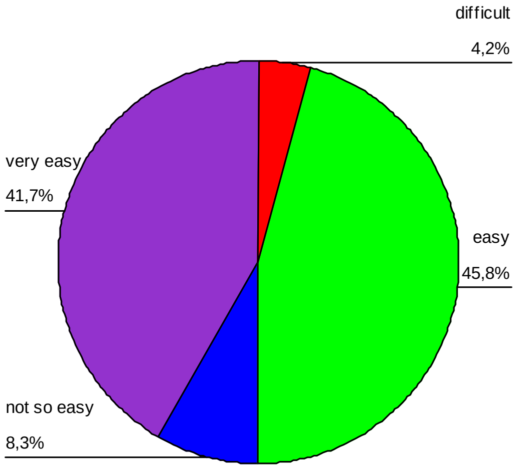


Figure 73 Use of specific (medical) terminology when writing in English

Appendix 8 Survey Content-Based Question 8: Informants' self-evaluation of their main problem areas when writing in English

GENERAL WORDS

		Frequency	Percent
Valid	very easy	4	16,7
	easy	6	25,0
	not so easy	11	45,8
	difficult	3	12,5
	Total	24	100,0

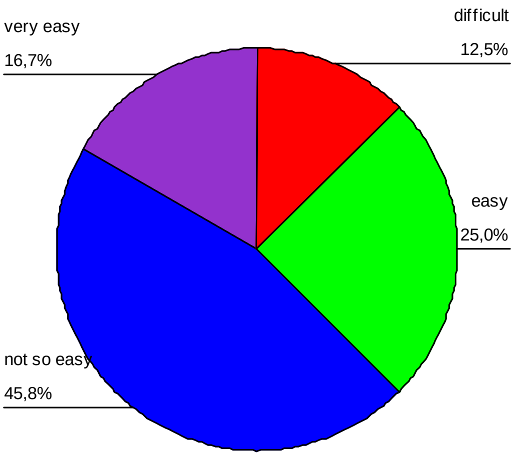


Figure 74 Use of general terms when writing in English

Appendix 8 Survey Content-Based Question 8: Informants' self-evaluation of their main problem areas when writing in English

SIMILAR PAIRS OF WORDS

		Frequency	Percent
Valid	easy	2	8,3
	not so easy	10	41,7
	difficult	11	45,8
	extremely difficult	1	4,2
	Total	24	100,0

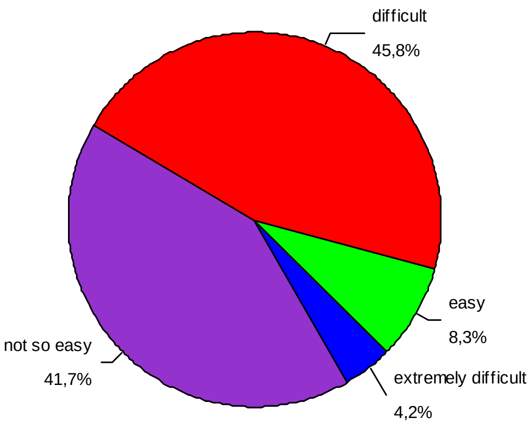


Figure 75 Distinction between similar pairs of words when writing in English

Appendix 8 Survey Content-Based Question 8: Informants' self-evaluation of their main problem areas when writing in English

PARAGRAPH DISTRIBUTION

		Frequency	Percent
Valid	very easy	1	4,2
	easy	7	29,2
	not so easy	13	54,2
	difficult	2	8,3
	Total	23	95,8
Missing	System	1	4,2
Total		24	100,0

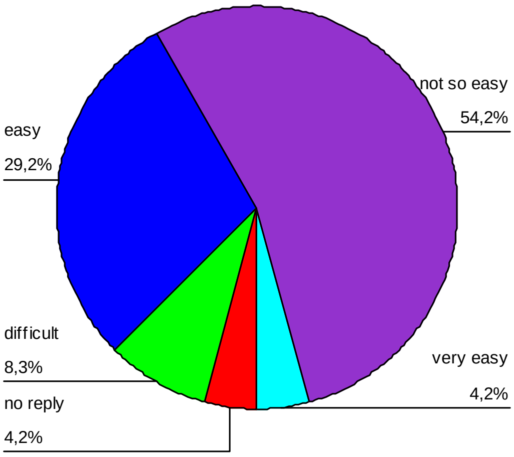


Figure 76 Paragraph distribution when writing in English

Appendix 8 Survey Content-Based Question 8: Informants' self-evaluation of their main problem areas when writing in English

GRAMMAR RULES

		Frequency	Percent
Valid	easy	4	16,7
	not so easy	6	25,0
	difficult	11	45,8
	extremely difficult	3	12,5
	Total	24	100,0

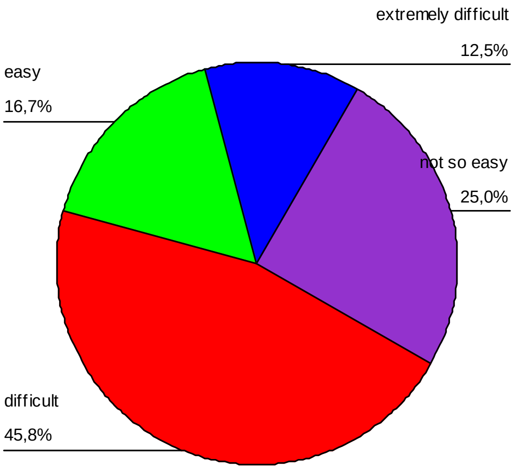


Figure 77 Grammatical rules when constructing sentences in English