UNIVERSIDAD DE MURCIA

Facultad de Letras

Departamento de Traducción e Interpretación

Tesis Doctoral

ESTUDIO DE LOS RASGOS LINGÜÍSTICOS DE LA MENTIRA EN EL MEDIO ESCRITO: UN ANÁLISIS CONTRASTIVO INGLÉS-ESPAÑOL

Ángela Almela Sánchez-Lafuente

Directores: Rafael Valencia García y Pascual Cantos Gómez

Noviembre 2012

UNIVERSITY OF MURCIA

Faculty of Arts

Department of Translation and Interpreting

Doctoral Thesis

FEATURING DECEPTION IN WRITTEN LANGUAGE:

A CONTRASTIVE STUDY OF ENGLISH AND SPANISH

Ángela Almela Sánchez-Lafuente

Supervisors: Rafael Valencia García and Pascual Cantos Gómez

November 2012

Dedication

To my late father, who sadly met his demise during the course of this project.

This doctoral thesis is dedicated to him, for living on most vividly in the realm of memory, and for teaching me how empty ears of grain stand up haughtily, whereas full ones bow down modestly.

Acknowledgements

During the writing of this dissertation I have been guided, supported and encouraged by many people who deserve my deepest gratitude.

Firstly, I wish to express my limitless appreciation to my supervisors, Dr Pascual Cantos Gómez and Dr Rafael Valencia García, for their sound advice during the completion of this dissertation. They have taught me so many valuable lessons since I was an undergraduate student, and have succeeded in bringing together the three fields of linguistics, computation and statistics.

I am also truly grateful to Fundación Séneca and the University of Murcia for awarding me a scholarship, by virtue of which I have had the chance to learn how to be a researcher.

I am very much indebted to the academic institutions that have hosted me as a doctoral student. First of all, Universitat de València, especially IULMA members Dr Francisca Suau Jiménez, Dr Miguel Fuster Márquez and Dr Julia Sanmartín Sáez; I thank them for their warm hospitality and for treating me like part of the team. I also appreciate the help received from Dr Ludwig J. Issing of Freie Universität Berlin, and our fruitful talks about psycholinguistic aspects of deception. Last but not least, I am more than grateful to Dr Carlo Strapparava, from the Fondazione Bruno Kessler, for kindly sharing his corpus –for which Dr Rada Mihalcea must be also mentioned–, for introducing me to Machine Learning foundations, and for always keeping his office door open.

Warm thanks also go to the LACELL research group for the help and support provided by its members.

Special mention must be made of Dr Gema Alcaraz for her invaluable assistance with my predoctoral enquiries and for the immense trust she has in me and in the youth of today.

I gratefully acknowledge Dr Mª Ángeles Orts Llopis for putting all her faith in me from the very moment I graduated, and for allowing me to conduct research with her when I still had so much to learn. I also thank her for managing to transcend the purely professional so subtly.

It gives me great pleasure to have met a black belt in deception detection like Dr Isabel Picornell, and I must extend my sincere gratitude to her for having always been willing to help me and to share her materials with me.

I would also like to extend my warmest thanks to Ms Lindsay Duffy, who has spent much of her time on reviewing my drafts and kindly offering me constructive suggestions.

As for my family, I owe a profound debt of gratitude to them, not only to my blood relatives but also to my in-laws, for their understanding and endless patience when faced with such time-consuming work. Special thanks go to my mother, who passed her sense of academic vocation on to me and taught me that, if necessary, a PhD dissertation can even be written at one's own kitchen table.

I am deeply indebted to Sam for his substantial help with the arduous task of data processing, his great understanding at challenging moments, and for being so genuinely delighted at my achievements, albeit modest.

Special mention must be made of Rocío, with whom I have identified so strongly throughout all these years of tough work, especially during the summer, when everybody was having so much fun around us. I thank her for helping me with all things forensic, for standing by me through thick and thin, and for thinking of me as a sister.

Sincere thanks go to all my other friends, for their enduring support and their constant words of encouragement, and because Aristotle was so right when he said that without them no one would choose to live, despite having all other goods.

I also have a warm feeling of gratitude towards Venus, who has been a silent witness to countless hours of work and a faithful companion during this lengthy process.

Last, but definitely not least, I am genuinely grateful to Antonio, my husband, for his enthusiastic and unwavering support at every step of the way towards my PhD degree. We have been together since the very beginning of our academic lives, and the worth of this work is certainly to be shared with him.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BoW: Bag-of-Words

CBCA: Criteria-based content analysis

CMC: Computer-mediated communication

DFA: Discriminant function analysis

IDT: Interpersonal deception theory

LIWC: Linguistic Inquiry and Word Count

ML: Machine learning

RM: Reality monitoring

SVA: Statement validity analysis

SVM: Support vector machine

RESUMEN EN ESPAÑOL

INTRODUCCIÓN

En el contexto de la comunicación humana, la mentira juega un papel activo. A este respecto, DePaulo et al. (1996) afirman que se suele contar de una a dos mentiras al día, ya sea en lenguaje oral o escrito. Por ello, la mentira se ha estudiado desde la perspectiva de varias disciplinas, como la psicología, la lingüística, la psiquiatría, y la filosofía (Granhag y Strömwall, 2004). Más recientemente, la condición de verdad de las opiniones vertidas a través de Internet ha suscitado un interés creciente en el campo de la minería de opiniones (Ott et al., 2011). Esta cuestión es particularmente compleja, ya que el investigador no dispone de más información que el propio lenguaje escrito, cuando la investigación en el área señala que el lenguaje no verbal es el que contiene mayor información sobre la mentira (Vrij, 2010).

En el ámbito lingüístico, la investigación en el área se limita casi exclusivamente a la lengua inglesa. Hay características verbales de la mentira que forman parte de herramientas para su detección utilizadas por profesionales e investigadores. Las técnicas lingüísticas automatizadas se utilizan para examinar los perfiles lingüísticos de la mentira en inglés. Más comúnmente, los investigadores han recurrido a las categorías de palabras definidas en *Linguistic Inquiry and Word Count*, conocido por sus siglas LIWC (Pennebaker et al., 2001); se trata de un programa de análisis de texto que clasifica las palabras en categorías significativas a nivel psicolingüístico. Comprende unas 2.200 palabras y raíces léxicas agrupadas en 75 categorías. Estas se han utilizado para estudiar cuestiones tales como la personalidad humana (Mairesse et al., 2007), los cambios psicológicos (Alpers et al., 2005), los juicios sociales (Leshed et al., 2007), y la salud mental (Rude et al., 2004). Además de ello, diversos trabajos avalan la utilidad de esta herramienta para la detección y caracterización de la mentira en el lenguaje. La validación del léxico contenido en su diccionario se ha obtenido a través de una comparación de la valoración de gran cantidad de textos escritos por parte de expertos y las puntuaciones obtenidas por medio de su análisis con LIWC.

La lengua española se encuentra muy limitada en cuanto a medios y herramientas informáticas para el análisis de la configuración lingüística de la mentira; hasta la fecha, el único estudio que explora la mentira en español escrito mediante este tipo de procedimientos es Masip et al. (2012). Por ello, el presente estudio se ocupa en parte de los mecanismos de la mentira en esta lengua. Además de ello, la lingüística contrastiva puede aportar una perspectiva transversal desde la que estudiar el fenómeno en cuestión. Por tanto, la presente tesis pretende llevar a cabo un estudio contrastivo de los rasgos lingüísticos de la mentira en el medio escrito en ambas lenguas, con la ayuda de herramientas informáticas de inteligencia artificial y métodos de clasificación estadísticos, explorando las variables contenidas en LIWC y otras no estudiadas anteriormente.

OBJETIVOS

El presente proyecto aborda, pues, el análisis de los perfiles lingüísticos de la mentira en inglés y en español mediante técnicas lingüísticas automatizadas. A este respecto, se plantean cinco preguntas de investigación:

1) ¿Cuál es la eficacia de las técnicas de clasificación propuestas en textos verdaderos y falsos en inglés y en español?

2) ¿Cuáles son los rasgos lingüísticos con mayor poder discriminatorio en el medio escrito a través de ambas lenguas?

3) ¿Cuáles son los indicadores más relevantes exclusivos del inglés?

4) ¿Cuáles son los indicadores más relevantes exclusivos del español?

5) ¿Hay algún indicador específico de ciertos temas?

METODOLOGÍA

El método describe las diferentes etapas en el desarrollo del presente estudio, incluyendo una introducción a la naturaleza del estudio, las variables, el proceso de recogida de datos, la descripción de los corpora empleados y las técnicas empleadas para el análisis de los datos.

Naturaleza del estudio y variables

Por su naturaleza, el estudio constituye una combinación de investigación primaria y secundaria. El capítulo 2, en el que se sientan las bases teóricas de este estudio, es una muestra de investigación secundaria que ofrece una síntesis de los trabajos más relevantes del área que nos ocupa. En el tercer capítulo da comienzo la descripción del estudio primario, esto es, la parte de investigación que trata con datos originales, su análisis y la discusión de los resultados. Además de ello, el presente estudio también puede clasificarse como cuasi-experimental.

Total de palabras	Afirmaciones	Amigos	Religión
Palabras por oración	Artículos	Familia	Muerte
Palabras > 6 caracteres	Preposiciones	Humanos	Estados corporales y síntomas
Punto	Números	Tiempo pasado del verbo	Sexualidad
Coma	Sentimientos positivos	Tiempo presente del verbo	Comer, beber, tomar
Dos puntos	Optimismo y energía	Tiempo futuro del verbo	Dormir, soñar
Punto y coma	Ansiedad o miedo	Arriba	Acicalarse
Oraciones interrogativas	Enfado	Abajo	Palabras malsonantes
Exclamaciones	Tristeza o depresión	Inclusivos	Ratio tipo/token estandarizada
Guión	Causa y efecto	Exclusiones	Longitud media de palabra
Comillas	Entendimiento	Movimiento	Oraciones/TP
Apóstrofe	Discrepancias	Escuela	Pal. de 1 letra/TP
Paréntesis	Inhibiciones	Trabajo	Pal. de 2 letras/TP
Otra puntuación	Tentativos	Logro	Pal. de 3 letras/TP
1ª persona del singular	Certeza	Hogar	Pal. de 4 letras/TP
1ª persona del plural	Ver	Deportes	Pal. de 5 letras/TP
2ª persona	Escuchar	Televisión y cine	Pal. de 6 letras/TP
3ª persona	Sentir	Música	Pal. de 7 letras/TP
Negaciones	Comunicación	Dinero y asuntos financieros	Pal. complejas/TP

Tabla 1. Variables independientes en el experimento

En lo que respecta a las variables, el estudio cuenta con una variable dependiente y un conjunto de 76 variables independientes. La primera corresponde al valor de verdad de las declaraciones, el cual se mide en una escala nominal binaria. Por su parte, las variables independientes serían las susceptibles de influir en la dependiente. La mayoría de ellas corresponden a categorías

LIWC (*Linguistic Inquiry and Word Count*)[1], mientras que las variables restantes corresponden a características estilométricas que no habían sido anteriormente testadas para el fin que nos ocupa; la presencia de este último grupo se ha obtenido mediante el software WordSmith Tools 5.0[2]. Conviene señalar que la totalidad de las variables independientes se han medido a través de una escala racional. La tabla 1 presenta todas las variables independientes.

Proceso de recogida de datos y descripción de los corpora

Al tratarse de un estudio de corte contrastivo, era necesario que el diseño de los corpora en ambas lenguas fuera similar. Primeramente, se tomó como punto de partida el corpus en inglés recopilado por Mihalcea y Strapparava (2009)[3], en el cual colaboraron 100 participantes hablantes nativos de inglés a través del servicio online Amazon Mechanical Turk[4]. Concretamente, todos ellos debían completar tres tareas, consistentes en escribir su opinión sobre el aborto, sobre la pena de muerte y sobre un buen amigo. A continuación, debían escribir una versión falsa de cada uno de los tres temas. Así pues, esta muestra comprende 300 contribuciones verdades y 300 falsas, que en total suman 51.204 palabras y cuya longitud media por contribución es de 85 palabras.

En lo que respecta al corpus en español, se tomaron 100 participantes hablantes nativos de español peninsular que debían completar tres tareas similares a las anteriores. El tema del buen amigo se mantuvo constante, mientras que para las otras dos tareas se escogieron dos temáticas de actualidad: la adopción homosexual y las corridas de toros. De este modo, esta muestra también

[1] Disponible en http://www.liwc.net/
[2] Disponible en http://wordsmith.org/
[3] Proporcionado por los investigadores durante la estancia predoctoral que la presente autora realizó en la Fondazione Bruno Kessler (Trento, Italia).
[4] Servicio disponible en https://www.mturk.com/mturk/welcome

comprende 300 contribuciones verdades y 300 falsas, que en total suman 56.882 palabras y cuya longitud media por contribución es de 94 palabras.

Técnicas de aprendizaje automático y estadísticas

Para el análisis de los datos, se ha desarrollado un marco de trabajo basado en un clasificador de máquinas de soporte vectorial (SVM). Estos algoritmos se habían aplicado ya con éxito en diversas tareas de clasificación de texto debido a sus múltiples ventajas: primero, son robustos en grandes espacios dimensionales; segundo, cualquier característica resulta relevante; tercero, son robustos cuando hay un conjunto escaso de muestras; finalmente, la mayoría de los problemas de categorización de texto son separables linealmente (Saleh et al., 2011). Para entrenar este clasificador, se han utilizado los valores de las categorías de LIWC agrupados por dimensiones –procesos lingüísticos estándares, procesos psicológicos, relatividad y asuntos personales–, así como las características estilométricas obtenidas con WordSmith Tools, constituyendo una quinta dimensión. Para cada uno de los clasificadores se ha efectuado una validación cruzada de diez iteraciones, arrojando una tasa de éxito de clasificación de los testimonios en forma de porcentaje. A continuación, se ha efectuado una comparación de dicha metodología con un sistema bag-of-words (BoW), en el cual el texto se representa como una colección desordenada de elementos al margen de factores lingüísticos como la gramática (Lewis, 1998).

Datos	Objetivo	Tipo de método	Metodología	Input
Corpora en inglés y en español	Identificación del poder clasificatorio de las dimensiones	Experimento de aprendizaje automático	Algoritmo SVM	Dimensiones LIWC
				Dimensión estilométrica
	Modelo BoW	Experimento de aprendizaje automático	Algoritmo SVM	Frecuencia léxica
	Identificación de predictores individuales	Estadística inferencial	Regresión logística binaria y análisis discriminante	Categorías LIWC
				Características estilométricas

Tabla 2. Metodologías y técnicas usadas en el análisis de datos

Por último, se ha evaluado la eficacia de las categorías individuales en la clasificación de los textos según su nivel de verdad por medio de dos técnicas estadísticas: un análisis discriminante para los corpora globales en ambas lenguas, y una regresión logística binaria para cada uno de los subcorpora organizados por temáticas. La tabla 2 ofrece un resumen de los métodos y las técnicas usados en el análisis de datos.

RESULTADOS

A continuación se ofrece un resumen de los resultados principales hallados tras el análisis de los datos:

1. Los experimentos de aprendizaje automático en ambas lenguas han obtenido una clasificación exitosa de los textos de acuerdo con su valor de verdad por medio de los clasificadores que combinan las cuatro dimensiones LIWC y la dimensión estilométrica propuesta por la presente autora (ver figuras 1 y 2). En general, la dimensión que muestra un mayor poder discriminante por

sí misma es la primera, procesos lingüísticos, mientras que la cuarta dimensión, relacionada con asuntos personales, es la que obtiene peores resultados. Conviene señalar que, en general, las tasas de éxito en español son mayores que en inglés. Además, los clasificadores que combinan varias dimensiones actúan de manera más satisfactoria que las categorías aisladas, y que la mejora derivada de la adición de la dimensión estilométrica es notable, especialmente en inglés. Estos experimentos también demuestran que los resultados de clasificación dependen en gran medida del tema tratado en los textos; concretamente, el tema del buen amigo es el que obtiene mejores resultados de clasificación en ambas lenguas.

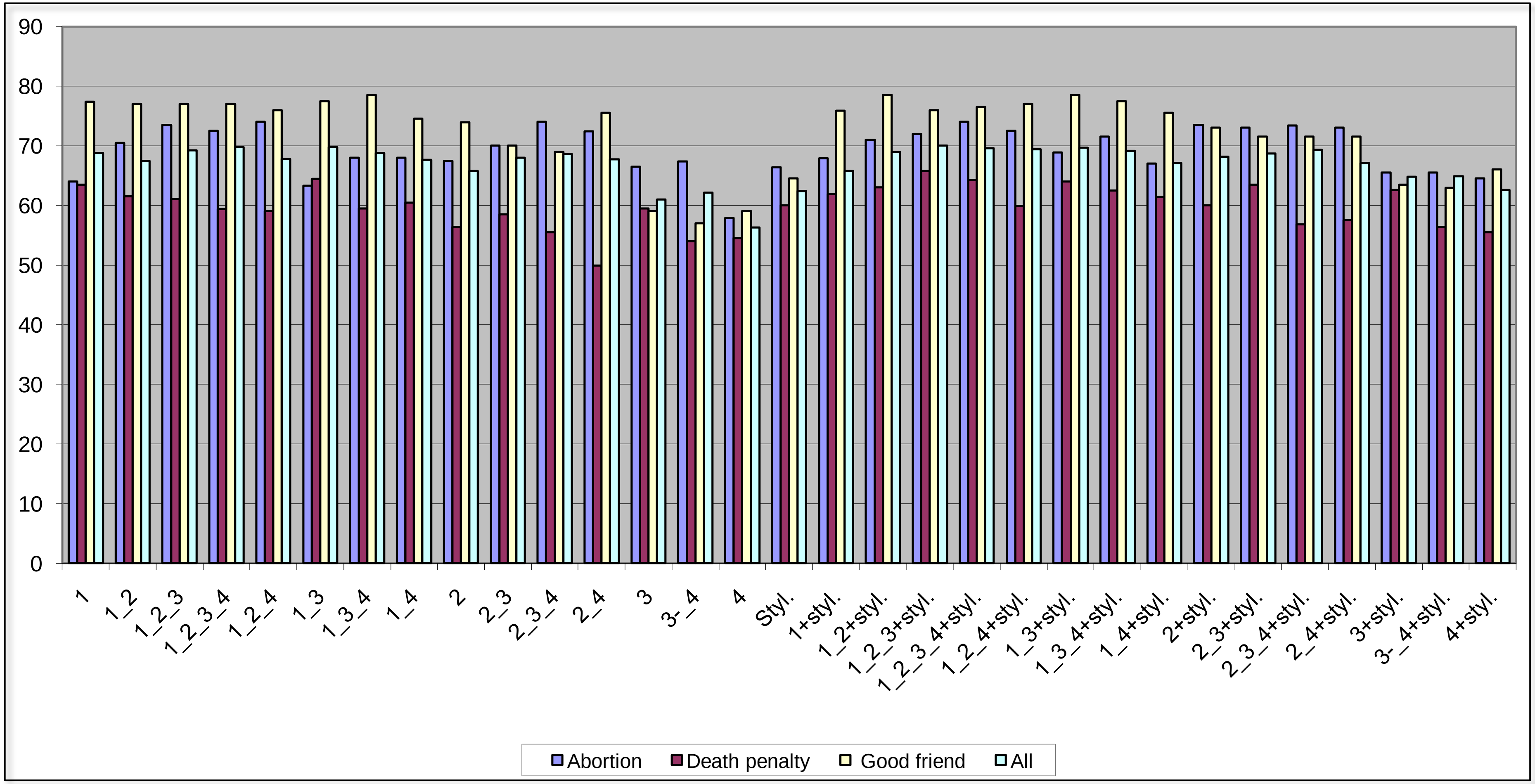

Figura 1. Resultados del experimento de aprendizaje automático para todos los subcorpora en inglés

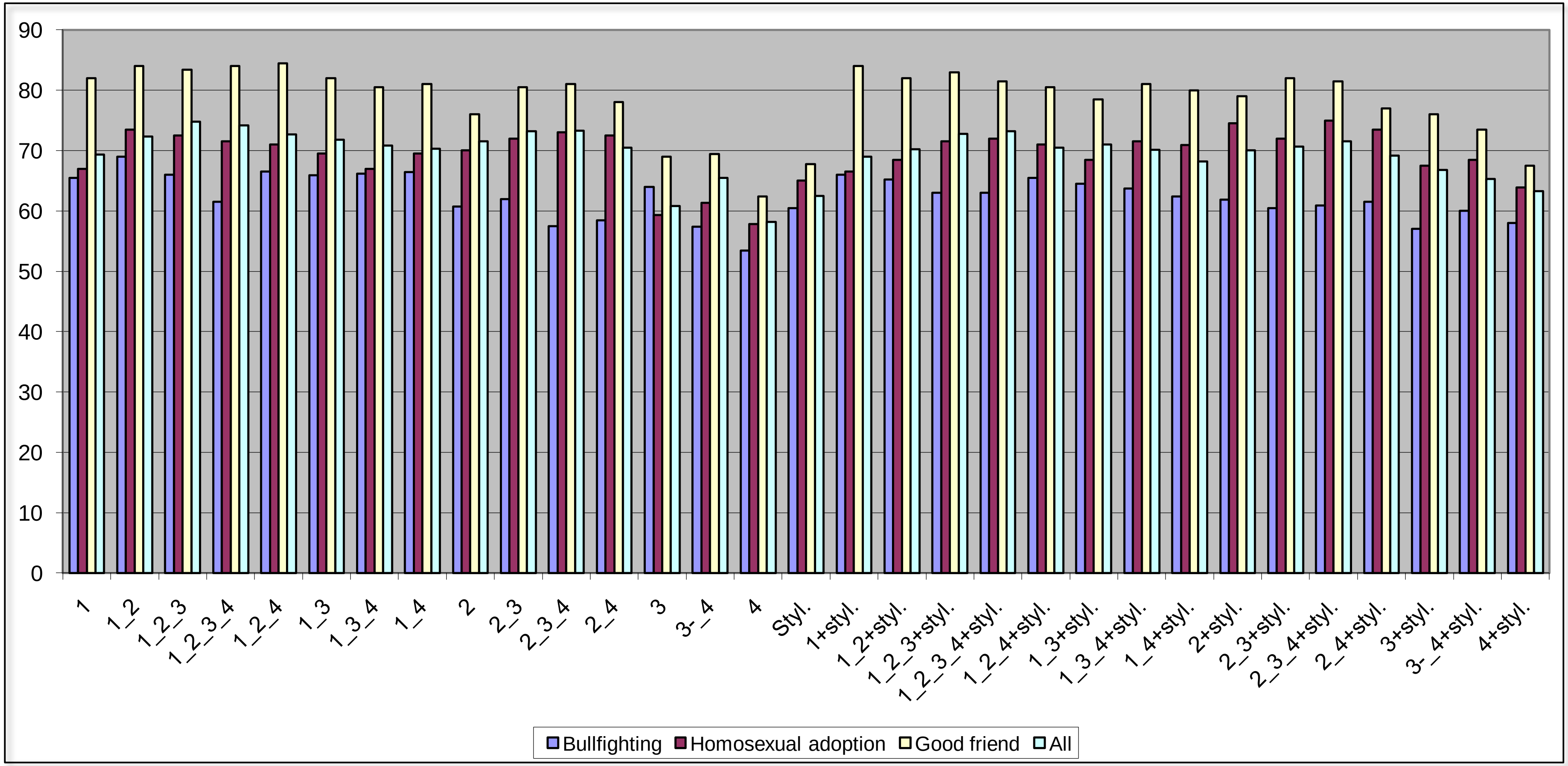

Figura 2. Resultados del experimento de aprendizaje automático para todos los subcorpora en español

2. El experimento de aprendizaje automático con el modelo BoW demuestra que sus resultados de clasificación son en todos los casos mejores que el azar, aunque en el tema de la pena de muerte la tasa es solo ligeramente mejor (figuras 3 y 4). A excepción de este subcorpus, al contrastar el modelo BoW con el experimento previo de aprendizaje automático, el comportamiento del clasificador es mejor en inglés que en español. Además de ello, se observa un paralelismo entre estos resultados y las tasas de éxito obtenidas previamente, excepto en el caso del subcorpus del buen amigo en inglés, que puntúa ligeramente peor que el del aborto.

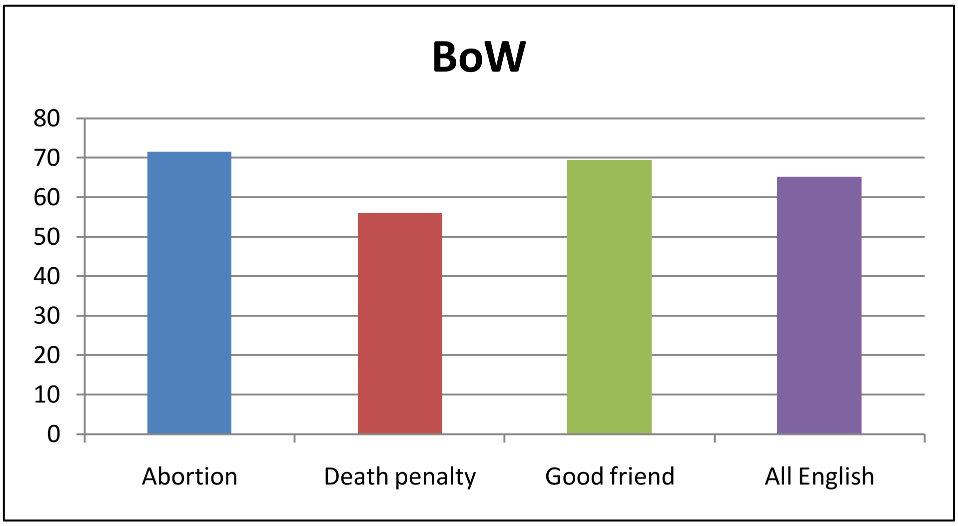

Figura 3. Resultados del modelo BoW en inglés

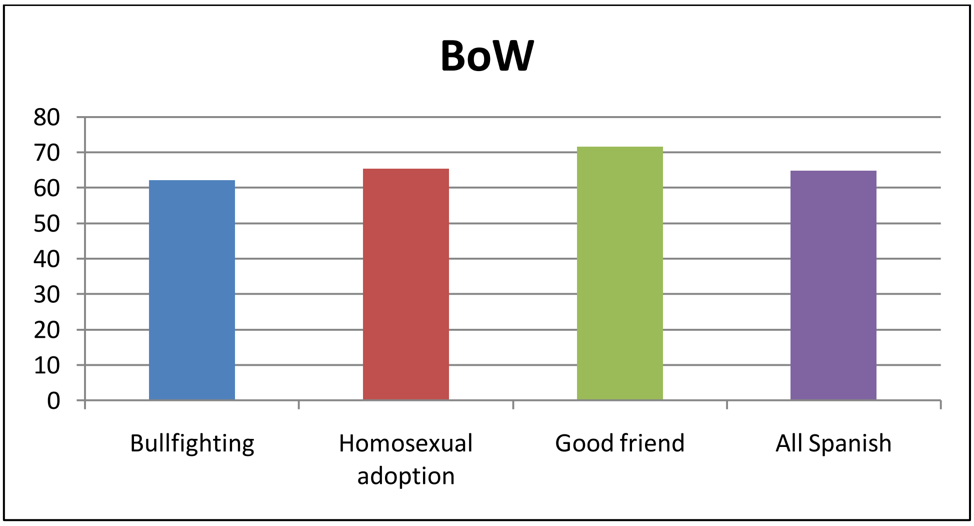

Figura 4. Resultados del modelo BoW en español

3. En general, las metodologías estadísticas de clasificación con categorías individuales obtienen mejores resultados que las técnicas de aprendizaje automático con dimensiones globales, excepto en el corpus general en español (ver figura 5). Además, la distribución de los resultados de clasificación es análoga a la resultante del experimento con las categorías globales.

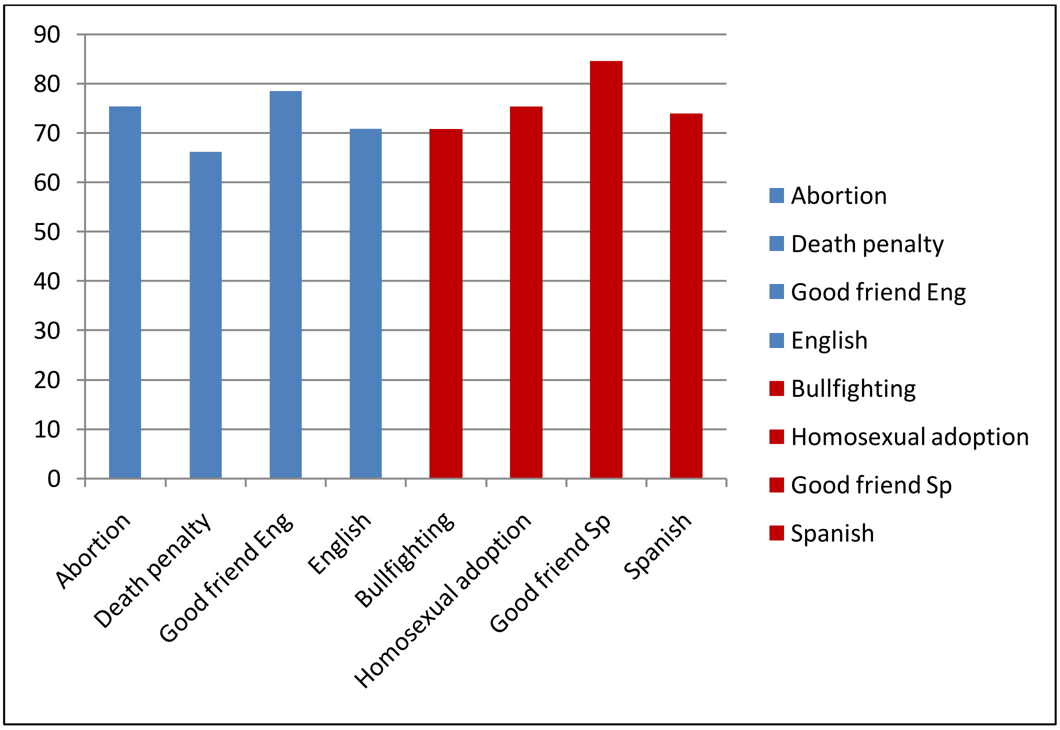

Figura 5. Resultados de clasificación estadística para todos los corpora

DISCUSIÓN

¿Cuál es la eficacia de las técnicas de clasificación propuestas en textos verdaderos y falsos en inglés y en español?

De acuerdo con lo comentado en el apartado de resultados, los experimentos de clasificación con las dimensiones LIWC y la dimensión estilométrica han resultado ser relativamente exitosos en ambas lenguas. La dimensión con mayor poder discriminante es la que engloba los procesos lingüísticos; parece lógico que así sea, dado el potencial de las palabras funcionales, las cuales constituyen una parte sustancial de esa primera dimensión. Su importancia ha sido ampliamente estudiada, no solo en lingüística computacional, sino también en el campo de la psicología. Se han asociado ciertas variaciones en su uso con el sexo, la edad, desórdenes mentales y la mentira. Por el contrario, como era de esperar a partir de la investigación anterior (Newman et al., 2003; Fornaciari y Poesio, 2011), la cuarta dimensión es la de menor poder discriminante, probablemente debido a su dependencia del tema. En general, los clasificadores que incluyen más de una dimensión funcionan mejor.

El modelo BoW muestra dos sublenguajes que se diferencian correctamente en términos de frecuencia léxica, a excepción del subcorpus de la pena de muerte. La menor distancia entre las tasas de éxito obtenidas con esta información básica y con las dimensiones en inglés puede deberse a la mayor idoneidad del contenido de LIWC en español. Esta idea la confirma la mayor cantidad de predictores incluidos en el modelo discriminante, que incluye 17 variables en español frente a las 9 del inglés.

Estos experimentos también demuestran que los resultados de clasificación dependen en gran medida del tema tratado. Resulta significativo el hecho de que los subcorpora del buen amigo obtienen las mayores tasas de éxito en ambas lenguas tanto en el experimento de aprendizaje automático como en la metodología estadística. Una posible explicación puede ser que al referirse a un bien amigo, los hablantes tienen más posibilidades de implicarse emocionalmente en el experimento; no están simplemente dando una opinión sobre un tema que les puede resultar ajeno, sino que se están refiriendo a una experiencia personal con un ser querido y mintiendo acerca de alguien por quien no sienten afecto. Es muy posible que, tal y como apuntan Newman et al. (2003), esta implicación personal se refleje en la expresión lingüística de la mentira.

¿Cuáles son los rasgos lingüísticos con mayor poder discriminatorio en el medio escrito a través de ambas lenguas?

Como se ha comentado anteriormente, la identificación de predictores ha sido más exitosa en los textos verdaderos, siendo uno de los principales la longitud textual. La tabla 3 muestra el inventario completo de predictores identificados para los textos verdaderos y falsos en todos los corpora.

	Inglés				Español			
	Aborto	Pena de muerte	Buen amigo	Global	Toros	Adopción homosex.	Buen amigo	Global
Total de palabras	V			V	V	V	V	V
1ª pers. sing.		V		V	V	V	V	V
1ª pers. plural	M							
2ª pers.			M	M			M	M
3ª pers.	M		M	M			M	M
Palabras > 6 caracteres								
Coma								
Dos puntos								V
Números							V	V
Ansiedad o miedo								V
Enfado	V							
Causa y efecto	M							
Entendimiento	V			V				V
Tristeza o depresión	V						V	
Amigos			V	V			V	V
Humanos			M			M		M
Familia			V					

Sentimientos positivos						V		
Certeza							M	M
Logro							M	
Inhibiciones							V	
Discrepancias				V				
Afirmaciones								M
Tentativos								V
Tiempo futuro del verbo								V
Tiempo pasado del verbo		V						V
Inclusivos	V			V		M		
Exclusiones		V		V	V			V
Comer, beber, tomar								
Sexualidad	M							V
Dinero y finanzas			M					
Movimiento						M		

Tabla 3. Predictores identificados para los textos verdaderos y falsos en todos los corpora

En lo que respecta a las diferencias entre ambas lenguas, en español suelen darse párrafos y respuestas más largas (Ramírez-Esparza et al., 2007), lo que se confirma en el presente experimento. Gran parte de la investigación previa en detección de la mentira ha concluido que los hablantes suelen ofrecer respuestas más cortas al mentir que al decir la verdad (DePaulo et al., 2003; Hartwig et al., 2006), y esto es precisamente lo que muestra el presente experimento en ambas lenguas. Ello puede deberse al hecho de que el sentimiento de culpa puede llevar a los mentirosos a ofrecer menos información para no incurrir en contradicción (Vrij, 2008).

Con respecto a la primera persona del singular, el otro predictor más común en el presente estudio, diversos investigadores como DePaulo et al. (2003), Mihalcea y Strapparava (2009) y Newman et al. (2003) aseguran que al decir la verdad el hablante se siente más cómodo con su discurso, y que por tanto tiende a identificarse con lo que dice a través del uso de esa persona gramatical. En este caso, esta variable ha demostrado ser significativa en todos los subcorpora en español y en la mitad de ellos en inglés. Ello puede deberse al poco uso de estos pronombres en español coloquial en comparación con el inglés.

Otros predictores significativos en los textos que contienen mentiras en ambas lenguas son las referencias a la tercera persona del singular, aunque ello es más evidente en inglés. Esta tendencia se había identificado ya en la investigación anterior sobre el tema (Burgoon et al., 2003; DePaulo et al., 1996; Hancock et al., 2004; 2005; 2008), al igual que ocurre con las referencias a la segunda persona (Mihalcea y Strapparava, 2009). Ello parece confirmar la tendencia del hablante a no identificarse con sus propias mentiras, distanciándose del discurso producido.

Por el contrario, una variable significativa en la identificación de textos verdaderos es la relativa al entendimiento. El papel de este proceso cognitivo no está claramente definido en el área de la detección de la mentira, ya que, como apunta Vrij (2010), la mayoría de investigadores no han encontrado diferencias estadísticamente significativas. Sin embargo, la tendencia encontrada en el presente estudio coincide con el planteamiento del Reality Monitoring, que predice que al decir la verdad se incluyen más operaciones cognitivas que al mentir. La categoría de exclusiones ha resultado ser también una variable significativa en los textos verdaderos, ya que suelen incluirse en explicaciones y descripciones de mayor complejidad, más frecuentes en el lenguaje carente de mentiras (Fuller et al., 2008; 2011; Newman et al., 2003).

Por último, conviene señalar la categoría amigos como último indicador común a ambas lenguas. Al igual que en el caso anterior, se trata de una variable característica de los textos verdaderos. Es esta una categoría perteneciente al grupo de humanos, al cual, de acuerdo con Mihalcea y Strapparava (2009), recurren los hablantes cuando se sienten cómodos con su discurso.

¿Cuáles son los indicadores más relevantes exclusivos del inglés?

Solo dos predictores han resultado significativos en inglés: discrepancias e inclusivos. Ambas categorías indican procesos cognitivos y resultan relevantes en los textos verdaderos, tal y como apuntan Granhag et al. (2001) o Vrij et al. (2001). El análisis contrastivo sobre las categorías LIWC realizado por Ramírez-Esparza et al. (2007) revela una mayor presencia de estas palabras en inglés que en español, lo que puede explicar los resultados del presente estudio. Los hablantes de inglés suelen incluir una mayor cantidad de este tipo de palabras, especialmente en lo concerniente a discrepancias y verbos modales, y un menor uso de las mismas parece estar asociado a la mentira.

¿Cuáles son los indicadores más relevantes en español?

En español, las categorías números y afirmaciones merecen especial atención. Con respecto a la primera, Ramírez-Esparza et al. (2007) encuentran frecuencias superiores en inglés; a pesar de ello, en el presente estudio se ha hallado una correlación positiva entre esta categoría y los testimonios verdaderos. Resulta interesante señalar que esta categoría no se había estudiado previamente. Este hallazgo puede estar relacionado de algún modo con los procesos cognitivos, así como los inclusivos y las exclusiones, ya que la expresión de cantidades requiere cierto grado de especificidad, de lo cual adolecen los relatos imaginados y las opiniones falsas. Por el contrario, las afirmaciones han resultado ser características de los textos falsos. Además, al igual que la categoría de números, de esta variable tampoco se había ocupado la literatura anterior sobre el tema. Las palabras contenidas en la categoría LIWC que implican afirmación están relacionadas con los términos que expresan certeza, los cuales también resultan relevantes en la clasificación de los textos falsos en español. Investigadores como Bond y Lee (2005) y Newman et al. (2003) también han encontrado esta correlación positiva, argumentando que probablemente se trate de la necesidad del hablante de utilizar palabras que aparenten reafirmar la verdad cuando mienten.

Existe otro proceso cognitivo que ha resultado ser significativo para el español, aunque esta vez está relacionado con la mentira: los términos tentativos. Esta categoría está relacionada con el entendimiento, el único proceso cognitivo que resultó significativo en los textos verdaderos en ambas lenguas. Como se ha comentado anteriormente, este tipo de términos suelen darse cuando la presencia de la verdad no requiere palabras relacionadas con la certeza a modo de reafirmación.

Otros dos predictores indican procesos psicológicos: ansiedad y humanos. La primera se ha asociado tradicionalmente a la mentira, especialmente en contextos forenses (Adams, 2001; Watson, 1981) o en experimentos donde el tema incomode a los participantes (DePaulo et al., 2003; Ekman, 1992).

Contrariamente a lo esperado, en el presente estudio esta categoría se encuentra asociada a los textos verdaderos. Una posible explicación es que las palabras relacionadas con la ansiedad en este tipo de estudios experimentales no se asocien a la expresión de la mentira, sino a la verdadera opinión de los participantes sobre un tema que les resulte especialmente polémico. Así pues, esos niveles más altos de ansiedad reflejados en el lenguaje podrían deberse a una vehemente defensa de sus ideales. En lo que respecta a la categoría humanos, los resultados de Mihalcea y Strapparava (2009) también la relacionan con la mentira, al igual que ocurría con los pronombres de segunda y tercera persona, debido a que las tres variables indican referencias a otros, esto es, la posibilidad de distanciarse de las mentiras de uno mismo.

La única categoría relacionada con la dimensión de asuntos personales que resulta discriminante en español, sexualidad, ofrece una contradicción con su papel en el subcorpus del aborto en inglés: mientras en el primero ayuda a discriminar los textos verdaderos, en el segundo es significativa en las opiniones falsas. Este es el único resultado contradictorio en el estudio, por lo que su interpretación entraña cierta dificultad. Como se mencionó anteriormente, la dimensión LIWC más dependiente del tema tratado es la cuarta, asuntos personales (Fornaciari y Poesio, 2011; Newman et al., 2003). En este caso, los subcorpora más íntimamente relacionados con la categoría sexualidad son el aborto y la adopción homosexual, pero curiosamente solo ha resultado ser un predictor en el primer tema en inglés y en el corpus global en español.

Por último, existe una característica estilométrica que ha resultado ser significativa en el modelo en español: los dos puntos. Aunque la longitud media de oración no aparece en ninguno de los dos modelos discriminantes, ambas variables están muy relacionadas. Las oraciones en español poseen una longitud media mayor que en inglés (Veiga, 2008), y ello se confirma en el presente análisis. Además, como se mencionó anteriormente, los participantes produjeron una mayor cantidad de palabras al elaborar textos verdaderos, especialmente en español, de ahí el poder discriminante de los dos puntos en esta lengua.

¿Hay algún indicador específico de ciertos temas?

En primer lugar, el modelo obtenido para el subcorpus del aborto incluye varios predictores únicos. Se ha encontrado una asociación entre la primera persona del plural y la mentira. Ello está en consonancia con los resultados presentados anteriormente, ya que, aunque tradicionalmente se haya estudiado como subcategoría del total de primera persona, en realidad se trata de una referencia a otros (Pennebaker et al., 2001). Por ello, puede utilizarse como un recurso lingüístico para distanciarse del yo, como la ssegunda o la tercera persona. En lo que respecta a los procesos psicológicos, dos emociones negativas contribuyen a discriminar los textos que contienen mentiras: el enfado y la tristeza. Tal y como se halló en el experimento llevado a cabo por Ali y Levine (2008), en este tipo de estudios experimentales este tipo de términos pueden derivar de la expresión de la verdadera opinión de los participantes sobre un tema que les resulta especialmente polémico. Ello está relacionado con el mayo nivel de ansiedad registrado en el corpus general en español. Además de ello, la categoría cognitiva de causa y efecto, que ha resultado ser un predictor de los textos falsos, tal y como ocurría en las investigaciones de Hartwig et al. (2006) y Vrij et al. (2008). Por el contrario, el subcorpus de la pena de muerte no incluye ningún predictor único.

La mayoría de variables significativas en el modelo del buen amigo en inglés están asociadas a los textos falsos, siendo una de ellas exclusiva: dinero y finanzas. Una posible explicación de la importancia de esta categoría está relacionada con el uso metafórico de palabras como costar, fortuna o ganancia, derivadas de la metáfora LA AMISTAD ES UN BIEN PRECIADO, estudiada por Kövecses (2000). De acuerdo con Gibbs (1994, el lenguaje metafórico suele utilizarse para delegar responsabilidades sobre el significado de lo comunicado, y esta puede ser la razón por la que la categoría de dinero y finanzas ha resultado significativa para los relatos inventados sobre la amistad. En cuanto a la categoría familia, esta ha resultado relevante para la clasificación de los textos verdaderos. La identificación de los verdaderos amigos, tema principal de estos textos, con parientes es bastante frecuente, de ahí la importancia de esta categoría en dicho subcorpus.

En lo que respecta a los subcorpora en español, no se ha hallado ningún predictor exclusivo para el de las corridas de toros. Por el contrario, el modelo para el subcorpus de la adopción homosexual incluye dos categorías exclusivas: sentimientos positivos y movimiento. El primero es indicativo de los textos verdaderos, lo que parece estar relacionado con la implicación personal del hablante comentada anteriormente; los resultados globales del presente estudio muestran así que todas las categorías relacionadas con las emociones están relacionadas con los textos verdaderos. Por otro lado, la categoría de movimiento, asociada con la mentira en este caso, había sido identificada también por Newman et al. (2003), quienes argumentan que los recursos cognitivos de los hablantes al producir mentiras pueden verse comprometidos por el esfuerzo de crear una historia creíble. Así pues, los verbos de movimiento son más simples y de acceso más inmediato que las palabras relacionadas con evaluaciones o juicios.

Finalmente, el modelo para el subcorpus del buen amigo en español incluye dos predictores exclusivos para la verdad y la mentira respectivamente: inhibiciones y logro. El primero está relacionado con el poder discriminatorio de los procesos cognitivos, mientras que la categoría logro puede relacionarse en cierto modo con los verbos de movimiento, los cuales, como se ha comentado anteriormente, están incluidos en el modelo correspondiente a la adopción homosexual. En la base de este paralelismo se encuentran la metáfora LA ESTRUCTURA DEL EVENTO y su submetáfora LOS PROPÓSITOS SON DESTINOS (Kövecses, 2000), por lo que la explicación esgrimida para el subcorpus anterior puede aplicarse también en este caso: la elaboración de una historia creíble o de una opinión falsa suele contener un mayor número de descripciones y expresiones concretas en detrimento de los marcadores de complejidad cognitiva.

CONCLUSIONES FINALES E INVESTIGACIÓN FUTURA

En general, los experimentos de clasificación llevados a cabo en el presente estudio han arrojado resultados satisfactorios, con una tasa de éxito máxima de un 78,5% en lengua inglesa y un 84,6% en lengua española. Los resultados confirman que hay una serie de características lingüísticas que contribuyen a los modelos de clasificación estadística en inglés y en español. Curiosamente, la identificación de predictores ha sido más exitosa a partir de los textos verdaderos. Las características lingüísticas comunes a ambas lenguas son longitud del texto, referencias propias, entendimiento, exclusiones y amigos. Por otro lado, la categoría referencias a otros ha resultado la más significativa para la identificación de la mentira. Hay también ciertas categorías específicas para cada lengua, así como diversas diferencias discursivas entre los temas, lo que confirma la importancia del estudio de la mentira dentro del contexto específico en el que se produce. Además de ello, cabe destacar que el conjunto de características estilométricas desarrollado para el presente estudio ha demostrado no ser significativo para el análisis individual, aunque ha mejorado el comportamiento del algoritmo de SVM, especialmente en inglés.

Como base para una posible investigación futura, la presente autora sugiere el análisis de palabras clave, el cual puede revelar algunas características no identificadas a través de un estudio más general basado en dimensiones léxicas o conceptuales. Ello supondría un tercer nivel de análisis jerárquicamente inferior a las dimensiones generales y a las categorías lingüísticas. Además de ello, una nueva línea de investigación en este campo podría profundizar en los mecanismos subyacentes a la mentira desde el punto de vista de la lingüística cognitiva, tal y como se ha esbozado en el presente trabajo. Puesto que el lenguaje metafórico se usa a menudo para delegar responsabilidades sobre el mensaje transmitido, resultaría ciertamente interesante explorar el papel que este juega en la producción de la mentira a nivel global.

En conclusión, el presente estudio pretende ser una contribución a la limitada exploración de la detección de la mentira en textos escritos, especialmente en lo que respecta a la lengua española. Además de ello, los hallazgos significativos obtenidos en relación con las peculiaridades discursivas no se habían apuntado antes en esta área de investigación, ni se había efectuado una descripción específica de los rasgos lingüísticos propios de la mentira a nivel contrastivo. En resumen, se espera que esta tesis doctoral contribuya a profundizar en el entendimiento humano de los mecanismos lingüísticos subyacentes al engaño.

CHAPTER 1

Introduction

Mundus vult decipi, ergo decipiatur

[The world wants to be deceived, so let it be deceived]

Latin proverb

1.1. Rationale for the Study

The distinction between truth and deception has garnered considerable attention from the domain of formal logic and psychological research. In the field of human kinetics, non-verbal communication has been claimed to play a key role in the detection of deception. More recently, verbal cues to deception have been also explored, but mostly in spoken language.

The popularity enjoyed by deception detection has transcended formal research, reaching popular culture and giving rise to several forms of entertainment such as “cheap” or “paperback” literature; it has even reached the television industry, becoming the central theme of different TV series[5] or of the well-known programme *Lie Detector,* where tools like the polygraph are used to check the veracity of the statements made by both the general public and important figures. However, this increasing appeal of deception detection has also fed widely held myths into the popular consciousness. For instance, most

[5] The most recent example is the TV series *Lie to Me*, created by Samuel Baum (Imagine Entertainment, 20th Century Fox Television).

people think that a liar can often be spotted just by observing their behaviour. Most relevantly, in the majority of cases, it is felt that, as Vrij (2010: 1) ironically puts it, "[f]ortunately, we are well protected against them, because professional lie catchers are good at spotting such liars." Nonetheless, this is just practitioners' and researchers' desideratum.

The investigation of linguistic cues to deception in written language is of utmost importance, not only in the forensic context with statements written by witnesses and people implicated in crimes, but also because of the increase seen by computer-mediated communication, where written texts constitute a fundamental element.

Nonetheless, the substantial body of literature on deception detection which is currently available should not mean that all is said in this field. Certain aspects of this area have not received enough attention. One of the gaps relates to the limitedexploration of the automated detection of deception in purely written texts, especially when it comes to the Spanish language. In addition, the discriminatory power of some stylometric features has not been studied yet. It is also worth noting the lack of research into this issue from a discursive perspective, exploring the potential linguistic variations according to the topic dealt with. Last but not least, as mentioned above, most machine learning approaches to the issue at hand have not given a comprehensive description of the linguistic cues to deception at a contrastive linguistics level, but have rather dealt with broader dimensions in a single language.

Extensive research has been conducted into the linguistic nature of deception, addressing the question of whether deceptive statements are deviant enough to betray insincere speakers. Experimental findings on the distinctive features seem to conflict, hence the scepticism towards this line of research voiced by certain scholars such as Vartapetiance and Gillam (2012). In the present study, we develop an approach in which deception in written language is not explored as a whole but taking into consideration the particular modality of the corpus and the discursive differences among topics. Our hypothesis is that the translation of linguistic cues to deception across languages could be an indication of their validity.

To summarise, the present PhD thesis intends to be a contribution to the study of computational linguistic tools as an aid to deception detection. By tackling this issue we hope to deepen our understanding of the linguistic mechanisms underlying deceit. It is hoped that forensic linguists, jurists, lawyers, computational linguists, and psychologists may benefit from this study.

1.2. Structure of the Thesis

1.2.1. Part one

The first part of this thesis consists of Chapter 2. It presents a theoretical review of the state of the art as regards linguistic cues to deception in written language. Chapter 2 provides the reader with the operational definition of the concept, an overview on several approaches to the study of deception and on previous research on its linguistic detection.

1.2.2. Part two

The second part of the thesis focuses on the empirical study itself. This covers Chapters 3, 4 and 5. Chapter 3 presents the research questions and the method which has been followed in order to carry out the study. The aim of the study is to explore the linguistic cues to deception in written language, performing a contrastive analysis between English and Spanish.

The results of the experiment can be found in Chapter 4, which also includes their discussion. The key findings reveal that the classification experiments perform efficiently, with a maximum success rate of 78.5% for English and 84.6% for Spanish. The results also confirm that, although there is a set of linguistic cues which contributes to the statistical classification models in English and in Spanish, there are some cues specific to each language.

Finally, having considered the obtained results, some final conclusions are drawn in Chapter 5. Furthermore, the major limitations of the present study are exposed and some plausible lines for further research are advanced.

CHAPTER 2

Deception, its Nature and its Detection

2.1. Shaping deception

In the context of human communication, deception plays an active role. Indeed, DePaulo et al. (1996) report that people tell an average of one to two lies a day, the context of mediated communication not being an exception (Hancock et al., 2004b). The philosophical discussion of lying to others and interpersonal deceiving mainly involves definitional questions. As Mahon (2008) puts it in the Stanford Encyclopedia of Philosophy, this group includes the questions of how lying and deceiving are to be defined; other questions relevant to the philosophical discussion of lying to others and other-deception are moral, including whether lying and deceiving are morally wrong. Due to the linguistic nature of this thesis, only questions of the first kind are considered here.

Although there is no universally accepted definition of lying (Kagan, 1998), Mahon (2008: 3) advances that the most common one is as follows: "to make a believed-false statement to another person with the intention that that other person believe that statement to be true". Accordingly, for lying there would be four necessary conditions: a statement given by a person; the untruthfulness condition of the statement; the requirement that the untruthful statement be made to another person or addressee; and the intention to deceive the addressee. Castelfranchi and Poggi (2002) insist on the last condition, stating that the deceiver's intention must be part of the definition. They clearly distinguish this purposeful act from misinformation which is perpetuated by mistake or not trying to mislead the recipient of the message, which is not to be deemed as deception. On the contrary, truthful messages intentionally conveyed in the belief that they are untruthful are usually considered to be a form of deception (DePaulo et al., 2003).

This is the definition of deception used in this work, since it necessarily involves the intention to communicate with another person by means of a statement, which is more than adequate for automated linguistic detection methods. Other forms of deception like omissions are more difficult to detect in this way. However, this does not imply that deception in the broadest sense of the term is to be identified with outright falsification. The former may be considered to be the hypernym and the latter just one of its potential hyponyms.

Methods of deception
Fabrications
Evasions
Equivocation
Concealments
Exaggerations
Omissions
Camouflage
Misdirection
Strategic ambiguity
Bluffs
Hoaxes
Tall tales
Charades
White lies
Sophistry
Half-truths

Table 2.1 List of methods of deception

This means in practice that it is possible to mislead others by means of further strategies like ambiguity or exaggeration (Burgoon et al., 1996). In this respect, Jensen (2007: 29) provides a comprehensive list of the methods used to deceive; Table 2.1 shows an adaptation.

2.1.1. Everyday lies

In everyday deception, which happens to be the most common type, especially conversation interaction, the methods in Table 2.1 are not mutually exclusive; quite the reverse, Burgoon et al. (1996) assure that deceivers tend to adapt their strategy during an interaction on the basis of the success of their deception. The communication modalities most suitable for this adaptation process are face-to-face oral interaction and synchronous computer-mediated communication (CMC, henceforth). In distance modalities such as asynchronous CMC or mail correspondence, immediate feedback is not that often available for the deceiver, thus the adaptation of their strategy is not so straightforward. In this respect, DePaulo et al. (1996) deem this kind of lying as an intrinsic part of everyday life, and address the categorisation of lying as follows: outright (total falsehoods), exaggerated (overstated facts and impressions), and subtle lies (evasions, omissions, and literal truths). A high percentage of lying occurs in social contexts and is low-consequence; as Picornell (2012: 19) puts it, in the event of detection “the consequences are little more than temporary slight embarrassment”.

2.1.2. High-stakes deception in forensic contexts

Everyday lies are contrasted with serious deception occurring in forensic contexts, which may lead to far-reaching consequences. In these situations, uncovering the full truth is essential for the clearing up of criminal and civil cases. Traditionally, a cue to deception considered sufficiently reliable in forensic investigations is nervousness or anxiety. An interrogated suspect displaying anxious behaviour is deemed likely to have something to hide. Nevertheless, some researchers have recently advanced that anxiety is not always an unequivocal symptom of deception. Specifically, Bull et al. (2006) insist that nervousness often arises from the stressful situation of a police interview. Most people are not used to police interrogations, and the mere fact of becoming a suspect makes them feel rather awkward. The authors also state that sometimes the police officers behave in an often accusatory way, which may intimidate the interviewees and lead to a leakage in behaviour, independent of the truth value of the statement, which is often interpreted erroneously.

In fact, these authors assure that some guilty suspects are able to manage their arousal and uneasiness far better than innocent ones, owing to the fact that they have become highly experienced in police interviews. As stated by Mann et al. (2002: 372),

> [d]ue to the large differences in people's attitudes, the content and consequences of their lies, their experience, and their ability to lie, there is never likely to be a less vague indicator than this [a change from normal behaviour within a particular individual].

In their study on high-stakes deception, the authors indeed highlight the basic differences in non-verbal behaviour during deceptive interactions between individuals. Furthermore, their findings reveal that anxiety does not play such a key role in deception as generally believed. This is monitored, among other non-verbal signs, through eye blinking, an increase of which had been previously associated with nervousness and anxiety. These researchers find that suspects blink less when lying, which they explain on the grounds of cognitive load. Significantly enough, deceivers also make longer pauses, which is attributed to the same reason by the researchers.

Both in common law and in civil law jurisdictions, the practice of deception is duly punished; accordingly, deceit is a tort and often a crime. The case *Derry v Peek*[6] is often quoted in the study of deceit within common law, since it is the first time deceit is defined, despite the fact that there was no sufficient evidence to confirm that deceit really occurred (Burdick, 1905: 373):

> The ground upon which an alleged belief was founded is a most important test of its reality… if I thought that a person making a false statement had shut his eyes to the facts, or purposely abstained from inquiring into them, I should hold that honest belief was absent…

In this respect, the intention to deceive the addressee condition is worth commenting on. In criminal law, intention or purpose is one of the four general types of *mens rea*[7] or the required mental state necessary to constitute a conventional as opposed to a strict liability crime, and it is said that it "requires a finding that the defendant has as a conscious objective to commit the act or result proscribed by the crime" (Strader, 2002: 9). The Spanish equivalent of this concept is *dolo* –literally, intention. This is certainly applicable to deceit within the forensic context, as shown in Green (as cited in Mahon, 2008: 21), who advances that lying is neither necessary nor sufficient for perjury in common law, since if a person under oath to testify, declare, depose, or certify truly,

[6] (1889) 14 App Cas 337
[7] Strader (2002) arranges the four types of *mens rea* from the most difficult to the least difficult to prove: intention or purpose, knowledge, recklessness, and negligence.

before a competent officer, wilfully makes an untruthful statement without the intention that any other person believe it to be true, the person is not lying, but they are considered to have committed perjury –a federal crime consisting of deliberately giving false, incomplete or inaccurate information. On the contrary, providing that a person is under oath and wilfully makes an allegedly untruthful statement with the intention that it is believed to be true, and the statement happens to be so, the person has not committed perjury.

Of interest, there is an exception to the commitment to tell only the whole truth in trials: in modern legal systems, no person is compelled to testify against themselves. In the context of Anglo-American common law, this originated in medieval England and Wales. This legal system has since then been in charge of the provision to individuals with the means to protect themselves from self-incrimination, guaranteeing fundamental fairness, justice, and liberty by means of the due process (Levy, 1969). The well-known Fifth Amendment to the United States Constitution enshrines the above-mentioned privilege against self-incrimination and the right to silence: "No person [...] shall be compelled in any criminal case to be a witness against himself, nor be deprived of life, liberty, or property, without due process of law."[8] This is crystallised in the Miranda warning, which is given by police in the United States to criminal suspects in police custody and during custodial interrogations. The rest of the countries deriving their laws as an extension of the English Common Law have equivalent rights, although some of them have their own peculiarities. Most notably, the English Criminal Justice and Public Order Act 1994 amended the right to silence by allowing the jury to draw inferences in the event of a suspect providing an

[8] Bill of Rights from Cornell University Law School, publicly available at http://www.law.cornell.edu/constitution/fifth_amendment

explanation after having refused to do so during custodial interrogation. Accordingly, although the jury is also free to not make such an inference, law supports their assumption of a defendant's alleged fabrication. This amendment has been rejected outright by most common law jurisdictions, on the grounds that an innocent person may also have a reason for not speaking freely to investigating police (Bagaric, 1997). However, in these jurisdictions a testifying defendant is sworn to tell the truth under oath and pain of perjury.

As regards civil law jurisdictions, these guarantees are taken a step forward. In the particular case of Spain, article 24 of the Spanish Constitution enshrines the right to effective protection of the court –*tutela judicial efectiva*–, which is a similar concept to the due process in common law. Two fundamental rights are specified in the second section of this article: the privilege against self-incrimination and the right not to plead guilty. According to Aparicio (2009), this has often been interpreted as having a positive side embodied in the so-called "right to lie". The author explains that this right is not explicitly acknowledged in Spanish legislation, but the aforementioned rights were established, as in common law, so as to create a safety net which prevents others from entering the private sphere of people when detrimental consequences may be derived for fundamental rights, which is the case when a person is involved in criminal proceedings. However, the Spanish Constitutional Court states that the defendant is not legally compelled to tell the truth, hence the relinquishment by the State to punish this behaviour. At this point, the main difference from common law jurisdictions is that a defendant does not swear an oath and is not technically a witness; thus, if they lie, they are not committing perjury.

Closely related to this different treatment of the defendant's statement is the concept of the plea, which is one of the major differences between both legal systems. In common law, a defendant may be sentenced immediately after their guilty plea, whereas civil law jurisdictions generally lack this concept. A confession by the defendant is treated like any other piece of evidence. Even if they provide a full confession, the trial is not prevented from occurring (Etienne, 2005). Probably due to the considerable relevance of the defendant's testimony and the checking of its veracity in common law jurisdictions, research on high-stakes deception is much more developed than in civil law tradition countries such as Spain, whose legal system cannot punish the defendant for lying if the lies have been told in the exercise of the privilege against self-incrimination.

2.1.3. Particular kinds of deception

There are some particular cases in which deception is considered to display special characteristics. For instance, children are usually lacking in credibility due to their overactive imagination (Bull, 1997), and their production of lies is usually studied apart. Furthermore, pathological lying has been deemed a special form of deception. A working definition of this phenomenon is provided by Healy and Healy (1969:1):

> Pathological lying is falsification entirely disproportionate to any discernible end in view, engaged in by a person who, at the time of observation, cannot definitively be declared insane, feebleminded, or epileptic. Such lying rarely, if ever, centers about an event; although exhibited in very occasional cases for a short time, it manifests itself most frequently by far over a period of years, or even a life time. It represents a trait rather than an episode.

Thus, this type of deception is characterised by a medical record of frequent lying for no apparent reason.

According to Hausman (2003), most scholarly literature references agree that pathological lying should be differentiated from other psychiatric conditions associated with different forms of deception. The most representative ones are psychopathy (Hare et al., 1989), antisocial personality disorder or ASPD (Ford, 1999), obsessive-compulsive disorder or OCD (Dike, 2008), narcissistic personality disorder or NPD (DSM IV-TR, 2000), histrionic personality disorder or HPD (DSM IV-TR, 2000), borderline personality disorder or BPD (Böhm & Steller, 2008), factitious or Münchausen disorder (Feldman et al., 1993), and Ganser syndrome (Carney et al., 1987). The particular case of confabulation is also worthy of attention. It is defined as a condition where patients try to plug memory gaps generated by organically derived amnesia with confabulated material (Dalla Barba, 1993). The special features of mental patients with these disorders make research on their untruthful speech a separate branch of study. For instance, ordinary lies are told with the intention to avoid punishment or to obtain an external benefit; on the contrary, pathological lying is rather purposeless and sometimes even self-incriminating.

2.2. The study of deception

The empirical study of deception in language dates at least from Undeutsch (1967), who firmly believed in the existence of certain criteria for the configuration of the truthfulness of statements. Some years later, he would formulate what is nowadays known as the *Undeutsch hypothesis*: “a statement derived from a memory of an actual experience differs in content and quality from a statement based on invention or fantasy” (Undeutsch, 1989: 102). Since then, this line of research has shown that deceivers somewhat differ in verbal, visual and physiological behaviour from truth-tellers. The identification of the key cues to deception has been commonplace in forensic linguistics, especially in the English-speaking world. Thus, several studies have been devoted to the assessment of human ability to detect lies. According to Vrij (2000), most pieces of research report accuracy rates ranging from 45% to 60%, with a mean accuracy rate of 56.6%, which shows that, in practice, it rarely performs above chance. More recently, researchers have become increasingly interested in the development and evaluation of automated tools for identifying lies, giving rise to automated and computer-aided deception detection. This discipline may benefit from several areas of natural language processing, such as automated text classification, opinion mining and sentiment analysis, as well as linguistic areas like discourse analysis, pragmatics and phonetics. The potential applications of automated and computer-aided deception detection cover the areas of law enforcement, advertising, computer-mediated communication, national security, and human resources, to name but a few.

Non-verbal language makes up about two-thirds of all human communication. Probably due to its importance, research on deception detection has mostly focused on physiological cues, which has led to the creation and use of several tools such as the polygraph. More recently, researchers have begun to explore the identification of verbal cues which are useful for separating truth from lies, mainly in oral language.

2.2.1. Theories of deception

Several theories of deception have provided the guiding foundation for methods of deception detection. In this section, the four major theories will be reviewed:

leakage theory, Zuckerman's four-factor model of deception, interpersonal deception theory, and the self-presentational view.

Ekman and Friesen (1969) first noted that deception could be made evident through unaware physical behaviour. According to the authors, lying can manifest itself through two different types of behavioural cues. First, deception may become obvious through physical leakage cues. This type of cue arises when deceivers attempt to conceal their spontaneous reactions and true emotions. This notion led to the study of micromomentary facial expressions as a method for lie detection (Jensen, 2007). Second, deception cues refer to the absence of natural movements commonly displayed by speakers during truthful interaction. In an attempt to control their non-verbal behaviour, deceivers usually show unnatural body stiffness which may itself be betraying. In this respect, DePaulo et al. (1988) explain that motivated deceivers trying to control their demeanour behave differently from those who do not try to cover it up. Thus, the least successful ones are motivated deceivers (usually in high-stakes situations) involved in oral interactions. This approach to deception is commonly known as leakage theory.

Some years later, Ekman (1985) first provided evidence to support Zuckerman's four-factor model of deception (1981). This model relies upon four different factors: arousal, emotion, cognitive effort, and attempted control of behaviour. These psychological areas are readily susceptible to variations during deceptive communication. According to this theory, higher levels of arousal and negative emotions are usually observed, increased cognitive load is shown, as well as a lack of spontaneity, resulting from the deliberate attempt to control non-verbal behaviour. However, this model has its detractors, especially concerning the first factor. As explained above, Mann et al. (2002) suggest that deceivers do not necessarily feel nervous when involved in high-stakes situations, since some of them happen to be highly experienced liars. Furthermore, an anxious interviewee is not always a deceiving one, and vice versa.

Buller and Burgoon (1996) discuss that the different ways of lying do not occur per se, but that the context where the lie is uttered plays a major role, formulating their interpersonal deception theory (IDT) within the context of interpersonal communication. They argue that lying is adaptable and that different issues should be considered, such as the deceiver’s motivation to lie; the consequences for liars in the event of being caught; the degree of formality of the situation; and the existing relationship between deceiver and deceived, to name but a few. Even personality plays a role in producing language, like in the dichotomy introversion-extraversion, hence its influence on the manner of deceiving and the major role played by idiolect (Campbell and Pennebaker, 2003; Pennebaker and King, 1999). Thus, this model emphasises the deceiver’s multiple roles of monitoring, interpreting and adapting untruthful messages taking into account the immediate feedback received.

It is worth noting that IDT presents deception as a strategic interaction among participants, resulting in a more holistic approach to deception than the previous ones. At a global level, the model considers that participants in the interaction bring with them cognitive and affective processes such as their expectations or detection apprehension, as well as their behavioural patterns –e.g. their skills. After an initial behavioural display, the deceiver tries certain strategies which may be modified depending on the feedback received from the recipient, hence the importance of their management of information, images and behaviour. This bidirectional feedback is obtained through the leakage of nonstrategic behaviours by both parties. Interestingly enough, the role of the recipient in this model is properly explored too, not only regarding the feedback provided, but also the elements leading to deception detection accuracy on their part. Accordingly, Buller and Burgoon (1996) assure that the recipient gets involved in interaction with a certain level of suspicion –although they allow for the possibility of level zero. Within their cognitions, credibility judgements may result in a modification of this level of suspicion. Regarding the recipient's behavioural patterns, they mainly include suspicion display and uncertainty management. In post-interaction, both parties will evaluate their success at their respective roles.

Finally, DePaulo et al. (2003) propound the self-presentational view of deception, on the grounds that deceivers and truth-tellers hold different beliefs, referred to by the authors as deception discrepancy. They explain the two main implications of this theory as follows: "First, deceptive self-presentations are often not as convincingly embraced as truthful ones. Second, social actors typically experience a greater sense of deliberateness when their performances are deceptive than when they are honest" (DePaulo et al., 2003: 77). This discrepancy notwithstanding, deceivers must project a convincing impression, which generates five categories of cues: they are expected to appear less communicative than truth-tellers; they are predicted to show a more negative attitude; they will probably display a higher level of anxiety; their accounts are expected to be less convincing than truthful ones, as well as to contain a limited amount of sensory details as compared to truthful accounts. Furthermore, the authors highlight the role of motivation to succeed.

2.2.2. Professional methods of deception detection

Nowadays, there are a considerable number of methods of deception detection, but only a few are regularly used by professional lie-catchers. Some others have been proposed by professionals and scholars, but their application to real world situations is virtually null. These methods are usually classified into two broad categories: physiological and behavioural methods (see Table 2.2). The former category includes polygraph, brain activity analysis, thermal analysis, and voice stress analysis. Statement validity assessment, linguistic analysis and behavioural analysis are comprised within the latter group. A basic description of these methods is provided in this section.

<table>
<tr><th>Physiological methods</th><th colspan="2">Behavioural methods</th></tr>
<tr><td>Polygraph</td><td colspan="2">Nonverbal assessment tools</td></tr>
<tr><td>Brain activity analysis</td><td rowspan="3">Verbal assessment tools</td><td>Statement Validity Assessment</td></tr>
<tr><td>Thermal analysis</td><td>Reality Monitoring</td></tr>
<tr><td>Voice stress analysis</td><td>Linguistic analysis</td></tr>
</table>

Table 2.2 Methods of deception detection

2.2.2.1. Physiological methods

The development of physiological methods relies on the assumption that arousal, emotions and cognitive changes associated with deception, relevant to Zuckerman’s four-factor model of deception, generate systematic physiological changes in aspects such as blood flow, hemo-oxygenation and neuronal activity (Vrij, 2000). This has resulted in a wide array of methods, the origins of which lie in the most rudimentary techniques, such as the holding of an ostrich’s egg during a suspect's confession –its breaking would be considered to be a reflection of the suspect's arousal, allegedly provoked by deception.

Nowadays, sophisticated technologies are available for the purpose of spotting liars. Probably the most popular physiological method is the polygraph, also known as a lie detector. This device is able to assess parameters such as palmar sweat or heart rate by virtue of sensors attached to the body, an increase of which is an alleged physical response to the human arousal generated by feelings of guilt (Vrij, 2000). Both the Control Question Test (CQT) and the Guilty Knowledge Test (GKT) –also known as the Concealed Information Test– are interviewing methods which involve the use of a polygraph. In their study on this tool and its limits, Faigman et al. (2003) explain that in the former method the functioning of the polygraph is adjusted by means of a set of control questions in which the subject is deliberately asked to lie. Thus, the level of arousal from these kind of questions is taken as a baseline to draw a comparison with real questions. The authors insist on the extensive criticism levelled at this interviewing method for being unreliable and non-scientific, despite its widespread use in the United States. Specifically, the major problem with this method seems to be the large proportion of false positives obtained. On the other hand, the GKT, as suggested by its name, "determines whether an interviewee has knowledge about a crime that would only be known to the perpetrator" (Jensen, 2007: 37). According to Faigman et al. (2003), this interviewing method is deemed more objective and scientific by experts in the field.

As shown in Table 2.2, in addition to the well-known polygraph, there are other physiological methods for lie detection, such as those relying on the analysis of brain activity. For instance, Farwell and Donchin (1991) report that the results from criminal applications involving electroencephalograms (EEGs) to measure event-related brain potentials (ERPs) are almost as accurate as the polygraph. Similarly, functional magnetic resonance imaging (fMRI) is also based on the analysis of brain activity, although nowadays its reliability is being questioned, which has led to the investigation of some alternatives to this technique, such as near-infrared spectroscopy (NIRS). In Jensen's words (2007: 38), "NIRS does not require people to remain stationary and uses optical technology to measure neuronal, metabolic, and hemodynamic changes that may indicate arousal associated with deception."

On the other hand, Pavlidis and Levine (2002) designed and tested a promising thermal image analysis method for polygraph testing. As the authors have it, this method "can serve as an additional channel for increasing the reliability and accuracy of traditional polygraph examination" (2002: 1). Specifically, it involves the extraction of subtle facial temperature fluctuation patterns through nonlinear heat transfer modelling, based on the assumption that an increase in the level of arousal caused by deception manifests itself through an instantaneous periorbital warming pattern.

Last but not least, voice stress analysis (VSA) assumes that the psychological stress experienced by deceivers forces certain changes in blood circulation, physiologically reflected by an elevated voice pitch (Vrij, 2010). Significantly enough, in most studies on this technique, researchers have been unable to prove its efficacy (Gamer et al., 2006), mainly due to the large proportion of false positives obtained.

2.2.2.2. Behavioural methods

As has been seen, physiological veracity assessment tools deal with the body functions of human beings. When it comes to behavioural methods, the object of study is the range of actions and mannerisms made in deceptive interactions. In this category, the term behaviour comprises two subcategories: the first one deals with the analysis of non-verbal language, whereas the second subcategory is concerned with the assessment of verbal communication. As shown in Table 2.2, the latter group comprises Statement Validity Assessment, Reality Monitoring and linguistic analysis.

The exploration of non-verbal cues has been widely used as a method of deception detection, either in isolation (e.g. Bond and Robinson, 1988; Burgoon et al., 1996; Ekman and Friesen, 1969; Meservy, 2007) or in combination with the role of verbal cues (e.g. Brownsell and Bull, 2011; Ebesu and Miller, 1994; Kraut, 1978; Vrij et al., 2000). Despite the extensive research conducted on non-verbal cues to deception, the lack of consistency in the findings hinders their application to professional practice.

Regarding verbal assessment tools, Statement Validity Assessment (SVA) is the most widely used of these tools, since, as Vrij puts it, it is "accepted as evidence in some North American courts and in criminal courts in several Western European countries, including Austria, Germany, Sweden, Switzerland and the Netherlands" (2010: 201). The SVA method is based on the Undeutsch hypothesis, which, as advanced in section 2.2, concerns the divergence in content and quality of outright fabrications from truthful statements. Specifically, it was designed to determine the credibility of children's statements in sexual offence trials (Trankell, 1972).

Within SVA, Criteria-Based Content Analysis (CBCA) is the core phase. It takes place after a case-file analysis and a semi-structured interview. In this phase, the interviewer assesses the quality of the statements in a systematic fashion by means of transcripts, scoring responses according to an inventory of 19 predefined criteria such as quantity of details, descriptions of interactions, related external associations, spontaneous corrections, and details characteristic of the offence (for a comprehensive list see Steller and Köhnken, 1989). As opposed to physiological methods of deception detection, interviewees' non-verbal behaviour is not evaluated. This allows the interviewer to give undivided attention to linguistic content, which is deemed convenient by most SVA experts (Vrij, 2010). Interestingly enough, all CBCA criteria indicate truth, thus, strictly speaking, it cannot be considered a deception detection method. In Ruby and Brigham's words, "the presence of one of these specific content characteristics indicates a truthful statement while the absence of them indicates nothing" (1994: 18).

A further verbal veracity assessment tool which uses a scoring mechanism to judge statements is Reality Monitoring (RM). Originally, it was mainly concerned with the different cognitive processes involved in the narration of perceived and imagined events (Sporer, 1997). Unlike SVA, professional practitioners do not use RM for lie detection. However, Vrij insists that this tool has become increasingly popular within the field, and that "it has attracted the attention of scientists worldwide, and to date researchers from Canada, Finland, France, Germany, Spain, Sweden, and the United Kingdom have published RM deception research" (2010: 261). Specifically, RM judges expect verbal recalls of actual events to include a greater deal of sensory, contextual, and affective information.

A feature shared by RM and CBCA is that both methods involve working with transcripts, hence the preclusion of non-verbal language in the analysis. Furthermore, despite the absence of a standardised set of RM criteria, some of them overlap with some CBCA criteria. Broadly speaking, RM experts expect certain criteria to occur more often in truthful statements, namely clarity, perceptual information, spatial information, temporal information, affect, reconstructability of the story, and realism. On the contrary, cognitive operations used to be considered characteristic of untruthful statements within this theory, although, as will be seen in subsequent chapters of this work, several studies have obtained the opposite results (e.g. Granhag et al., 2001; Memon et al., 2010).

Last but not least, linguistic analysis is also based on the premise that fabricated messages qualitatively differ from truthful ones. Nonetheless, a crucial difference from the previous methods is worth highlighting: linguistic analyses operate independently of message meaning. Jensen (2007) establishes two broad categories in linguistic analysis: message feature mining and speech act profiling. Firstly, researchers dealing with message feature mining attempt to determine the truth value of the statements by virtue of objective features extracted from the text. Often, this kind of research is not enshrined in any preconceived deception theory, and makes use of certain software applications specifically developed for its purposes or subsequently adapted. By and large, empirical deception studies which apply an automated method for detecting linguistic cues in text independent of context fall within this group, including the experiments performed in the present work. On the other hand, speech act profiling is a method of conversation classification devised by Twitchell et al. (2004). Thus, the main purpose of this method is to make clear the speaker's intention, so as to explore the role of pragmatics in the configuration of linguistic deception.

2.2.3. Research on verbal cues to deception: state of the art

In recent decades, there has been a considerable amount of studies exploring verbal cues to deception, both in oral and in written language. Vrij (2010) summarises the findings of each of the 69 studies which together make up all of the studies published in English, where verbal behaviour of adult truth-tellers and liars has been compared. As he puts it, "a verbal cue uniquely related to deception, akin to Pinocchio's growing nose, does not exist. However, some verbal cues can be viewed as weak diagnostic indicators of deceit" (Vrij, 2010: 103). Accordingly, scholars have traditionally studied a cluster of verbal cues rather than each cue individually. Despite the successful classifications obtained on this basis –ranging from a 67% to 80% success rate– Vrij remains pessimistic about the discriminatory potential of individual verbal cues; in his own words, "it is unknown to what extent a certain cluster that works in one situation or one group of participants also works in another situation or with another group of participants" (2010: 108).

In this section, an overview of these verbal cues is provided, firstly explaining the major cues –organised into linguistic and psychological– and then commenting on some further cues also present in previous literature.

2.2.3.1. Major linguistic cues

The two linguistic elements most widely proved as cues to deception are overall production of words or text length and pronouns. Their measuring is straightforward, since the first category simply involves the counting of all the words making up a text, and the second one the raw frequency –which may be subsequently computed into relative frequency if needed– of 1st, 2nd and 3rd person pronouns.

2.2.3.1.1. Overall production of words

As regards the first parameter, previous research has yielded varying results depending on the language modality. Concerning oral communication, where time to plan the responses is limited, research tends to suggest that deceivers give shorter answers and accounts than truth-tellers (DePaulo et al., 2003; Hartwig et al., 2006; Vrij, 2008). According to Vrij (2008), liars may prefer not to give too much information about a situation they have not really experienced so as not to be caught deceiving. In this respect, DePaulo et al. (2003) highlight that the real reason behind this is that fabricated events are more cognitively demanding than real ones. As for high-stakes situations, deceivers may prefer not to provide a great deal of oral information in order to avoid leakages caused by nervousness which may betray them (Savitsky and Gilovich, 2003). A sense of guilt is more commonly found in law enforcement contexts, where the lies are frequently told so as to protect the speaker from punishment.

However, when it comes to the written medium –most commonly explored through CMC– it has been observed that liars tend to use more words (Hancock, 2004, 2005, 2008; Zhou et al., 2004b; 2004c). Burgoon et al. (2001) explain that liars may use more words to manage information flow and to decrease suspicion, and Zhou and Zhang (2007) explain that deceivers may rely on greater detail to successfully persuade their addressee when leakages do not have so severe consequences as in forensic contexts. At this point, it is interesting to determine whether the language produced by deceivers in the experiments is mainly devoted to lies or to truths. Anolli et al. (2002) and Picornell (2012) tackle this issue, concluding that despite the larger amount of words found in the allegedly deceptive narratives, a mass of truthful information frequently surrounds fabricated events. On the other hand, Vrij et al. (2007) found no difference as far as this parameter is concerned.

2.2.3.1.2. Personal pronoun use

The patterns of pronoun usage associated with deception are worth exploring. As regards 1st person singular, several researchers such as DePaulo et al. (2003), Mihalcea and Strapparava (2009), and Newman et al. (2003) assure that truth-tellers show an increased use of 1st person singular pronouns, since they are more prone to identify with their statements than deceivers. Although another study has found just a weak relationship between this cue and deception (see Vrij, 2008), broadly speaking it has been deemed a powerful predictor in English. When it comes to Spanish, in Masip et al. (2012) 1st person singular pronouns do not prove significant for either truthful or untruthful statements.

On the other hand, references to others normally include 2nd and 3rd person pronouns. Concerning the former, it is not so commonly associated with the truth value of the statements, although some studies such as Mihalcea and Strapparava (2009) have identified it as a predictor of deception. As regards the latter, previous research has generally found that the 3rd person pronoun is more frequently found in untruthful statements (Burgoon et al., 2003; DePaulo et al., 1996; Hancock et al., 2004; 2005; 2008; Knapp et al., 1974; Knapp and Comadena, 1979; Kuiken, 1981; Vrij, 2000; Weiner and Mehrabian, 1968; Zhou et al., 2004a). DePaulo et al. (2003) explain that this cue entails detachment from the self, which is a general tendency in deceivers. Some contradictory findings are obtained in Bond et al. (2005), Newman et al. (2003), and Zhou et al. (2004b), where there is a reduction in the use of the 3rd person. A plausible explanation is advanced by Newman et al. (2003): they assert that the topic involved in the experiment (abortion) may lead deceivers to include specific proper nouns in their narratives to make them more credible, to the detriment of 3rd person pronouns.

2.2.3.2. Major psychological cues

Psychological categories are not as straightforward as linguistic ones. The selection of the words making up these categories is a rather subjective process, which usually entails two or more human judges. As will be explained below, such is the case of the second LIWC dimension. The two psychological elements most commonly identified as cues to deception are emotion and cognitive complexity.

2.2.3.2.1. Emotion words

Previous research suggests that increased levels of negative emotion, embodied in words like *grief*, *hate*, or *afraid*, are observed during deceptive communication (Newman et al., 2003; Zhou et al., 2004b). Therefore, liars are expected to produce more negative terms during deception. According to Burgoon et al. (2003a), these terms are a reflection of their feelings of guilt about their lying, which seems to be closely related to the detachment from the self reflected in the deceivers' preference for references to others instead of 1st person singular pronouns. In this respect, DePaulo et al.'s (2003) meta-analysis revealed that the effects of negative emotions leaking through to their speech could be the result of liars finding it too difficult to come up with a plausible answer.

Nonetheless, some contradictory findings have been published in this respect. For instance, Hancock et al. (2004; 2005) do not find any correlation between negative emotion and deception, and this is also the case of Masip et al. (2012) in the only study conducted on a Spanish corpus. The latter authors suggest that the reason may lie with the topic, since the accounts provided by the participants were about a trip, which is neither a controversial topic nor likely to cause uneasiness.

2.2.3.2.2. Cognitive complexity

The increased mental loading experienced by deceivers has been hypothesised by several researchers. Specifically, Vrij et al. (2008) conducted a major study on this issue, advancing six aspects of lying which may have an influence on increased cognitive complexity. First, the formulation of the lie itself is cognitively demanding. Second, deceivers are more prone to control their behaviour so as to pretend to be honest –previously suggested by DePaulo and Kirkendol (1989). Third, apart from monitoring their own demeanour, deceivers may attempt to control their interviewer's behaviour, analysing the feedback obtained so as to assess the success of their deception; this aspect is also highlighted by Buller and Burgoon (1996) concerning their IDT. Fourth, they face the complexity arising from constantly remembering the role-playing in which they are involved as deceivers (DePaulo et al., 2003). Furthermore, the added complexity of the suppression of the truth and the deliberate activation of the lie (Gilbert, 1991), which constitute the fifth and the sixth aspects, must be borne in mind. It is worth noting that the second and the third aspects are specific to conversational interactions.

The role of cognitive processes as a cue is worth considering, since it is surrounded by considerable controversy. In his meta-analysis, Vrij (2010) finds that most researchers do not reveal significant differences as regards this cue. Such is the case of Alonso-Quecuty (1996), Bond and Lee (2005), and Sporer (1997), to name but a few. On the other hand, other researchers find an increased presence of cognitive processes in deceptive statements (Hartwig et al., 2006; Vrij et al., 2008), while the opposite is observed by Granhag et al. (2001), Memon et al. (2010) and Vrij et al. (2000).

Due to the breadth of this category, it is essential that researchers delimit and operationalise it before performing the count of the elements (Sporer, 2004). Most commonly, this cue has been explored as defined in the RM approach, in which a positive correlation between cognitive processes and truth-telling is expected. The main difference between this approach and others like the one taken by Vrij's team is that the former only rate operations made at the time of the participant's recalling, whereas the latter takes into account the ones made at the time of the event too. Furthermore, the automated coding of the elements constituting this cue normally involves a broader definition of cognitive operations. As will be explained below, this automated coding has most often been performed by means of the software Linguistic Inquiry and Word Count (LIWC, henceforth), developed by Pennebaker et al. (2001). Vrij (2010: 273) explains it as follows:

> For example, the LIWC cognitive mechanism category includes words such as think. Thus, the sentence "I think she had dark hair" would produce a hit in the LIWC cognitive mechanism category. By comparison, human RM coders do not count this as a cognitive operation.

There are a series of cognitive operations which have been widely explored in deception detection. Such is the case of insight, the process exemplified by Vrij above, and tentative words, e.g. *guess*, *maybe, perhaps*. The vagueness conveyed by the latter category has been associated with untruthfulness in Adams and Jarvis (2006), Bond and Lee (2005), and Newman et al. (2003). Furthermore, distinction markers have proved the lexical correlates of cognitive complexity (Tausczik and Pennebaker, 2010), hence their strong presence in truthful accounts in comparison with untruthful ones. Specifically, exclusive words (e.g. *except*, *without*, *but*) are most commonly used during truthful communication (Fuller et al., 2008; 2011; Mihalcea and Strapparava, 2009; Newman et al., 2003), owing to the cognitive complexity involved in them. The difficulty in managing exclusive elements experienced by deceivers crops up from the possibility of contradiction, since their successful use requires a higher level of precision. Similarly, inclusive words (e.g. *with, and, include*) have also proved significant as a cue to truthful language on the same grounds (Granhag et al. (2001); Höfer et al. (1996); Memon et al. (2010); Vrij et al. (2000). Thus, it is generally agreed that liars tend to use fewer distinction markers that delimit what is in their story and what is not (Newman et al., 2003). Despite this fact, Hancock et al. (2008) have not found any effect of this category of words. Instead, they have observed that liars used fewer causation words (e.g. *because, effect, hence*), which also entails a higher degree of specificity, hence their common avoidance by deceivers.

These difficulties commonly experienced during deceptive communication are closely related to working memory capacity. According to Suengas and Johnson (1988), invented events are hard to keep in memory after 24 hours, in contrast to real events[9]. Accordingly, some deception researchers have used this information, creating situations that strain deceivers' cognition. Such is the case of police interviewers requiring suspects to relate events in reverse order, which increases their cognitive load and makes deception cues more evident. One of the above-mentioned studies in which cognitive processes proved more frequent in deceivers tests this technique: Vrij et al. (2008). Interestingly enough, Vrij's team had previously found this category to be a cue to truthfulness (Vrij et al., 2000), which is indicative of the increased complexity involved in the task undertaken by deceivers and of its magnifying effect. In fact, Hartwig et al. (2006) observe a higher presence of contradiction in the deceivers who verbalised more cognitive operations.

2.2.3.3. Further cues

In addition to the main verbal cues explained above, there are some others which, despite not having been so consistently explored, are also worthy of attention. They have been classified into content-independent and content-dependent ones. The former group comprises those cues not dependent on speech content, which contain one or more elements which are easily quantifiable, even by means of computerised coding. On the contrary, the latter cues are more dependent on speech content and are more appropriately identified by a trained expert.

2.2.3.3.1. Content-independent cues

Firstly, the study of negation as a cue to deception is worth delving into. Negative statements are defined by Vrij as "statements indicating aversion towards an object, person or opinion, such as denials, disparaging statements and

[9] The following quotation by Mark Twain best sums up this idea: "If you tell the truth, you don't have to remember anything."

statements indicating a negative mood" (2010: 101). This definition certainly reflects the traditional approach to the study of negation and denials, since these elements have been most often included in the category of negative emotion. However, some other scholars have considered them to be independent categories, following the psycholinguistic reasoning of classifying the denials as a linguistic process and disparaging statements and indications of negative mood as a psychological process (Newman et al., 2003; Pennebaker and King, 1999).

The use of linguistic denials and negative sentences (e.g. *no, not, never*) in deceptive communication has received certain attention in previous literature. As stated by Picornell, "the easiest way to lie is to deny something" (2012: 28). In this respect, findings in Hancock et al. (2005) suggest that the elaboration of the lie is directly proportional to the level of motivation involved, the less complex form of deception, simple negations, being used by senders in the low motivation condition. In addition to this study, Adams and Jarvis, Newman et al. (2003), and Toma and Hancock (2010) also identified denials and negative sentences as a cue to deception.

A further type of negation worth exploring is equivocal negation, defined by Picornell as "a form of evasive negation that creates ambiguity without commitment to any actual information" (2012: 28). This kind of negation seems to be especially relevant in law enforcement situations, tackled in section 2.1.2. Owing to the privilege against self-incrimination and the right to silence, suspects may resort to statements containing this mechanism, such as *I don't know* or *I don't remember*. As commented on above, in certain jurisdictions the use of this kind of negation or silence may look suspicious to the jury's eye, but it is not thought as improper as a more elaborated lie.

It has frequently been hypothesised that deceivers would resort to generalising terms, since they raise the possibility of a speech lacking in detail behind which the speaker can be hidden. Generalising terms provide the recipient with general references difficult to verify. This seems to be especially true in high-stakes situations, where liars may experience a lack of embracement deriving from their lack of conviction when making their statements. This idea is closely related to the arousal of anxiety in law enforcement contexts commented on above. However, as has been shown, fear and anxiety are not only felt as a consequence of lying, but also due to the unpleasant situation addressed by the interviewee. Thus, Vrij advances that "liars may also lack conviction because they have not personally experienced the claims they make. This lack of personal experience may result in using more generalising terms" (2010: 105). This plausible hypothesis has been examined by several researchers, although only some have confirmed it (Cody et al., 1989; Knapp et al., 1974). It is worth noting that not all scholars agree on the elements to be considered generalising terms. Most commonly, categories such as indefinite pronouns have been comprised within this group, including negative (e.g. *nobody, neither*), universal (e.g. *everyone, everything*), assertive existential (e.g. *someone, somebody*), and elective existential pronouns (e.g. *anybody, anything*). Similarly, time adverbs with a generalising effect like *never* and *always* have also been embraced by this category. Nevertheless, certain scholars have adopted alternative approaches to this category; such is the case of Picornell (2012), who makes a selection of universal and assertive indefinite pronouns and determiners, redistributing other terms to further categories, like negative indefinite pronouns. In other cases, researchers in the field try a broader approach: for instance, Zhou et al. consider generalising terms to be any word referring to "a person (or object) as a class of persons or objects that includes the person (or object)" (2004a: 94). Thus, words such as *men* or *animals* would be comprised within this category. As will be seen below, LIWC categories allow for this approach to be adopted, since a great deal of general categories is included in the default dictionary. Significantly enough, Mihalcea and Strapparava (2009) find the category humans, included in social processes, to be indicative of deception; these researchers explain it on the grounds that it represents references to others and hence detachment from the self, similar to the use of the 2nd and 3rd person.

A further content-independent cue is lexical diversity, understood as the total amount of types –different words in a text– divided by the total amount of tokens. This measure, known as type/token ratio, has been tested in deception research by several authors, yielding disparaging results. Most significantly, Colwell et al. (2002) and Dulaney (1982) find that deceivers display a higher degree of lexical diversity, whereas studies by Burgoon and Qin (2007), Knapp et al. (1974) and Zhou et al. (2004a) show quite the opposite. In Vrij's words, these contradictory findings are "not surprising because the attempted control approach could predict either an increase or a decrease of lexical diversity as a consequence of lying" (2010: 107). Nonetheless, the reason may lie in the fact that the results are not very reliable, since type/token ratio has proved to be too size-dependent as an index of lexical richness (Chipere et al., 2004).

Finally, motion words have proved indicative of deception in Newman et al. (2003). These researchers state that unpublished findings from their labs reveal a negative relationship between cognitive complexity and motion verbs (e.g. *walk, go, carry*), hence their hypothesis that if deceptive communication is less complex at the cognitive level, liars will produce more statements expressing motion. They explain it on the grounds that motion verbs are not cognitively demanding as compared to cognitive processes such as insight or inhibition.

2.2.3.3.2. Content-dependent cues

The cues contained in this section are directly related to speech content. They have been less systematically studied in recent literature, but they have still received certain attention. First, Burgoon et al. (1996) and Ebesu and Miller (1994) find immediacy, understood as the use of relevant and clear responses instead of evasive statements, to be negatively correlated to deception, which is in line with the equally negative correlation found for the 1st person singular. Furthermore, plausible answers, that is to say, accounts that make sense and appear reasonable, are also expected to appear more frequently in truthful communication, as shown in Kraut (1978), Riggio and Friedman (1983), and Stiff and Miller (1986).

There are two further cues related to speech content which could be deemed a priori indicative of the truth value of the statement: contradictions –within a statement or between two or more statements– and consistency. Nonetheless, no empirical evidence has been obtained for or against this hypothesis as regards the former cue (Granhag et al., 2003). Concerning consistency, which is understood as “the number of details that are repeated in two different statements about the same subject” (Vrij, 2010: 107), no link with deception has been found when it is judged intra-individual. However, when it comes to the assessment of the cue in pair interactions, certain evidence emerges that deceivers show more consistency than truth-tellers. This is explained by Vrij as follows:

> Two liars may discuss amongst themselves what they are going to say and subsequently recall this agreed scripted story. Their stories are then likely to be similar. In contrast, two truth tellers both use their memory of the event as the source of their recall. Their memories may differ slightly, or the truth tellers may differ in which details of the event are worth recalling. Less consistent recall is then the result (2010: 107).

2.2.4. Linguistic Inquiry and Word Count: An automatic tool for deception detection in language

Linguistic analysis can be automated so as to detect certain cues to deception, which would otherwise become a time-consuming task. There is a substantial body of studies using automatic tools for deception detection, using computer programs such as General Architecture for Text Extraction (GATE), Agent99-Analyzer, and CohMetrix (Hauch et al., 2012). However, the most widely used software for this purpose is Linguistic Inquiry and Word Count (LIWC, henceforth). Accordingly, a comprehensive review of previous research is provided here.

Significantly enough, LIWC was first tested by Newman et al. (2003) on a corpus of college students' deceptive and truthful written and spoken language purposely produced. Furthermore, in this process of data collection the participants are not biased towards the concealment of the lies, which, according to Bull et al. (2006: 76), is highly frequent among *professional* liars. The cost of the lies being detected would not be high in this case, opposite to what happens in high-stakes situations.

One of the key issues in psycholinguistics is the reflection of the emotional and cognitive frames of humans on the oral and written language they produce. Early approaches to psycholinguistic concerns involved almost exclusively qualitative philosophical analyses. More modern research in this field provides empirical evidence on the relation between language and the state of mind of subjects, or even their mental health (Rosenberg and Tucker, 1978; Stiles, 1992). In this regard, further studies such as Pennebaker and Francis (1996) and Pennebaker et al. (1997) have dealt with the therapeutic effect of verbally expressing emotional experiences and memories. Linguistic Inquiry and Word Count (LIWC, henceforth) was developed precisely for providing an efficient method for studying these psycholinguistic concerns, and has been considerably improved since its first version (Francis and Pennebaker, 1993). An updated revision of the original application was presented in Pennebaker et al. (2001), namely LIWC2001. This software application provides an effective tool for studying the emotional, cognitive, and structural components contained in language on a word by word basis, working out the percentage of words which fall into the four broad dimensions: standard linguistic processes, psychological processes, relativity, and personal concerns. That is to say, the program maps each word against a dictionary containing a series of words and the psychologically meaningful categories to which each word is assigned. It is worth noting that, as Tausczik and Pennebaker (2010) put it, the selection of words attached to language categories in LIWC has been made after hundreds of studies on psychological behaviour.

2.2.4.1. LIWC in Spanish

After assessing the reliability of LIWC2001 in the context of the English language, Ramírez-Esparza et al. (2007) developed the Spanish version of the dictionary. Certain studies have been carried out in order to analyse the equivalence of this version of the program to the English one, by means of translation from English to Spanish, and by testing the equivalence between categories in the English LIWC and the corresponding categories in the Spanish version. The process of transferring the categories involved four main steps:

(1) Literal translation and direct placing in the equivalent categories in the target language. For instance, pronoun “I” was translated as “yo” and was kept in its corresponding categories: 1 (Pronoun), 2 (First person singular), and 4 (Total first person).

(2) Some words were not kept in the same categories as in the original, because with some words there are many connotations which do not translate across languages.

(3) Verbs were conjugated for tense (present, past and future) and number.

(4) Finally, it was confirmed that the categories originally assigned in English made sense in Spanish, including the appropriate modifications. In this respect, the translation of certain words required conformance to the linguistic norms of Spain and South and Central America. Thus, for instance, “take” was translated as “tomar”, “agarrar” and “coger”.

The Spanish dictionary comprises the same categories as the English dictionary. In order to check the equivalence among both dictionaries, Ramírez-

Esparza et al. (2007) analysed English texts with their corresponding translations into Spanish; this enabled the researchers to work with texts containing the same information in both languages. 83 English texts were analysed with their translations. Topics lay within the domains of politics, health, marketing, advertisements, songs, reflections, recipes, poems, and prayers, to name but a few. The authors found that, for most categories, the correlations among the Spanish and English dictionaries were strong and significant (Study 1), and that the same differentiations can be made with the Spanish and English dictionaries (Study 2). However, the authors highlight that some categories are not strictly comparable among languages, mainly due to grammar issues. First of all, they observe the differences in 1^{st} and 3^{rd} person, owing to the constant presence of the personal pronoun in verbal expressions in English and not in Spanish. Furthermore, the distribution of verbal tenses also shows significant differences in both languages, as well as in the categories friends and humans; the authors explain these divergences on the grounds that in Spanish there are more verbal conjugations and a larger amount of human-related terms marked for genre. Thus, they warn that these differences are to be considered in cross-cultural studies involving both languages. Nonetheless, this asymmetry does not pose major problems for the present study, since, as will be explained below, its aim is not to draw a direct comparison between both languages, but to contrast the potential differences between the sublanguages of deception and truth both in English and in Spanish.

2.2.4.2. LIWC2001 categories

LIWC2001 versions were used in both languages for the study, since, according to Ramírez-Esparza, they have been more properly validated than LIWC2007 (personal communication, September 23, 2011). The LIWC2001 internal dictionary comprises 2,300 words and word stems classified under the four broad dimensions mentioned above. In the tables below, classes in bold indicate categories comprising a group of subcategories; for instance, ‘sensory and perceptual processes’ consists of ‘seeing’, ‘hearing’ and ‘feeling’. Apart from this, underlined classes are subcategories containing further subcategories; such is the case of ‘positive emotions’, which is the sum of ‘positive feelings’ and ‘optimism and energy’.

Within the first dimension, namely standard linguistic processes, most categories involve function words and grammatical information; thus, the selection of words is straightforward, as in the case of articles, which are made up of three words in English –*a, an* and *the*– and of nine words in Spanish –*el, la, los, las, uno, un, una, unos* and *unas.* A comprehensive list of the standard linguistic categories is provided in Table 2.3.

Dimension	English		Spanish	
	Abbrev.	Examples	Abbrev.	Examples
I. STANDARD LINGUISTIC DIMENSIONS				
Word count	WC	-	CP	-
Words per sentence	WPS	-	PPO	-
Sentences ending with “?”	Qmarks	-	Signos?	-
Unique words (type/token ratio)	Unique	-	Únicas	-
% words captured, dictionary words	Dic	-	Dic	-
and words longer than 6 letters	Sixltr	-	Seisltr	-
Total pronouns	Pronoun	*I, our, they, you're*	Pronom	*Yo, nosotros, tu*
Total first person	Self	*I, we, me*	Unomismo	*Yo, nosotros, mío*
1st person singular	I	*I, my, me*	Yo	*Yo, mío*
1st person plural	We	*we, our, us*	Nosotros	*Nosotros, nuestro*
Total second person	You	*you, you'll*	Tú	*Tu, ustedes*
Total third person	Other	*she, their, them*	Otro	*Ella, él, ellos*
Negations	Negate	*no, never, not*	Negación	*No, nunca*
Assents	Assent	*yes, OK, mmhmm*	Afirma	*Sí, claro*
Articles	Article	*a, an, the*	Artículo	*El, la, los, las*
Prepositions	Preps	*on, to, from*	Prepo	*A, ante, bajo*
Numbers	Number	*one, thirty, million*	Número	*Uno, dos, tres*

Table 2.3 LIWC2001 standard linguistic dimensions

On the other hand, the second and fourth dimensions are more subjective, especially those denoting emotional processes within the second dimension. These categories indeed required human judges to make the lexical selection. For all subjective categories, an initial list of word candidates was compiled from dictionaries and thesauruses, being subsequently rated by groups of three judges working independently. Categories making up the second dimension, psychological processes, are shown in Table 2.4.

Dimension	English		Spanish	
	Abbrev.	Examples	Abbrev.	Examples
II. PSYCHOLOGICAL PROCESSES				
Affective or Emotional Processes	Affect	*happy, ugly, bitter*	Afectivo	*Feliz, feo, amargado*
Positive Emotions	Posemo	*happy, pretty, good*	Emopos	*Feliz, bonito, bueno*
Positive feelings	Posfeel	*happy, joy, love*	Sentpos	*Feliz, felicidad, amor*
Optimism and energy	Optim	*certainty, pride, win*	Optime	*Certeza, orgullo, ganar*
Negative Emotions	Negemo	*hate, worthless, enemy*	Emoneg	*Odio, enemigo, feo*
Anxiety or fear	Anx	*nervous, afraid, tense*	Ansiedad	*Nervioso, miedo, tenso*
Anger	Anger	*hate, kill, pissed*	Enojo	*Odiar, matar, enojo*
Sadness or depression	Sad	*grief, cry, sad*	Tristeza	*Luto, llorar, tristeza*
Cognitive Processes	Cogmech	*cause, know, ought*	Meccog	*Causa, saber, debería*
Causation	Cause	*because, effect, hence*	Causa	*Porque, efecto, por*

Insight	Insight	*think, know, consider*	Insight	*Pensar, saber, considerar*
Discrepancy	Discrep	*should, would, could*	Discrep	*Debería, podría*
Inhibition	Inhib	*block, constrain*	Inhib	*Bloquear, obligar, forzar*
Tentative	Tentat	*maybe, perhaps, guess*	Tentat	*Quizá, creo, supongo*
Certainty	Certain	*always, never*	Certeza	*Siempre, nunca*
Sensory and Perceptual Processes	Senses	*see, touch, listen*	Sentidos	*Ver, tocar, escuchar*
Seeing	See	*view, saw, look*	Ver	*Ver, vista, mirada*
Hearing	Hear	*heard, listen, sound*	Oir	*Oído, sonido, escuchar*
Feeling	Feel	*touch, hold, felt*	Sentir	*Tocar, sostener, sentir*
Social Processes	Social	*talk, us, friend*	Social	*Hablar, nosotros, amigos*
Communication	Comm	*talk, share, converse*	Comu	*Hablar, compartir, conversar*
Other references to people	Othref	*1st, 2nd, 3rd pers prns*	Refotro	*1era plural, 2nda, 3era persona*
Friends	Friends	*pal, buddy, coworker*	Amigos	*Compañero, amigo, colega*
Family	Family	*mom, brother, cousin*	Familia	*Mamá, hermano, primo*
Humans	Humans	*boy, woman, group*	Humanos	*Niño, mujer, grupo*

Table 2.4 LIWC2001 psychological processes

Similar to the first dimension, the third dimension, relativity (see Table 2.5), comprises a category concerning time which is quite clear-cut: past, present, and future tense verbs. Within the same dimension, this is also the case of the category space, in which spatial prepositions and adverbs have been included.

Dimension	English		Spanish	
	Abbrev.	Examples	Abbrev.	Examples
III. RELATIVITY				
Time	Time	*Hour, day, oclock*	Tiempo	*Hora, día, noche*
Past tense verb	Past	*walked, were, had*	Pasado	*Caminé, fue, tuve*
Present tense verb	Present	*Walk, is, be*	Presente	*Ando, es, tengo*
Future tense verb	Future	*will, might, shall*	Futuro	*Será, haría*
Space	Space	*around, over, up*	Space	*Alrededor, arriba, abajo*
Up	Up	*up, above, over*	Arriba	*Arriba, encima, alto*
Down	Down	*down, below, under*	Abajo	*Debajo, abajo*
Inclusive	Incl	*With, and, include*	Incl	*Con, y, incluyendo*
Exclusive	Excl	*but, except, without*	Excl	*Pero, sin, excepto*
Motion	Motion	*Walk, move, go*	Moción	*Andar, ir*

Table 2.5 LIWC2001 relativity categories

Finally, the fourth dimension involves word categories related to personal concerns intrinsic to the human condition (see Table 2.6). As explained below, this dimension has often been excluded in deception detection studies, on the basis that it is too content-dependent (Newman et al. 2003; Hancock et al., 2004, 2008).

Dimension	English		Spanish	
	Abbrev.	Examples	Abbrev.	Examples
IV. PERSONAL CONCERNS				
Occupation	Occup	*work, class, boss*	Ocupa	*Trabajar, clase, jefe*
School	School	*class, student, college*	Escuela	*Clase, estudiante, colegio*
Job or work	Job	*employ, boss, career*	Trabajo	*Empleado, jefe, carrera*
Achievement	Achieve	*try, goal, win*	Logro	*Intentar, ganar, objetivo*
Leisure activity	Leisure	*house, TV, music*	Leisure	*TV, música, películas*
Home	Home	*house, kitchen, lawn*	Casa	*Casa, cocina, refrigerador*
Sports	Sports	*football, game, play*	Deportes	*Fútbol, juego, jugar*
Television and movies	TV	*TV, sitcom, cinema*	TV	*Televisión, telenovela, programa*
Music	Music	*tunes, song, CD*	Música	*Música, canciones, CD*
Money and financial issues	Money	*cash, taxes, income*	Dinero	*Cambio, dinero, ganancia*
Metaphysical issues	Metaph	*God, heaven, coffin*	Metafi	*Dios, cielo, ataúd*
Religion	Relig	*God, church, rabbi*	Relig	*Dios, cielo, iglesia*
Death and dying	Death	*dead, burial, coffin*	Muerte	*Muerte, entierro, ataúd*

Physical states and functions	Physcal	*ache, breast, sleep*	Físico	*Dolor, pecho, dormir*
Body states, symptoms	Body	*ache, heart, cough*	Cuerpo	*Dolor, corazón, toser*
Sex and sexuality	Sexual	*lust, penis, suck*	Sexual	*Lujuria, pene, sexo*
Eating, drinking, dieting	Eating	*eat, swallow, taste*	Comer	*Comer, tragar, probar*
Sleeping, dreaming	Sleep	*Asleep, bed, dreams*	Dormir	*Dormitar, cama, sueño*
Grooming	Groom	*wash, bath, clean*	Asearse	*Lavar, baño, limpiar, ducha*

Table 2.6 LIWC2001 personal concerns

Furthermore, LIWC2001 includes an appendix comprising some experimental categories, namely swear words, nonfluencies, and fillers. While the first category is useful in the analysis of both written and oral texts, the other two are exclusive to oral language.

2.2.4.3. LIWC and deception detection

This software application has been used to study issues like personality (Mairesse et al., 2007), psychological adjustment (Alpers et al., 2005), social judgments (Leshed et al., 2007), tutoring dynamics (Cade et al., 2010), and mental health (Rude et al., 2004). The validation of the lexicon contained in its dictionary has been performed by means of comparing human ratings of a large number of written texts with the ratings obtained through their LIWC-based analyses. LIWC was first used by Pennebaker’s team for a number of studies on the language of deception, the results of which were published in Newman et al. (2003). For their purposes, they collected a corpus of true and false statements through five different studies. In the first three tests, the participants expressed their true opinions on abortion, as well as the opposite of their point of view. The first study dealt with oral language, hence the videotaping of the opinions, whereas in the second and the third ones the participants were respectively asked to type and handwrite their views. In the fourth study, the subjects orally expressed true and false feelings about friends, and the fifth one involved a mock crime in which the participants had to deny any responsibility for a fictional theft.

As explained above, certain theoretical perspectives such as the RM approach predict linguistic markers of deception (Masip et al., 2005; Sporer, 2004), although traditionally they have been applied manually. Some of these markers are contained in the software application Linguistic Inquiry and Word Count or LIWC (Pennebaker et al., 2001), and for this reason it has been used to automatically explore verbal deception cues (Ali and Levine, 2008; Bond and Lee, 2005; Fuller et al., 2006; Hancock et al., 2008; Hirschberg et al., 2005; Newman et al., 2003; Vrij et al., 2007).

Extensive research has been conducted into the linguistic nature of deception, addressing the question of whether deceptive statements are deviant enough to betray insincere speakers. Experimental findings on the distinctive features seem to conflict, hence the scepticism towards this line of research voiced by certain scholars such as Vartapetiance and Gillam (2012).

The texts were analysed using the 29 variables of LIWC selected by the authors. Of the 75 categories considered by the program, they excluded the categories reflecting essay content, any linguistic variables used at low rates, and those unique to one form of communication (spoken vs written language). Table

2.7 shows the list of variables used in their analysis grouped into the three broad dimensions.

Standard linguistic dimensions	**Psychological processes**	**Relativity**
1. Word Count	11. Affective or emotional processes	22. Space
2. % words captured by the dictionary	12. Positive emotions	23. Inclusive
3. % words longer than six letters	13. Negative emotions	24. Exclusive
4. Total pronouns	14. Cognitive processes	25. Motion verbs
5. First-person singular	15. Causation	26. Time
6. Total first person	16. Insight	27. Past tense verb
7. Total third person	17. Discrepancy	28. Present tense verb
8. Negations	18. Tentative	29. Future tense verb
9. Articles	19. Certainty	
10. Prepositions	20. Sensory and perceptual processes	
	21. Social processes	

Table 2.7 List of variables used in Newman et al. (2003)

The values for these 29 variables were standardised by converting the percentages to z scores so as to enable comparisons across studies with different subject matters and modes of communication. For predicting deception, a logistic regression was trained on four of the five subcorpora and testing on the fifth, which entailed a fivefold cross validation. The authors obtained a correct classification of liars and truth-tellers at a rate of 67% when the topic was constant and a rate of 61% overall. However, in two of the five studies, the performances were not better than chance. Finally, the variables that were significant predictors in at least two studies were used to simultaneously evaluate the five tests, namely self-reference words, other reference words, exclusive words, negative emotion words and motion words. The reason for the poor performance in some of the studies may lie in the mixing of modes of communication, since, as stated by Picornell (2011) and mentioned above, the verbal cues to deception in oral communication do not translate across into written deception and vice versa.

As of this study, LIWC has been used in the forensic field mainly for the investigation of deception in spoken language. There are some early studies along these lines which are concerned with the usefulness of this software application as compared to the Reality Monitoring technique (RM, henceforth). First, Bond and Lee (2005) applied LIWC to random samples from a corpus comprising lie and truth oral statements by sixty-four prisoners, only taking into consideration the variables selected by Newman et al. (2003) for the global evaluation. Overall, the results show that deceivers score significantly lower than truth-tellers as regards sensory details, but outstandingly higher in spatial aspects. The latter finding goes against previous research in RM theory; such is the case of Newman et al. (2003), where these categories did not produce considerable results. Apart from this difference, both studies share common ground: despite considering RM theory, the authors did not perform manual RM coding on their data. Thus, they do not draw a direct comparison between the effectiveness of automated RM coding through LIWC software and manual RM coding.

This gap in research was covered by Vrij et al. (2007). Their hypothesis predicts that LIWC coding is less successful than manual RM coding in discriminating between deceivers and truth-tellers. In order to test this theory, they collected a corpus of oral interviews of 120 undergraduate students. Half the participants were given the role of deceivers, having to lie about a staged event, whereas the remainder had to tell the truth about the action. The analysis revealed that RM distinguished between truth-tellers and deceivers better than Criteria-Based Content Analysis. In addition, manual RM coding offered more verbal cues to deception than automated coding of the RM criteria. There is a second experiment in this study assessing the effects of three police interview styles on the ability to detect deception, but the results will not be presented here because the subject lies outside the scope of this work.

More recently, Fornaciari and Poesio (2011) conducted a study on a corpus of transcriptions of oral court testimonies. This work presents two main novelties: first, the object of study is a sample of spontaneously produced language instead of statements uttered ad-hoc or laboratory-controlled; moreover, it deals with a language other than English, namely Italian. The authors resume Newman et al.'s (2003) idea of a method for classifying texts according to their truth value instead of simply studying the language in descriptive terms, the utterance being their analysis unit instead of the text. Their ultimate aim is a comparison between the efficiency of the content-related features of LIWC and surface-related features, including the frequency and use of function words or of certain n-grams of words or parts-of-speech. They used five kinds of vectors, taking the best features from their own experiment, from that of Newman et al. (2003), and all LIWC categories. The latter results in a slightly better performance than the former, but they do not obtain a statistically significant difference.

The design of this experiment has its drawbacks, on the one hand because, as stated by Coulthard (2004), forensic transcripts in court are not always verbatim, and they do not contain written deception but oral language filtered by a person who is not the author of the lies. On the other hand, the language of deception is taken as a whole, ignoring the particular features which may distinguish one speaker from the others, assuming that everybody lies similarly. Instead of comparing each individual sample of deceptive language to its corresponding control text, the whole set of statements labelled as "false" is contrasted with the set comprising "true" statements. It is worth noticing that the main disadvantage of a corpus of "authentic" language is precisely the difficulty to obtain a control sample of language in which the same speaker tells the truth for the sake of idiolectal comparison.

LIWC has also been used for the investigation of deception in written language. Curiously enough, research along these lines has been approached by computational linguists and not from the point of view of forensic science. First, Mihalcea and Strapparava (2009) used LIWC for post-hoc analysis, measuring several language dimensions on a corpus of 100 false and true opinions on three controversial topics –the design of the questionnaire is indeed similar to Newman et al.'s (2003). As a preliminary experiment, they used two ML classifiers: Naïve Bayes and Support Vector Machines, using word frequencies for the training of both algorithms. They achieved an average classification performance of 70%, which is significantly higher than the 50% baseline. On the basis of this information, they carry out the calculation of a dominance score associated with a given word class inside the collection of deceptive texts as a measure of saliency. They then compute word coverage, which is the weight of the linguistic item in the corpora. Thus, they identify some distinctive characteristics of deceptive texts, but merely in descriptive terms.

In this strand of research, Ott et al. (2011) use the same two ML classifiers. For their training, apart from comparing lexically-based deception classifiers with a random guess baseline, the authors additionally evaluate and compare two other computational approaches: genre identification through the frequency distribution of part-of-speech (POS) tags, and a text categorisation approach which allows them to model both content and context with n-gram features. Their ultimate aim is deceptive opinion spam, which is qualitatively different from deceptive language itself. Findings reveal that n-gram-based text categorisation is the best detection approach; however, a combination of LIWC features and n-gram features perform marginally better.

These studies deal with written language as used in an asynchronous means of communication. In contrast, Hancock and his team explore deceptive language in synchronous computer-mediated communication (CMC), in which all participants are online at the same time (Bishop, 2009). Specifically, they use chat rooms. In their first study using LIWC, Hancock et al. (2004) explore differences between the sender’s and the receiver’s linguistic style across truthful and deceptive communication. For the analysis, they select the variables deemed relevant to the hypotheses, namely word counts, pronouns, emotion words, sense terms, exclusive words, negations, and question frequency. Results show that, overall, when participants (n=66) told lies, they used more words, a larger amount of references to others, and more sense terms. Hancock et al. (2008) report rather similar results from a comparable experiment. In addition to this, they introduce the element of motivation, and observe that motivated liars tended to avoid causal terms, while unmotivated liars increased their use of negations.

As has been seen, the classification rates using some of the LIWC categories have been significantly greater than chance. However, this line of research has come under moderate criticism (Vartapetiance and Gillam, 2012; Vrij, 2010) on the grounds that some categories which have proved discriminant in some studies (e.g., Bond and Lee, 2005) have not done so in others (e.g., Vrij et al., 2007). In addition, Masip et al. (2012) highlight the opposing results found for some categories in different studies. For this reason, these authors question the usefulness of LIWC to detect deception. Nonetheless, it could be that they try a too general approach to deception detection, attempting to find the key to identifying lies in any modality and within any context. As commented on above, the present study attempts a model in which deception in written language is not explored as a whole, but taking into consideration the particular modality of the corpus and the discursive differences among topics.

CHAPTER 3

The Study: Research Questions and Method

3.1. Area of research

As mentioned in the previous chapter, the main area of research of the present PhD dissertation is the study of language by means of corpus-based and computational techniques. Nevertheless, like any interdisciplinary subject, the detection of deception in written language involves the combining of two or more academic fields in the pursuit of a common task, crossing traditional boundaries between linguistics, computation and psychology.

Specifically, the main aim of this research is to explore the linguistic cues to deception in written language both in English and Spanish, performing a contrastive analysis between both languages.

3.2. Research Questions

The following five questions need to be addressed:

(1) How successful are LIWC dimensions, the Bag-of-Words model, and the further stylometric dimensions in written deception classification in English and in Spanish?

(2) Which are the most consistent linguistic cues to written deception across the whole corpora?

(3) Which are the relevant predictors specific to English?

(4) Which are the relevant predictors specific to Spanish?

(5) Are there any linguistic cues to deception specific to certain topics?

3.3. Method

The method addresses the different stages in the development of the present study. This section comprises five main parts: an introduction to the nature of the study, the variables, the sample, the corpora description, and the data collection.

The first part describes the nature of the research carried out in the study. The second part introduces the basis for this study: the dependent and independent variables used for exploring the features of truthful and untruthful language. The third part of the method describes the sample of participants and the description of the corpora used. Finally, an account of the data collection process is provided.

3.3.1. Nature of the study

The current study involves a combination of primary and secondary research. As Brown and Rodgers (2002) put it, primary research deals with original data, whereas secondary research is based on bibliographical resources such as scientific books and papers. The previous chapter, where the theoretical basis for this study is formed, provides a sample of secondary research. A piece of primary research may be found in Chapters 3, 4 and 5. Specifically, the present chapter offers detailed information about the development of the study.

Furthermore, the study may be classified as quasi-experimental. Quasi-experiments resemble quantitative and qualitative experiments, but they lack random assignment of groups or proper controls (Shadish et al., 2002). This feature has been seen as an inherent weakness, especially from the viewpoint of experimental purists in the natural sciences. However, this is a very useful design for measuring social variables, since it is not always possible to accomplish a purely random allocation of groups when dealing with human subjects. Thus, the present research takes advantage of the possibilities of this experimental design, by comparing two groups of participants under similar circumstances. As explained below, an inter-group comparison has been drawn, delving into the similarities and differences of the linguistic profiling of deception in written language across languages. In addition, an intra-group assessment has been undertaken in order to explore differences across topics, using the truthful statements as the control subcorpus against which the untruthful data set is compared. Due to the quasi-experimental nature of the study, the intention is not to generalise the inferences drawn from the data analysis, but they are to be treated provisionally.

The nature of the linguistic data of the corpora is also worth commenting on. Much has been discussed about the importance of deception in language being naturally produced. Laboratory-produced lies have been criticised in forensic literature for not being very reliable:

> A major criticism of almost all published studies involving professionals is that the video clips shown to them have not been of people lying in real-life, high stakes situations (but usually of students lying for the purposes of the experiment) (Bull et al., 2006: 77).

Similarly, Miller and Stiff (1993) question the value of applying laboratory research results to field settings, and Sporer (1997) suggests that further research should involve retrospective studies in law enforcement settings, to study realistic responses with known outcomes. Nevertheless, it presents some practical difficulties. First of all, the veracity or deception of the narratives has to be already determined, and this is not always an easy task. Generally speaking, the accounts are not to be taken as truthful or untruthful blocks of information, but rather as sets of truthful and untruthful statements. Thus, their categorisation becomes a complex task. On the contrary, this problem does not exist in a laboratory-controlled set of data, where the corresponding accounts are untruthful as a whole.

One of the major criticisms of those studies not dealing with real-life situations is the participants’ absence of anxiety. This emotion is considered to arise from the guilt experienced by an offender making a false statement, as well as from the possible consequences of being caught lying. Obviously, in a laboratory-controlled situation this element is absent, or minimised at best. However, this may be considered to be not so negative, since, as Bull et al. (2006: 79) put it, one of the main beliefs falsely associated with deceivers is that “people who look nervous are liars (when they are probably just socially anxious or introverted)”. This anxiety is often provoked by police interviewers behaving in an accusatory way, which may be transmitted to interviewees, resulting in an increase in their nervousness, whether lying or telling the truth. In this way, the language of deception in this corpus is not spoilt by a pressing situation. Furthermore, professional and recidivist liars do not experience anxiety when deceiving, so it has been deemed interesting to neutralise this factor. In addition, it especially influences spontaneously produced oral language, where the suspects do not have enough time to plan and organise their speech; however, this is not the case with written statements, since the amount of time available is usually enough to carefully plan the language and thus to control the initial anxiety.

The strength of this kind of data is the possibility for controlling variables and attributes so that the conclusions drawn are valid. What remains constant during this experiment are the participants and the topics on which they write, which allows the researcher to avoid confounding variables and to focus on deception in opinions and memories as the only plausible causal factor. Put another way, providing that some variation is observed regarding the dependent variables analysed, this scientific control will allow the present author to assure that the participants' situations were identical until they were asked to lie, and so the potentially new outcome may be attributed to the independent variable. The usefulness of this kind of corpus has indeed been proved in the forensic context, as shown in Chapter 2 (e.g. Fitzpatrick and Bachenko, 2010). Apart from research on verbal cues to deception, some tools for detecting deception on a physiological basis rely on 'on-purpose lying'. Such is the case of the functioning of a polygraph, which is adjusted by means of some preliminary questions in which the subject is deliberately asked to lie (Arce and Fariña, 2006b). Specifically, the subject is usually asked to choose from any two numbers, for instance 1 and 6. He or she must then try to deny this selection when offered between 1 and 10. After the responses, the subject is informed that a change has been observed with respect to the other numbers, and is warned to be honest, otherwise their perjury will be detected (Bradley and Janisse, 1981). This application in a realistic setting shows the validity of on-purpose lying.

Last but not least, the differences between the situation of forensic linguistics in the United Kingdom and Spain are worth noting. As Bull et al. (2006) assure, there is an ever-growing respect between British police, criminal psychologists and linguists, probably because of the long tradition of these disciplines in their country. However, in Spain these areas are in their infancy, hence the difficulty when it comes to securing comprehensive assistance to conduct realistic lie detection studies in Spanish.

3.3.2. Variables

Variables are the operationalised way in which the attributes of objects are represented for further data processing (Babbie, 2009). Values of each variable are statistically distributed across the domain, which is the set of all possible values that a variable is allowed to have. The values must be defined for each variable, since domains can range from dichotomous or binary variables to multi-way variables, with a higher level of measurement.

The variables explored in the present study are classified into a dependent variable and a set of independent variables.

3.3.2.1. Dependent variable

The dependent variable is the main variable to be measured or observed. In the present study, it corresponds to a narrative being veracious or untruthful. The scoring of veracity or deception is derived from the coding of narratives determined to be truthful or untruthful by participants. The results of the research reflect the accuracy of the predicted veracity and deception of narratives when compared with actual veracity and deception.

3.3.2.2. Independent variables

The independent variable is the element that is believed to relate to, or influence, the dependent variable. In this case, a set of independent variables have been taken in view of the previous research on deception detection, commented on in

Chapter 2. Specifically, the categories within the four broad LIWC dimensions have been considered, in addition to some further stylometric variables comprised within the stylistic profiling of the corpus.

3.3.2.2.1. LIWC variables

First of all, most of the basic psychologically meaningful categories contained in LIWC (Pennebaker et al., 2001) and described in Chapter 2 have been used. Certain selected categories are not included in the software application by default, namely period, comma, colon, semicolon, exclamation, dash, quote, apostrophe, parenthesis, and other punctuation. It is worth noting that all the variables selected from LIWC reflect the percentage of total words, with three exceptions: raw word count, words per sentence, and percentage of interrogative sentences.

As commented on in Chapter 2, the LIWC dictionary generally arranges categories hierarchically. Thus, some of the categories are the sum of others. For example, the category “Total pronouns” comprises “1st person singular”, “1st person plural”, “Total 1st person”, “Total 2nd person”, and “Total 3rd person”. The categories “1st person singular” and “1st person plural”, in turn, are both subsumed under “Total 1st person”. Some previous studies such as Newman et al. (2003) and Fornaciari and Poesio (2011) explore categories from different levels in the hierarchy using the same experiment, which can be considered a methodological flaw. In ML classification and statistical techniques, this would result in redundancy, which may yield misleading results. As suggested by Picornell (2012), in this case results might be skewed by counting those variables twice. In order to avoid this, there are two options: remove the hierarchically superior categories, or keep them and leave the inferior categories out. In this case, the first option has been selected so as to keep the most specific information. Table 3.1 shows the LIWC categories removed and their correspondences. The first column contains the highest categories, the second one the subcategories, and the third one the subcategories of the previous subcategories –it is worth noticing that the categories which involve no complexity are not contained in this table. Categories in red are the most general ones, which have been altogether removed. These categories may comprise either categories in purple, which in turn comprise other lower categories, or just green ones, which are the terminal part of the sequence. Only the latter have been kept.

I. Linguistic dimensions		
Total pronouns	Total 1st person	1st person singular
		1st person plural
	Total 2nd person	-
	Total 3rd person	-
II. Psychological processes		
Affective or emotional processes	Positive emotions	Positive feelings
		Optimism and energy
	Negative emotions	Anxiety or fear
		Anger
		Sadness or depression
Cognitive processes	Causation	-
	Insight	-
	Discrepancy	-
	Inhibition	-
	Tentative	-
	Certainty	-
Sensory and perceptual processes	Seeing	-
	Hearing	-
	Feeling	-
Social processes	Communication	-
	Other references to people	1st person plural
		Total 2nd person
		Total 3rd person
	Friends	-
	Family	-
	Humans	-
III. Relativity		
Time	Past tense verb	-
	Present tense verb	-
	Future tense verb	-
Space	Up	-
	Down	-
	Inclusive	-
	Exclusive	-
IV. Personal concerns		
Occupation	School	-
	Job or work	-
	Achievement	-
Leisure activity	Home	-
	Sports	-
	Television and movies	-
	Music	-
	Money and financial issues	-
Metaphysical issues	Religion	-
	Death and dying	-
Physical states and functions	Body states, symptoms	-
	Sex and sexuality	-
	Eating, drinking, dieting	-
	Sleeping, dreaming	-
	Grooming	-
	Swearing	-

Table 3.1 Selection of LIWC categories for the experiment

3.3.2.2.2. Further stylometric variables

There are some linguistic features not included in LIWC standard linguistic dimensions, which has been deemed relevant for the present study. To the best of my knowledge, they had not been explored for deception detection. As regards their computation, they have been obtained by means of the statistics worked out by Wordsmith Tools 5.0[10].

The first of these variables is standardised type/token ratio; as commented on above, the non-standardised version of this ratio was included in LIWC standard linguistic dimensions, but it has been proved to be too size-dependent as an index of lexical richness (Chipere et al., 2004). Thus, the discriminatory power of the original version of the ratio may be greater due to the disparities among the values for the different texts. However, it is not as reliable a measure as the standardised version. On the other hand, word length has also been considered. Despite the fact that a category similar to “complex words” was already included in LIWC, namely “Sixltr”, all words longer than 6 letters were comprised. Since the general agreement in corpus linguistics is that complex words should include any word consisting of 8 or more letters, their frequency has been used for the calculation of one of the independent variables: the ratio of complex words to the number of tokens. Similarly, the ratios of the total amount of 1, 2, 3, 4, 5, 6, and 7-letter words to the number of tokens have been worked out. In addition to this, average word length (in characters) and average text length (in sentences) have also been considered in this section. A summary of all the variables is provided in Table 3.2.

[10] Commercially available at www.lexically.net/wordsmith/

Variables	Type	Scale	Class
Likelihood of veracity or deception	Dependent	Nominal (binary)	-
Word count	Independent	Ratio	LIWC
Words per sentence	Independent	Ratio	LIWC
Words longer than 6 letters	Independent	Ratio	LIWC
Period	Independent	Ratio	LIWC
Comma	Independent	Ratio	LIWC
Colon	Independent	Ratio	LIWC
Semicolon	Independent	Ratio	LIWC
Sentences ending with “?”	Independent	Ratio	LIWC
Exclamation	Independent	Ratio	LIWC
Dash	Independent	Ratio	LIWC
Quote	Independent	Ratio	LIWC
Apostrophe	Independent	Ratio	LIWC
Parenthesis	Independent	Ratio	LIWC
Other punctuation	Independent	Ratio	LIWC
1st person singular	Independent	Ratio	LIWC
1st person plural	Independent	Ratio	LIWC
2nd person	Independent	Ratio	LIWC
3rd person	Independent	Ratio	LIWC
Negations	Independent	Ratio	LIWC
Assents	Independent	Ratio	LIWC
Articles	Independent	Ratio	LIWC
Prepositions	Independent	Ratio	LIWC
Numbers	Independent	Ratio	LIWC
Positive feelings	Independent	Ratio	LIWC
Optimism and energy	Independent	Ratio	LIWC
Anxiety or fear	Independent	Ratio	LIWC
Anger	Independent	Ratio	LIWC
Sadness or depression	Independent	Ratio	LIWC

Causation	Independent	Ratio	LIWC
Insight	Independent	Ratio	LIWC
Discrepancy	Independent	Ratio	LIWC
Inhibition	Independent	Ratio	LIWC
Tentative	Independent	Ratio	LIWC
Certainty	Independent	Ratio	LIWC
Seeing	Independent	Ratio	LIWC
Hearing	Independent	Ratio	LIWC
Feeling	Independent	Ratio	LIWC
Communication	Independent	Ratio	LIWC
Friends	Independent	Ratio	LIWC
Family	Independent	Ratio	LIWC
Humans	Independent	Ratio	LIWC
Past tense verb	Independent	Ratio	LIWC
Present tense verb	Independent	Ratio	LIWC
Future tense verb	Independent	Ratio	LIWC
Up	Independent	Ratio	LIWC
Down	Independent	Ratio	LIWC
Inclusive	Independent	Ratio	LIWC
Exclusive	Independent	Ratio	LIWC
Motion	Independent	Ratio	LIWC
School	Independent	Ratio	LIWC
Job or work	Independent	Ratio	LIWC
Achievement	Independent	Ratio	LIWC
Home	Independent	Ratio	LIWC
Sports	Independent	Ratio	LIWC
Television and movies	Independent	Ratio	LIWC
Music	Independent	Ratio	LIWC
Money and financial issues	Independent	Ratio	LIWC
Religion	Independent	Ratio	LIWC
Death and dying	Independent	Ratio	LIWC
Body states, symptoms	Independent	Ratio	LIWC

Sex and sexuality	Independent	Ratio	LIWC
Eating, drinking, dieting	Independent	Ratio	LIWC
Sleeping, dreaming	Independent	Ratio	LIWC
Grooming	Independent	Ratio	LIWC
Swearing	Independent	Ratio	LIWC
Standardised type/token ratio	Independent	Ratio	Styl.
Mean word length	Independent	Ratio	Styl.
Sentences/WC	Independent	Ratio	Styl.
1-letter words/WC	Independent	Ratio	Styl.
2-letter words/WC	Independent	Ratio	Styl.
3-letter words/WC	Independent	Ratio	Styl.
4-letter words/WC	Independent	Ratio	Styl.
5-letter words/WC	Independent	Ratio	Styl.
6-letter words/WC	Independent	Ratio	Styl.
7-letter words/WC	Independent	Ratio	Styl.
Complex words/WC	Independent	Ratio	Styl.

Table 3.2 Variables in the experiment

3.3.3. Sample

The corpus used for English was collected by Mihalcea and Strapparava (2009)[11]. Their sample comprised 100 participants whose contributions were gathered through Amazon Mechanical Turk (MTurk)[12], one of the suites of Amazon Web Services. It is a crowdsourcing Internet marketplace that allows computer programmers and researchers in general to coordinate the use of human intelligence to perform tasks that computers are unable to perform. The requesters can post Human Intelligence Tasks (HITs) to be fulfilled by workers.

[11] The researchers readily accepted to share their corpus for the purposes of the present study during the period I spent as a visiting scholar at Fondazione Bruno Kessler in Trento (Italy).
[12] Service available at https://www.mturk.com/mturk/welcome

Its reliability as a source of data has been assessed in previous research (Snow et al., 2008). All the participants were native speakers of English, but no information is provided concerning either their variety of English or their age or sex.

In order to offer a direct comparison with the existing set of data for English, the design of the Spanish corpus was similar, as commented on below, and the sample of participants provides its basis. Thus, the sample also comprised 100 participants and 600 contributions. All of them were university students, native speakers of Peninsular or European Spanish. Thus, the task was assigned as an exercise for extra credit, and sent back via e-mail. Personal information such as age and sex has not been taken into account, since it has been considered irrelevant to the present analysis.

It was deemed of utmost importance to avoid overfitting, which may occur when a sample size is too small in relation to the number of variables used, since this could lead to over-optimistic results. It is generally agreed that, for this kind of analysis, it is necessary that the number of cases be twice the number of variables, expressed as N=2k (Guilford, 1954; Kline, 1986). As noted in the previous section, in the present study a set of 76 independent variables is used; thus, in principle a minimum of 152 contributions would be required. In this case, every data set comprises at least 200 contributions –in the case of the subcorpora organised by topics. The two data sets for English and Spanish include 600 cases each, and the global corpus totals 1,200. As it stands, statistical overfitting should not be a problem in subsequent analyses.

3.3.4. Corpora description

The instruments used in the present study can be classified as follows: (1) English corpus; (2) Spanish corpus.

3.3.4.1. English corpus

As commented on above, to study the distinction between truthful and untruthful statements, a corpus with explicit labelling of the truth value associated with each statement was required. For the design of the questionnaire, the authors focused on three different topics: opinions on abortion, opinions on the death penalty, and feelings about one's best friend.

For the first two topics, the authors provided instructions that asked the contributors to imagine they were taking part in a debate, and had 10-15 minutes to express their opinion about the topic. First, they were asked to prepare a brief speech expressing their true opinion on the topic. Next, they were asked to prepare a second brief speech expressing the opposite of their opinion, thus lying about their true beliefs. Participants were told that the content of the messages needed to be unambiguously truthful or deceptive, and, in both cases, the guidelines asked for at least 4-5 sentences in as much detail as possible.

For the other topic, the contributors were first asked to think about their best friend, including facts and anecdotes considered relevant to their relationship. Thus, in this case, they were asked to tell the truth about how they felt. Next, they were asked to think about a person they could not stand, and describe it as if he or she were their best friend. In this second case, they had to lie about their feelings toward this person. As before, the guidelines asked for at least 4-5 detailed sentences.

They collected 100 true and 100 false statements for each topic, making a total of 600 contributions, with an average of 85 words per statement and a total of 51,204 words. Each verbal sample was entered into a separate text file; they made a manual verification of the contributions and misspellings were corrected. Figure 3.1 provides a sample of this corpus.

TRUTH	LIE
ABORTION	
I am against abortion. I feel that to get an abortion done is a crime equivalent to murder. You are killing an unborn child and taking away his chance of coming into this world. The innocent soul to whom God has given the chance to come in human form is being deprived of this golden opportunity. Spirituality says that it is only in human form that one can realize oneself to be a soul and know God.	I think abortion is one's own personal decision and nobody should interfere in that. People must be allowed to decide for themselves whether they want to have the child or not. There can be several reasons why the person would want to abort the child. So only the parents know better whether it's the right time for them or not, they will be able to support the child or not, etc, etc.
DEATH PENALTY	
I believe the death penalty should be abolished. As a country that was founded on Christian principles, this is one area of the law that has strayed far off the path. God should be the only giver of life and death. To end someone's life for any reason is sinful. It is just and right to punish those who commit crimes and keep them segregated from society for the safety of others.	The death penalty should stand as it is. It is a necessary part of our criminal justice system. There are many criminals who would recidivate if given the chance and therefore cannot be put back into society. However, for those who commit the most heinous of crimes, it is not feasible in terms of space or finances to keep them in prison for their entire lives. Our prisons would be filled to the max in no time.
BEST FRIEND	
My best friend is so warm and inviting. The first time I met her I felt like I had known her forever. I told her my life story, that's how comfortable I felt talking to her. She has the nicest smile and the funniest laugh! She has been there for me during good times and bad...she held my heart when it was broken when my son died and then again when I found out I had cancer.	This girl is sweet but doesn't want you to know it! She has a huge smile, I wish she would use it more. We hang out and have fun together. My life is better for having known her.

Figure 3.1 Sample of truthful and untruthful statements in English (Mihalcea and Strapparava, 2009)

3.3.4.2. Spanish corpus

The design of the questionnaire for the compilation of the corpus was similar to that used by Mihalcea and Strapparava (2009). It focused on three different topics: opinions on homosexual adoption, opinions on bullfighting, and feelings about one's best friend.

The first two topics (homosexual adoption and bullfighting) are controversial and sensitive subjects, which cause people to entertain a personal opinion on them. Specifically, the participants received instructions to imagine they were taking part in a debate, and had 10-15 minutes to express their opinion about the topic. First, they were asked to prepare a brief speech expressing their true opinion on the topic. Next, they were asked to prepare a second brief speech expressing the opposite of their opinion, thus lying about their true beliefs. In both cases, the guidelines asked for at least 4-5 sentences in as much detail as possible.

For the other topic, the contributors were asked to think about a good friend of theirs, including facts and anecdotes considered relevant to their relationship. This topic was selected so as to offer a counterpart to the previous topics, since they entailed less emotional involvement. Thus, in this case, they were asked to tell the truth about how they felt. Next, they were asked to think about a person they could not stand, and describe it as if he or she were their best friend, and the same with a bad teacher. In this second case, they had to lie about their feelings toward these people. As before, the guidelines asked for at least 4-5 detailed sentences. It is worth noting that time restrictions were not imposed for either language.

In line with the English corpus, 600 contributions were collected –100 true and 100 false statements for each topic–, with an average of 94 words per statement and a total of 56,882 words. A manual verification of the quality of the contributions was made, and each verbal sample was entered into a separate text file, misspellings being corrected. Figure 3.2 shows a sample of truthful and untruthful language for each of the three topics.

TRUTH	LIE
HOMOSEXUAL ADOPTION	
Para mí no está clara la repercusión que tendría sobre los niños el hecho de que las parejas homosexuales adopten. Sería necesario un estudio previo de las posibles consecuencias o secuelas psicológicas, o de la ausencia de ellas, en el mejor de los casos.	La familia es y ha sido siempre la formada por un hombre y una mujer. No debemos cambiar esto, pues es un claro síntoma de la degeneración de la sociedad. Hemos de defender las tradiciones que llevan funcionando bien durante miles de años.
BULLFIGHTING	
Es una salvajada. Regodearse en el sufrimiento de un animal, disfrutar viendo cómo realiza sus últimos movimientos, agotado y herido. ¿Cómo puede ser un arte esto? Sin duda hay muchas personas que están familiarizadas con las corridas de toros. Es para ellos una situación normal.	Los espectáculos relacionados con los toros son una tradición antiquísima y un arte. Es más, los toros de lidia se pasan la vida al aire libre y son bien mimados por sus criadores, disfrutando así de una vida muchísimo mejor que la que se les ofrece a los animales de granja.
BEST FRIEND	
Cuando conocí a José María pensé que era uno más, que incluso no nos podríamos llevar bien. Qué equivocación más grande, ¡y qué afortunada! Es hoy uno de mis mejores amigos, que me encontré de casualidad en una de mis muchas andanzas por el mundo.	Sergio es un chaval inteligente, que sabe lo que quiere. Es realmente una buena persona, con la que puedes contar para todo. Su principal cualidad es su simpatía y amabilidad con todos, no importa que no te conozca de nada, siempre te da una oportunidad.

Figure 3.2 Sample of truthful and untruthful statements in Spanish

3.3.5. Data collection

As mentioned above, 600 truthful and 600 untruthful statements for each topic were collected for the two languages involved. The data collection processes have been described in section 3.3.5.

It is worth noting that, although the language in the corpus object of study has not been produced in a realistic setting, medium motivation is involved in the present study, as suggested by Hancock et al. (2010) and Ott et al. (2011). During the collection of both data sets, the participants were told that they had to make sure they were able to convince the readers about the topics on which they were lying. In addition, incentives were offered to increase external validity in both cases. The incentive provided by Mihalcea and Strapparava was monetary, since, as commented on above, the participants were Amazon Mechanical Turk workers. Furthermore, the requesters had the option to reject the results submitted by the workers, which would reflect on their reputation, hence the strengthening of their motivation. As regards the data set in Spanish, the students participating in the project were awarded extra credit which might improve the final grade of the course.

Besides this, the effect of the observer's paradox (Labov, 1972) or the Hawthorne effect has been minimised by not explicitly telling the subjects the ultimate aim of the research, since they might modify the aspects of their behaviour experimentally measured simply in response to the fact that they know they are being observed and studied (McCarney et al., 2007).

3.4. Machine Learning techniques and statistical methodologies

The techniques used to examine the questions raised above are advanced in this section. These techniques and methodologies fall within two broad areas: machine learning (ML, henceforth) and statistics. The field of ML is concerned with the automated training and learning of machines. According to Rätsch (2004), learning in this context refers to inductive inferences, since it deals with examples that provide incomplete information about some statistical phenomena.

This learning process may be unsupervised or supervised. The former typically concerns the uncovering of hidden regularities or the detection of abnormalities in the data, whereas in the latter a label which is considered an answer to a question is associated with each example. In this sense, Rätsch holds that

> [i]f the label is discrete, then the task is called *classification problem* – otherwise, for real-valued labels we speak of a *regression problem*. Based on these examples (including the labels), one is particularly interested to *predict* the answer for other cases before they are explicitly observed. Hence, learning is not only a question of remembering but also of *generalization to unseen cases* (2004: 1).

Specifically, in the first experiment conducted here a classification problem is addressed, since there is a discrete label related to the truth condition of the statements, as will be explained in 3.4.2.

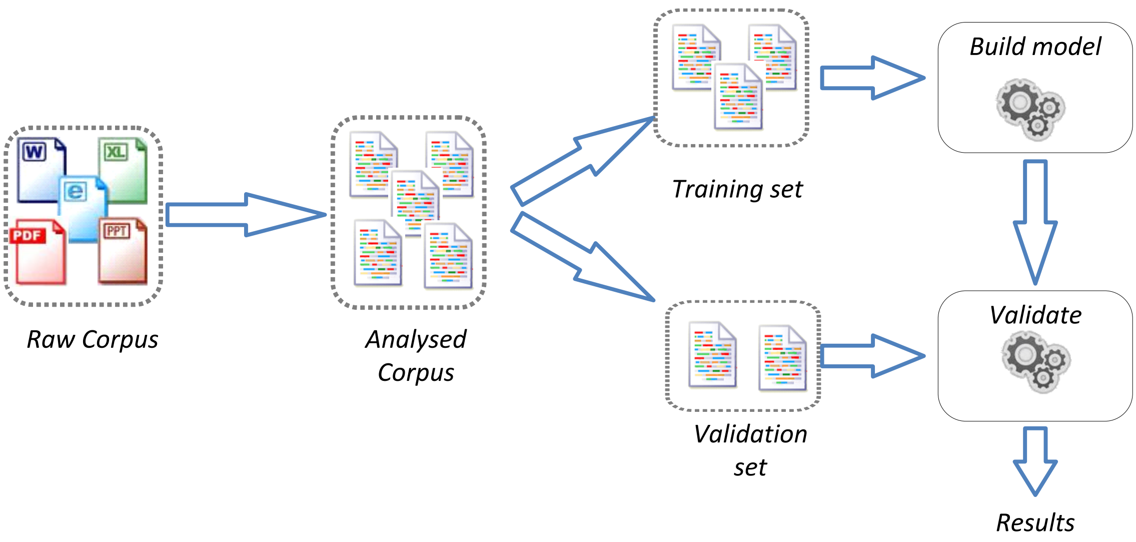

Figure 3.3 ML experiment

Broadly speaking, the experiments involved the analysis of both corpora with two software tools, as outlined in 3.4.1. With this text analysis, a set is built in order to train an ML algorithm which will be subsequently validated with the remaining subset; specifically, a ten-fold cross validation has been applied (Figure 3.3).

As regards the statistical methodologies, a discriminant function analysis and several logistic regressions have been performed so as to assess the discriminant power of the independent variables individually, instead of testing the dimensions as a whole.

3.4.1. Text analysis with LIWC and WordSmith Tools 5.0

As mentioned above, the LIWC[13] program (Linguistic Inquiry and Word Count) has been used to obtain the values for the vast majority of the categories for the subsequent training of the ML classifier and the identification of the predicting categories. Each of the 1200 text files was analysed using LIWC to create the samples. It is worth noting that the version used was LIWC2001, since this is the one which has been fully validated for Spanish. As explained in section 3.3.2.2.1., the whole LIWC output has not been used for the experiment, since the categories which comprise other subcategories have been left out. In addition, the two categories classified as experimental dimensions (Pennebaker et al., 2001), namely nonfluencies (e.g. *er, hm, umm*) and fillers (e.g. *blah, Imean, youknow*), have not been considered for analysis either, since they are exclusive to spoken language. The remaining experimental dimension, swear words, has been included for our purposes in the first dimension, linguistic processes, since this is the case for LIWC2007.

[13] Commercially available at www.liwc.net

On the other hand, WordSmith Tools 5.0 has been used for the analysis of the further stylometric features commented on in section 3.3.2.2.2.

3.4.2. ML experiments

For the ML classification experiments, a linear Support Vector Machine (SVM, henceforth) function was used to develop several classifiers of truth or deception on the input data (see Figure 3.4); it was selected based on its performance and diversity of learning methodologies. SVMs have been applied successfully in many text classification tasks due to their main advantages: first, they are robust in high dimensional spaces; second, any feature is relevant; third, they are robust when there is a sparse set of samples; and finally, most text categorisation problems are linearly separable (Rushdi-Saleh et al., 2011). Furthermore, the SVM approach has been previously used to map the solution to high dimensional space to find the greatest separation between liars and truth tellers among the predictor variables in the models (Elkins, 2011; Fornaciari and Poesio, 2011; Mihalcea and Strapparava, 2009). Unlike other classifiers such as decision trees or logistic regressions, SVM assumes no linearity and can be difficult to interpret outside of its accuracy values (Chen and Lin, 2006, Efron et al., 2004).

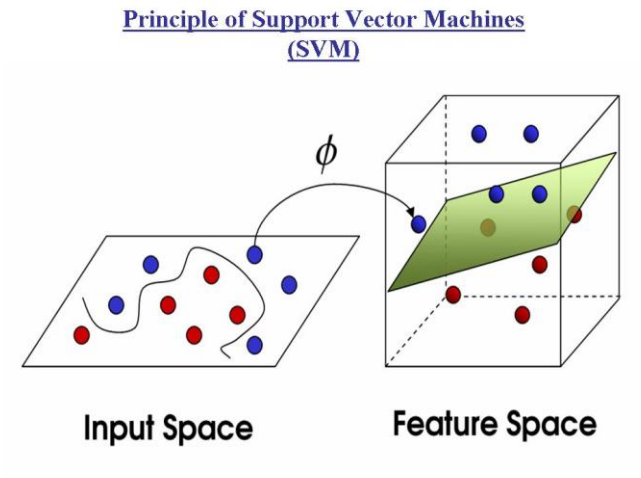

Figure 3.4 Visual representation of Support Vector Space[14]

Several scripts written in Perl were developed in a format appropriate for Weka[15] (Waikato Environment for Knowledge Analysis). It is a suite of ML software written in Java, developed at the University of Waikato (New Zealand), and is free software available under the GNU General Public License (Bouckaert et al., 2010). It was used extensively to evaluate the classification success of the various combinations of variables. The system is written in Java, and is thus portable across all major platforms. A linear SVM was applied using the default configuration set by the tool, which implies PUK, a universal Pearson VII function based kernel. It was proposed by Üstün et al. (2006) to solve SVM-based regression problems, as this kernel can be an alternative to the linear, polynomial and radial basis function kernels.

3.4.2.1. ML experiment with dimensions

Several classifiers were obtained by using all the possible combinations of the LIWC dimensions and the further stylometric features commented on in section

[14] Retrieved from http://www.imtech.res.in/raghava/rbpred/svm.jpg
[15] Freely available at www.cs.waikato.ac.nz/ml/weka/

3.3.2.2, in order to test the classifying potential of the dimensions, both individually and in combination. As shown in Table 3.3, a total of 31 classifiers were obtained, 4 of them involving the four LIWC dimensions individually, and 11 combinations of these dimensions. In addition, a classifier was obtained from the further stylometric features added, and 15 classifiers accounted for all the possible combinations of this group of variables and LIWC dimensions.

Classifier	Explanation
1	LIWC dimension 1
2	LIWC dimension 2
3	LIWC dimension 3
4	LIWC dimension 4
1_2	Combination of LIWC dimensions 1 and 2
1_3	Combination of LIWC dimensions 1 and 3
1_4	Combination of LIWC dimensions 1 and 4
2_3	Combination of LIWC dimensions 2 and 3
2_4	Combination of LIWC dimensions 2 and 4
3_4	Combination of LIWC dimensions 3 and 4
1_2_3	Combination of LIWC dimensions 1, 2 and 3
1_2_4	Combination of LIWC dimensions 1, 2 and 4
1_3_4	Combination of LIWC dimensions 1, 3 and 4
2_3_4	Combination of LIWC dimensions 2, 3 and 4
1_2_3_4	Combination of LIWC dimensions 1, 2, 3 and 4
Styl.	Further stylometric variables
1+ styl.	Combination of LIWC dimension 1 and further stylometric variables
2+ styl.	Combination of LIWC dimension 2 and further stylometric variables
3+ styl.	Combination of LIWC dimension 3 and further stylometric variables
4+ styl.	Combination of LIWC dimension 4 and further

	stylometric variables
1_2+ styl.	Combination of LIWC dimensions 1 and 2 and further stylometric variables
1_3+ styl.	Combination of LIWC dimensions 1 and 3 and further stylometric variables
1_4+ styl.	Combination of LIWC dimensions 1 and 4 and further stylometric variables
2_3+ styl.	Combination of LIWC dimensions 2 and 3 and further stylometric variables
2_4+ styl.	Combination of LIWC dimensions 2 and 4 and further stylometric variables
3_4+ styl.	Combination of LIWC dimensions 3 and 4 and further stylometric variables
1_2_3+ styl.	Combination of LIWC dimensions 1, 2 and 3 and further stylometric variables
1_2_4+ styl.	Combination of LIWC dimensions 1, 2 and 4 and further stylometric variables
1_3_4+ styl.	Combination of LIWC dimensions 1, 3 and 4 and further stylometric variables
2_3_4+ styl.	Combination of LIWC dimensions 2, 3 and 4 and further stylometric variables
1_2_3_4+ styl.	Combination of LIWC dimensions 1, 2, 3 and 4 and further stylometric variables

Table 3.3 Combinations of LIWC dimensions and the further stylometric features classifiers

For each classifier, a ten-fold cross-validation has been done, all sets having an equal distribution between truthful and untruthful statements. This technique is used to evaluate how the results of a statistical analysis would generalise to an independent data set. Since the aim of this experiment is the prediction of the truth condition of the texts, a cross-validation is applied in order to estimate the accuracy of the predictive models. It involves partitioning a sample of data into complementary subsets, performing an analysis on the training set and validating the analysis on the testing or validation set (Kohavi, 1995).

In addition to this, a Bag-of-Words (BoW) representation has been offered to provide a basis for comparison with the other classifiers employed.

3.4.2.2. Bag-of-Words (BoW) model

In this model, a text is represented as an unordered collection of words, disregarding any linguistic factor such as grammar, semantics or syntax (Lewis, 1998). It has been successfully applied to a wide variety of NLP tasks such as document classification (Joachims, 1998), spam filtering (Provost, 1999), and opinion mining (Dave et al., 2003). However, its basis is not too sophisticated, hence its simple implementation (see Figure 3.5).

In the experiment, the feature vector of each text is represented as a vector with the number of occurrences of the stemmed words in the text. In this context, a stem is the part of the word which never changes even when morphologically inflected –the difference from lemma is worth highlighting, since the latter term denotes the base form of the word (Mitkov, 2003). For this purpose, Snowball[16] has been used, a framework for writing stemming algorithms which has given rise to an improved English stemmer together with stemmers for several other languages like Spanish.

[16] Freely available at http://snowball.tartarus.org/

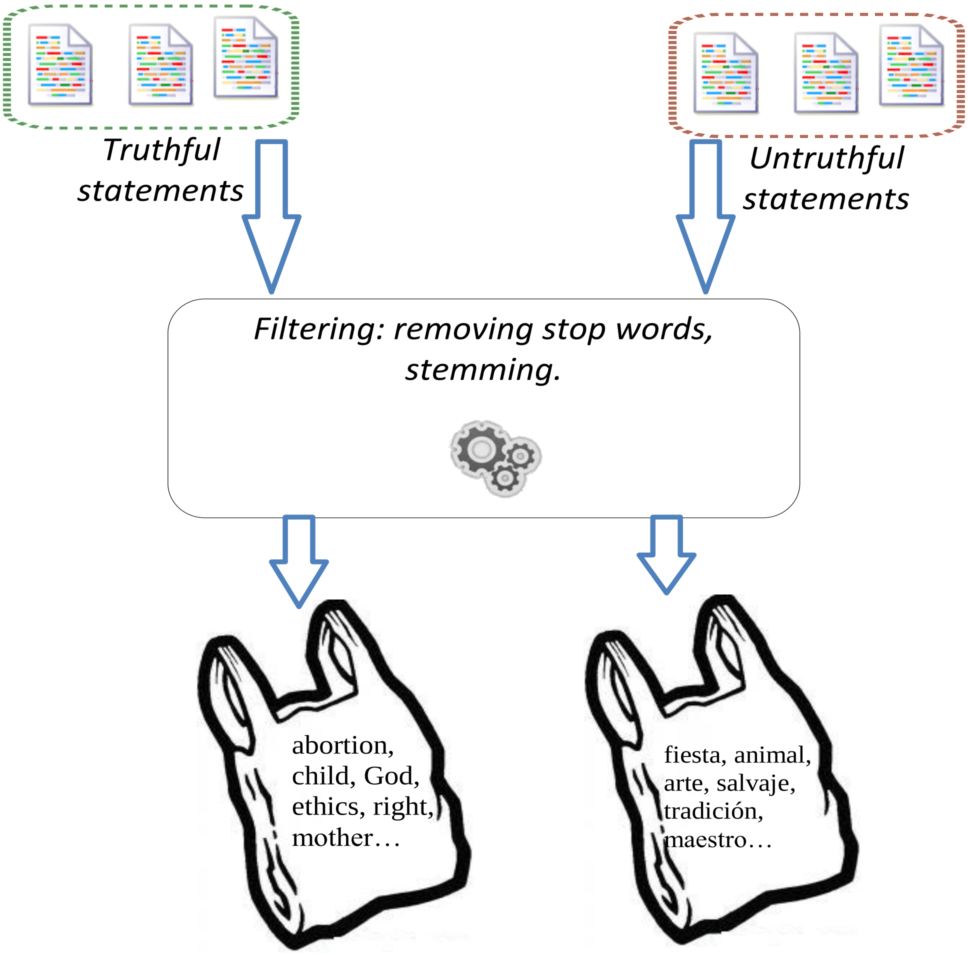

Figure 3.5 Bag-of-Words model

3.4.3. Statistical techniques

In order to gain a deeper understanding of the specific categories which best discriminate between both sublanguages, a test for predicting category membership on the basis of the independent variables proposed has been conducted. Two classification methods have been used, binary logistic regression and discriminant function analysis (DFA, hereafter), depending on how well the underlying data meet their statistical requirements. The latter makes more demanding requirements on the data, since it assumes that the dependent variable is categorical –which happens to be the case in the present study– and shares all the usual assumptions of correlation, requiring linear and homoscedastic relationships –homogeneity of variances– and normal distribution of the interval or continuous data. Nevertheless, when the condition of normal distribution is not met, the central limit theorem (CLT) applies providing that the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed (Rice, 1995). It also justifies the approximation of large-sample statistics to the normal distribution in controlled experiments. Thus, its application on the global corpora in English and in Spanish is justified, since they involve samples of 600 cases each for measuring the 76 independent variables. As far as the subcorpora are concerned, a one sample Kolmogorov-Smirnov test provided evidence against the null hypothesis, implying that the samples had not been drawn from a normal population. The distributions of the variables were found to be significant, thus a binary logistic regression was conducted instead.

As regards the calculations involved, DFA is broken down into a two-step process: first, a test is used to check whether the discriminant model as a whole is significant. In this case, Wilks' Lambda (λ) has been applied as a multi-variable measure of group means. Providing that the test reveals significance, the individual independent variables are then assessed to see which differ significantly in mean by group and these are used to classify the dependent variable. As a result, this test has enabled the evaluation of the categories at a global level. This statistical analysis predicts a categorical dependent variable, the grouping variable –in this case, likelihood of veracity or deception– by one or more continuous or binary independent variables, namely the predictor variables. DFA is useful in determining whether a set of variables is effective in predicting category membership, since it assigns individuals, for whom several variables have been measured, to certain groups already identified in the sample (Cantos, 2012). It answers the question of whether a combination of variables can be used to predict group membership. Thus, it determines the variables which discriminate between two or more naturally occurring groups, that is to say, truthful and untruthful accounts as regards the present study. The main difference with cluster analysis is that DFA is used for verifying that apparent clusters are real and for deciding to which cluster a new individual should be assigned.

From the set of independent variables used for separating cases, DFA creates new variables based on linear combinations which separate the groups as far apart as possible. The performance of the model is usually reported in terms of classification accuracy, with a percentage figure indicating how many cases would be correctly assigned to their groups using the new variables from DFA. The new variables can also be used to classify a new set of cases. It is worth noting that this analysis performs a previous significance test which checks whether the discriminant model as a whole is significant. If this is the case, then the individual independent variables are assessed to see which differ significantly in mean by group and these are used to classify the dependent variable.

Like DFA, logistic regression is a technique in which a set of predictors is used to determine group membership, but, as mentioned above, does not impose restrictive normality assumptions on the predictors. For this reason, it has been considered more appropriate for the analysis of the individual subcorpora, as only a few variables met the requirements of normality and involved just 200 cases. Two methods of binary logistic regression have been used for the sake of a double-check process: a backward stepwise method and a forward stepwise method. The former appears to be the preferred method in exploratory analyses (Hosmer and Lemeshow, 2000). It begins with a saturated model, and variables are eliminated from the model in an iterative process. The fit of the model is tested after the elimination of each variable to ensure that the model still adequately fits the data, and the analysis is completed when no more variables can be eliminated from the model. This type of regression calculates the probability of success over failure, and the results are in the form of an odds ratio. A Wald test has been used to assess the statistical significance of each coefficient (β) in the model, obtaining a Wald statistic with a chi-square distribution. This has enabled the preliminary selection of predictors. However, authors like Menard (1995) and Cohen et al. (2002) have identified problems with the use of the Wald statistic, mainly related to large or extremely low coefficients; they often tend to have associated inflated standard errors, which increases the probability of a Type-II error. For instance, Table 3.4 shows in green the variables with very high odd ratios and standard errors higher than 1, the results corresponding to the topic of good friend in English:

On the other hand, forward stepwise methods start with a model which does not include any of the predictors. Gradually, among the variables with a significance value lower than 0.05, the one with the largest score is selected and added to the model. In this case, a likelihood ratio method has been adopted, since, as Hosmer and Lemeshow (2000) explain it, the change in a -2 log-likelihood is generally more reliable than the Wald statistic. This guarantees that the variables chosen by both methods provide a good model; see Table 3.5 for the predictors finally included in the model after applying both methods. The easiest value to interpret is Exp(B), which represents the ratio change in the odds of the event of interest for a one-unit change in the predictor (Cohen et al., 2002).

	B	E.T.	Wald	d.f.	Sig.	Exp(B)	C.I. 95.0% for exp(B)	
							Lower	Upper
Sixltr	-.567	.166	11.657	1	.001	.567	.410	.785
Dash	-.977	.278	12.365	1	.000	.377	.219	.649
Parenth	-39.428	18551.472	.000	1	.998	.000	.000	.
I	-.245	.101	5.932	1	.015	.782	.642	.953
You	.680	.239	8.073	1	.004	1.974	1.235	3.155
Other	.852	.160	28.473	1	.000	2.345	1.715	3.208
Negate	.659	.212	9.635	1	.002	1.933	1.275	2.930
Preps	.214	.092	5.355	1	.021	1.238	1.033	1.484
Posfeel	.819	.273	9.005	1	.003	2.268	1.329	3.873
Optim	-.746	.251	8.797	1	.003	.474	.290	.777
Discrep	-.291	.145	4.036	1	.045	.747	.563	.993
Friends	-1.062	.281	14.241	1	.000	.346	.199	.600
Family	-1.616	.446	13.113	1	.000	.199	.083	.477
Humans	.733	.253	8.376	1	.004	2.082	1.267	3.421
School	.630	.202	9.675	1	.002	1.877	1.262	2.791
Achieve	.665	.210	9.994	1	.002	1.944	1.288	2.937
Money	1.214	.418	8.444	1	.004	3.366	1.485	7.633
Sexual	-1.250	.510	6.005	1	.014	.286	.105	.779
mean_word_length	13.213	3.217	16.868	1	.000	547552.1	999.872	3E+008
sentences	55.691	16.238	11.762	1	.001	2E+024	2E+010	1E+038
one_letterW	42.906	13.802	9.664	1	.002	4E+018	7681314	2E+030
two_letterW	23.391	8.236	8.067	1	.005	1E+010	1407.368	1E+017
seven_letterW	26.494	14.314	3.426	1	.064	3E+011	.210	5E+023

Table 3.4 Example of the variables in the equation using a backward stepwise method

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 9	Dash	-.284	.147	3.743	1	.053	.753
	I	-.152	.052	8.637	1	.003	.859
	Other	.333	.065	26.485	1	.000	1.395
	Friends	-.539	.139	15.130	1	.000	.583
	Family	-.843	.274	9.498	1	.002	.430
	Humans	.453	.165	7.560	1	.006	1.573
	Money	.492	.259	3.616	1	.057	1.636
	Sexual	-.644	.300	4.626	1	.031	.525

Table 3.5 Example of the variables in the equation using a forward stepwise method

In order to obtain more reliable classification results, a random sample of cases has been automatically generated for each topic to create a logistic regression model, setting the remaining contributions aside to validate the analysis. The reason for using a validation set is that classifications based upon the cases used to create the model tend to yield an inflated rate; thus, subset validation tends to be more reliable (Effron et al., 2004). It is worth noting that a Bernoulli distribution has been used to randomly generate the values of the variable *validate* with a probability parameter of 0.70, this variate taking values of 0 and 1. That is to say, approximately 70 percent of the truthful and untruthful statements will have a validate value of 1, being used to create the model, whereas the remaining statements will be used to validate the model results.

3.4.4. Summary

For the sake of clarity, Table 3.6 provides a summary of the classification techniques applied to the corpora:

Data	Aim	Type of method	Methodology	Input
English and Spanish corpora	Identification of predicting dimensions	ML experiment	SVM algorithm	LIWC dimensions
				Further stylometric features dimension
	BoW model	ML experiment	SVM algorithm	Stems frequency
	Identification of individual predictors	Inferential statistics	Binary logistic regression and DFA	LIWC categories
				Further stylometric features

Table 3.6 Methodologies and techniques used in the data analysis

3.5. Final remarks

The present chapter addresses two main areas, namely the research questions which have been raised and the method which is followed in order to conduct the study. The type of research on which the study is based responds to a quasi-experimental design.

The dependent variable has been identified as the likelihood of a narrative being veracious or untruthful. A set of 76 independent variables have been selected for their testing as potential discriminators in relation to the dependent variable, firstly using the grouping into LIWC dimensions and a new group comprising the further stylometric categories in the ML experiment, and then testing the variables individually by means of inferential statistics. The last part of the chapter centres on the description of the instruments, data collection, and the techniques used in the data analysis, which will be discussed in the next chapter.

CHAPTER 4

Data Analysis and Discussion

4.1. Introduction

The present chapter deals with the analysis of the results given by the methods that have been used in the study. These results are presented, analysed and evaluated in order to answer the research questions raised in the previous chapter. Finally, the main limitations of the analysis are exposed.

4.2. Results

This section addresses the results yielded by the application of the ML techniques, including both the experiment with whole dimensions and with the BoW model, and the results from the statistical classification techniques implemented.

4.2.1. ML experiment

Tables 4.1 and 4.2 show the results from the ML experiment conducted on LIWC and further stylometric dimensions (the whole output can be found in Appendix II.1). In the first column, the number of dimensions used for each classifier is indicated. For example, 1_2_3_4 indicates that all the dimensions have been used in the experiment, and 1_2 indicates that only the categories of dimensions 1 and 2 have been used to train the classifier. The scores provided stand for the F-measure, which is the weighted harmonic mean of precision and recall. The rest of the results can be found in Appendixes II.2 and II.3.

4.2.1.1. Dimension classifiers for English

The findings in Table 4.1 and Figure 4.1 reveal that the dimension which performs best overall is the first one, linguistic processes, the case of abortion excepted; this dimension proves especially successful with the good friend topic. The second one, psychological processes, shows a relatively high performance, except for the death penalty subcorpus, where the third dimension, relativity, is more successful. Interestingly enough, the dimension comprising further stylometric features is the next one in terms of discriminatory power. The fourth dimension, personal concerns, is the least discriminant on its own irrespective of the topic, whereas in the good friend subcorpus its performance is similar to the third dimension. Moreover, when the classifier is trained with certain combinations of dimensions, its performance improves noticeably. In this way, it seems clear that, in general terms, a combination of dimensions is more effective than in isolation, although the results from the classification with these dimensions are strongly dependent on the topics of each subcorpus.

	Abortion	Death penalty	Good friend	All
1	64.0	63.5	77.4	68.8
1_2	70.5	61.5	77.0	67.5
1_2_3	73.5	61.1	77.0	69.2
1_2_3_4	72.5	59.4	77.0	**69.8**
1_2_4	**74.0**	59.0	76.0	67.8
1_3	63.3	**64.4**	77.5	**69.8**
1_3_4	68.0	59.5	**78.5**	68.8
1_4	68.0	60.5	74.5	67.6
2	67.5	56.4	73.9	65.8
2_3	70.0	58.5	70.0	68.0
2_3_4	**74.0**	55.5	69.0	68.6
2_4	72.4	49.9	75.5	67.7
3	66.5	59.5	59.0	61.0
3_4	67.4	54.0	57.0	62.1
4	57.9	54.5	59.0	56.3
Styl.	66.4	60.0	64.5	62.4
1+styl.	67.9	61.9	75.9	65.8
1_2+styl.	71.0	63.0	**78.5**	69.0
1_2_3+styl.	72.0	**65.8**	76.0	**70.0**
1_2_3_4+styl.	**74.0**	64.3	76.5	**69.6**
1_2_4+styl.	72.5	59.9	77.0	69.4
1_3+styl.	68.9	64.0	78.5	**69.7**
1_3_4+styl.	71.5	62.5	77.5	69.1
1_4+styl.	67.0	61.4	75.5	67.1
2+styl.	73.5	60.0	73.0	68.2
2_3+styl.	73.0	63.5	71.5	68.7
2_3_4+styl.	73.4	56.8	71.5	69.3
2_4+styl.	73.0	57.5	71.5	67.1
3+styl.	65.5	62.6	**63.5**	64.8
3_4+styl.	65.5	56.4	**62.9**	64.9
4+styl.	64.5	55.5	**66.0**	62.6

Table 4.1 Results from the selected categories grouped into dimensions in English

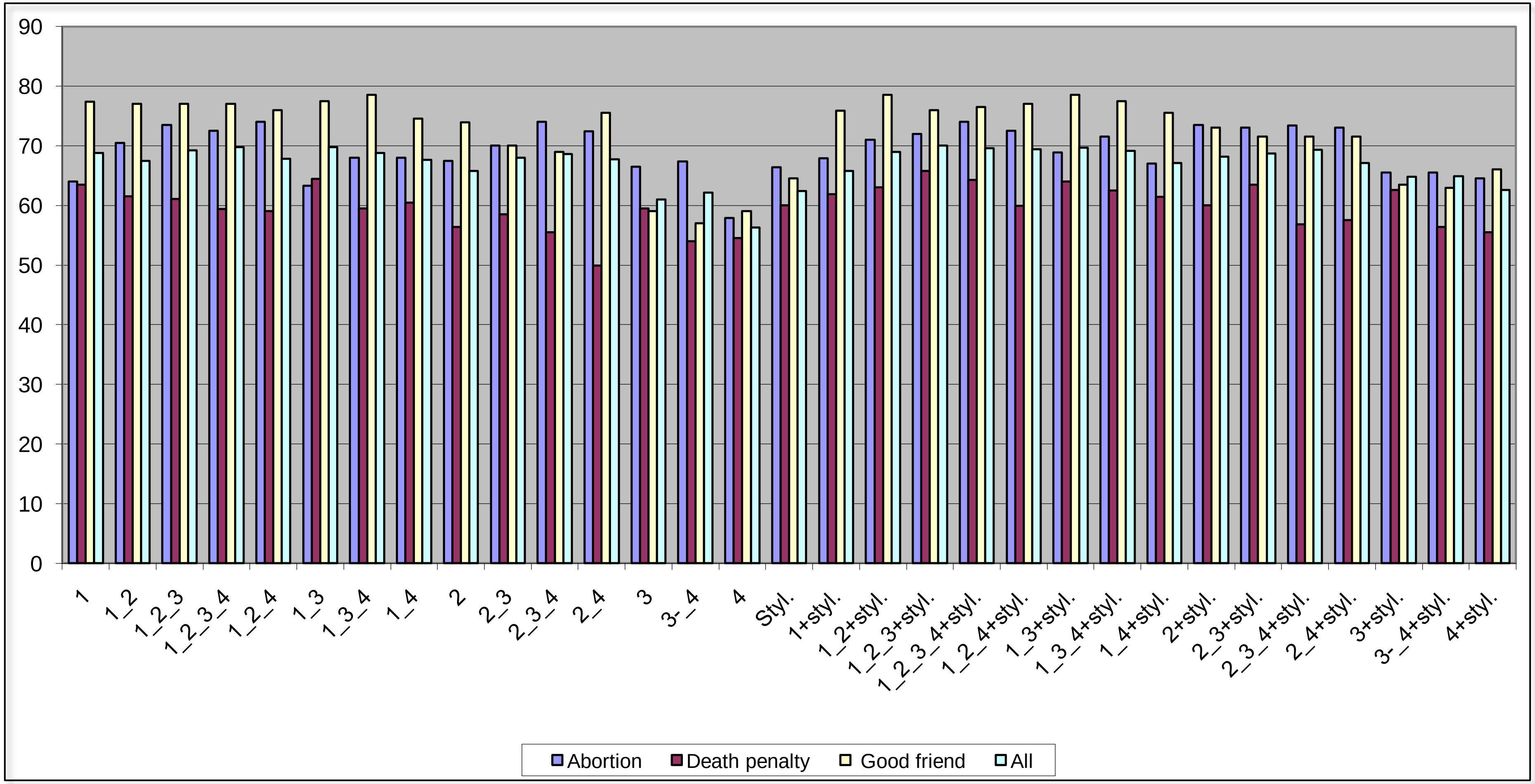

Figure 4.1 Results from the ML experiment for all the topics in English

As can be seen in Figure 4.2, the most remarkably successful LIWC combinations in the abortion subcorpus are the groupings 1_2_4 and 2_3_4. The combination of all LIWC dimensions and the further stylometric features is equally successful, which means that this classifier contains redundant information. Significantly enough, the addition of the stylometric dimension improves most combinations. On the other hand, the worst combination for this topic is 1_3.

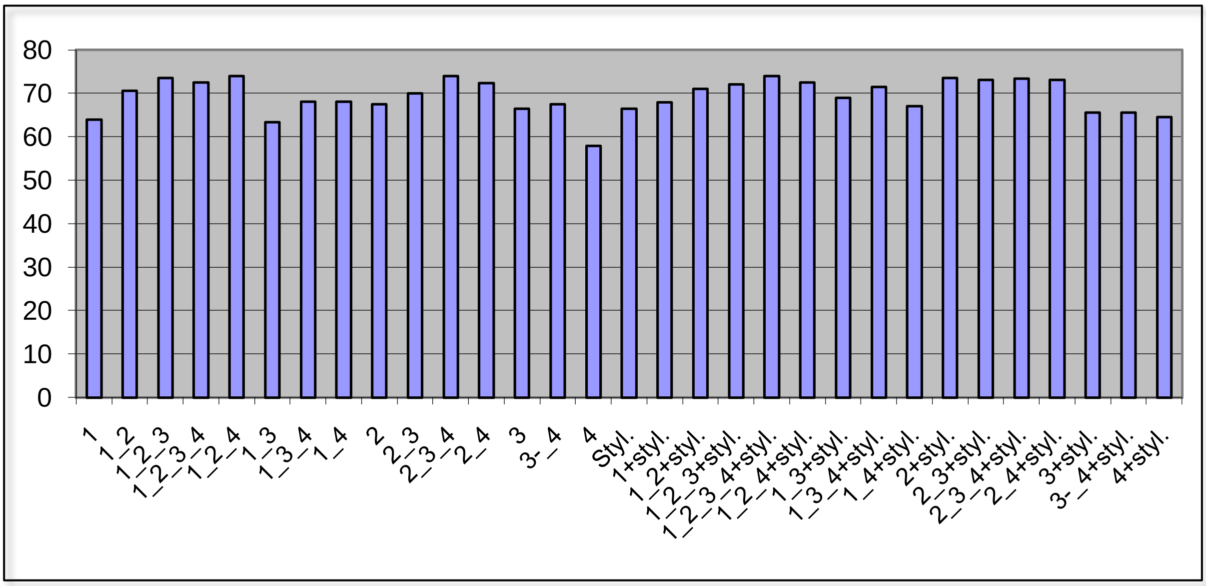

Figure 4.2 Results from the ML experiment for the abortion topic in English

The death penalty subcorpus is the only one in which a classifier is no better than chance: 2_4 (see Figure 4.3). As commented on above, these dimensions have a low performance on their own, hence the bad results from their combination. These are not the only poor classifiers within this subcorpus, since there are others whose rate falls below 60%, namely 1_2_3_4, 1_2_4, 1_3_4, 2_3, 2_3_4, 3, and 3_4. That is to say, virtually all the classifiers comprising dimension 3 and/or 4 are not a great deal better than chance. This is especially the case with the addition of the fourth dimension; it seems to be counterproductive in all the combinations, since it makes the scores worse instead of improving them, probably due to its production of noise. On the other hand, the best rates for this subcorpus are achieved by the classifiers 1_3 and 1_2_3 with further stylometric features. Moreover, this last dimension exerts a positive effect on the classifiers, since it improves all of them except for three, namely 1, 1_3, and 1_4.

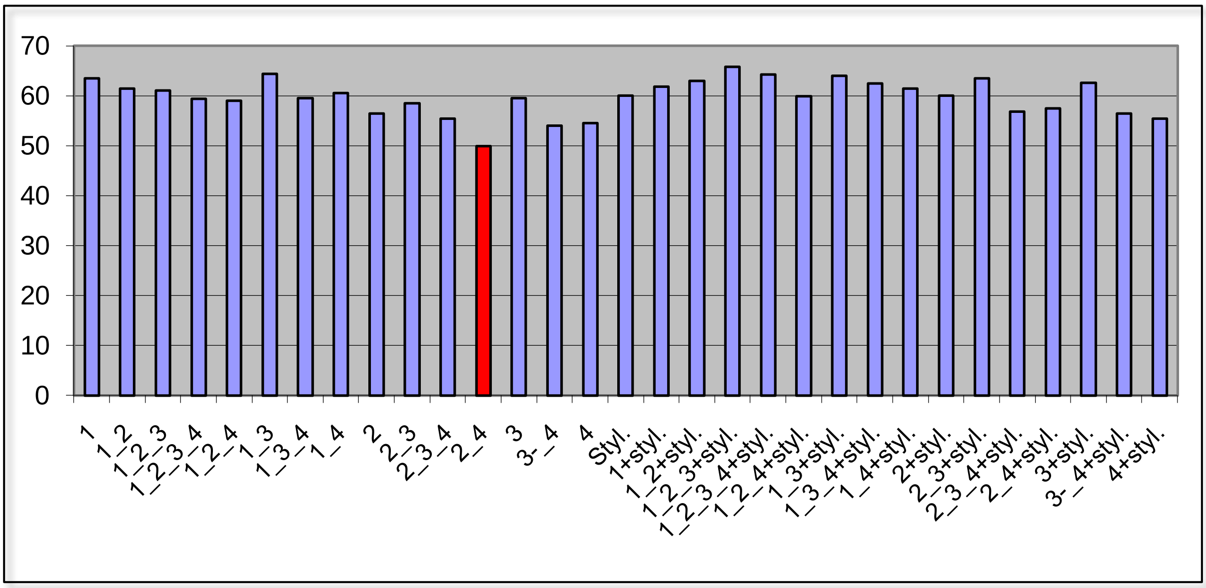

Figure 4.3 Results from the ML experiment for the death penalty topic in English

Broadly speaking, the best results from the ML experiment are obtained with the good friend subcorpus. Specifically, the first dimension seems to have great discriminatory power, since all the classifiers containing it score above 74.5% (see Figure 4.4), the combinations 1_2+styl. and 1_3_4 being the best performers. It is worth noting that, in the rest of combinations, the third and fourth dimensions perform particularly poorly, with the success rates of this subcorpus ranging widely.

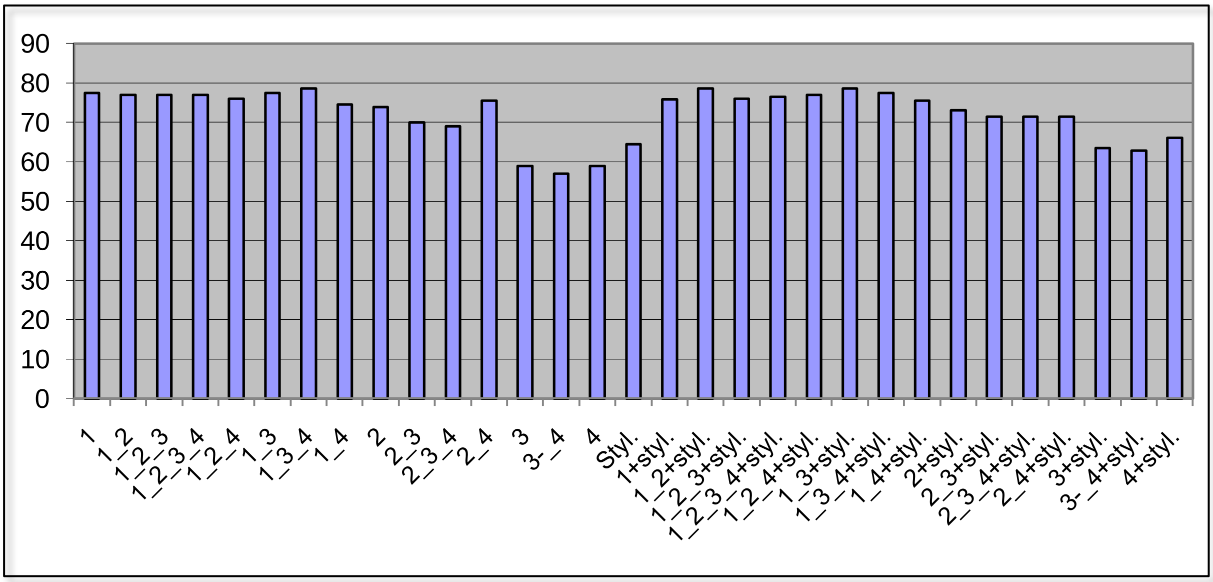

Figure 4.4 Results from the ML experiment for the good friend topic in English

Finally, the experiment on the whole English corpus yields rather regular results, the fourth dimension on its own excepted (see Figure 4.5). Interestingly enough, a parallel may be drawn with the abortion subcorpus, since classifier 4 also performs atypically low when compared to the rest of the classifiers within their class (see Figure 4.1). The best classifier is 1_2_3+styl., and 1_2_3_4 and 1_3 also perform strongly.

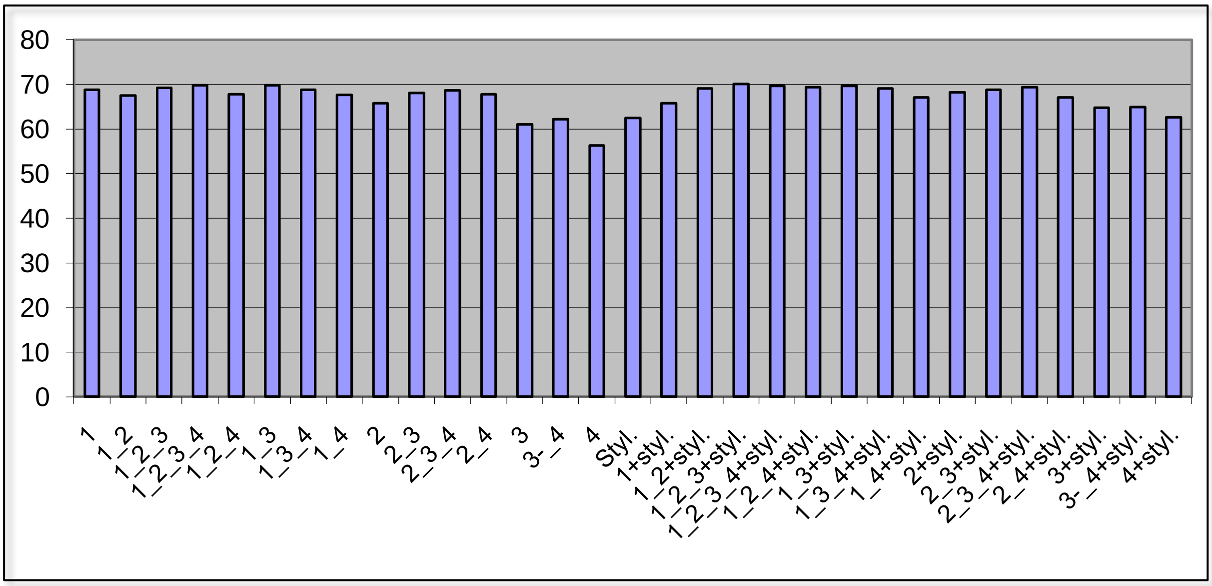

Figure 4.5 Results from the ML experiment for the whole corpus in English

4.2.1.2. Dimension classifiers for Spanish

As shown in Table 4.2 and Figure 4.6, the performance of the individual dimensions across the Spanish subcorpora fluctuates more than in English. For instance, dimension 1 is the most discriminating on its own in the subcorpora bullfighting and good friend, whereas in the case of homosexual adoption and the whole corpus, dimension 2 is more effective. In these cases, the next one in importance is dimension 1, followed by the stylometric dimension and finally the third and the fourth ones. The worst classifier in all cases is indeed 4. It is worth noting that, in general terms, the stylometric dimension in combination does not improve success rates as much as in the case of the English subcorpora, especially when it comes to the whole Spanish corpus, where it only improves the results in the classifiers 3+styl. and 4+styl.

	Bullfighting	Homosexual adoption	Good friend	All
1	65.5	67.0	82.0	69.3
1_2	**69.0**	73.5	**84.0**	72.3
1_2_3	66.0	72.5	83.4	**74.8**
1_2_3_4	61.5	71.5	**84.0**	**74.2**
1_2_4	66.5	71.0	**84.5**	72.7
1_3	65.9	69.5	82.0	71.8
1_3_4	66.2	67.0	80.5	70.8
1_4	66.4	69.5	81.0	70.3
2	60.7	70.0	76.0	71.5
2_3	62.0	72.0	80.5	**73.2**
2_3_4	57.5	73.0	81.0	**73.3**
2_4	58.4	72.5	78.0	70.5
3	64.0	59.3	69.0	60.8
3_4	57.4	61.3	69.4	65.5
4	53.4	57.8	62.4	58.2
Styl.	60.5	65.0	67.8	62.5
1+styl.	66.0	66.5	**84.0**	69.0
1_2+styl.	65.2	68.5	82.0	70.2
1_2_3+styl.	63.0	71.5	83.0	72.8
1_2_3_4+styl.	63.0	72.0	81.5	**73.2**
1_2_4+styl.	65.5	71.0	80.5	70.5
1_3+styl.	64.5	68.5	78.5	71.0
1_3_4+styl.	63.7	71.5	81.0	70.1
1_4+styl.	62.4	70.9	80.0	68.2
2+styl.	61.9	**74.5**	79.0	70.0
2_3+styl.	60.5	72.0	82.0	70.7
2_3_4+styl.	60.9	**75.0**	81.5	71.5
2_4+styl.	61.5	73.5	77.0	69.2
3+styl.	57.0	67.5	76.0	66.8
3_4+styl.	60.0	68.5	73.5	65.3
4+styl.	58.0	63.9	67.5	63.3

Table 4.2 Results from the selected categories grouped into dimensions in Spanish

As commented on above, dimension 1 is the most discriminating on its own in the subcorpus bullfighting (see Figure 4.7). The next one in importance is not dimension 2, as might be expected from previous findings, but rather dimension 3, relativity. The psychological processes dimension is comparatively less successful, although not as poor as dimension 4, which yields a result not much better than chance. This dimension does not exert a positive effect in either combination. On the other hand, the best classifier for this topic is 1_2, whereas the worst one is 3+styl., which differs from classifier 3 by 7 points. Overall, this subcorpus is the one which yields the worst classification results.

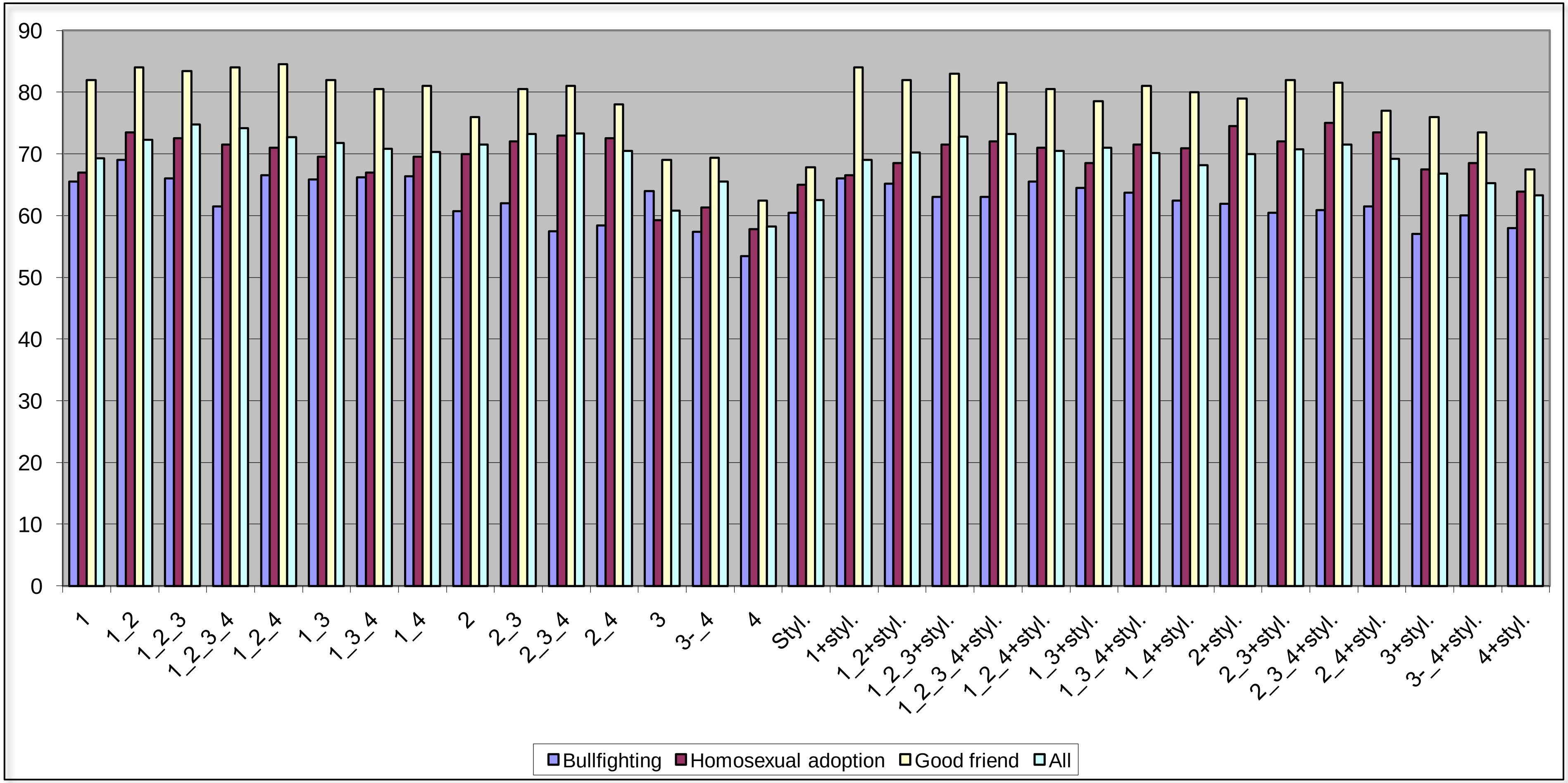

Figure 4.6 Results from the ML experiment for all the topics in Spanish

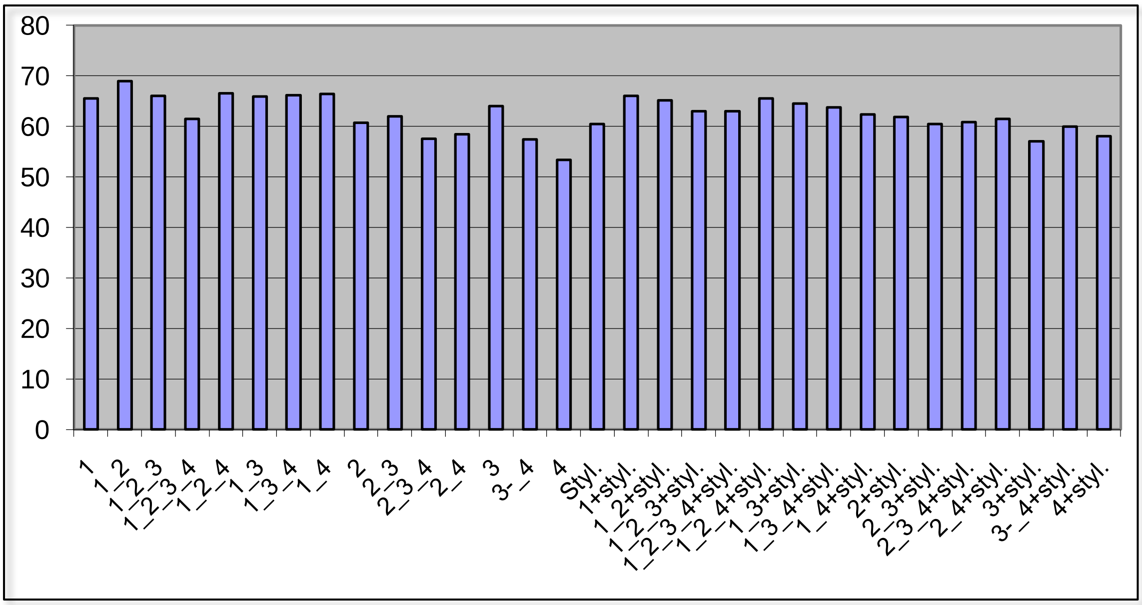

Figure 4.7 Results from the ML experiment for the bullfighting topic in Spanish

In the case of homosexual adoption, dimension 2 is not only the most effective on its own, but also in combination, since, as shown in Figure 4.8, it improves all the scores of the classifiers to which it is added. Specifically, 2+styl. and 2_3_4+styl. are the best combinations. Apart from sharing similarities with the whole Spanish corpus mentioned above, it is worth noting that their success rates range across the same scores.

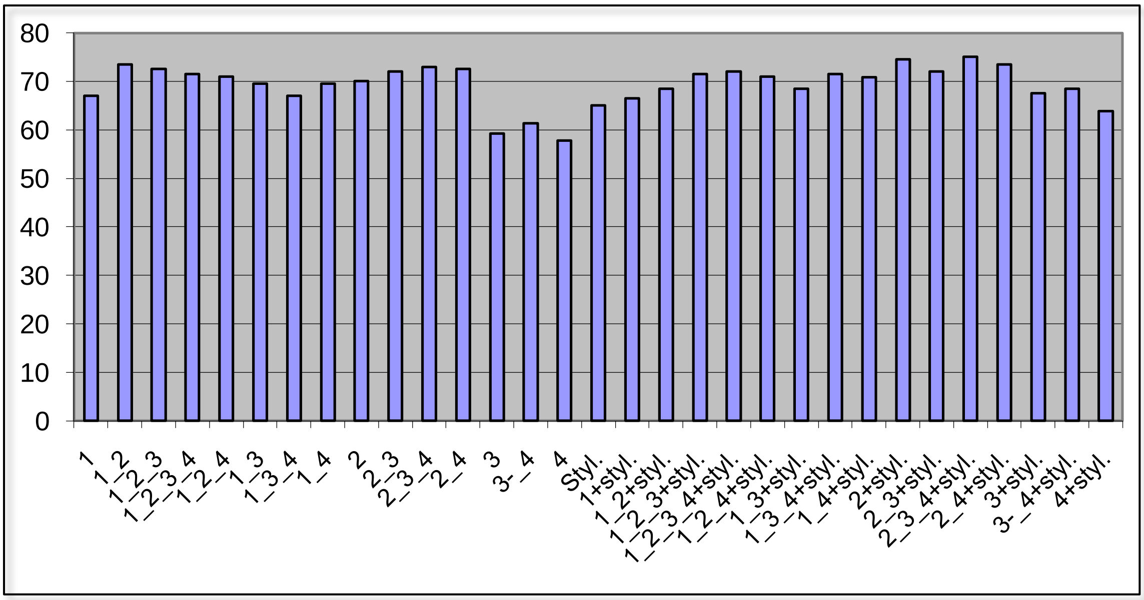

Figure 4.8 Results from the ML experiment for the homosexual adoption topic in Spanish

As regards the good friend topic, it is worth noting that the classifiers achieve the best results not only in Spanish, but in both languages (see Figure 4.9). The best classifier is 1_2_4, and, interestingly enough, 1+styl. is remarkably successful, obtaining the same results as 1_2 and 1_2_3_4.

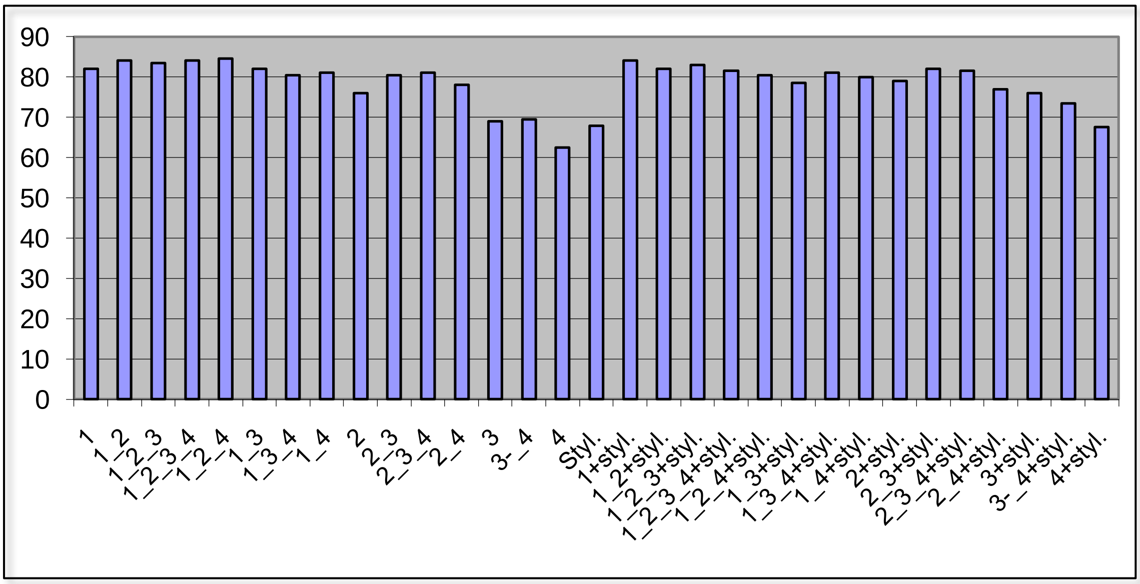

Figure 4.9 Results from the ML experiment for the good friend topic in Spanish

Finally, the best classifier for the whole Spanish corpus is 1_2_3, and the grouping of all LIWC dimensions is also remarkably successful, as in the case of the good friend subcorpus (see Figure 4.10). As mentioned above, the stylometric dimension is more successful on its own than dimensions 3 and 4, but their performance in combinations is only positive in classifiers 3+styl. and 4+styl. Broadly speaking, these classification results are better than with the English corpus.

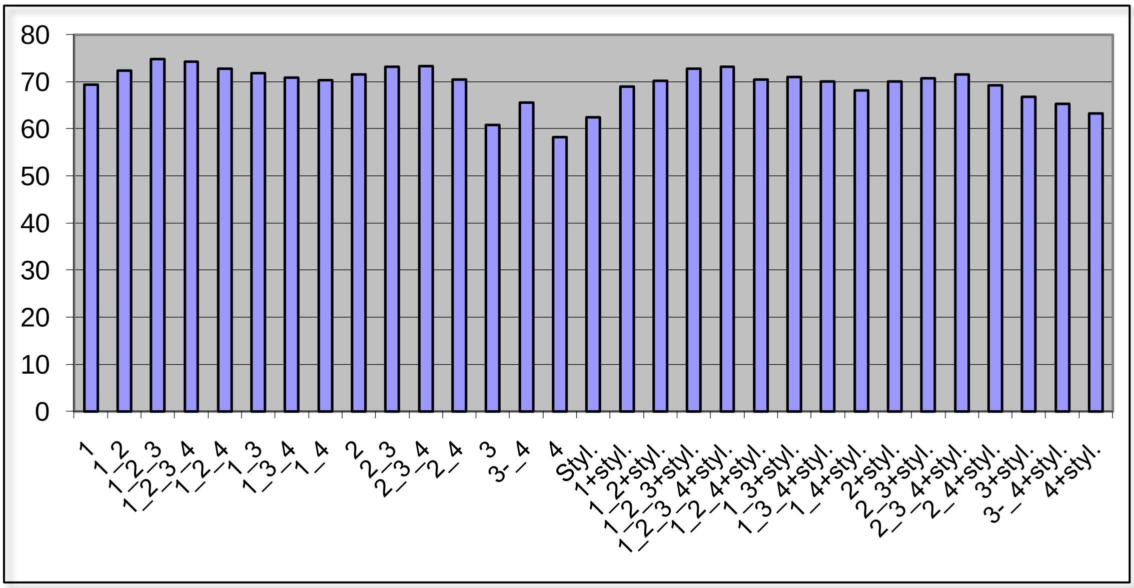

Figure 4.10 Results from the ML experiment for the whole corpus in Spanish

4.2.1.3. Bag-of-Words (BoW) model in English

The results from the Bag-of-Words experiment in English, in which feature vectors include stem frequency, are shown in Table 4.3 and Figure 4.11. The success rates resemble the distribution across topics found in the previous experiment, albeit just partially. The death penalty subcorpus obtains the poorest rate, as it did with the dimension classifiers; the model applied to the whole corpus in English yields average results, and the experiment on the abortion subcorpus proves slightly more successful, as might be expected from previous findings. However, against all odds, the good friend subcorpus classification is a little less successful.

Abortion	Death penalty	Good friend	All English
71.5	55.9	69.3	65.1

Table 4.3 Results from the BoW model in English

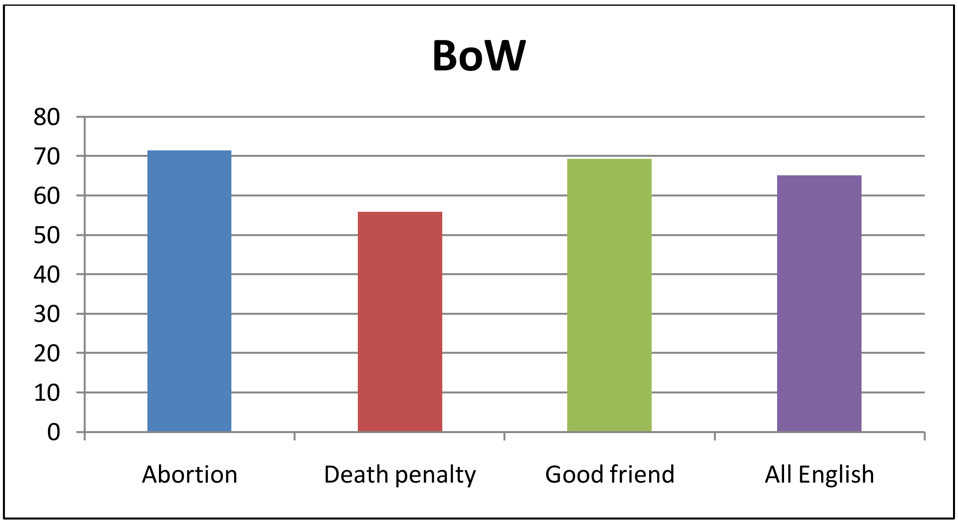

Figure 4.11 Results from the BoW model in English graph

4.2.1.4. Bag-of-Words (BoW) model in Spanish

On the other hand, Table 4.4 and Figure 4.12 show how the distribution of the results from the BoW model in Spanish strongly resembles that of the previous ML test. Accordingly, bullfighting statements are the most difficult to tell apart by means of stem frequencies, whereas the good friend subcorpus responds to the model most adequately. As regards the experiment on the whole corpus, it is almost equally successful as with homosexual adoption, yielding average values.

Bullfighting	Homosexual adoption	Good friend	All Spanish
62.20	65.40	71.50	64.80

Table 4.4 Results from the BoW model in Spanish

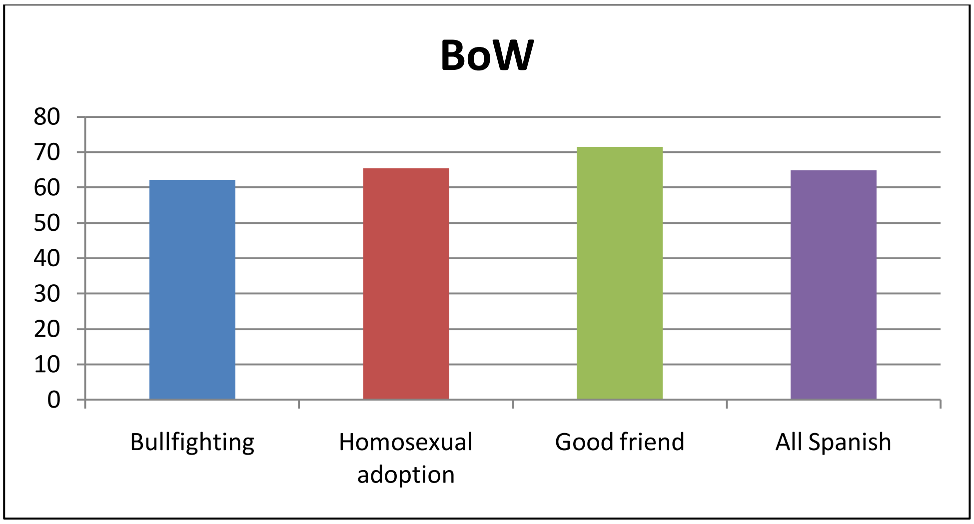

Figure 4.12 Results from the BoW model in English graph

4.2.2. Statistical techniques

As commented on in the previous chapter, the statistical classification techniques implemented help gain a deeper understanding of the specific categories which best discriminate between both sublanguages. First, the results are presented for English, Spanish and the combination of both corpora. Then, the results from the analysis of the individual subcorpora are evaluated.

4.2.2.1. Prediction of membership with individual categories in the English corpus

Since the central limit theorem (CLT) justifies the approximation of large-sample statistics to the normal distribution in controlled experiments (Rice, 1995), DFA has been applied to the English corpus, comprising 600 cases.

First of all, as shown in Table 4.5, Wilks' lambda confirms that the variables in combination successfully discriminate between truthful and untruthful statements (Wilks' $\lambda = 0.756$, $\chi^2 = 166.3$, $p = 0.000$). Smaller values of Wilks' lambda indicate greater discriminatory ability of the function. In addition, the associated chi-square statistic tests the hypothesis that the means of the functions listed are equal across groups. Table 4.5 displays the results of a one-way ANOVA for the independent variable using the grouping variable as the factor, and, if the significance value is lower than 0.10, the variable remarkably contributes to the model. Thus, according to the results, every variable in Table 4.6 is significant. Ranking the F-ratios –equality of group means– identifies word count as the best single predictor. Figure 4.13 reflect the importance of other predictors from a distance of 8 points, namely insight, 3rd person, friends, 1st person singular, exclusive words, 2nd person, inclusive words and discrepancy.

Test of function(s)	Wilks' lambda	Chi-square	df	Sig.
1	.756	166.341	9	.000

Table 4.5 Wilks' lambda for English

Predictors	F	Sig.
WC	44.473	.000
Insight	36.562	.000
Other	35.865	.000
Friends	32.265	.000
I	31.928	.000
Excl	28.640	.000
You	25.481	.000
Incl	22.985	.000
Discrep	21.206	.000

Table 4.6 F-ratios for English

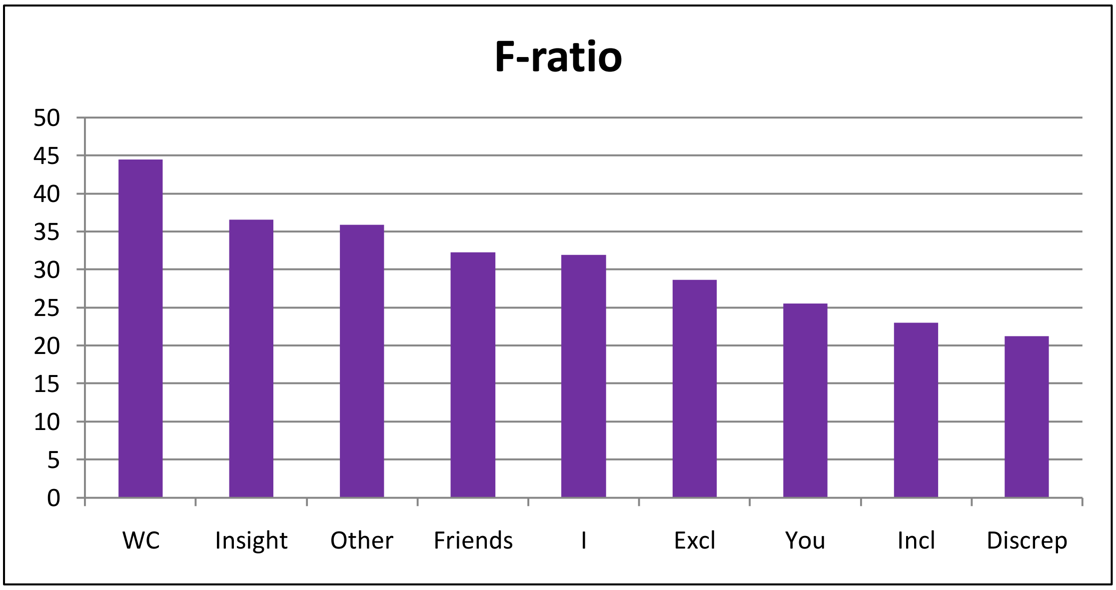

Figure 4.13 F-ratios for English graph

These results give relevant information concerning the most important predictors for the English corpus. Nonetheless, this information would be incomplete without the Fisher linear discriminant functions (see Table 4.7), which broaden knowledge on the variables substantially contributing to the ascription of statements to the group of truthful or untruthful texts in the model.

Specifically, the discriminant model assigns the case to the group whose classification function obtained the highest score. In this respect, the categories that contribute the most to deception detection are 2nd and 3rd person. On the contrary, a strong presence of word count –that is to say, longer statements–, 1st person singular, words related to insight, discrepancy, friends, and inclusive and exclusive words are fairly characteristic of truthful texts.

	Deception	
	No	Yes
WC	.074	.057
I	.096	.005
You	.193	.321
Other	.082	.276
Insight	.829	.605
Discrep	.798	.701
Friends	.717	.335
Incl	.915	.830
Excl	.816	.678
(Constant)	-11.443	-8.608

Table 4.7 Fisher linear discriminant functions for English

		Deception	Predicted group membership		Total
			No	Yes	1
Original(a)	Count	No	219	81	300
		Yes	85	215	300
	%	No	73.0	27.0	100.0
		Yes	28.3	71.7	100.0
Cross-validated (b)	Count	No	212	88	300
		Yes	87	213	300
	%	No	70.7	29.3	100.0
		Yes	29.0	71.0	100.0

(a) 72.3% of original grouped cases correctly classified.

(b) 70.8% of cross-validated grouped cases correctly classified.

Table 4.8 Classification results from DFA for English

This DFA model successfully classified 72.3% of the original grouped cases (see Table 4.8), and cross-validation was similarly successful, with the leave-one-out classification method seeing 70.8% of the statements correctly classified. It is worth noting that the percentage of truthful and untruthful statements correctly classified in the cross-validation was remarkably similar (70.7% vs 71.0%). This means in practice that 212 truthful and 213 untruthful texts were classified as such (see Appendix II.4 for a comprehensive list of the cases classified successfully and misclassified in the full model).

4.2.2.2. Prediction of membership with individual categories in the Spanish corpus

Much like the English corpus, a DFA has been applied to the Spanish corpus. Table 4.9 shows a successful discrimination between both kinds of statements (Wilks' $\lambda = 0.699$, $\chi^2 = 210.7$, $p = 0.000$). Again, text length proves to be the best single predictor, as shown in Table 4.10 and Figure 4.14. Curiously enough, in this case the difference between this predictor and the next one in importance is 20 points, which is more than twice the difference observed in the English corpus. Despite this fact, the F-ratio for the next predictor, 1st person singular, is still rather high. There are some other variables identified as predictors shared with the English corpus, namely 2nd person, friends, insight, exclusive words, and 3rd person. The remaining predictors for the Spanish corpus are words related to certainty, humans, sexual, number, anger, semicolon, past, assent, future, and tentative words. Figure 4.15 shows the distribution of the predictor categories in both corpora.

Test of function(s)	Wilks' lambda	Chi-square	df	Sig.
1	.699	210.704	17	.000

Table 4.9 Wilks' lambda for Spanish

Predictors	F	Sig.
WC	69.812	.000
I	49.259	.000
Certain	39.199	.000
You	33.516	.000
Friends	30.167	.000
Humans	27.682	.000
Insight	25.708	.000
Excl	23.601	.000
Sexual	21.871	.000
Number	20.568	.000
Anger	19.397	.000
SemiC	18.329	.000
Other	17.495	.000
Past	16.643	.000
Assent	15.909	.000
Future	15.239	.000
Tentat	14.709	.000

Table 4.10 F-ratios for Spanish

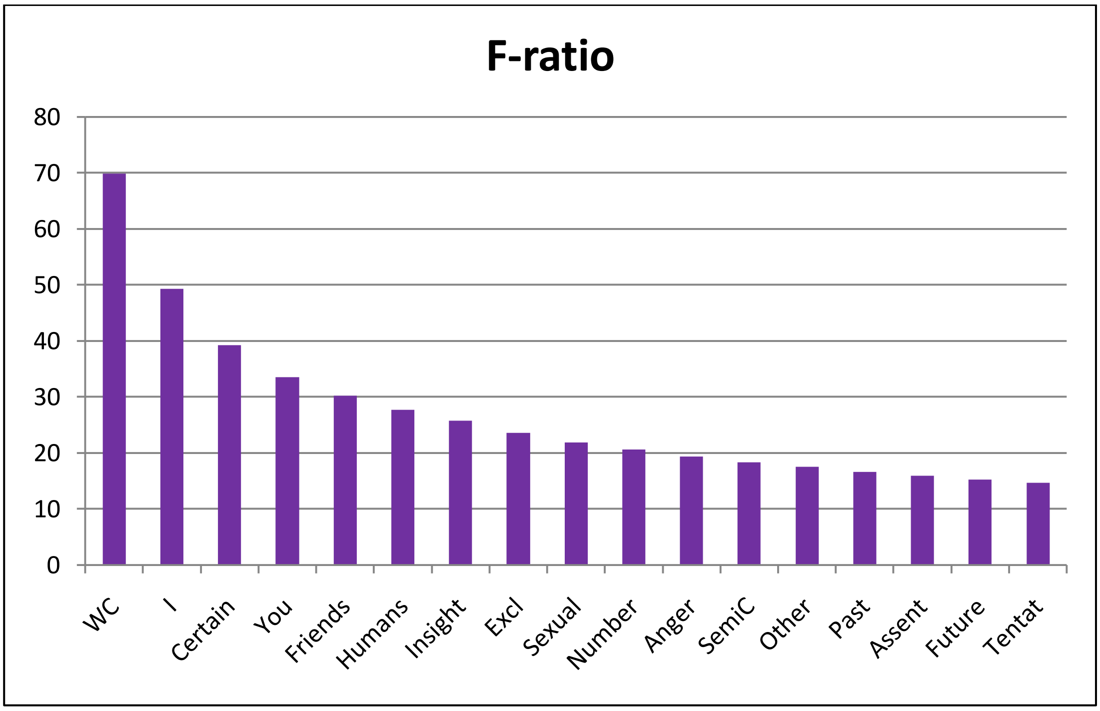

Figure 4.14 F-ratios for Spanish graph

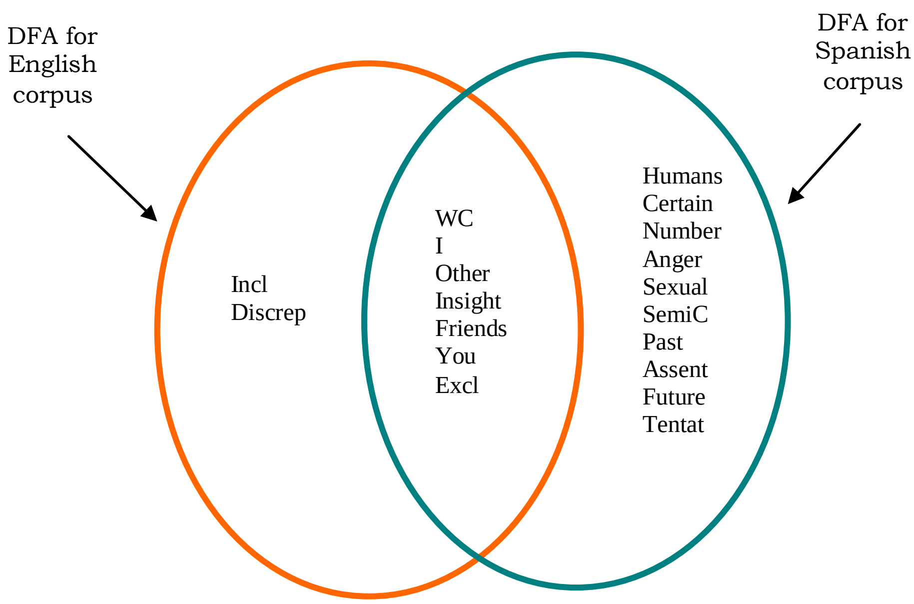

Figure 4.15 Venn diagram of the predictor categories identified by the DFA in the English and Spanish corpora

On the other hand, the Fisher linear discriminant functions show that the categories that contribute the most to the model concerning untruthfulness are 2nd and 3rd person –predictors shared by both corpora–, certainty, humans, and future, whereas a strong presence of exceptional text length, semicolon, 1st person singular, words indicating assent, number, anxiety, insight, tentative, friends, past, sexuality, and exclusive words typically characterises truthful statements in Spanish (see Table 4.11).

	Deception	
	No	Yes
WC	.096	.077
SemiC	.675	-.055
I	.435	.222
You	.169	.397
Other	.842	.907
Assent	-.439	-.156
Number	.459	.279
Anx	1.350	1.051
Insight	.901	.712
Tentat	.682	.557
Certain	-.007	.194
Friends	.642	.361
Humans	.549	.736
Past	.607	.485
Future	-.261	.029
Excl	.837	.615
Sexual	.657	.400
(Constant)	-13.647	-11.167

Table 4.11 Fisher linear discriminant functions for Spanish

		Deception	Predicted group membership		Total
			No	Yes	1
Original (a)	Count	No	233	67	300
		Yes	75	225	300
	%	No	77.7	22.3	100.0
		Yes	25.0	75.0	100.0
Cross-validated (b)	Count	No	227	73	300
		Yes	83	217	300
	%	No	75.7	24.3	100.0
		Yes	27.7	72.3	100.0

(a) 76.3% of original grouped cases correctly classified.
(b) 74.0% of cross-validated grouped cases correctly classified.

Table 4.12 Classification results from DFA for Spanish

In this case, the DFA model is slightly more successful than with the English corpus: 76.3% of the original grouped cases were correctly classified (see Table 4.12), and the leave-one-out classification method has achieved a success rate of 74.0%. As regards the percentage of truthful and untruthful statements correctly classified in the cross-validation, the former is slightly more successful than the latter (75.7% vs 72.3%). Specifically, there is a difference of ten more statements correctly classified.

4.2.2.3. Prediction of membership with individual categories in the subcorpora in English

As commented on in the previous chapter, a one-sample Kolmogorov-Smirnov test provided evidence against the null hypothesis for the three subcorpora in English (see Appendix II.5), implying that the samples had not been drawn from a normal population. The distributions of the variables were found to be significant, thus binary logistic regressions were conducted instead of DFA.

4.2.2.3.1. Prediction of membership with individual categories in the abortion subcorpus in English

From the two methods of binary logistic regression used, the backward stepwise method begins with a saturated model, and variables are eliminated from the model in an iterative process. The fit of the model is tested after the elimination of each variable to ensure that the model still adequately fits the data, and the analysis is completed when no more variables can be eliminated from the model. The variables kept in the last step are shown in Table 4.13.

	B	S.E.	Wald	df	Sig.	Exp(B)
WC	-.044	.012	12.979	1	.000	.957
Sixltr	-2.532	.598	17.904	1	.000	.080
Comma	-.326	.168	3.780	1	.052	.722
Dash	-1.429	.525	7.426	1	.006	.239
Quote	-.322	.164	3.869	1	.049	.725
Apostro	1.403	.411	11.654	1	.001	4.068
We	1.080	.362	8.921	1	.003	2.944
Other	.413	.183	5.097	1	.024	1.512
Number	-2.132	.635	11.281	1	.001	.119
Anger	-2.895	.648	19.946	1	.000	.055
Sad	-1.449	.599	5.857	1	.016	.235
Cause	.951	.360	6.984	1	.008	2.589
Insight	-1.697	.407	17.353	1	.000	.183
Incl	-.736	.221	11.038	1	.001	.479
Money	2.222	.725	9.407	1	.002	9.227
Death	1.595	.418	14.577	1	.000	4.926
Sexual	.536	.191	7.842	1	.005	1.709

Table 4.13 Variables in the last step of the equation in the backward stepwise logistic regression for the abortion topic

As can be seen, a Wald test has been used to assess the statistical significance of each coefficient (β) in the model, so as to make the preliminary selection of predictors. Following Menard (1995) and Cohen et al. (2002), large and extremely low coefficients with associated inflated standard errors must be discarded, since they increase the probability of a Type-II error; this usually includes variables with standard errors higher than 1 (see Appendix II.6 for a comprehensive list). Subsequently, a forward stepwise binary logistic regression has been performed on the preselected variables. In terms of significance, this procedure reports the Hosmer-Lemeshow goodness-of-fit statistics (see Table 4.14), which is useful to determine whether the built model reasonably approximates the behaviour of the data. It indicates a poor fit if the significance value is less than 0.05. Here, the model adequately fits the data (p=.899).

Step	Chi-square	df	Sig.
1	8.937	6	.177
2	9.355	8	.313
3	4.954	8	.763
4	3.046	8	.931
5	6.943	8	.543
6	10.953	8	.204
7	3.591	8	.892
8	9.948	8	.269
9	3.498	8	.899

Table 4.14 Hosmer-Lemeshow statistic in the forward stepwise logistic regression for abortion topic

Following Hosmer and Lemeshow (2000), if there is a statistically significant relationship, the pattern of significance in the individual Wald statistics is potentially useful to interpret the role of the variable in predicting membership in dependent variable categories. However, the authors identify the likelihood ratio method as more effective in identifying relationships than the Wald statistics for the individual logistic regression equations. That is to say, the change in a -2 log-likelihood is generally more reliable than the Wald statistic. Table 4.15 and Figure 4.16 show the predictors identified by this method in the abortion subcorpus:

Predictors	Change in the -2 log-likelihood
Insight	42.582
WC	35.086
We	11.279
Anger	9.719
Cause	9.013
Other	8.302
Incl	7.294
Sad	7.208
Sexual	5.481

Table 4.15 Changes in the -2 log-likelihood for the abortion topic

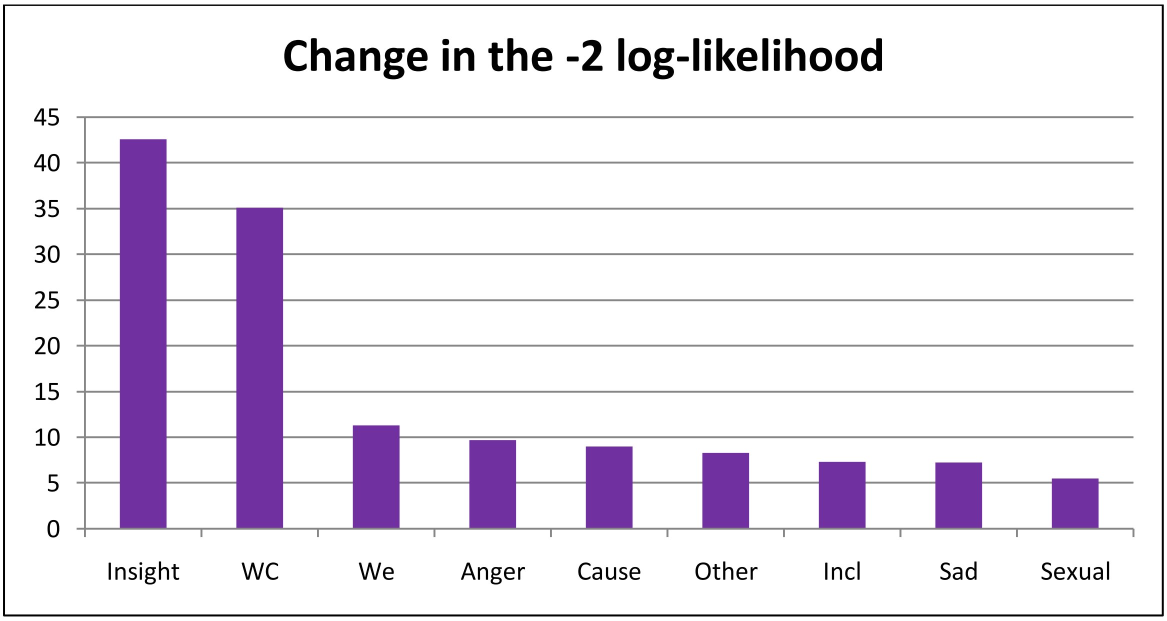

Figure 4.16 Changes in the -2 log-likelihood for the abortion topic graph

Table 4.16 shows the predictors finally included in the model after applying both methods with their corresponding coefficients. The easiest value to interpret is Exp(B), which represents the ratio change in the odds of the event of interest for a one-unit change in the predictor (Cohen et al., 2002). As in the previous test, large and extremely low coefficients with associated inflated standard errors must be discarded, since they increase the probability of a Type-II error (Cohen et al., 2002; Menard, 1995). However, the coefficients and the standard errors of all the predictors kept in the last step prove to be adequate. As regards the coefficient (β) of logistic regression, it is worth noting that it does not have the same straightforward interpretation as it does with linear regression (Efron et al., 2004), but its sign gives information on the truth value of the statements in the classification experiment. Specifically, positive values are indicative of predictors of untruthful statements, namely 1st person plural, 3rd person, causal words, and words related to sex, whereas negative values here are associated with truthful statements, in this case text length, words related to anger, sadness, insight, and inclusive words.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 9	WC	-.047	.011	19.664	1	.000	.954
	We	.854	.302	7.983	1	.005	2.350
	Other	.334	.124	7.282	1	.007	1.396
	Anger	-.627	.212	8.777	1	.003	.534
	Sad	-1.205	.483	6.221	1	.013	.300
	Cause	.742	.268	7.670	1	.006	2.100
	Insight	-1.064	.219	23.615	1	.000	.345
	Incl	-.284	.111	6.594	1	.010	.752
	Sexual	.289	.130	4.947	1	.026	1.335

Table 4.16 Variables in the last step of the equation in the forward stepwise logistic regression for the abortion topic

Once a reliable set of predictors has been selected, the classification results are to be explored. As explained above, a random sample of cases has been automatically generated for each topic to create a logistic regression model, setting the remaining contributions aside to validate the analysis and obtain a more reliable classification rate. By means of random generation and of a Bernoulli distribution, approximately 70 percent of the truthful and untruthful statements have been used to create the model, whereas the remaining statements have been used to validate the model results. As can be seen in Table 4.17, the model is successful in the classification of the original cases (69.6%), although it is remarkably more successful in the validation subset, its rate being 75.4%, with 75.8% of truthful statements correctly classified and 75% of untruthful statements classified as such –there is just one less statement correctly classified.

Observed			Predicted					
			Selected cases			Unselected cases		
			Deception		Percent. Correct	Deception		Percent. Correct
			No	Yes		No	Yes	
Step 9	Deception	No	48	19	71.6	25	8	75.8
		Yes	22	46	67.6	8	24	75.0
	Overall Percentage				69.6			75.4

Table 4.17 Classification results in the forward stepwise logistic regression for the abortion topic

4.2.2.3.2. Prediction of membership with individual categories in the death penalty subcorpus in English

In this case, the amount of valid variables preselected by the backward stepwise method is slightly larger than in the previous topic (see Table 4.18). In order to reduce the probability of a Type-II error, the large and extremely low coefficients with associated inflated standard errors have been discarded.

	B	S.E.	Wald	df	Sig.	Exp(B)
WC	-.025	.010	6.342	1	.012	.976
WPS	.115	.050	5.242	1	.022	1.122
Sixltr	-.440	.239	3.384	1	.066	.644
Period	-1.085	.374	8.410	1	.004	.338
Dash	-.932	.474	3.871	1	.049	.394
Apostro	.776	.343	5.106	1	.024	2.173
OtherP	-1.191	.563	4.486	1	.034	.304
I	-1.374	.271	25.735	1	.000	.253
Negate	-.347	.148	5.495	1	.019	.706
Posfeel	1.097	.489	5.035	1	.025	2.995
Tentat	-.423	.143	8.773	1	.003	.655
Certain	.364	.215	2.857	1	.091	1.439
Feel	.866	.404	4.589	1	.032	2.378
Past	-.635	.186	11.638	1	.001	.530
Present	.190	.088	4.649	1	.031	1.209
Future	.439	.166	6.987	1	.008	1.551
Excl	-.315	.111	8.064	1	.005	.730
Job	-.865	.319	7.343	1	.007	.421
Body	-.847	.295	8.213	1	.004	.429
STTR	-.322	.131	6.069	1	.014	.725

Table 4.18 Variables in the last step of the equation in the backward stepwise logistic regression for the death penalty topic

The subsequent forward stepwise binary logistic regression has been performed on the preselected variables. As shown in Table 4.19, the Hosmer-Lemeshow goodness-of-fit statistics indicate that the model adequately fits the data (p=.519), that is to say, the built model reasonably approximates the behaviour of the data.

Step	Chi-square	df	Sig.
1	10.836	8	.211
2	5.625	8	.689
3	7.166	8	.519

Table 4.19 Hosmer-Lemeshow statistic in the forward stepwise logistic regression for the death penalty topic

As far as the change in the -2 log-likelihood is concerned, in the death penalty subcorpus it has proved significant with three predictors: 1st person singular, past and exclusive words (see Table 4.20 and Figure 4.17).

Predictors	Change in the -2 log-likelihood
I	18.491
Past	15.357
Excl	14.705

Table 4.20 Changes in the -2 log-likelihood for the death penalty topic

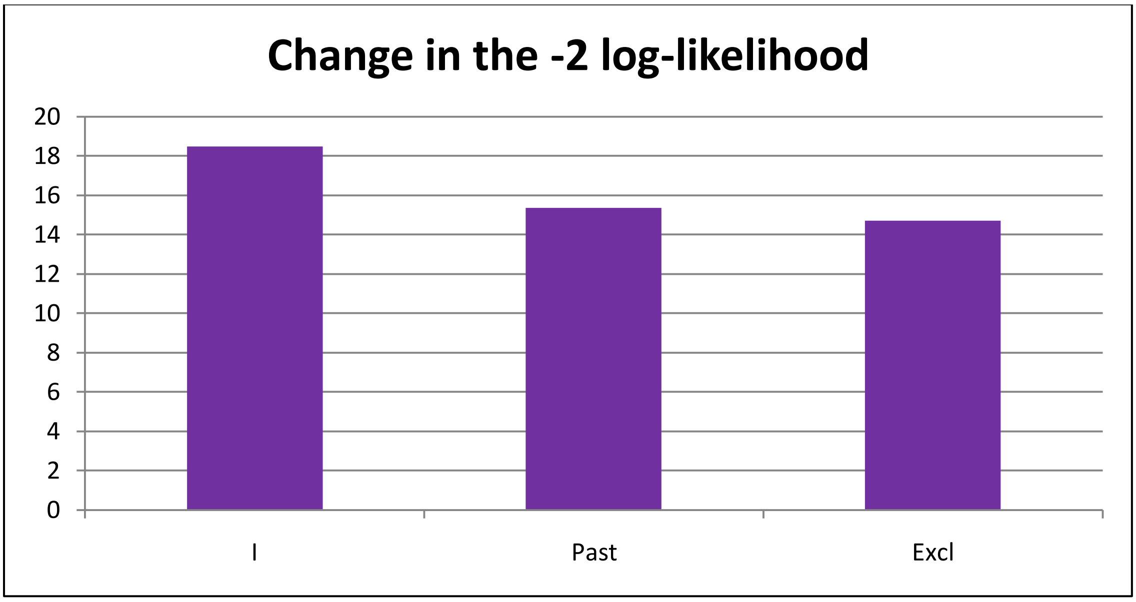

Figure 4.17 Changes in the -2 log-likelihood for the death penalty topic graph

Curiously enough, all of these predictors have obtained negative coefficients in the last step of the equation in the forward stepwise logistic regression, which means that they are indicative of truthful statements (see Table 4.21).

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 3	I	-.522	.136	14.690	1	.000	.593
	Past	-.537	.154	12.142	1	.000	.584
	Excl	-.348	.101	11.828	1	.001	.706

Table 4.21 Variables in the last step of the equation in the forward stepwise logistic regression for the death penalty topic

With these predictors, the model has been able to successfully classify 71.6% of truthful statements and 75% of untruthful ones of the original cases, with a global success rate of 73.3%, as shown in Table 4.22. With the validation subset, the model classifies 66.2% of correct cases. Most interestingly, the success rate in the case of the untruthful statements is significantly higher than with the truthful ones (78.1% vs 54.4%). It is worth noting that this subcorpus, the death penalty, has registered the lowest success rate in English.

Observed			Predicted					
			Selected cases			Unselected cases		
			Deception		Percent. Correct	Deception		Percent. Correct
			No	Yes		No	Yes	
Step 3	Deception	No	48	19	71.6	18	15	54.5
		Yes	17	51	75.0	7	25	78.1
	Overall Percentage				73.3			66.2

Table 4.22 Classification results in the forward stepwise logistic regression for the death penalty topic

4.2.2.3.3. Prediction of membership with individual categories in the good friend subcorpus in English

In line with the abortion topic, a total of 17 valid variables out of the set of 76 have been preselected by the backward stepwise method (see Table 4.23), leaving aside the discarded variables on the grounds of abnormal coefficients and standard errors.

	B	S.E.	Wald	df	Sig.	Exp(B)
Sixltr	-.567	.166	11.657	1	.001	.567
Dash	-.977	.278	12.365	1	.000	.377
I	-.245	.101	5.932	1	.015	.782
You	.680	.239	8.073	1	.004	1.974
Other	.852	.160	28.473	1	.000	2.345
Negate	.659	.212	9.635	1	.002	1.933
Preps	.214	.092	5.355	1	.021	1.238
Posfeel	.819	.273	9.005	1	.003	2.268
Optim	-.746	.251	8.797	1	.003	.474
Discrep	-.291	.145	4.036	1	.045	.747
Friends	-1.062	.281	14.241	1	.000	.346
Family	-1.616	.446	13.113	1	.000	.199
Humans	.733	.253	8.376	1	.004	2.082
School	.630	.202	9.675	1	.002	1.877
Achieve	.665	.210	9.994	1	.002	1.944
Money	1.214	.418	8.444	1	.004	3.366
Sexual	-1.250	.510	6.005	1	.014	.286

Table 4.23 Variables in the last step of the equation in the backward stepwise logistic regression for the good friend topic in English

Table 4.24 shows how the model built by the forward stepwise logistic regression adequately fits the data (p=.194). This goodness-of-fit statistic describes a model comprising the six significant predictors shown in Table 4.25 and Figure 4.18, namely 3rd person, words related to friendship and humans, 2nd person, and words concerning money and family. Like in the previous cases, their significance is assessed in terms of the changes in the -2 log-likelihood.

Step	Chi-square	df	Sig.
1	6.834	8	.555
2	6.861	8	.552
3	8.921	8	.349
4	9.439	8	.307
5	9.504	8	.302
6	11.133	8	.194

Table 4.24 Hosmer-Lemeshow statistic in the forward stepwise logistic regression for the good friend topic in English

Predictors	Change in the -2 log-likelihood
Other	30.071
Friends	11.410
Humans	10.504
You	8.186
Money	4.919
Family	4.575

Table 4.25 Changes in the -2 log-likelihood for the good friend topic in English

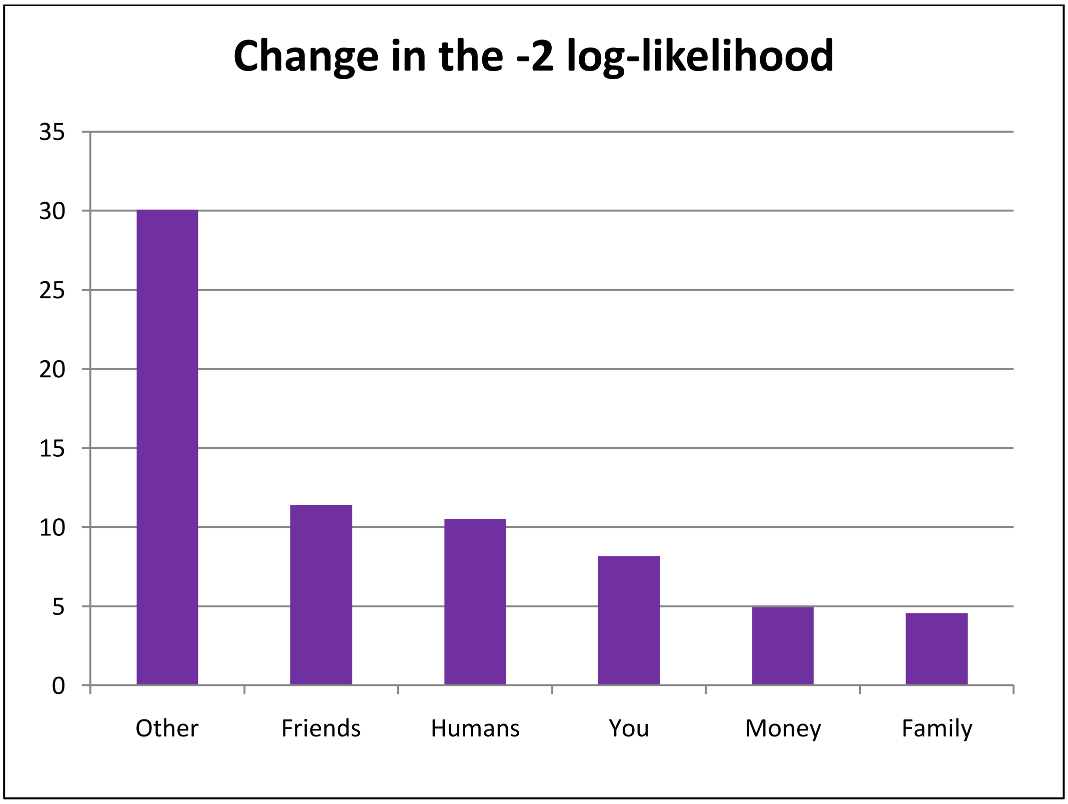

Figure 4.18 Changes in the -2 log-likelihood for the good friend topic in English graph

Table 4.26 shows the predictors finally included in the model after applying both methods with their corresponding coefficients. Positive values, indicative of predictors of untruthful statements, correspond to 2nd and 3rd person, as well as words related to humans and money. On the other hand, terms concerning friendship and family are associated here with truthful statements.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 6	You	.387	.139	7.769	1	.005	1.472
	Other	.325	.071	20.983	1	.000	1.384
	Friends	-.509	.161	10.040	1	.002	.601
	Family	-.631	.300	4.419	1	.036	.532
	Humans	.584	.191	9.378	1	.002	1.793
	Money	.558	.268	4.344	1	.037	1.747

Table 4.26 Variables in the last step of the equation in the forward stepwise logistic regression for the good friend topic in English

Regarding the success rate of this model, it is 77% with the original cases and 78.5% with the validation subset, which is the highest rate obtained for English. In the first case, the proportion of untruthful statements correctly classified is slightly higher than the proportion of truthful ones (77.9% vs 76.1%), whereas the situation with the validation subset is exactly the opposite (78.1% vs 78.8%), as shown in Table 4.27. These are the best success rates in English.

Observed			Predicted					
			Selected cases			Unselected cases		
			Deception		Percent. Correct	Deception		Percent. Correct
			No	Yes		No	Yes	
Step 8	Deception	No	51	16	76.1	26	7	78.8
		Yes	15	53	77.9	7	25	78.1
	Overall Percentage				77.0			78.5

Table 4.27 Classification results in the forward stepwise logistic regression for the good friend topic in English

4.2.2.4. Prediction of membership with individual categories in the subcorpora in Spanish

In line with the English subcorpora, the distributions of the variables in the Spanish subcorpora were found to be significant by means of a one-sample Kolmogorov-Smirnov test, hence the performance of logistic regressions on the data over DFA.

4.2.2.4.1. Prediction of membership with individual categories in the bullfighting subcorpus in Spanish

First of all, the amount of valid variables preselected by the backward stepwise method is similar to the English subcorpora (see Table 4.28). Variables with abnormal coefficients were also discarded for this and the rest of the Spanish subcorpora (see Appendix II.6).

In this case, Table 4.29 shows how the Hosmer-Lemeshow goodness-of-fit statistic proves that the model built by the forward stepwise regression adequately fits the data (p=.755). This model comprises just three predictors, as shown in Table 4.30 and Figure 4.19. Interestingly enough, the model is rather similar to that built on the death penalty subcorpus, since it also included just three significant variables. Two of them were common to both subcorpora, namely 1st person and exclusive words; the other is text length.

	B	S.E.	Wald	df	Sig.	Exp(B)
WC	-.030	.007	16.578	1	.000	.970
Period	-.984	.376	6.858	1	.009	.374
QMark	-1.455	.505	8.295	1	.004	.233
I	-1.024	.217	22.192	1	.000	.359
We	.674	.203	11.058	1	.001	1.962
Other	-.182	.098	3.428	1	.064	.834
Optim	-.486	.244	3.961	1	.047	.615
Tentat	-.452	.137	10.929	1	.001	.636
See	-.667	.279	5.725	1	.017	.513
Hear	.771	.449	2.955	1	.086	2.163
Comm	-.444	.248	3.196	1	.074	.641
Excl	-.683	.177	14.901	1	.000	.505
Job	-.780	.304	6.601	1	.010	.458
Achieve	.513	.255	4.063	1	.044	1.671
TV	2.666	.738	13.047	1	.000	14.376
Money	1.022	.446	5.257	1	.022	2.779
Eating	-1.001	.271	13.663	1	.000	.368

Table 4.28 Variables in the last step of the equation in the backward stepwise logistic regression for the bullfighting topic

Step	Chi-square	df	Sig.
1	5.898	5	.316
2	10.115	8	.257
3	5.020	8	.755

Table 4.29 Hosmer-Lemeshow statistic in the forward stepwise logistic regression for the bullfighting topic

Predictors	Change in the -2 log-likelihood
I	24.768
WC	11.681
Excl	7.200

Table 4.30 Changes in the -2 log-likelihood for the bullfighting topic

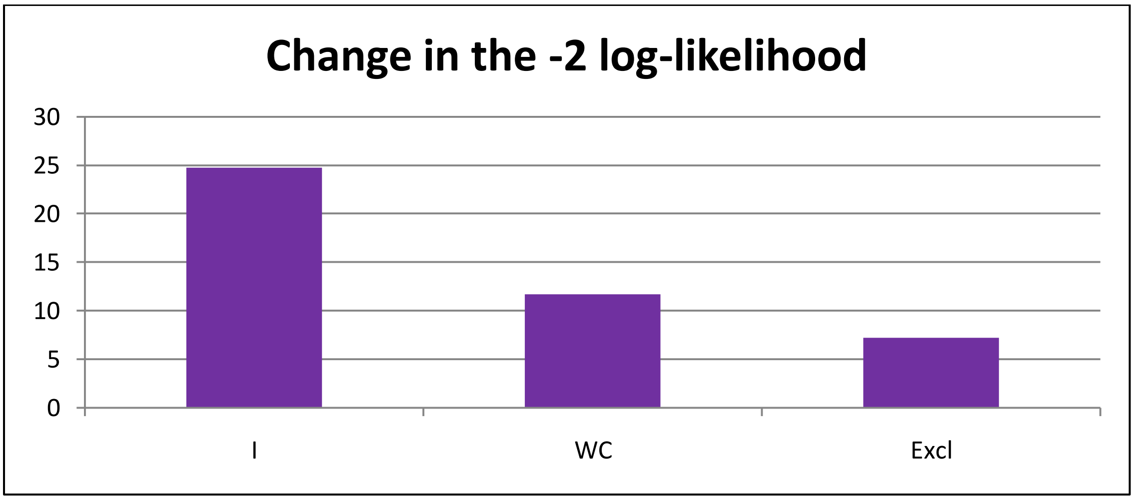

Figure 4.19 Changes in the -2 log-likelihood for the bullfighting topic graph

Furthermore, as can be seen in Table 4.31, the three variables included in the model are predictors of untruthful statements, like with the death penalty subcorpus:

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 3	WC	-.019	.006	10.320	1	.001	.981
	I	-.786	.184	18.293	1	.000	.456
	Excl	-.387	.152	6.470	1	.011	.679

Table 4.31 Variables in the last step of the equation in the forward stepwise logistic regression for the bullfighting topic

Finally, another parallel with the death penalty subcorpus is that it is the one with the lowest success rate in its language. Specifically, 75.8% of the truthful statements and 65.6% of the untruthful ones were correctly classified within the validation subset, with an overall percentage of 70.8. As can be seen in Table 4.32, so far as the original cases are concerned, the overall success rate is 78.5%, the classification success being notable both with truthful statements –76.1%– and with untruthful ones –80.9%.

Observed			Predicted					
			Selected cases			Unselected cases		
			Deception		Percent. Correct	Deception		Percent. Correct
			No	Yes		No	Yes	
Step 3	Deception	No	51	16	76.1	25	8	75.8
		Yes	13	55	80.9	11	21	65.6
	Overall Percentage				78.5			70.8

Table 4.32 Classification results in the forward stepwise logistic regression for the bullfighting topic

4.2.2.4.2. Prediction of membership with individual categories in the homosexual adoption subcorpus in Spanish

As shown in Table 4.33, a total of 22 valid variables out of the set of 76 have been preselected by the backward stepwise method, leaving aside the discarded variables on the grounds of abnormal coefficients and standard errors, which is the largest number in all the subcorpora.

In the forward stepwise logistic regression, the goodness-of-fit statistic shows that the model fits the data (p=.100), although its significance is not as evident as in the other subcorpora (see Table 4.34).

	B	S.E.	Wald	df	Sig.	Exp(B)
WC	-.038	.010	15.362	1	.000	.963
Apostro	1.138	.352	10.453	1	.001	3.122
I	-.861	.355	5.891	1	.015	.423
You	.758	.255	8.854	1	.003	2.134
Negate	.518	.203	6.470	1	.011	1.678
Article	.347	.133	6.831	1	.009	1.415
Posfeel	-.894	.295	9.200	1	.002	.409
Anx	1.451	.744	3.805	1	.051	4.267
Anger	-2.088	.748	7.794	1	.005	.124
Hear	1.255	.560	5.022	1	.025	3.509
Feel	-.575	.226	6.459	1	.011	.563
Family	.427	.150	8.057	1	.005	1.532
Humans	.281	.143	3.852	1	.050	1.325
Future	.591	.338	3.051	1	.081	1.806
Incl	.739	.213	11.991	1	.001	2.094
Excl	-1.102	.329	11.230	1	.001	.332
Motion	1.303	.618	4.445	1	.035	3.680
School	-.626	.316	3.933	1	.047	.534
Achieve	-.818	.329	6.207	1	.013	.441
Home	-.833	.449	3.439	1	.064	.435
Sexual	-.642	.233	7.560	1	.006	.526
STTR	.231	.067	11.841	1	.001	1.260

Table 4.33 Variables in the last step of the equation in the backward stepwise logistic regression for the homosexual adoption topic

Step	Chi-square	df	Sig.
1	6.015	8	.646
2	5.476	8	.706
3	5.063	8	.751
4	7.512	8	.482
5	14.015	8	.081
6	13.365	8	.100

Table 4.34 Hosmer-Lemeshow statistic in the forward stepwise logistic regression for the homosexual adoption topic

Six predictors have been identified as significant on the grounds of the changes in the -2 log-likelihood (see Table 4.35 and Figure 4.20), namely 1st person singular, text length, positive feelings, inclusive words, and terms related to humans and motion. As can be appreciated in Table 4.36, the first three are predictors of truthful statements, and the remainder of untruthful ones, although motion is only near to significant (p=.057).

Predictors	Change in the -2 log-likelihood
I	16.239
WC	13.870
Posfeel	8.786
Incl	6.227
Humans	5.264
Motion	4.231

Table 4.35 Changes in the -2 log-likelihood for the homosexual adoption topic

With these predictors, the model has been able to successfully classify 73.1% of truthful statements and 75% of untruthful ones of the original cases, with an overall success rate of 74.1% (see Table 4.37). With the validation subset, the model classifies 75.4% of correct cases. Interestingly enough, the success rate in the case of the untruthful statements is significantly higher than with the truthful ones (81.3% vs 69.7%), which is a similar situation with the death penalty subcorpus.

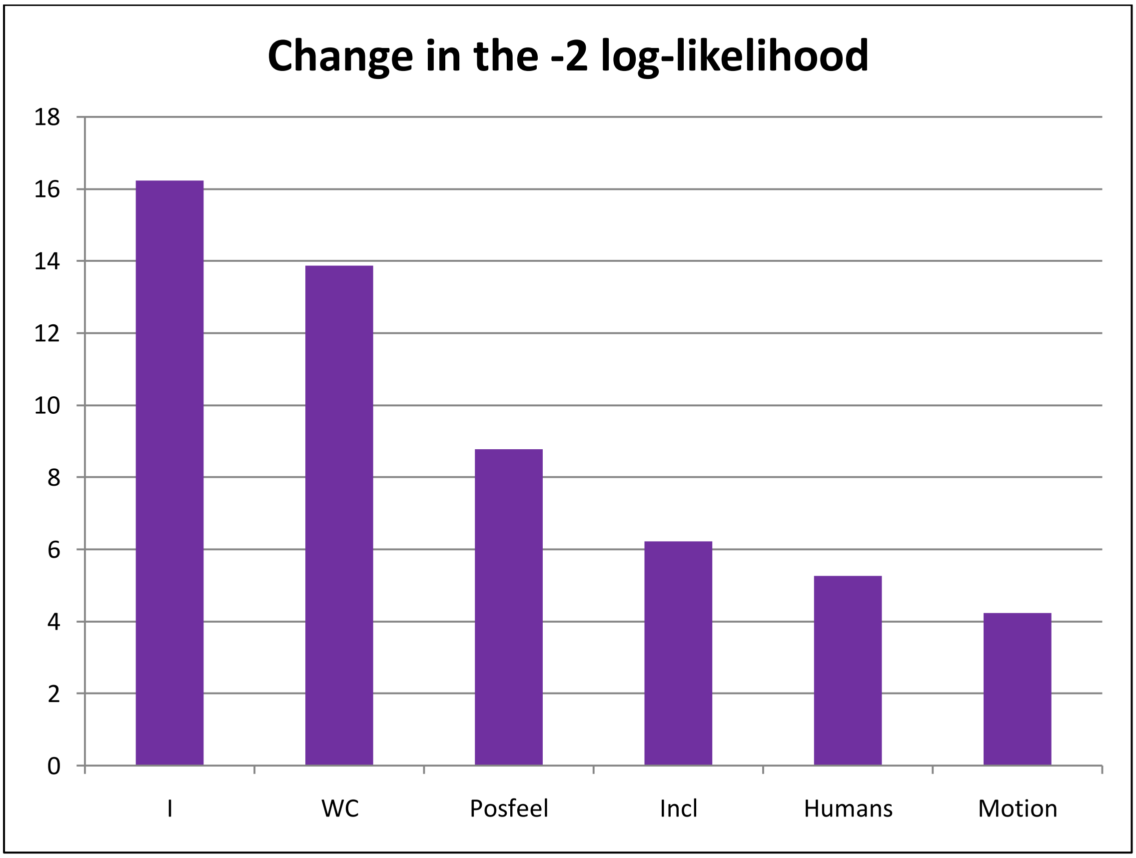

Figure 4.20 Changes in the -2 log-likelihood for the homosexual adoption topic graph

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 6	WC	-.022	.006	11.275	1	.001	.979
	I	-.845	.241	12.335	1	.000	.429
	Posfeel	-.537	.192	7.798	1	.005	.584
	Humans	.240	.109	4.857	1	.028	1.272
	Incl	.284	.121	5.558	1	.018	1.329
	Motion	.688	.362	3.619	1	.057	1.990

Table 4.36 Variables in the last step of the equation in the forward stepwise logistic regression for the homosexual adoption topic

Observed			Predicted					
			Selected cases			Unselected cases		
			Deception		Percent. Correct	Deception		Percent. Correct
			No	Yes		No	Yes	
Step 6	Deception	No	49	18	73.1	23	10	69.7
		Yes	17	51	75.0	6	26	81.3
	Overall Percentage				74.1			75.4

Table 4.37 Classification results in the forward stepwise logistic regression for the homosexual adoption topic

4.2.2.4.3. Prediction of membership with individual categories in the good friend subcorpus in Spanish

For the last subcorpus, the preliminary selection of valid variables made by the backward stepwise regression is shown in Table 4.38. When it comes to the forward stepwise method, it can be stated that the model adequately fits the data (p=.905). In fact, this is the best value for the Hosmer-Lemeshow statistic not only for Spanish but for the whole set of subcorpora (see Table 4.39).

This model is the one comprising the highest amount of predictors (see Table 4.40 and Figure 4.21). Specifically, on the grounds of the changes in the -2 log-likelihood, certainty is the category with the greatest discriminatory power; other relevant predictors are number, 2nd person, text length, 1st person singular, 3rd person, quotation punctuation, words related to achievement, friendship, sadness, and inhibition.

	B	S.E.	Wald	df	Sig.	Exp(B)
WC	-.040	.010	15.513	1	.000	.961
Period	-.381	.145	6.911	1	.009	.683
SemiC	-1.842	.551	11.162	1	.001	.158
Quote	-1.302	.593	4.826	1	.028	.272
I	-.454	.133	11.686	1	.001	.635
We	-.319	.154	4.273	1	.039	.727
You	.726	.259	7.883	1	.005	2.068
Other	.288	.096	9.020	1	.003	1.334
Number	-1.433	.386	13.757	1	.000	.239
Posfeel	.443	.190	5.441	1	.020	1.557
Sad	-2.031	.710	8.171	1	.004	.131
Inhib	-1.605	.899	3.187	1	.074	.201
Certain	.663	.184	13.050	1	.000	1.941
Friends	-.699	.232	9.068	1	.003	.497
Future	1.245	.501	6.164	1	.013	3.471
Achieve	.823	.285	8.310	1	.004	2.277
STTR	.038	.018	4.299	1	.038	1.039

Table 4.38 Variables in the last step of the equation in the backward stepwise logistic regression for the good friend topic in Spanish

Step	Chi-square	df	Sig.
1	9.996	8	.265
2	5.312	8	.724
3	6.448	8	.597
4	10.701	8	.219
5	3.805	8	.874
6	4.328	8	.826
7	3.603	8	.891
8	10.311	8	.244
9	7.162	8	.519
10	4.107	8	.847
11	3.421	8	.905

Table 4.39 Hosmer-Lemeshow statistic in the forward stepwise logistic regression for the good friend topic in Spanish

Predictors	Change in the -2 log-likelihood
Certain	20.318
Number	16.850
You	16.755
WC	12.775
I	11.551
Other	11.546
Quote	10.989
Achieve	10.080
Friends	10.037
Sad	5.304
Inhib	4.629

Table 4.40 Changes in the -2 log-likelihood for the good friend topic in Spanish

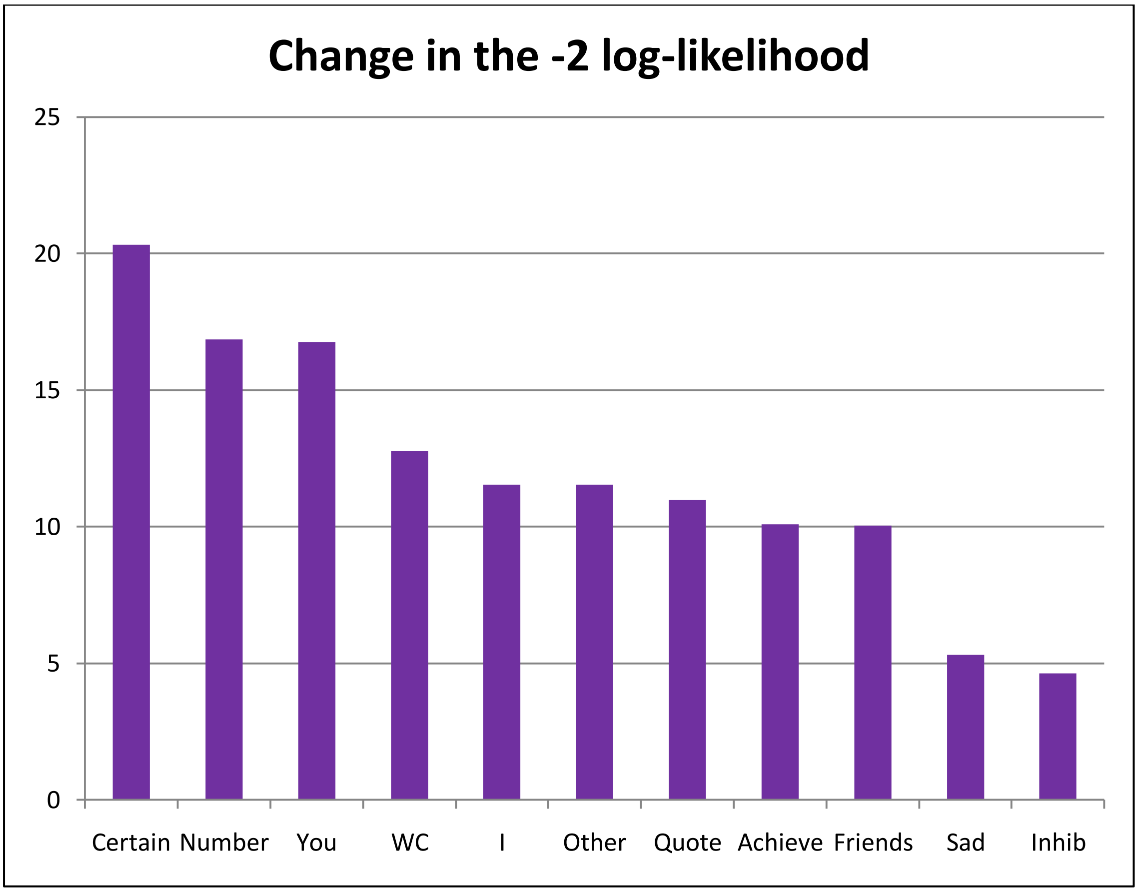

Figure 4.21 Changes in the -2 log-likelihood for the good friend topic in Spanish graph

As shown in Table 4.41, this is the first time there is a variable with a relevant change in the -2 log-likelihood which proves not significant in the last step of the equation in the forward stepwise regression: quotation punctuation (highlighted in green). Apart from this, positive values, indicative of predictors of untruthful statements, correspond to 2nd and 3rd person, and words concerning certainty and achievement, whereas text length, 1st person singular, number, terms related to sadness, inhibition, and friendship are associated in this case with truthful statements.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 11	WC	-.038	.013	9.144	1	.002	.962
	Quote	-13.620	6559.151	.000	1	.998	.000
	I	-.443	.144	9.531	1	.002	.642
	You	1.044	.339	9.476	1	.002	2.840
	Other	.323	.109	8.795	1	.003	1.381
	Number	-1.551	.485	10.230	1	.001	.212
	Sad	-1.656	.759	4.767	1	.029	.191
	Inhib	-1.943	.941	4.265	1	.039	.143
	Certain	.770	.217	12.597	1	.000	2.160
	Friends	-.748	.261	8.219	1	.004	.473
	Achieve	.940	.343	7.525	1	.006	2.561

Table 4.41 Variables in the last step of the equation in the forward stepwise logistic regression for the good friend topic in Spanish

Observed			Predicted: Selected cases: Deception No	Predicted: Selected cases: Deception Yes	Predicted: Selected cases: Percent. Correct	Predicted: Unselected cases: Deception No	Predicted: Unselected cases: Deception Yes	Predicted: Unselected cases: Percent. Correct
Step 7	Deception	No	58	9	86.6	29	4	87.9
		Yes	6	62	91.2	6	26	81.3
	Overall Percentage				88.9			84.6

Table 4.42 Classification results in the forward stepwise logistic regression for the good friend topic in Spanish

Finally, the remarkable success rates in this subcorpus are worth noting, since it is 88.9% with the original cases and 84.6% with the validation subset (see Table 4.42). In the first case, the proportion of untruthful statements correctly classified is slightly higher than the proportion of truthful ones (91.2% vs 86.6%), whereas the situation with the validation subset is exactly the opposite (81.3% vs 87.9%). These are the best success rates in both languages.

4.2.2.5. Summary of the overall results from the statistical techniques in both languages

In order to present a comprehensive picture of the effectiveness of the statistical classification methods employed, a summary of the success rates is provided in Table 4.43 and a visual representation in Figure 4.22. The experiments conducted on the good friend subcorpora, both in English and Spanish, yield the best results within the set of each respective language. This is especially so in Spanish, where there is a difference of more than 9 points with the previous subcorpus in terms of success, homosexual adoption (84.6% vs 75.4%). In English, the good friend subcorpus, despite being the best one in this language, differs notably from the homologous subcorpus in Spanish; specifically, it scores 78.5%.

Subcorpus		Overall percentage correct
English	Abortion	75.4
	Death penalty	66.2
	Good friend	78.5
	All	70.8
Spanish	Bullfighting	70.8
	Homosexual adoption	75.4
	Good friend	84.6
	All	74

Table 4.43 Classification results for all corpora

On the other hand, the worst rate in English is obtained by the death penalty, where a difference of more than 9 points exists with the next subcorpus, abortion (66.2% vs 75.4%), with the worst score in Spanish corresponding to bullfighting (70.8%). As regards the global corpora, the classification method performs better with the Spanish corpus than with the English one (74% vs 70.8%).

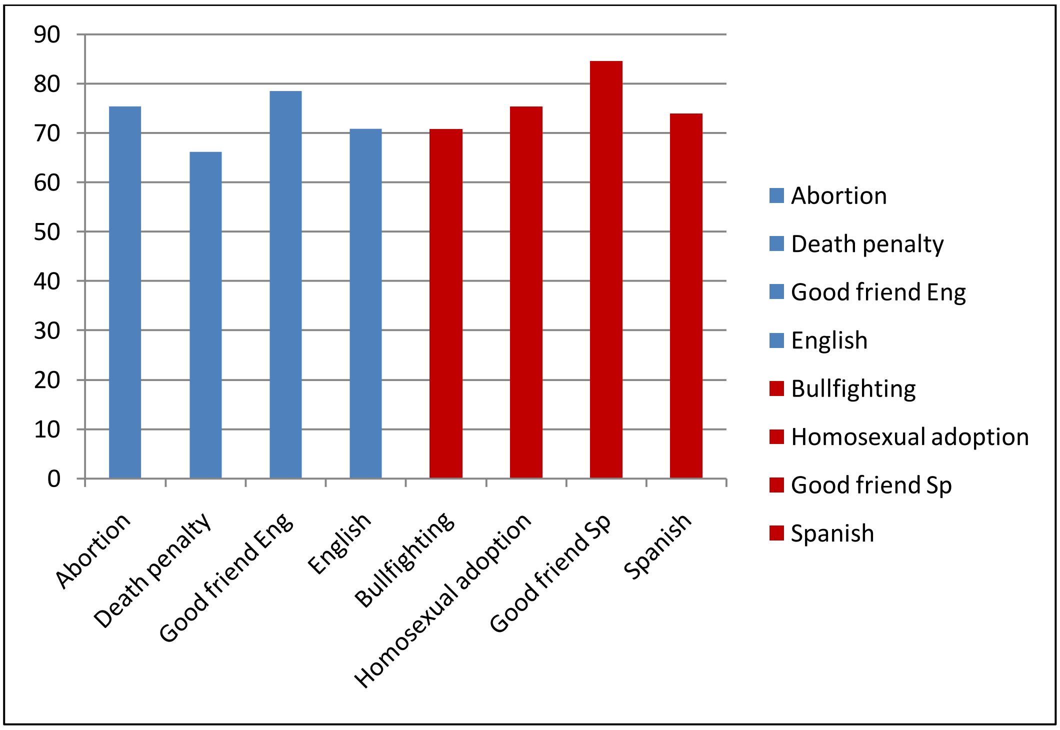

Figure 4.22 Classification results for all corpora

Furthermore, Table 4.44 shows a collection of the predictors identified for truthful –in blue– and untruthful –in orange– statements across the examined corpora. A blank cell means that the category has not proved a significant predictor.

	English				Spanish			
	Abortion	Death p.	Friend	All	Bullfight.	Homos. ad.	Friend	All
WC	True			True	True	True	True	True
1^{st} p. sing.		True		True	True	True	True	True
1^{st} p. plur.	Lie							
2^{nd} p.			Lie	Lie			Lie	Lie
3^{rd} p.	Lie		Lie	Lie			Lie	Lie
ComplexW								
Comma								
SemiC								True
Number							True	True
Anx								True
Anger	True							
Cause	Lie							
Insight	True			True				True
Sad	True						True	
Friends			True	True			True	True
Humans			Lie			Lie		Lie
Family			True					
Posfeel						True		
Certain							Lie	Lie
Achieve							Lie	
Inhib							True	
Discrep				True				
Assent								Lie
Tentat								True
Future								True
Past		True						True
Incl	True			True		Lie		
Excl		True		True	True			True
Eating								
Sexual	Lie							True
Money			Lie					
Motion						Lie		

Table 4.44 Predictors identified for truthful and untruthful statements in all corpora

4.2.3. Overall results

This section offers a visual comparison of the overall results from the classification methods. Firstly, Figures 4.23, 4.24, 4.25 and 4.26 show that ML experiments conducted on subcorpora in English show quite a successful classification of the statements into truthful or untruthful. Overall, the most discriminating dimension on its own is linguistic processes, whereas the dimension related to personal concerns yields the poorest results. Furthermore, classifiers combining several dimensions perform more satisfactorily than isolated categories, and the improvement resulting from the addition of the stylometric dimension is notable in English. As regards the Bag-of-Words model, its classification results are in all cases better than chance. When contrasting the BoW model with the previous ML experiment, the performance is comparatively better in this language. Finally, statistical classification methodologies with individual categories perform better than ML techniques with whole dimensions.

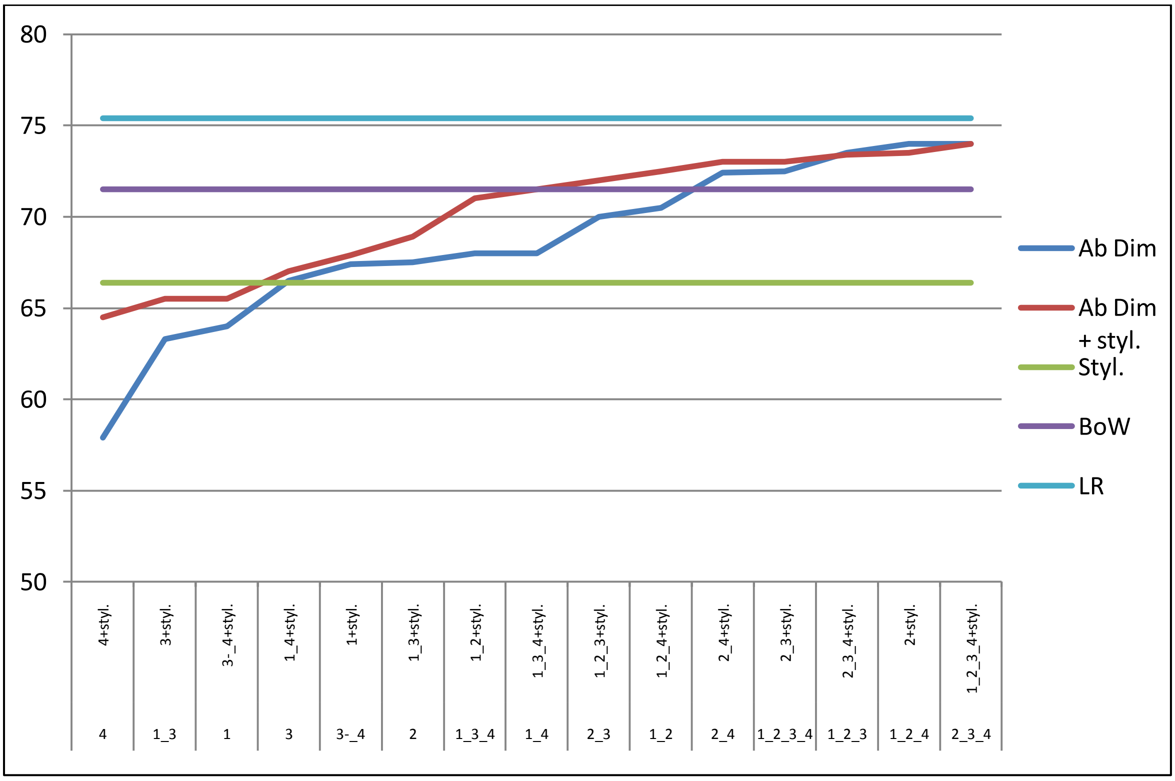

Figure 4.23 Comparison of methods in English abortion

Apart from the broad tendency displayed by general results, it is worth considering the performance of each method as applied to the different subcorpora. As commented on above, broadly speaking, a parallel may be drawn between the results from the ML experiments with whole dimensions and with the BoW model. Nonetheless, Figure 4.23 shows that this trend is not followed by the abortion subcorpus, since BoW performs strongly as compared to both the first experiment –2.5 points– and to the logistic regression model –barely 4 points–. Moreover, the ML experiment with whole dimensions shows the greatest difference with the logistic regression model in English.

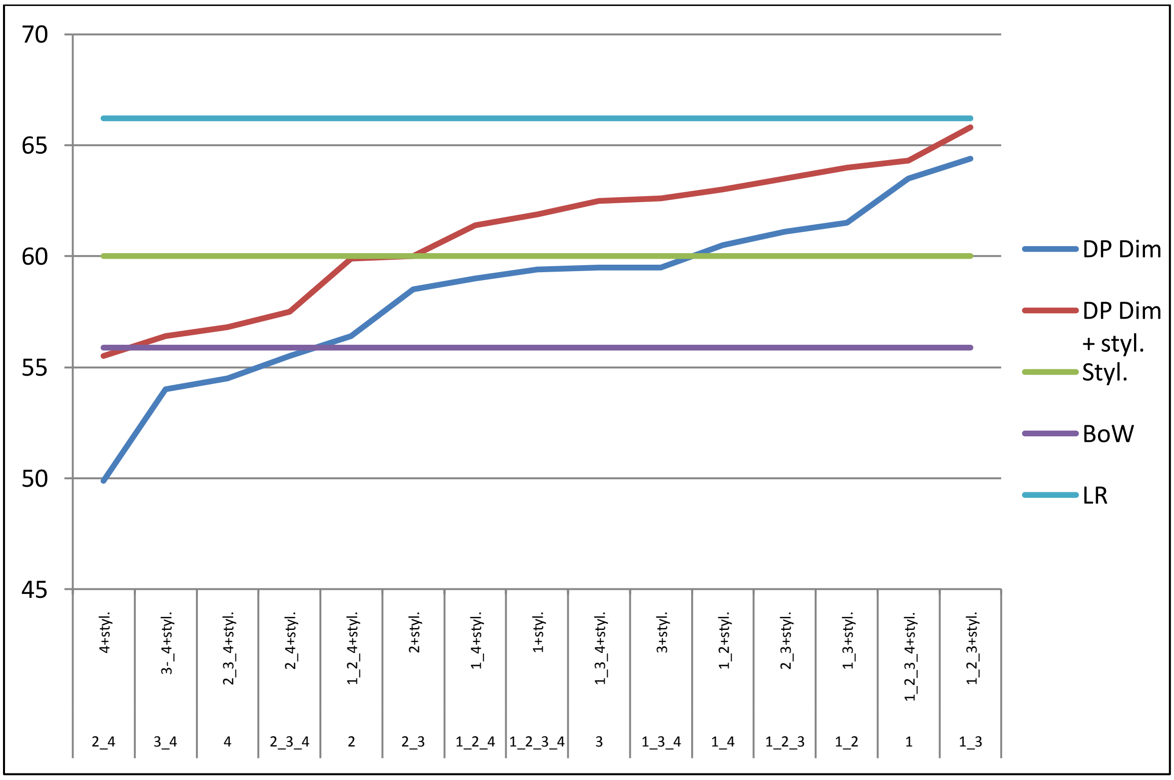

Figure 4.24 Comparison of methods in English death penalty

As can be seen in Figure 4.24, results for the death penalty subcorpus show the worst results in English. Most significantly, this is the only case in which a classification rate worse than chance occurs. As regards the BoW model, the rate is just marginally better than chance, and its performance is poor even when compared to the logistic regression model for the same topic. On the other hand, the best classification rates obtained with the ML experiment with whole dimensions and the statistical procedure are very similar –only a difference of 1.4 points.

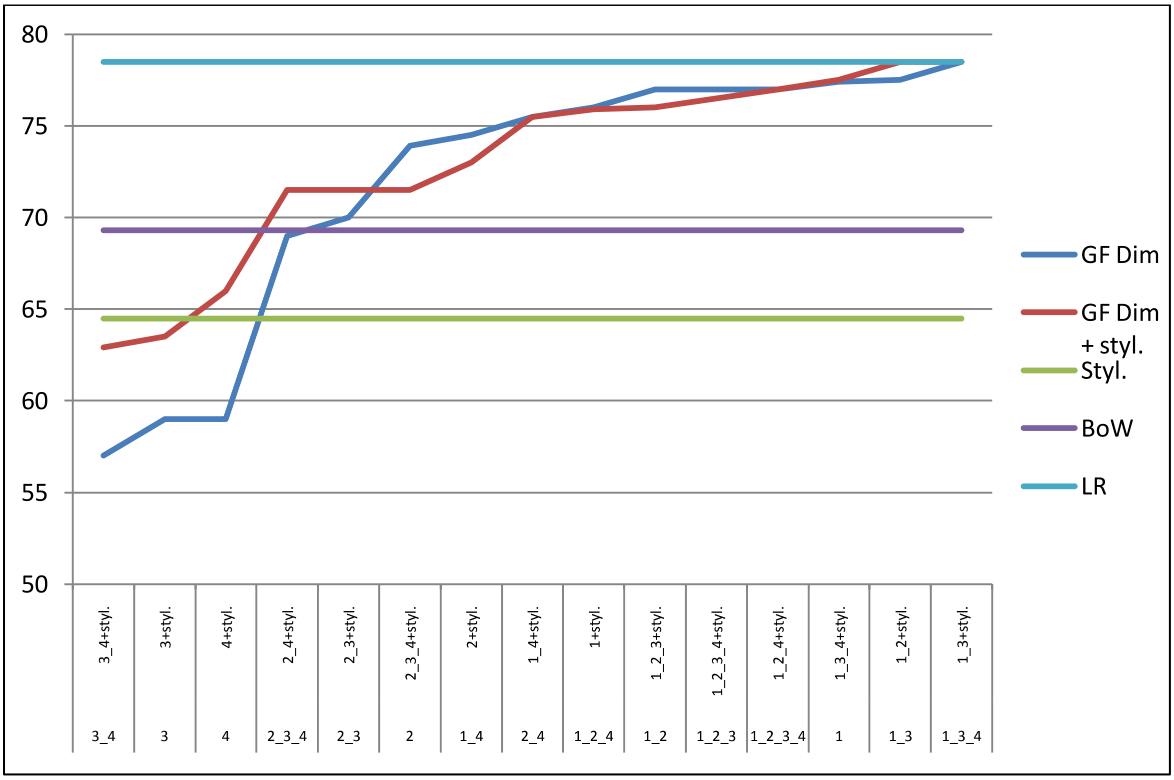

Figure 4.25 Comparison of methods in English good friend

Interestingly enough, the best ML rate for the good friend subcorpus in English equals the logistic regression model (see Figure 4.25). Nevertheless, the stylometric dimension on its own performs particularly poorly in comparison with the other methods.

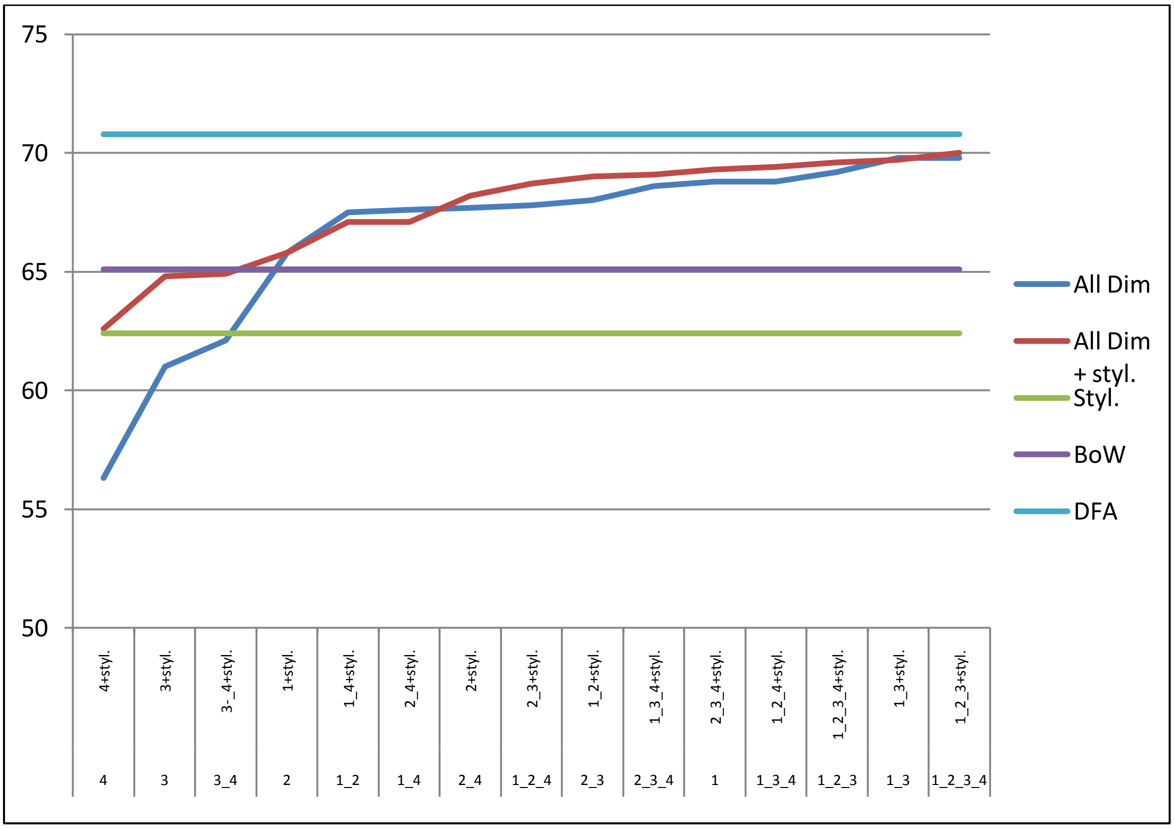

Figure 4.26 Comparison of methods in English

Similarly, Figures 4.27, 4.28, 4.29 and 4.30 provide a contrast of the results from the classification methods for Spanish. Overall, it can be observed how in this language the rates are higher than in English. The ML classifiers perform adequately and, likewise with English, classifiers combining several dimensions perform more satisfactorily than isolated categories. Although the addition of the stylometric dimension improves the success rates in some cases, the enhancement is not as obvious as in English. In this case, the Bag-of-Words model demonstrates that its classification results are in all cases better than chance, although in comparison with the previous ML experiment its performance is also poorer than in English. In addition, in all cases a parallel may be drawn between these results and the previous rates.

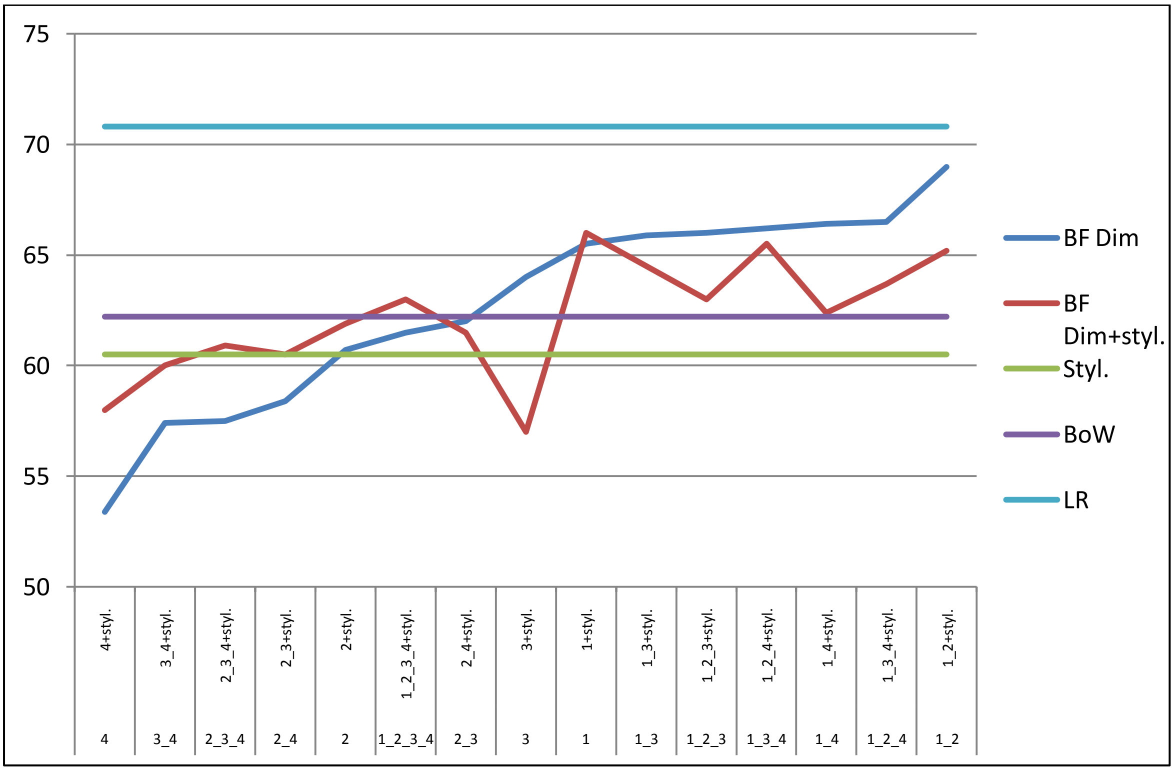

Figure 4.27 Comparison of methods in Spanish bullfighting

Concerning the bullfighting subcorpus, the ML experiment with whole dimensions shows the greatest difference with the logistic regression model in Spanish (Figure 4.27). Furthermore, the graph shows how the addition of the stylometric dimension is not only non-beneficial, but it gives the most differing results when compared to the rates obtained without that dimension.

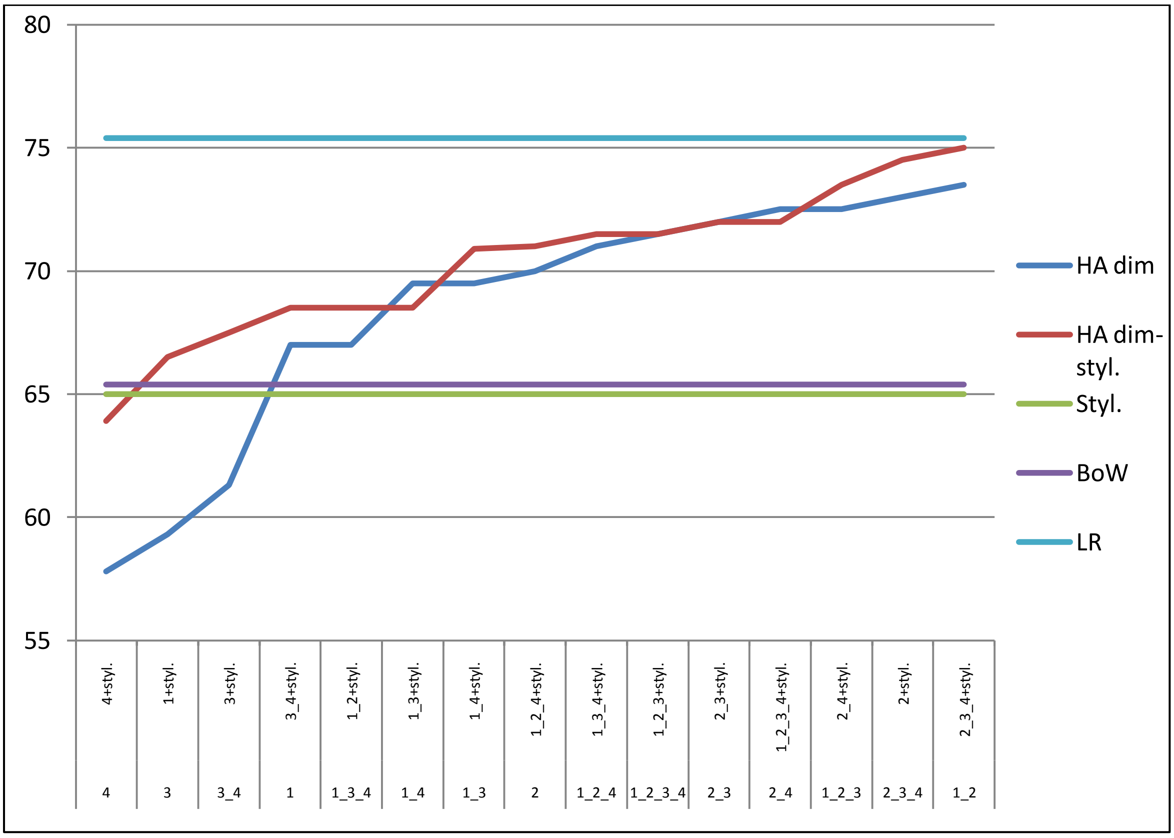

Figure 4.28 Comparison of methods in Spanish homosexual adoption

The most striking finding regarding the homosexual adoption subcorpus is the similar results obtained with the stylometric dimension on its own and the BoW model, as shown in Figure 4.28. In the other Spanish subcorpora, the latter performs better than the former by at least 1.7 points.

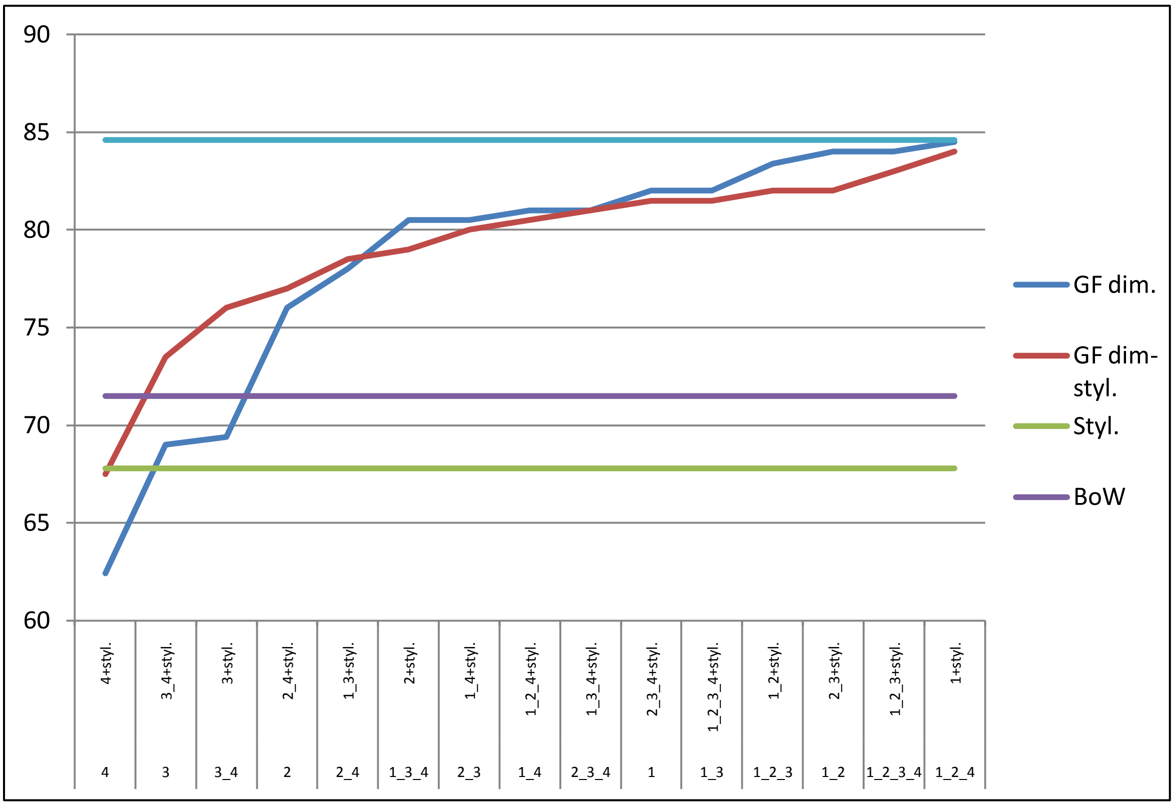

Figure 4.29 Comparison of methods in Spanish good friend

Similar to the good friend subcorpus in English, the success rates obtained for this subcorpus in Spanish are the best ones. In addition, Figure 4.29 shows a minimum difference between the ML experiment with whole dimensions and statistical techniques, whereas the distance from the rates obtained with the stylometric dimension on its own and the BoW model is the greatest one for Spanish.

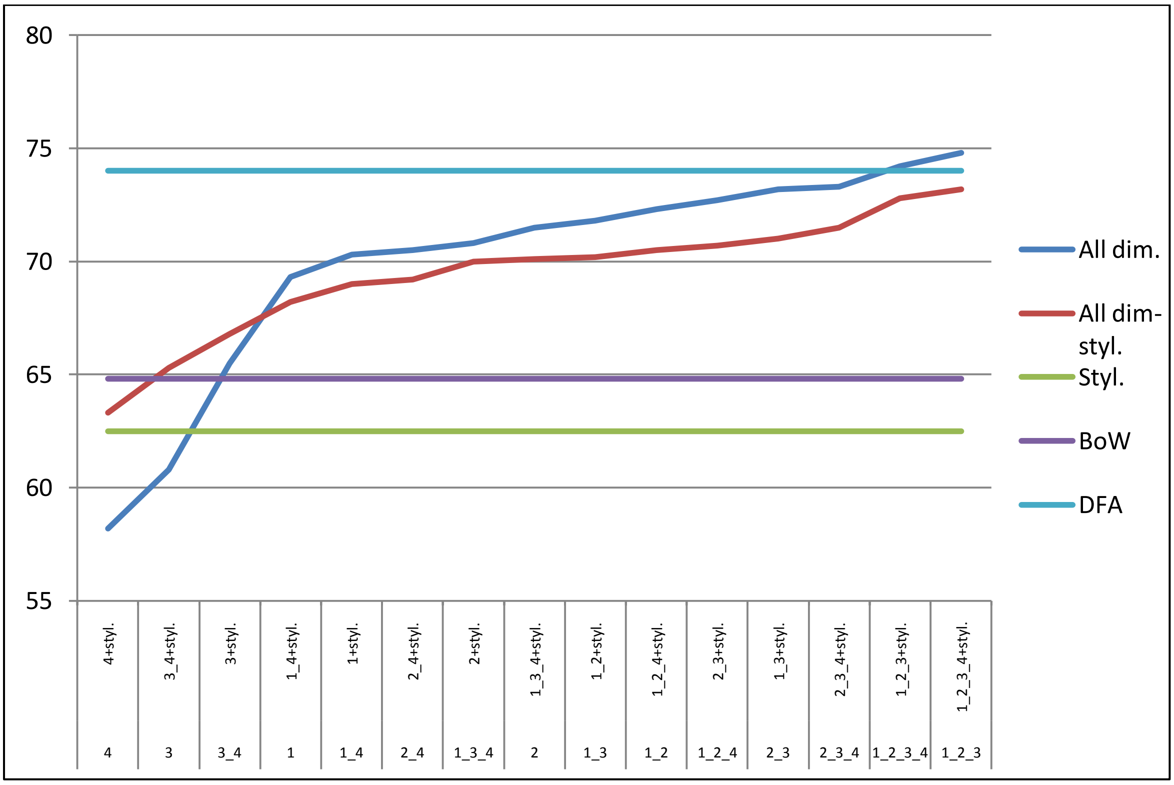

Figure 4.30 Comparison of methods in Spanish

Last but not least, the only exception to the superiority of statistical classification methodologies is found in the whole Spanish corpus, where two ML classifiers perform better (see Figure 4.30).

4.3. Summary of results

The key findings of this study are summarised as follows:

1. ML experiments conducted on subcorpora both in English and in Spanish show quite a successful classification of the statements into truthful or untruthful based on the four LIWC dimensions and the stylometric dimension proposed by the present author. Overall, the most discriminating dimension on its own is linguistic processes, whereas the dimension related to personal concerns yields the poorest results. It is worth noting that, broadly speaking, the rates in Spanish

are higher than in English. Furthermore, classifiers combining several dimensions perform more satisfactorily than isolated categories, and the improvement resulting from the addition of the stylometric dimension is notable, especially in English. These experiments also prove that classification results are largely dependent on the subject dealt with; specifically, the subject of good friend obtains the highest success rates in both languages.

2. The ML experiment with the Bag-of-Words model demonstrates that its classification results are in all cases better than chance, although in the case of the death penalty the rate is just marginally better. This subcorpus excepted, when contrasting the BoW model with the previous ML experiment, the performance is comparatively better in English than in Spanish. Furthermore, a parallel may be drawn between these results and the rates previously obtained, except for the good friend subcorpus in English, which scores slightly worse than the abortion one.

3. Overall, statistical classification methodologies with individual categories perform better than ML techniques with whole dimensions, except for the whole Spanish corpus. Furthermore, the distribution of the classification results parallels that from the experiment with the whole categories.

4. The identification of predictors has proved more successful at pinpointing categories indicative of truthful statements, the most widely shared among subcorpora being text length and 1st person singular. On the other hand, the strongest predictors for untruthfulness are 2nd and 3rd person.

4.4. Discussion of results

The findings reported above reveal that certain statistically significant differences among the independent variables under examination are associated with the likelihood of veracity or deception of the statements. Put another way, when significant differences in certain variables exist, the automatic classification into truthful and untruthful statements is more successful, both with ML techniques and with statistical methodologies. The key issue, though, is to pinpoint the most discriminating predictors of veracity and deception across subjects and languages. Thus, the results presented above are discussed further as they relate to each research question posed.

4.4.1. How successful are LIWC dimensions, the Bag-of-Words model, and the further stylometric dimensions in written deception classification in English and in Spanish?

ML classification experiments with LIWC dimensions and the further stylometric dimension are overall successful in both languages. The most discriminating dimension on its own is linguistic processes; it is natural that it should be so, bearing in mind the considerable potential of function words, which constitutes a substantial part of standard linguistic dimensions. The prime importance of these grammatical elements has been widely explored, not only in computational linguistics but also in psychology. As Chung and Pennebaker (2007: 344) have it, these words “can provide powerful insight into the human psyche”. Variations in their usage have been associated with sex, age, mental disorders such as depression, status, and deception. On the contrary, and as could be expected from previous research (Newman et al., 2003; Fornaciari and Poesio, 2011), the fourth dimension is the least discriminant on its own, probably due to its dependence on the subject matter. Broadly speaking, classifiers including more than one dimension perform better, and the improvement resulting from the addition of the stylometric dimension is notable in English, where there is more variation in parameters like mean word length, mean sentence length and STTR, although the subsequent statistical classification has revealed that their discriminant power in isolation is limited.

The Bag-of-Words model shows two sublanguages which are properly told apart in terms of lexical frequency, the death penalty subcorpus excepted. The fact that there is less distance between the rates obtained with this basic information and the experiment involving the dimensions in English may derive from the greater suitability of LIWC content in the Spanish. This finding is also supported by the larger amount of predictors included in the discriminant model –17 variables in Spanish versus 9 in English.

These experiments also prove that classification results are largely dependent on the subject dealt with; significantly enough, the subject of good friend obtains the highest success rates in both languages in the ML experiment and the statistical methodology. A plausible explanation may be that when speakers refer to a good friend, they are more likely to be emotionally involved in the experiment; they are not just giving an opinion on a topic which is alien to them, but relating their personal experience with a dear friend and lying about a person they really dislike. This personal involvement is probably reflected on the linguistic expression of deception, as suggested by Newman et al. (2003).

4.4.2. Which are the most consistent linguistic cues to written deception across the whole corpora?

The identification of predictors has proved more successful at pinpointing categories indicative of truthful statements, one of the most widely shared among subcorpora being text length. As regards differences between English and Spanish, longer paragraphs and responses in general are frequently found in the latter language (Ramírez-Esparza et al., 2007), which is confirmed in the present study. Previous research on deception detection has found that, broadly speaking, deceivers provide shorter oral responses compared to truth tellers (DePaulo et al., 2003; Hartwig et al., 2006; Vrij, 2008). This is also the case with participants in synchronous CMC, where time to plan the responses is limited, almost like in oral communication. This is explained by the fact that guilt possibly leads deceivers to avoid providing much information on fabricated facts so as not to contradict themselves (Vrij, 2008). As commented on above, the suffering of guilt is more commonly found in high-stakes deception, in which defendants or witnesses tell lies in order to protect themselves from punishment or conviction. Nonetheless, for the present research, where the object of study is sanctioned deception, the explanation advanced by DePaulo et al. (2003) is more plausible: creating and managing misinformation is more cognitively demanding than telling the plain truth.

Despite the general agreement, there are some other findings which suggest that the length of truthful and untruthful statements may be related to the mode of production. By and large, text length does not seem to translate across communication media, since deceivers are generally less verbose in oral language but produce a higher amount of words when using the written medium.

In this respect, Zhou et al. (2004b) state that several studies dealing with written language in CMC report that deceivers provide longer responses and resort to expressive language, probably in an attempt to make their lies more credible and specific by means of a great deal of detail and embellishment. Such is the case of Hancock et al. (2004 and 2005), Zhou et al. (2004c) and Zhou and Zhang (2007), who advance the explanation that providing greater detail may help persuade the addressee in a situation in which the consequence of being caught lying is not as severe as in law enforcement contexts. An interesting finding concerning this issue is reported by Anolli et al. (2002) and Picornell (2012): it is probable that a larger amount of words are found in the allegedly deceptive narratives, but truthful information may be surrounding untruthfulness. These researchers indeed find it in their corpora, and they also check whether deceivers devote a larger amount of words to truthful information. In Picornell's own words, "support is found for the hypothesis that deceivers use more words; however, there is no support for the hypothesis that deceivers use more words when lying" (2012: 125). Thus, the identification of text length as a predictor of truth in the present study is supported by previous findings, since the statements labelled as untruthful do not contain truthful information.

Concerning 1st person singular, the other most common predictor of truth in the present study, several researchers such as DePaulo et al. (2003), Mihalcea and Strapparava (2009) and Newman et al. (2003) assure that truth tellers are more comfortable with their speech and, thus, tend to identify with what they are saying by means of 1st person singular pronouns. However, in his second review of deception studies in English, Vrij (2008) finds just a weak relationship between this cue and untruthfulness. Differences in 1st person singular pronouns between general corpora in both languages had been reported by Ramírez-

Esparza et al. (2007), since their frequency in English is higher in a direct comparison due to the need to explicitly include the pronoun so as to indicate the verbal person. As Spanish verbal inflection does not require this, these pronouns were found to be less frequent by Ramírez-Esparza and her team. Nonetheless, as mentioned in Chapter 3, the present study does not involve a direct comparison between relative frequencies in one language against the other, but rather a contrast between the statistically significant differences between the sublanguages of truth and deception in English and the same phenomenon in Spanish. Specifically, in the present study 1st person singular happens to be a significant predictor of truth across all subcorpora in Spanish and in half the data sets in English. This difference may be due precisely to the low everyday usage of these pronouns in Spanish as compared to English, since a fall in the number of pronouns in untruthful statements makes for a more significant difference across all subcorpora in the former language. The only study conducted in Spanish (Masip et al., 2012) does not find a significant correlation with this feature. Nonetheless, the authors advance that the communication topic might make a difference, since their participants write about trips, which is unlikely to generate guilt, preoccupation or remorse. On the contrary, the controversial topics dealt with in the present study are more likely to arouse these feelings, despite not being a high-stakes situation.

On the other hand, a significant 3rd person orientation exists in untruthful statements in most data sets in both languages, although it is more evident in English. This is clearly in line with previous research, since 3rd person references are more frequently found in deceivers' speech (Burgoon et al., 2003; DePaulo et al., 1996; Hancock et al., 2004; 2005; 2008; Knapp et al., 1974; Knapp and Comadena, 1979; Kuiken, 1981; Vrij, 2000; Weiner and Mehrabian, 1968; Zhou et al., 2004a). This cue entails detachment from the self when providing false or imprecise information, indicating the leading role of non-immediacy in deception (DePaulo et al., 2003). However, there are some studies in which a decreased use of 1^{st} person singular is also associated with a reduction in the use of the 3^{rd} person. Such is the case of Newman et al. (2003), who ascribe it to the topic involved in the experiment; they suggest that opinions on abortion may involve the use of specific references to people instead of 3^{rd} person pronouns as a way to add credibility to their statements. Curiously enough, the opposite results have been obtained with the abortion subcorpus in the present study, thus the previous findings are difficult to justify. A similar case is that of Zhou et al. (2004b), since in their written computer messages they also found a higher amount of the 3^{rd} person. Finally, a study on oral language where a similar finding has been obtained is Bond and Lee (2005), who explain the constant presence of 3^{rd} person pronouns in truthful accounts on the grounds that their participants talk about what they had watched in a video recording. However, this should apply to all the accounts, hence the inconsistency of the explanation.

There is also a significant 2^{nd} person orientation in untruthful statements (as in Mihalcea and Strapparava, 2009), which is the only group of pronouns among which Ramírez-Esparza et al. (2007) do not find significant differences in both languages. Interestingly enough, it has proved a predictor of deception in the subcorpora of good friend and in the whole corpora, confirming the preference of deceivers for non-immediacy. In previous literature, 2^{nd} person pronouns are comprised within the group of other references (Vrij, 2000) together with the 3^{rd} person, although the latter has often proved more discriminant as far as deception detection is concerned. It is worth noting that the 2^{nd} person pronouns found in these corpora are most likely to refer to the generic or indefinite *you*, since the experiments deal with informal writing instead of oral interviews in which a direct recipient may be addressed.

A further predictor of truth indicates a cognitive process: insight. This type of cue is worth delving into, since prolonged controversy surrounds it. In this respect, Vrij (2010) explains how most researchers have not found statistically significant differences as far as cognitive processes are concerned, such as Alonso-Quecuty (1996), Bond and Lee (2005), Sporer (1997), or Strömwall and Granhag (2005). Anolli et al. (2002) explain these findings on the grounds that in low-consequence deception, the same cognitive mechanisms operate as in truth-telling. However, in other studies on oral language, cognitive verbs have proved to positively correlate with deception (Hartwig et al., 2006; Vrij et al., 2008), whereas Granhag et al. (2001), Höfer et al. (1996), Memon et al. (2010) and Vrij et al. (2000) find a stronger presence of cognitive verbs in truthful accounts. According to Sporer (2004), discrepancy may arise from the definition of cognitive processes and from the terms included in the inventory of this category. Most of these studies measure this cue as a criterion of the Reality Monitoring approach, which predicts that truth-tellers include more cognitive operations than deceivers, and have probably made a manual codification of the items comprised in the category. In this respect, Vrij (2010) acknowledges that his definition of cognitive operations is broader than others, since he rates inferences that participants make both at the time of the event and at the time of recalling the event as cognitive operations. He explains this as follows:

> [T]he sentence "She looked polite" contains a cognitive operation in our coding system, so do the sentences "She seemed quite clever" and "He looked tall for his age". Those inferences are all made at the time of recalling the event and are unlikely to be coded as cognitive operations by

> other researchers. Our broader definition of cognitive operations works well and appears to discriminate between truth tellers and liars (Vrij, 2010: 171).

Even broader is the definition included in the LIWC cognitive process insight, since it includes verbs such as *think* or its Spanish equivalent *pensar*, which produces a hit not only in the sense included in RM –to use one's mind actively– but also meaning to hold a particular opinion. With this configuration of the category, it appears to discriminate appropriately in both languages. This finding is supported by Newman et al.'s idea that "the process of creating a false story should consume cognitive resources, leading liars to tell less complex stories" (2003: 666).

As regards exclusive words (e.g. *except, without, but*), a predictor for truthful statements belonging to the dimension of relativity, there seems to be widespread agreement in previous literature on deception detection (Fuller et al., 2008; 2011; Mihalcea and Strapparava, 2009; Newman et al., 2003, to name but a few). Researchers have usually found that truth-tellers use comparatively more exclusive words, as they imply cognitive complexity and deceivers may find them difficult to manage. They imply signalling more complex explanations of what occurred or of the entertained opinions, which requires that deceivers be careful if they want to avoid being caught out in a contradiction.

Worthy of special attention is the last common predictor of truth, that is to say, the category friends. It happens to be another human-related group of words which, according to Mihalcea and Strapparava (2009), is used when the speaker is comfortable with identifying themselves with their statements. This category and social processes in general have been less widely explored than the previous cues, since they are not included in approaches such as RM or CBCA. In the present study, friends proves to be a relevant category for the whole corpora in both languages. Furthermore, as could be expected, it has also proved to be a predictor of truthful statements for the good friend subcorpora in both languages. This is an indicator that participants giving truthful accounts about real good friends are more likely to use friendship-related lexis such as *fellow, colleague* and *mate*, to name but a few.

4.4.3. Which are the relevant predictors specific to English?

There are just two predictors which have proved significant only for English: discrepancy and inclusive words. The first category, discrepancy, includes words such as *should, would, could* or *debería, podría*, and indicates cognitive processes. On the other hand, inclusive words –which come under the relativity dimension– are, for instance, *with, and*, or *include, con, y* or *incluyendo*. Both categories involve cognitive complexity and happen to be indicators of truthful language, which is in line with the findings presented in Granhag et al. (2001), Höfer et al. (1996), Memon et al. (2010) and Vrij et al. (2000), as discussed above.

The contrastive analysis on LIWC categories by Ramírez-Esparza et al. (2007) reveals a stronger presence of words indicating discrepancy and inclusiveness in English than in Spanish, which may help explain the results obtained in the present study. English speakers usually include a larger amount of these kinds of words in their speech –especially concerning discrepancy and modal verbs–, and a decrease in their usage, along with the features shared with Spanish, is indicative of the untruthfulness of their statements.

4.4.4. Which are the relevant predictors specific to Spanish?

As regards the linguistic categories relevant only for the discrimination in Spanish, numbers and assents are worth highlighting. The former includes not only cardinal and ordinal numbers, but also some expressions of quantity such as *doble (double)*, *infinidad (countless)* and *mitad (half)*. Despite the fact that Ramírez-Esparza et al. (2007) find this category more frequent in English, the present study shows a significant positive correlation with truthful language only in Spanish. Interestingly enough, previous literature on deception detection has not dealt with this feature. The stronger presence of expressions of quantity in truthful accounts may be loosely related to cognitive processes, as well as inclusive and exclusive words, since the expression of quantity requires certain specificity, which is usually absent in imagined memories and fake opinions. On the contrary, the category assents is usually more frequent in Spanish (Ramírez-Esparza et al., 2007), and this is precisely the case in the present experiment. Specifically, this is a cue to untruthfulness which, along with numbers, had not been studied in previous research. Words indicating assertion like *acceder (to agree)*, *aceptar* or *admitir (to accept)*, and *conceder (to admit)* fall within this category. They are closely related to words indicating certainty, which has also been identified as a cue to untruthfulness for Spanish. It is considered to be a cognitive process, and is conveyed in words such as *asegurar (to assure)*, *certeza (certainty)* and *siempre (always)*. Previous research also highlights this category as dominant in deceptive communication, probably due to the speaker's need to explicitly use truth-related words in an attempt to conceal the lies (Bond and Lee, 2005; Newman et al., 2003).

There is a further cognitive process which happens to be a predictor in Spanish, although in this case it is positively correlated with truth: tentative words, such as *quizá (maybe)* and *suponer (guess)*. This category is largely related to insight, the only cognitive process revealed as indicative of truth for both languages. As commented on above, these kinds of words are frequently used where the presence of the naked truth does not require certainty-related words for reassurance. However, previous studies like Adams and Jarvis (2006) and Newman et al. (2003) have found a positive correlation between tentative words and vague terms in general, and deception.

Two further predictors indicate psychological processes: anxiety and humans. The former is a subcategory of negative emotional processes which has been traditionally associated with deception. As explained in previous chapters, in law enforcement contexts, deception is usually related to the covering up of disruptive or even criminal behaviour, the uncovering of which may result in legal consequences for the deceiver. In these cases, high levels of anxiety are expected (Adams, 2001; Watson, 1981). Furthermore, in low-consequence deception, participants may also feel guilty about the topic being discussed and reflect this in language, as in DePaulo et al. (2003), Ekman (1992), Knapp and Comadena (1979), Knapp et al. (1974) and Vrij (2000). In this respect, Masip et al. (2012) do not find any correlation between negative emotion and deception, probably due to the topic involved in the experiment –participants were asked to write an account of a trip. Nonetheless, previous literature has dealt with negative emotion as a global category rather than with words concerning anxiety, usually due to the low frequency of this isolated lexical group (Newman et al. 2003). In the present study, this class of words has unexpectedly proved discriminant for truthful statements in Spanish; as has been seen, this is inconsistent with previous literature. A plausible explanation may be that the generation of anxiety by the subject matter in lower-stakes situations does not only apply to deception, but to language in general, thus it is natural that it is significantly present in truthful accounts. This supports the preliminary findings in Ali and Levine (2008). It may be that the identification of speakers with real opinions on a controversial topic in laboratory-controlled contexts leads to higher levels of anxiety in their commitment to certain ideals. On the other hand, the category humans, included in social processes, has not been widely explored in literature, although Mihalcea and Strapparava (2009) identify it as dominant in deception, advancing that it represents references to others and hence detachment from the self, as well as the use of the 2nd and 3rd person. Such is also the case in the present study.

Apart from exclusive words, which, as mentioned above, are a common predictor of truth in both languages, there are two further categories from relativity which help discriminate truthful statements in Spanish: past and future. It is worth noting that previous literature on deception detection does not tackle these cues, but they seem to provide valuable information for the discriminant model. The reason may lie with the cognitive complexity involved in the use of verbal tenses other than the present (Smith, 2005). As might be expected, in terms of frequency the most favoured tense across corpora in both languages is the present, since it is very common in the voicing of opinions and in descriptions. Precisely for this reason, accounts in which participants do not have to create fabricated stories or fake opinions seem more conducive to any deviation from the normal temporal paradigm.

The only category related to personal concerns which is discriminant in Spanish, sexuality, offers a contradiction with its role in the abortion subcorpus in English; in the former it is a predictor of truth, whereas in the latter it predicts untruthfulness. This is the only contradictory finding obtained in the analysis, and is certainly difficult to interpret. As mentioned above, personal concerns is the LIWC dimension most subject dependent (Fornaciari and Poesio, 2011; Newman et al., 2003) and their categories have usually been discarded in previous research. In this case, the subcorpora most closely related to this category are abortion and homosexual adoption. Curiously enough, the category does not prove itself a predictor for the latter topic, whereas in Spanish it is only reflected in the global corpus.

Finally, one stylometric feature has proved significant for the model in Spanish: the semicolon. Although average sentence length does not appear in any of the discriminant models, both variables are integrally related. Spanish sentences are, on average, longer than English ones (Veiga, 2008), and this is indeed the case in all the present analyses. Furthermore, as explained above, participants produced a larger amount of words when telling the truth, especially the Spanish ones, hence the discriminant power of the semicolon in this language.

4.4.5. Are there any linguistic cues to deception specific to certain topics?

First of all, the model obtained for the abortion subcorpus includes several unique predictors. Concerning the linguistic dimension, 1^{st} person plural has been identified as a predictor of untruthfulness. Despite having been usually studied as a subcategory of the total 1^{st} person (e.g. Newman et al., 2003), from a psycholinguistic perspective it falls within the category of references to others

(Pennebaker et al., 2001). Thus, it indicates detachment from the self, like the 2nd and 3rd person or humans, commented on in previous sections. As regards psychological processes, two negative emotions have also proved significant for this model: anger and sadness. This finding is in line with the discriminant function of anxiety in the whole Spanish corpus, thus the same explanation applies in this case. Although the experiment deals with low-consequence deception, abortion is a topic about which participants may feel guilty. In line with Ali and Levine (2008), it may be that the identification of speakers with their real opinions on so controversial a topic leads to higher levels of anger and sadness, successfully discriminating truthful statements. Last but not least, a cognitive process worth mentioning in this model is causation. As has been seen, the remaining cognitive processes involved in the global discriminant models had proved significant for truth, especially those involving cognitive verbs. However, in this case the category causation is a good predictor of untruthfulness, in line with Hartwig et al. (2006) and Vrij et al. (2008). This apparent contradiction confirms the idea that exploring cognitive processes separately is more useful than taking them as a whole category.

Regarding the model for the death penalty subcorpus, it does not include any unique predictors, but three cues discriminating truth which are common to either both languages –1st person singular and exclusive words– or to the Spanish corpus –past tense. It is worth noting that there is no predictor for untruthfulness. On the contrary, most predictors in the good friend subcorpus in English are significant for untruthfulness, two of them being unique: money and financial issues, and family. A plausible explanation for the weight of the former category as to the subject of good friend is the metaphorical usage of words such as *costar (to cost), deuda (debt), fortuna (fortune),* and *ganancia (profit),* derived from the metaphor FRIENDSHIP IS A VALUABLE COMMODITY, studied by Kövecses (2000). According to Gibbs (1994), metaphorical language is frequently used to avoid responsibility for the significance of what is communicated, and this may be the reason why it is significant for untruthful accounts on friendship. The other unique predictor in this model, family, belongs to social processes, and is relevant to the classification of truthful statements. The usual identification of real friends with relatives seems fairly frequent, hence the significance of the category in this subcorpus.

As far as the Spanish subcorpora are concerned, bullfighting seems to have a parallel in the death penalty subcorpus, since only three common predictors for truth have been identified and no unique cues have been found. On the contrary, two unique cues have proved relevant in the homosexual adoption model: positive feelings and motion. The former, which happens to be a positive emotional process, has proved a predictor of truth. This finding is in line with the negative emotional processes previously discussed, which were also significant for the discrimination of truthful statements. Thus, an alternative reason to that proposed by Newman et al. (2003) is suggested here. As explained above, in their study they find that negative emotion words are positively correlated with deception, but no relation is found with positive emotion. They explain the former finding on the grounds of the subject matter involved in the experiment, which may produce guilt and a sense of unease. Nevertheless, the present results reveal that all the categories related to emotion are positively correlated with truth, thus it may be that the identification of speakers with their real opinions on so controversial topics and their commitment to certain ideals lead to higher levels not only of anxiety, anger or sadness, but of emotion in general. Regarding motion words, they have proved significant for the classification of untruthful statements in this model. This is certainly in line with Newman et al., who explain it as follows: “Because liars’ stories are by definition fabricated, some of their cognitive resources are taken up by the effort of creating a believable story. Motion verbs (e.g. *walk*, *go*, *carry*) provide simple, concrete descriptions and are more readily accessible than words that focus on evaluations and judgments (e.g. *think*, *believe*)” (2003: 672).

Finally, the model for the good friend subcorpus in Spanish has identified two unique predictors for truth and untruthfulness respectively: inhibition and achievement. According to the Oxford Dictionary[17], the former entails a cognitive process by means of which there is a restraint on the direct expression of an instinct, expressed by words such as *abstenerse (refrain)*, *detener (halt)* and *desistir (desist)*. Although this specific process has not been explored as such in deception detection literature, this is in line with the findings discussed above on the discriminant power of cognitive processes. On the other hand, together with the other specific cues related to personal concerns, achievement has proved significant for the classification of untruthful statements. A close parallel may be drawn between achievement and motion, which proved successful in the classification of untruthful statements in the homosexual adoption model. Lakoff’s EVENT STRUCTURE metaphor and its submetaphor PURPOSES ARE DESTINATIONS (as cited in Kövecses, 2000: 53) is at the basis of this parallel. Thus, the same explanation advanced above applies in this case: the fabrication of a believable story or of a fake opinion is more easily performed by means of concrete descriptions and expressions to the detriment of markers of cognitive complexity.

[17] Available at http://oxforddictionaries.com/

CHAPTER 5

Final Conclusions, Limitations and Further Research

5.1. Conclusions

Overall, the key findings reveal that the classification experiments perform efficiently, with a maximum success rate of 78.5% for English and 84.6% for Spanish. The results also confirm that there is a set of linguistic cues which contributes to the statistical classification models in English and in Spanish. Curiously enough, the identification of predictors has proved more successful at identifying linguistic cues to truth instead of deception. Specifically, the shared categories are text length, self-references, insight, exclusive words, and friends. On the other hand, the other-references predictor has proved the most powerful one for deception. There are also certain cues specific to each language, and, most significantly, several discursive differences among topics, which confirms the importance of the study of deception within the specific context in which it is produced. In addition to this, the new set of stylometric features first tested in the present study has not proved significant for the individual analysis, although it has improved the performance of the SVM algorithm, especially in English. Other promising variables, like lexical richness and average word length, have not proved significant for the discrimination between both sublanguages.

These findings, as stated by Mann et al., show considerable differences in deceptive behaviour and "they challenge the simplistic view, even expressed by professional lie catchers (Ekman, 1992; Vrij, 2000), that a typical of deceptive behaviour exists" (2002: 372). This statement is closely related to the significance of idiolects for establishing a baseline in behaviour against which any subsequent finding may be contrasted. These individual differences have often been neglected in previous research on deception detection, hence the need for further insight.

To summarise, the present study has addressed the gap related to the limited exploration of the automated detection of deception in purely written texts, especially when it comes to the Spanish language. Furthermore, the significant findings obtained concerning discursive peculiarities are unprecedented in this line of research, as well as the comprehensive description of the linguistic cues to deception at a contrastive linguistics level. Thus, it is hoped that the present PhD thesis deepen human understanding of the linguistic mechanisms underlying deceit.

5.2. Limitations

It may be argued that the main limitation of the present study is precisely the very nature of the deceptive language compiled for the corpus, since it is not spontaneously produced language, which does not seem the ideal condition for the projection of the results on a real life sample of language. Participants were perfectly aware of the fact that their interlocutor knew that they were telling lies, and for this reason they were supposedly not that interested in convincing anyone of a fake truth. In addition, the original motivation behind their lying was not a real world one, like rejecting a charge. However, participants had to make sure that they were able to convince their recipient on the topics that they were lying about. Furthermore, as has been seen, criminal proceedings do not always offer a verbatim transcript of the liar's words; thus, the only available option for the study of high-stakes deception seems to be forensic written statements. In civil law jurisdictions, this is not the usual procedure, hence the insurmountable difficulty of finding this kind of material.

Furthermore, a proper control of any intervening variables in the experiment would have entailed two data sets with exactly the same topics in both languages. In the present study, just one of the subject matters is shared, namely good friend, whereas the other two topics in Spanish, bullfighting and homosexual adoption, differ from the original English corpus. The two new subjects were selected because they were highly topical and controversial, and they were considered to actively involve participants in their discussion. Nonetheless, it must be acknowledged that two suitably comparable corpora would have enabled a direct inter-language comparison, as it is the case with the English and Spanish good friend subcorpora.

5.3. Further research

The preliminary findings presented may be useful for establishing a new line in the exploration of deception in the Spanish language, which has received very limited attention, especially in the written medium, and which corresponds to the type of data most frequently found in computational contexts. In addition, this study, the first to perform a contrastive analysis in the field of deception detection, may be the starting point for further comparisons between other pairs of languages, in order to identify possible structural and lexical differences between the linguistic expression of deceit across two languages.

Interestingly enough, the analysis of outstanding keywords may reveal some characteristics not identified through a more general study based on broader dimensions. This would entail a third level of analysis hierarchically inferior to general dimensions as well as to linguistic cues; it would explore the individual constituents of these cues. This method would enable an in-depth analysis of specific contributions, even allowing for a more careful exploration of idiolectal peculiarities.

Finally, a new line of research could delve into the cognitive linguistic aspects of deception, as suggested by certain findings commented on in the present study. Since metaphorical language is often used to disclaim responsibility for the conveyed message, it would certainly be interesting to explore the extent to which it is used in deceptive communication.

REFERENCES AND BIBLIOGRAPHY

Abe, N., Okuda, J., Suzuki, M., Sasaki, H., Matsuda, T., Mori, E., Tsukada, M. and Fujii, T. (2008). Neural correlates of true memory, false memory, and deception. *Cerebral Cortex*, 13, 830–836.

Adams, S. H. (2002). *Communication under stress: indicators of veracity and deception in written narratives.* Unpublished PhD dissertation, Virginia Polytechnic Institute and State University, Falls Church, Virginia.

Adams, S. H. and Jarvis, J. P. (2006). Indicators of veracity and deception: an analysis of written statements made to police. *Speech, Language and the Law*, 13(1), 1–22.

Adenzato, M. and Bucciarelli, M. (2008). Recognition of mistakes and deceits in communicative interactions. *Journal of Pragmatics,* 40, 608–629.

Ainsworth, P. G. (1981). Incident Perception by British Police Officers. *Law and Human Behavior*, 5.

Ali, M. and Levine, T. (2008). The language of truthful and deceptive denials and confessions. *Communication Reports,* 21, 82–91.

Almela, A., Valencia-García, R. and Cantos, P. (2012). Detectando la mentira en lenguaje escrito [Detecting deception in written language]. *Procesamiento de Lenguaje Natural*, 48, 65–72.

Almela, A., Valencia-García, R. and Cantos, P. (2012). Seeing through deception: A computational approach to deceit detection in written communication. In E. Fitzpatrick, J. Bachenko and T. Fornaciari (Eds.), *Proceedings of the Workshop on Computational Approaches to Deception Detection of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 15–22.

Alonso-Quecuty, M. L. (1996). Detecting fact from fallacy in child and adult witness accounts. In G. Davies, S. Lloyd-Bostock, M. McMurran and C. Wilson (Eds.), *Psychology, Law, and Criminal Justice: International developments in research and practice* (pp. 74–80). Berlin: de Gruyter.

Alpers, G. W., Winzelberg, A. J., Classen, C., Roberts, H., Dev, P. and Koopman, C. (2005). Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior*, 21, 361–376.

Anolli, L., Balconi, M. and Ciceri, R. (2002). Deceptive Miscommunication Theory (DeMiT): A new model for the analysis of deceptive communication. In L. Anolli, R. Ciceri and G. Riva (Eds.), *Say Not to Say: New Perspectives in Miscommunication* (pp. 73–100). IOS Press.

Anolli, L. and Ciceri, R. (1997). The voice of deception: Vocal strategies of naive and able liars. *Journal of Nonverbal Behavior*, 21, 259–284.

Ansarra, R., Colwell, K., Hiscock-Anisman, C., Hines, A., Fleck, R., Cole, L. and Belarde, D. (2011). Augmenting ACID with affective details to assess credibility. *The European Journal of Psychology Applied to Legal Context*, 3(2), 141–158.

Aparicio Díaz, L. (2009, August 5). La mentira, ¿un derecho del imputado? [Lies, a defendant's right?] *El Periódico*. Retrieved from http://www.elperiodico.com/default.asp?idpublicacio_PK=46andidioma=CASandidnoticia_PK=635055andidseccio_PK=1008 [31/08/11]

Arce, R. and Fariña, F. (2005). Peritación psicológica de la credibilidad del testimonio, la huella psíquica y la simulación: el sistema de evaluación global (SEG) [Psychological report on the credibility of the testimony, psychic trace and simulation: the global assessment system]. *Papeles del Psicólogo*, 92, 59–77.

Arce, R. and Fariña, F. (2006a). Psicología del testimonio: Evaluación de la credibilidad y de la huella psíquica en el contexto penal [The psychology of testimony: Assessment of the credibility and the psychic trace in the criminal context]. In Consejo General del Poder Judicial (Ed.), *Psicología del testimonio y prueba pericial* (pp. 39–103). Madrid: Consejo General de Poder Judicial.

Arce, R. and Fariña, F. (2006b). Psicología del testimonio y evaluación cognitiva de la veracidad de testimonios y declaraciones [The psychology of testimony and the cognitive assessment of the veracity of testimonies and statements]. In J. C. Sierra, E. M. Jiménez and G. Buela-Casal (Eds.), *Psicología forense: Manual de técnicas y aplicaciones* (pp. 563–601). Madrid: Biblioteca Nueva.

Arce, R. and Fariña, F. (2007). La posible simulación de la víctima. Evaluación psicológico-forense de la credibilidad y daño psíquico mediante el sistema de evaluación global [The possible simulation of the victim. Psychological and forensic assessment of the credibility and psychic damage by means of the global assessment system]. In P. Rivas and G. L. Barrios (Eds.), *Violencia de género: Perspectiva multidisciplinar y práctica forense* (pp. 357–367). Navarra: Thomson Aranzadi.

Arciuli, J., Villar, G., Mallard, D. (2010). Lies, lies and more lies. In S. Ohlsson and R. Catambrone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2329–2334). Austin: Cognitive Science Society.

Arehart-Treichel, J. (2005, November 18). Can pathological liars blame it on the brain? *Psychiatry News, Clinical and Research News*. Retrieved from http://psychnews.psychiatryonline.org/newsarticle.aspx?articleid=109507&RelatedNewsArticles=true [15/09/11]

Babbie, E. R. (2009). *The Practice of Social Research* (12th ed.). Belmont, California: Wadsworth Publishing.

Bachenko, J., Fitzpatrick, E. and Schonwetter, M. (2008). Verification and implementation of language-based deception indicators in civil and criminal narratives. In D. Scott and H. Uszkoreit (Eds.), *Proceedings of the 22nd International Conference on Computational Linguistics* (pp. 41–48). Manchester, UK.

Bagaric, M. (1997). The Diminishing “Right” of Silence. *Sydney Law Review,* 19, 366.

Bedwell, J. S., Gallagher, S., Whitten, S. N. and Fiore, S. M. (2011). Linguistic correlates of self in deceptive oral autobiographical narratives. *Consciousness and Cognition*, 20(3), 547–555.

Bembibre, J. and Higueras, L. (2010). A vueltas con el error de Otelo: aplicación del modelo de control de fuentes a la credibilidad del testimonio y su afectación por la carga emocional [Struggling with Othello's error: Application of the model of sources control to the credibility of testimony and its affectation by the emotional load]. *Psicothema*, 22(1), 125–130.

Bensi, L., Gambetti, E., Nori, R. and Giusberti, F. (2009). Discerning truth from deception: the sincere witness profile. *The European Journal of Psychology Applied to Legal Context*, 1 (1), 101–121.

Biber D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press

Biros, D., Daly, M. and Gunsch, G. (2004). The influence of task load and automation trust on deception detection. *Group Decision and Negotiation*, 13(2), 173–189.

Bishop, J. (2009). Enhancing the understanding of genres of web-based communities: The role of the ecological cognition framework. *International Journal of Web-Based Communities*, 5(1), 4–17.

Blandón-Gitlin, I. (2005). *Detecting Deception: Content Analysis of Accounts of True and Suggested Events.* Unpublished PhD dissertation, Claremont Graduate University, CA.

Böhm, C. and Steller, M. (2008). Testimonio y trastorno límite de personalidad [Testimony and borderline personality disorder]. In F. J. Rodríguez, C. Bringas, F. Fariña, R. Arce and A. Bernardo (Eds.), *Psicología Jurídica. Entorno judicial y delincuencia*, 135–147. Oviedo: Ediciones de la Universidad de Oviedo.

Bond, C. F. and DePaulo, B. M. (2006). Accuracy of Deception Judgments. *Personality and Social Psychology Review*, 10(3), 214–234.

Bond, C. F., and Robinson, M. (1988). The evolution of deception. *Journal of Nonverbal Behavior*, 12, 295–307.

Bond, C. F. and Uysal, A. (2007). On Lie Detection "Wizards". *Law and Human Behavior*, 31(1), 109–115.

Bond, G. D. and Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology*, 19, 313–329.

Bouckaert, R. R., Frank, E., Hall, M. A., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. (2010). WEKA–experiences with a java open–source project. *Journal of Machine Learning Research*, 11, 2533–2541.

Bradley, M. T and Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal and papillary measures. *Psychophysiology*, 18, 307–315.

BRITISH NATIONAL CORPUS Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium.

Brown, J. D. and Rodgers, T. (2002). *Doing Applied Linguistics Research*. Oxford: Oxford University Press.

Brownsell, A. and Bull, R. (2011). Magistrates' beliefs concerning verbal and non-verbal behaviours as indicators of deception. *The European Journal of Psychology Applied to Legal Context*, 3(1), 29–46.

Brunet, M. K. (2009). *Why bullying victims are not believed: Differentiating between children's true and fabricated reports of stressful and non-stressful events* (Unpublished Master's thesis), University of Toronto, Toronto.

Bull, R. (1997). Entrevistas a niños testigos [Interviews to child witnesses]. In F. Fariña and R. Arce (Eds.), *Psicología e investigación judicial,* 19–38. Madrid: Fundación Universidad Empresa.

Bull, R., Cook, C., Hatcher, R., Woodhams, J., Bilby, C. and Grant, T. (2006). *Criminal Psychology: A Beginner's Guide*. Oxford: Oneworld Publications.

Buller, D. B. and Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, 6, 203–242.

Buller, D. B., Burgoon, J. K., Buslig, A. and Roiger, J. (1996). Testing interpersonal deception theory: the language of interpersonal deception. *Communication Theory*, 6, 268–289.

Burdick, F. M. (1905/2000). *The Law of Torts: A Concise Treatise on Civil Liability at Common Law and Under Modern Statutes for Actionable Wrongs to Person and Property*. Washington: BeardBooks.

Burgoon, J. K., Blair, J. P., Qin, T. and Nunamaker, J. F. (2003a). Detecting deception through linguistic analysis. *Intelligence and Security Informatics*, 2665, 91–101.

Burgoon, J. K., Buller, D. B., Blair, J. P. and Tilley, P. (2006). Sex differences in presenting and detecting deceptive messages. In D. J. Canary and K. Dindia, *Sex Differences and Similarities in Communication*. Mahwah, NJ: Taylor and Francis Group, Communication Series.

Burgoon, J. K., Buller, D. B. and Floyd, K. (2001). Does participation affect deception success? A test of the interactivity principle. *Human Communication Research*, 27, 503–534.

Burgoon, J. K., Buller, D. B., White, C., Afifi, W. and Buslig, A. (1999). The role of conversational involvement in deceptive interpersonal interactions. *Personality and Social Psychology Bulletin*, 25, 6, 669–685.

Burgoon, J. K., Buller, D. B. and Woodall, W. G. (1996). *Nonverbal communication: The unspoken dialogue.* New York, NY: The McGraw-Hill Companies, Inc.

Burgoon, J. K., Chen, F. and Twitchell, D. (2010). Deception and its detection under synchronous and asynchronous computer-mediated communication. *Group Decision and Negotiation*, 19(4), 345–366.

Burgoon, J. K. and Nunamaker, J. F. (2004). Toward computer-aided support for the detection of deception. *Group Decision and Negotiation*, 13(1), 1–4.

Burgoon, J. K. and Nunamaker, J. F. (2004). Introduction: Computer-aided support of the detection of deception. *Group Decision and Negotiation*, 13(2), 107–110.

Burgoon, J. K. and Nunamaker, J. F. (2010). Toward Computer-Aided Support for Detection of Deception. *Group Decision and Negotiation*, 19(4): 323–416.

Burgoon, J. K. and Qin, T. (2007). The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology*, 25(1): 76–96.

Burgoon, J. K., Stoner, G. M, Bonito, J. and Dunbar, N. E. (2003b). Trust and deception in mediated communication. *Proceedings of the annual meeting of the Hawaii International Conference on Computers and System Science*, Kona, HI.

Cade, W. L., Lehman, B. A. and Olney, A. (2010). An exploration of off topic conversation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 669–672). Association for Computational Linguistics.

Campbell, R. S. and Pennebaker, J. W. (2003). The secret life of pronouns: flexibility in writing style and physical health. *Psychological Science*, 14(1), 60–65.

Cantos, P. (2012). *Statistical Methods in Language and Linguistic Research*. Sheffield: Equinox.

Carlson, J. R. and George, J. F. (2004). Media appropriateness in the conduct and discovery of deceptive communication: The relative influence of richness and synchronicity. *Group Decision and Negotiation*, 13(2), 191–210.

Carlson, J. R., George, J. F., Burgoon, J. K., Adkins, M. and White, C.H. (2004). Deception in computer-mediated communication. *Group Decision and Negotiation*, 13(1), 5–28.

Carney, M. W., Chary, T. K., Robotis, P., Childs, A. (1987). Ganser syndrome and its management. *The British Journal of Psychiatry: the journal of mental science,* 151(5): 697–700.

Carofiglio, V., de Rosis, F. and Castelfranchi, C. (2001). Ascribing and Weighting Beliefs in Deceptive Information Exchanges. In M. Bauer, P. J. Gmytrasiewicz and J. Vassileva (Eds.), *UM '01 Proceedings of the 8th International Conference on User Modeling* (pp. 222–224).

Castelfranchi, C. and Poggi, I. (2002). *Bugie, finzioni, sotterfugi [Lies, fictions, subterfuges]*. Roma: Carocci.

Chau, R. T. (2011). *Linguistic markers of deception: um and like*. Unpublished MA dissertation, University of Hong Kong, China.

Chen, M. (2009). *Laser Doppler vibrometry measures of physiological function: Evaluation of biometric capabilities*. Unpublished doctoral dissertation, Washington University, Saint Louis, MO.

Chen, Y. and Lin, C. (2006). Combining SVMs with various feature selection strategies. *Studies in Fuzziness and Soft Computing*, 207, 315.

Chipere, N., Malvern, D. and Richards, B. J. (2004). Using a corpus of children's writing to test a solution to the sample size problem affecting Type-Token Ratios. In G. Aston, S. Bernardini and D. Stewart (Eds.), *Corpora and Language Learners* (pp. 139–147). Amsterdam: John Benjamins.

Chung, C. K. and Pennebaker, J. W. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social communication* (pp. 343–359). New York: Psychology Press.

Churyk, T., Lee, C. and Clinton, B. D. (2008). Can we detect fraud earlier? *Strategic Finance*, 12, 51–55.

Cody, M. J., Lee, W. S. and Chao, E. Y. (1989). Telling lies: Correlates of deception among Chinese. In J. P. Forgas and J. M. Innes (Eds.), *Recent advances in social psychology: An international perspective* (pp. 359–368). North Holland, the Netherlands: Elsevier.

Cohen, J., Cohen, P., West, S. G. and Aiken, L. S. (2002). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). London: Routledge.

Colwell, K., Hiscock, C. K. and Memon, A. (2002). Interview techniques and the assessment of statement credibility. *Applied Cognitive Psychology,* 16, 287–300.

Colwell, K., Hiscock-Anisman, C.K., Miller, J., Ansarra, R., Hines, A. and Memon, A. (2008). Reliability and validity of Reality Monitoring criteria in credibility assessment of suspects from a staged theft. Paper presented at the *2008 European Association for Psychology and Law*, Maastricht, Netherlands.

Cooper, J. E. (2008). *Using natural language processing to identify the rhetoric of deception in business and competitive intelligence email communications*. Unpublished doctoral dissertation, Robert Morris University, Illinois.

Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, 25(4), 431–447.

Dalla Barba, G. (1993). Confabulation: knowledge and recollective experience. *Cognitive Neuropsychology*, 10(1), 1–20.

Dave, K., Lawrence, S. and Pennock, D. M. (2003). Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web (WWW '03)* (pp. 519–528). ACM, New York, USA.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M. and Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70, 979–995.

DePaulo, B. M. and Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 51–70). Dordrecht, the Netherlands: Kluwer.

DePaulo, B. M., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K. and Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74–118.

Derrick, D., Meservy, T., Burgoon, J. and Nunamaker, J. (2012). An experimental agent for detecting deceit in chat-based communication. In *Proceedings of the Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium* (pp. 5–13).

DSM IV–TR: *Diagnostic and Statistical Manual of Mental Disorders* (2000), 4th ed. American Psychiatric Association.

Dike, C. C. (2008, June 1). Pathological Lying: Symptom or Disease. *Psychiatric Times*, 25(7). Retrieved from http://www.psychiatrictimes.com/display/article/10168/1162950 [25/07/11]

Dilmon, R. (2009). Between thinking and speaking. Linguistic tools for detecting a fabrication. *Journal of Pragmatics*, 41(6), 1152–1170.

Dulaney, E. F. (1982). Changes in language behavior as a function of veracity. *Human Communication Research*, 9, 75–82.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19, 61–74.

Duran, N. D., Crossley, S., Hall, C., McCarthy, P. and McNamara, D. S. (2009). Expanding a catalogue of deceptive linguistic features with NLP technologies. In *Proceedings of the 22nd Annual Conference of the International Florida Artificial Intelligence Research Society*, Sanibel Island, FL (pp. 243–248).

Duran, N. D., McCarthy, P. M., Hall, C. and McNamara, D. S. (2010). The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics*, 31, 439–462.

Dzindolet, M. T. and Pierce, L. G. (2004). A computerized text analysis can detect deception. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 635–639). Santa Monica, CA: Human Factors and Ergonomics Society.

Dzindolet, M. T. and Pierce, L. G. (2005). Using a linguistic analysis tool to detect deception. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 49 (pp. 563–567).

Ebesu, A. S. and Miller, M. D. (1994). Verbal and nonverbal behaviours as a function of deception type. *Journal of Language and Social Psychology*, 13, 418–442.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32, 407–451.

Ekman, P. (1985). *Telling lies*. New York: W. W. Norton and Company.

Ekman, P. (1992). *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (2nd ed.). New York: Norton.

Ekman, P. and Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32, 88–105.

Elkins, A. C. (2011). *Vocalic Markers of Deception and Cognitive Dissonance for Automated Emotion Detection Systems*. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Enos, F. (2009). *Detecting Deception in Speech*. Unpublished doctoral dissertation, Columbia University, NY.

Enos, F., Benus, S., Cautin, R., Graciarena, M., Hirschberg, J. and Shriberg, E. (2006). Personality Factors in Human Deception Detection: Comparing Human to Machine Performance. In *ICSLP Proceedings* (pp. 813–816). Pittsburgh, USA.

Enos, F., Shriberg, E., Graciarena, M., Hirschberg, J. and Stolcke, A. (2007). Detecting Deception Using Critical Segments. In *Interspeech Proceedings* (pp. 2281–2284). Antwerp, Belgium.

Eriksson, A. and Lacerda, F. (2007). Charlantry in forensic speech science: A problem to be taken seriously. *International Journal of Speech Language and the Law*, 14.

Etienne, M. (2005). The Ethics of Cause Lawyering: An Empirical Examination of Criminal Defense Lawyers as Cause Lawyers. *The Journal of Criminal Law and Criminology*, 95, 1195–1260.

Evans, A. D., Brunet, M. K., Talwar, V., Bala, N., Lindsay, R. C. L. and Lee, K. (2011). The effects of repetition on children's true and false reports. *Psychiatry, Psychology and Law*, 1–13.

Faigman, D. L., Fienberg, S. E. and Stern, P. C. (2003). Limits of the Polygraph. *Issues in Science and Technology*, Fall, 40-46.

Farwell, L. A. and Donchin, E. (1991). The truth will out: interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology*, 28, 531–47.

Feeley, T. H. and DeTurck, M. A. (1995). Global cue usage in behavioral lie detection. *Communication Quarterly*, 43, 420–430.

Feeley, T. H. and DeTurck, M. A. (1998). The behavioral correlates of sanctioned and unsanctioned deceptive communication. *Journal of Nonverbal Behavior*, 22, 189–204.

Feldman, M. D., Ford, C. V. and Reinhold, T. (1993). *Patient or Pretender: Inside the Strange World of Factitious Disorders*. John Wiley and Sons Inc.

Fitzpatrick, E. and Bachenko, J. (2008). Testing language-based indicators of deception on a corpus of legal narratives. Paper presented at the *2008 Conference of the American Association for Corpus Linguistics (AACL 2008)*, Brigham Young University, Provo, Utah.

Fitzpatrick, E. and Bachenko, J. (2010). Building a forensic corpus to test language-based indicators of deception. In S. T. Gries, S. Wulff and M. Davies (Eds.), *Language and Computers, special issue Corpus–linguistic Applications: Current Studies, New Directions* (pp. 183–196).

Ford, C. V. (1999). Lies! Lies! Lies! The Psychology of Deceit, Relation of Neurological Dysfunction to Deceit. *Pseudologia fantastica*, American Psychiatric Pub. Retrieved from http://www.omnimedicalsearch.com/conditions-diseases/pathological-lying-differential-diagnosis.html [20/07/2011]

Ford, C. V., King, B. K. and Hollander, M. H. (1988). Lies and liars: Psychiatric aspects of prevarication. *American Journal of Psychiatry*, 145, 554–562.

Fornaciari, T. and Poesio, M. (2011). Lexical vs. Surface Features in Deceptive Language Analysis. In A. Wyner and K. Branting (Eds.), *Proceedings of the ICAIL 2011 Workshop Applying Human Language Technology to the Law.* Retrieved from http://wyner.info/research/Papers/AHLTL2011Papers.pdf [13/07/11]

Fornaciari, T. and Poesio, M. (2012). Sincere and deceptive statements in Italian criminal proceedings. In *Proceedings of the International Association of Forensic Linguists Tenth Biennial Conference* (pp. 126–138).

Francis, M. E. and Pennebaker, J. W. (1993). *LIWC: Linguistic Inquiry and Word Count. Technical Report.* Dallas, TX: Southern Methodist University.

Frank, M. G. (1989). *Human Lie Detection Ability as a Function of the Liar's Motivation*. Unpublished doctoral dissertation, Cornell University, Ithaca.

Frank, M. G., Menasco, M. A. and O'Sullivan, M. (2008). Human behavior and deception detection. In J. G. Voeller (Ed.), *Wiley Handbook of Science and Technology for Homeland Security*. John Wiley and Sons, Inc.

Frank, M. G., Paolantinio, N., Feeley, T. H. and Servoss, T. J. (2004). Individual and small group accuracy in judging truthful and deceptive communication. *Group Decision and Negotiation*, 13(1), 45–59.

Fu, G., Brunet, M. K., Lv, Y., Ding, X., Heyman, G. D., Cameron, C.A. and Lee, K. (2010). Chinese Children's moral evaluation of lies and truths –roles of context and parental individualism-collectivism tendencies. *Infant and Child Development*, 19(5), 498–515.

Fuller. C. M. (2008). *High-stakes, Real–world Deception: An Examination of the Process of Deception and Deception Detection Using Linguistic-based Cues.* Unpublished doctoral dissertation, Oklahoma State University, OK.

Fuller, C. M., Biros, D. P., Burgoon, J. K., Adkins, M. and Twitchell, D. P. (2006). An analysis of text-based deception detection tools. In *Proceedings of the Twelfth Americas Conference on Information Systems* (p. 418).

Fuller, C. M., Biros, D. P. and Delen, D. (2008). Exploration of feature selection and advanced classification models for high-stakes deception detection. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences* (pp. 1–8).

Fuller, C. M., Biros, D. P. and Delen, D. (2011). An investigation of data and text mining methods for real world deception detection. *Expert Systems with Applications*, 38(7), 8392–8398.

Fuller, C. M., Biros, D. P. and Wilson, R. L. (2009). Decision Support for Determination of Veracity Using Linguistic–Based Cues. *Decision Support Systems*, 46(3), 695–703.

Fuller, C. M., Marett, K. and Twitchell, D. P. (2012). An Examination of Deception in Virtual Teams: Effects of Deception on Task Performance, Mutuality, and Trust. *IEEE Transactions on Professional Communication*, 55(1), 20–35.

Fuller, C. M., Twitchell, D. P. and Biros, D. P. (2009). Deception and the impact of complexity on a group decision making task. In M. Jensen, T. Meservy, J. Burgoon and J. Nunamaker (Eds.), *Proceedings of the HICSS–42 Symposium on Credibility Assessment and Information Quality in Government* (pp. 29–35).

Furner, C. P. (2008). Cultural Determinants of Media Choice for Deception. *Electronic Theses, Treatises and Dissertations*, paper 4373. Retrieved from http://diginole.lib.fsu.edu/etd/4373 [29/09/2011]

Gamer, M., Rill, H. G., Vossel, G. and Gödert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *International Journal of Psychophysiology*, 60, 76–87.

Gibbs, R. (1994). *The Poetics of Mind: Figurative Thought, Language and Understanding*. New York: Cambridge University Press.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, 46, 107–119.

Gottschalk, L.A. and Gleser, G.C. (1969). *The measurement of psychological states through the content analysis of verbal behavior.* Berkeley: University of California Press.

Graciarena, M., Shriberg, E., Stolcke, A., Enos, F., Hirschberg, J. and Kajarekar, S. (2006). Combining prosodic, lexical and cepstral systems for deceptive speech detection. In *IEEE ICASSP Proceedings*, vol. 1 (pp. 1033–1036). Toulouse, France.

Granhag, P. A., and Strömwall, L. A. (2004). *The detection of deception in forensic contexts*. Cambridge, UK: Cambridge University Press.

Granhag, P. A., Strömwall, L. A. and Jonsson, A. C. (2003). Partners in crime: How liars in collusion betray themselves. *Journal of Applied Social Psychology*, 33, 848–868.

Granhag, P. A., Strömwall, L. A. and Olsson, C. (2001). Fact or fiction? Adults' ability to assess children's veracity. Paper presented at the *11th European Conference on Psychology and Law*, Lisbon, Portugal.

Grazioli, S. (2004). Where did they go wrong? An analysis of the failure of IT–knowledgeable Internet consumers to detect deception over the Internet. *Group Decision and Negotiation*, 13(2), 149–172.

Grubin, D. (2005). Getting at the Truth about Pathological Lying. *Journal of the American Academy of Psychiatry and the Law Online*, 33(3), 350–353.

Guilford, J. P. (1954). *Psychometric Methods*. New York: McGraw–Hill.

Gupta, S. (2007). *Modelling Deception Detection in Text*. Master's thesis, School of Computing, Queen's University.

Gupta, S., Sakamoto, K. and Ortony A. (in press). Telling it like it isn't: a comprehensive approach to analyzing verbal deception. In F. Paglieri, L. Tummolini, R. Falcone and M. Miceli (Eds.), *The goals of cognition: Festschrift for Cristiano Castelfranchi.* London: College Publications.

Hall, J. A. and Knapp, M. L. (in press). The role of nonverbal communication in detecting and telling lies. In M. G. Frank and E. Svetieva (Eds.), *Nonverbal Behavior and Deception*. Berlin: Mouton de Gruyter.

Hancock, J. T., Curry, L., Goorha, S. and Woodworth, M. (2004a). Lies in conversation: an examination of deception using automated linguistic analysis. *Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Taylor and Francis Group, Psychology Press.

Hancock, J. T., Curry, L., Goorha, S. and Woodworth, M. (2005). An automated linguistic analysis of deceptive and truthful synchronous computer–mediated communication. In *Proceedings, 38th Hawaii International Conference on System Sciences* (pp. 1–10).

Hancock, J. T., Curry, L. E., Goorha, S. and Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer–mediated communication. *Discourse Processes*, 45, 1–23.

Hancock, J. T., Birnholtz, J., Bazarova, N., Guillory, J., Perlin, J. and Amos, B. (2009). Butler lies: awareness, deception and design. In *ACM CHI Proceedings* (pp. 517–526).

Hancock, J. T., Thom–Santelli, J. and Ritchie, T. (2004b). Deception and design: The impact of communication technologies on lying behavior. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2004)*, 6 (pp. 130–136). New York, ACM.

Hancock, J. T., Woodworth, M. and Goorha, S. (2010). See No Evil: The effect of communication medium and motivation on deception detection. *Group Decision and Negotiation*, 19, 327–343.

Hancock, J. T., Woodworth, M. T. and Porter, S. (2011). Hungry like the wolf: A word-pattern analysis of the language of psychopaths. *Legal and Criminological Psychology 2011*, 1–13.

Hancock, J. T., Woodworth, M. T., Morrow, R., McGillivray, H., Boochever, R. (2012). Assessing Credibility through Text: A Preliminary Analysis for Identifying Psychopathy. In *Proceedings of the Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium* (pp. 83–90).

Hare, R. D., Forth, A. E. and Hart, S. D. (1989). The psychopath as prototype for pathological lying and deception. In J. C. Yuille (Ed.), *Credibility assessment*, 25–49. Dordrecht: Kluwer.

Hartwig, M., Granhag, P.A., Stromwall, L.A. and Kronvist, O. (2006). Strategic use of evidence during police interviews: when training to detect deception works. *Law and Human Behavior*, 30, 603–619.

Hauch, V., Blandón-Gitlin, I., Masip, J. and Sporer, S. L. (2012). Linguistic Cues to Deception Assessed by Computer Programs: A Meta-Analysis. In. E. Fitzpatrick, J. Bachenko and T. Fornaciari (Eds.), *Proceedings of the Workshop on Computational Approaches to Deception Detection of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 1–4).

Hausman, K. (2003). Does Pathological Lying Warrant Inclusion in DSM? *Psychiatric News, Clinical and Research News*, 38(1): 24. American Psychiatric Association.

Healy, W. and Healy, M. T. (1969). *Pathological Lying, Accusation, and Swindling: A Study in Forensic Psychology*. Montclair, NJ: Paterson Smith.

Hirschberg, J., Benus, S., Brenier, J. M., Enos, F., Friedman, S., Gilman, S., Girand, C., Graciarena, M., Kathol, A., Michaelis, L., Pellom, B., Shriberg, E. and Stolcke, A. (2005). Distinguishing deceptive from non-deceptive speech. In *Proceedings of Interspeech 2005* (pp. 1833-1836).

Hiscock-Anisman, C. K., Colwell, K., Hazelett, G. and Morgan, C. A. (2011). Using ACID for Credibility Assessment through an Interpreter with an English Speaking Interviewer and Arabic Speaking Suspect. Paper presented at the *Biennial International Congress of Psychology and Law* in Miami, Florida.

Höfer, E., Akehurst, L. and Metzger, G. (1996). Reality Monitoring: A chance for further development of CBCA? Paper presented at the *Annual meeting of the European Association on Psychology and Law* in Siena, Italy.

Hogan, K. and Stubbs, R. (2003). Can't get Through 8 Barriers to Communication. Grenta. LA: Pelican Publishing Company.

Hollien, H. F. (1990). *The Acoustics of Crime: The New Science of Forensic Phonetics*. New York: Plenum.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). Wiley.

Humpherys, S. L. (2010). *A System of Deception and Fraud Detection Using Reliable Linguistic Cues Including Hedging, Disfluencies, and Repeated Phrases*. Unpublished PhD dissertation, University of Arizona, Tucson, AZ.

Humpherys, S. L., Moffit, K. C., Burns, M. B., Burgoon, J. K. and Felix, W. L. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50 (3), 585–594.

Jensen, M. L. (2007). *The Effects of an Expert System on Novice and Professional Decision Making with Application in Deception Detection*. Unpublished PhD dissertation, University of Arizona, Tucson, AZ.

Jensen, M. L., Bessarabova, E., Adame, B., Burgoon, J. K. and Slowik, S. (2011a). Deceptive language by innocent and guilty criminal suspects: the influence of dominance, question, and guilt on interview responses. *Journal of Language and Social Psychology,* 1–19.

Jensen, M. L., Burgoon, J. K. and Nunamaker, J. F. (2010a). Judging the Credibility of Information Gathered from Face-to-Face Interactions. *ACM Journal of Data and Information Quality*, 2(1), 1–23.

Jensen, M. L., Lowry, P. B., Burgoon, J. K. and Nunamaker, J. F. (2010b). Technology dominance in complex decision making: The case of aided credibility assessment. *Journal of Management Information Systems*, 27(1), 175–202.

Jensen, M. L., Lowry, P. B. and Jenkins, J. L. (2011). Effects of automated and participative decision support in computer–aided credibility assessment. *Journal of Management Information Systems,* 28(1), 203–236.

Jensen, M. L., Meservy, T. O., Burgoon, J. K. and Nunamaker, J. F. (2009). Automatic, Multimodal Evaluation of Human Interaction. In J. K. Burgoon and J. F. Nunamaker Jr., *Group Decision and Negotiation Special Issue: Toward Computer–Aided Support for the Detection*, 19(4), 367–389.

Joachims, T. (1998). Text categorization with support vector machines: learning with many relevant features. In *Proceedings of ECML–98* (pp. 137–142).

Kagan, S. (1998). *Normative Ethics*. Boulder: Westview Press.

Keila, P. S. and Skillicorn, D. B. (2005). *Detecting Unusual and Deceptive Communication in E–mail*. External technical report, School of Computing, Queen's University, Kingston, Ontario, Canada.

Klaver, J. R., Lee, Z., Spidel, A. and Hart, S. D. (2009). *Legal and Criminological Psychology*, 14(1), 171–182.

Kline, P. (1986). *A Handbook of Test Construction*. New York: Methuen.

Knapp, M.L. (2006). Lying and deception in close relationships. In A. Vangelisti and D. Perlman (Eds.), *Handbook of Personal Relationships* (pp. 517–532). NY: Cambridge University Press.

Knapp, M. L. (2007). *Lying and Deception in Human Interaction*. Saddle River, N.J.: Pearson Custom Publishing.

Knapp, M. L. and Comadena, M. A. (1979). Telling It Like It Isn't: A Review of Theory and Research on Deceptive Communications. *Human Communication Research*, 5, 270–285.

Knapp, M. L., Hart, R. P. and Dennis, H. S. (1974). An Exploration of Deception as a Communication Construct. *Human Communication Research*, 1, 15–29.

Kohavi, R. (1995). A study of cross–validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence,* 2(12) (pp. 1137–1143). Morgan Kaufmann, San Mateo, CA.

Kövecses, Z. (2000). *Metaphor and Emotion: Language, Culture, and Body in Human Feeling*. Cambridge: Cambridge University Press,

Koyanagi, J. and Blandón–Gitlin, I. (2011). Analysis of Children's Deception with the Linguistic Inquiry and Word Count Approach. Poster session presented at the *4th International Congress on Psychology and Law / 2011 Annual Meeting of the American Psychology-Law Society*, Miami, Florida.

Kraut, R. E. (1978). Verbal and nonverbal cues in the perception of lying. *Journal of Personality and Social Psychology*, 36, 380–391.

Kuiken, D. (1981) Non-immediate language style and inconsistency between private and expressed evaluations. *Journal of Experimental Social Psychology*, 17, 183–196.

Labov, W. (1972). *Sociolinguistic Patterns*. Oxford: Blackwell.

Leach, A. M., Lindsay, R. C. L., Koehler, R., Beaudry, J. L. and Bala, N. C. The Reliability of Lie Detection Performance. *Law and Human Behavior*, 33(1), 96–109.

Leal, S., Vrij, A., Mann, S. and Fisher, R. P. (2010). Detecting true and false opinions: The Devil's Advocate approach as a lie detection aid. *Acta Psychologica*, 134, 323–329.

Lee, C. C., Welker, R. B. and Odom, M. D. (2009). Features of computer-mediated, text-based messages that support automatable, linguistics-based indicators for deception detection. *Journal of Information Systems*, 23(1), 5–24.

Lemos, S. (2005). Simulación, engaño y mentira [Pretence, deception and lie]. *Papeles del Psicólogo*, 26, 57–58.

Leshed, G., Hancock, J. T., Cosley, D., McLeod, P. L. and Gay, G. (2007). Feedback for guiding reflection on teamwork practices. In *Proceedings of the GROUP'07 conference on supporting group work* (pp. 217–220). New York: Association for Computing Machinery Press.

Levy, L. W. (1969/1986). *Origins of the Fifth Amendment*. New York: Macmillan.

Lewis, D. D. (1998). Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval. In *Proceedings of ECML–98, 10th European Conference on Machine Learning* (pp. 1–12). Heidelberg: Springer Verlag.

Li, L. (2011). *Sex Differences in Deception Detection*. Unpublished Master's Thesis, University of Miami.

Little, A. (2007). *Defining a Signature for Deception in Court Testimony*. Master's thesis, Queen's University.

Little, A. and Skillicorn, D. B. (2008). Detecting Deception in Testimony. In *Proceedings of the IEEE International Conference on Intelligence and Security Informatics* (pp. 13–18). Taiwan.

Liu, X., Hancock, J. T., Zhang, G., Xu, R., Markowitz, D., and Bazarova, N. (2012). Exploring linguistic features for deception detection in unstructured text. In *Forty–Fifth Hawaii International Conference on System Sciences (HICSS45)* (pp. 14–23).

Mahon, J. E. (2008). The Definition of Lying and Deception. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition). Retrieved from http://plato.stanford.edu/archives/fall2008/entries/lying–definition/ [02/01/2012]

Mairesse, F., Walker, M. A., Mehl, M. R. and Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30(1), 457–500.

Mann, S., Vrij, A. and Bull, R. (2002). Suspects, lies, and videotape: An analysis of authentic high-stake liars. *Law and Human Behavior*, 26, 365–376.

Mann, S., Vrij, A. and Bull, R. (2004). Detecting true lies: Police offers' ability to detect deceit. *Journal of Applied Psychology*, 89, 137–149.

Marett, L. K. and George, J. F. (2004). Deception in the case of one sender and multiple receivers. *Group Decision and Negotiation*, 13(1), 29–44.

Masip, J. (2005). ¿Se pilla antes a un mentiroso que a un cojo? Sabiduría popular frente a conocimiento científico sobre la detección no-verbal del engaño. *Papeles del Psicólogo*, 25, 78–91.

Masip, J., Bethencourt, M., Lucas, G., Sánchez–San Segundo, M. and Herrero, C. (2012). Deception detection from written accounts. *Scandinavian Journal of Psychology*, 53, 103–111.

Masip, J., Garrido, E. and Herrero, C. (2004). Defining deception. *Anales de Psicología*, 20, 147–171.

Masip, J., Sporer, S. L., Garrido, E. and Herrero, C. (2005). The detection of deception with the Reality Monitoring Approach: a review of the empirical evidence. *Psychology, Crime and Law,* 11(1), 99–122.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M. and Fisher, P. (2007). The Hawthorne Effect: a randomised, controlled trial. *BMC Medical Research Methodology*, 7, 30.

McMenamin, G. R. (2002). *Forensic Linguistics: Advances in Forensic Stylistics*. CRC Press.

Mehrabian, A. (1971). Nonverbal Betrayal of Feeling. *Journal of Experimental Research in Personality*, 5, 64–73.

Meibauer, J. (2005). Lying and falsely implicating. *Journal of Pragmatics,* 37, 1373–1399.

Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences*, 20, 91–136.

Memon, A., Fraser, J., Colwel, K., Odinot, G. and Mastroberardino, S. (2010). Distinguishing truthful from invented accounts using Reality Monitoring criteria. *Legal and Criminological Psychology*, 15, 177–194.

Menard, S. (1995). *Applied Logistic Regression Analysis. Sage Publications Series: Quantitative Applications in the Social Sciences*, 07/106. Thousand Oaks: Sage.

Meservy, T. (2007). *Augmenting Human Intellect: Automatic Recognition of Nonverbal Behavior with Application in Deception Detection*. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Meservy, T. O., Jensen, M. L., Kruse, J., Burgoon, J. K. and Nunamaker, J. F. (2005). Automated extraction of deceptive behavioral cues from video. In P. Kantor (Ed.), *Intelligence and Security Informatics*, 198–208. Berlin: Springer.

Mihalcea, R. and Strapparava, C. (2009). The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In *Proceedings of the Association for Computational Linguistics, ACL-IJCNLP 2009* (pp. 309–312), Singapore.

Miller, G. R. and Stiff, J. B. (1993). *Deceptive Communication*. Newbury Park, CA: Sage.

Mitkov, R. (2003). *Oxford Handbook of Computational Linguistics*. Oxford University Press.

Morgan, C. A., Colwell, K. and Hazlett, G. A. (2011). Efficacy of Forensic Statement Analysis in Distinguishing Truthful from Deceptive Eyewitness Accounts of Highly Stressful Events. *Journal of Forensic Sciences*, 56(5), 1227–34.

Morgan, C. A., Rabinowitz, Y., Christian, J., Hazlett, G. A. (2009). Detecting deception in Vietnamese: efficacy of forensic statement analysis when interviewing via an interpreter. *Journal of Intelligence Community Research and Development*, 1, 1–22.

Newman, M., Pennebaker, J., Berry, D. and Richards, J. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 665–675.

Norusis, M. (2004). *SPSS 13.0 Guide to Data Analysis*. Upper Saddle-River, N.J.: Prentice Hall, Inc.

Orn, M. T., Holland, C. H. (1968). On the ecological validity of laboratory deceptions. *International Journal of Psychiatry*, 6, 282–293.

Osman, M., Channon, S. and Fitzpatrick, S. (2009). Does the truth interfere with our ability to deceive? *Psychonomic Bulletin and Review*, 16(5), 901–906.

Ott, M., Cardie, C. and Hancock, J. T. (2012). Estimating the prevalence of deception in online review communities. In *Proceedings of WWW 2012 Fraud and Bias in User Ratings* (pp. 201–210), Lyon, France.

Ott, M., Choi, Y., Cardie, C. and Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of ACL* (pp. 309–319).

Pal, M. (2009). Kernel Methods in Remote Sensing: A Review. *ISH Journal of Hydraulic Engineering*, 15(1), 194–215.

Parker, A. D. and Brown, J. (2000). Detection of deception: Statement Validity Analysis as a means of determining truthfulness or falsity of rape allegations. *Legal and Criminological Psychology*, 5, 237–259.

Pavlidis, I. and Levine, J. (2002). Thermal Image Analysis for Polygraph Testing. *IEEE Engineering in Medicine and Biology*, November/December, 56–64.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science*, 8, 162–166.

Pennebaker, J. W. and Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion*, 10, 601–626.

Pennebaker, J. W., Mayne, T., and Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology*, 72, 863–871.

Pennebaker, J. W. and King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77, 1296–1312.

Pennebaker, J. W., and Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science*, 10, 90–93.

Pennebaker, J. W., Francis, M. E. and Booth, R. J. (2001). *Linguistic Inquiry and Word Count.* Mahwah, NJ: Erlbaum Publishers.

Pennebaker J. W., Mehl M. R. and Niederhoffer K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547–77.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A. L. and Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net.

Pezdek, K., Morrow, A., Blandon–Gitlin, I., Goodman, G. S., Quas, J. A., Saywitz, K. J., Bidrose, S., Pipe, M. E., Rogers, M. and Brodie, L. (2004). Detecting deception in children: event familiarity affects criterion-based content analysis ratings. *Journal of Applied Psychology*, 89, 119–126.

Piantadosi, J., Howlett, P. and Boland, J. (2007). Matching the grade correlation coefficient using a copula with maximum disorder. *Journal of Industrial and Management Optimization*, 3 (2), 305–312.

Picornell, I. (2003). It is not what we say but how we say it. *The White Paper*, Association of Certified Fraud Examiners in Austin, Texas (May/June issue). Retrieved from http://www.qed.info/whitepaper.html [15/05/11]

Picornell, I. (2009). A Strategy for Deception Detection. Paper presented at the *International Association of Forensic Linguists Ninth Biennial Conference*, Amsterdam, The Netherlands.

Picornell, I. (2012a). *Cues to Deception in a Textual Narrative Context: Lying in Written Witness Statements.* Unpublished PhD dissertation, Aston University, Birmigham (UK).

Picornell, I. (2012b). The Rake's Progress: Mapping deception in written witness statements. *Proceedings of the International Association of Forensic Linguists Tenth Biennial Conference*, 153–168.

Porter, S. and ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology*, 15, 57–75.

Porter, S., ten Brinke, L. and Wilson, K. (2009). Crime profiles and conditional release performance of psychopathic and non–psychopathic sexual offenders. *Legal and Criminological Psychology*, 14, 109–118.

Porter, S. and Woodworth, M. (2007). I'm sorry I did it . . . but he started it: A comparison of the official and self–reported homicide descriptions of psychopath and non–psychopaths. *Law and Human Behavior*, 31, 91–107.

Provost, J. (1999). Naive-Bayes vs. rule-learning in classification of e–mail. *Technical Report AI–TR–99–284*, University of Texas at Austin, Artificial Intelligence Lab.

Qin, T. and Burgoon, J. (2007). An investigation of heuristics of human judgment in detecting deception and potential implications in countering social engineering. In *Proceedings of the International Conference on Intelligence and Security Informatics* (pp. 152–159).

Qin, T., Burgoon, J. K. and Nunamaker, J. F. (2004). An exploratory study on promising cues in deception detection and application of decision tree. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences* (pp. 23–32).

Qin, T., Burgoon, J. K., Blair, J. P. and Nunamaker, J. F. (2005). Modality effects in deception detection and applications in automatic deception-detection. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (p. 23).

Ramírez-Esparza, N., Pennebaker, J. W. and García, F.A. (2007). La psicología del uso de las palabras: Un programa de computadora que analiza textos en español [The psychology of word use: A computer program that analyzes texts in Spanish]. *Revista Mexicana de Psicología*, 24, 85–99.

Rätsch, G. (2004). A brief introduction into Machine Learning. In *Proceedings of 21st Chaos Communication Congress* (pp. 1–6).

Rayson, P., Berridge, D. and Francis, B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. Paper presented at the *7th International Conference on Statistical Analysis of Textual Data*, Louvain–la–neuve, Belgium.

Reinhard, M. A., Dickhäuser, O., Marksteiner, T. and Sporer, S. L. (2011). The case of Pinocchio: teachers' ability to detect deception. *Social Psychology of Education*, 14, 299–318.

Reynolds, L., Gillette, S., Marder, J., Miles, Z., Vodenski, P., Weintraub, A., Birnholtz, J. and Hancock, J. T. (2011). Contact stratification and deception: blackberry messenger versus SMS use among students. In *Proceedings of the ACM conference on Computer–Supported Cooperative Work, CSCW 2011* (pp. 221–224).

Rice, J. (1995). *Mathematical Statistics and Data Analysis* (2nd ed.). Boston: Duxbury Press.

Riggio, R. E. and Friedman, H. S. (1983). Individual differences and cues to deception. *Journal of Personality and Social Psychology*, 45, 899–915.

Rodríguez, J. I. (1996). Deceptive communication from collectivistic and individualistic perspectives. *Intercultural Communication Studies*, VI(2), 1996-7.

Rosenberg, S. D. and Tucker, G. J. (1978). Verbal behavior and schizophrenia: The semantic dimension. *Archives of General Psychiatry*, 36, 1331–1337.

Rowe, K. and Blandón-Gitlin, I. (2008). Discriminating true, suggested, and fabricated statements with the Linguistic Inquiry and Word Count approach. Poster session presented at the *Annual Meeting of the American Psychology–Law Society*, Jacksonville, Florida.

Rubin, V. L. (2010). On deception and deception detection: Content analysis of computer-mediated stated beliefs. In *Proceedings of the American Society for Information Science and Technology*, 47 (pp. 1–10).

Rubin, V. L. and Conroy, N. (2012). Discerning truth from deception: Human judgments and automation efforts. *First Monday*, 17, 3.

Ruby, C. L. and Brigham, J. C. (1994). *The Criteria-Based Content Analysis and its Utility in Distinguishing between Truthful and Fabricated Criminal Allegations: A Critical Review.* DTIC report. Department of the Air Force – Florida State University.

Rude, S. S., Gortner, E. M. and Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18, 1121–1133.

Rushdi-Saleh, M., Martín-Valdivia, M. T., Montejo, A. and Ureña, L. A. (2011). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications*, 38(12), 14799–14804.

Savitsky, K. and Gilovitch, T. (2003). The illusion of transparency and the alleviation of speech anxiety. *Journal of Experimental Social Psychology*, 39, 618–625.

Schafer, J. R. (2007). *Grammatical Differences between Truthful and Deceptive Written Narratives.* Unpublished doctoral dissertation, Fielding Graduate University, Santa Barbara, CA

Schelleman-Offermans, K. and Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling*, 7, 247–260.

Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston/New York: Houghton Mifflin Company.

Shuy, R. (1998). *The language of confession, interrogation, and deception*. Thousand Oaks/London/Delhi: Sage.

Skillicorn, D. (2005). Beyond keyword filtering for message and conversation detection. In *Proceedings of IEEE International Conference on Intelligence and Security Informatics, ISI2005* (pp. 231–243).

Skillicorn, D. and Little, A. (2010). Patterns of Word Use for Deception in Testimony. In C. C. Yang et al. (Eds.), *Security Informatics, Annals of Information Systems*, 9, (pp. 25–39).

Smith, C. A. P., Johnston, J. and Paris, C. (2004). Decision support for air warfare: Detection of deceptive threats. *Group Decision and Negotiation*, 13(2), 129–148.

Smith, C. S. (2005). Time with and without Tense. Paper presented at *The International Round Table on Tense and Modality*, Paris.

Smith, N. (2001). *Reading between the lines: An evaluation of the Scientific Content Analysis Technique (SCAN). Police research series paper 135.* London: UK Home Office, Research, Development and Statistics Directorate.

Snow, R., O'Connor, B., Jurafsky, D. and Ng, A. (2008). Cheap and fast – but is it good? Evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1–10).

Spence, S. A., Hunter, M. D., Farrow, T. F., Green, R. D., Leung, D. H., Hughes, C. J. and Ganesan, V. (2004). A cognitive neurobiological account of deception: Evidence from functional neuro-imaging. *Philosophical Transactions of the Royal Society of London*, 359, 1755–1762.

Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, 11, 373–397.

Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P. A. Granhag and L. Stromwall (Eds.), *Deception Detection in Forensic Contexts* (pp. 64–102). Cambridge: Cambridge University Press.

Sporer, S. L. and Sharman, S. J. (2006). Should I believe this? Reality Monitoring of accounts of self-experiences and invented recent and distant autobiographical events. *Applied Cognitive Psychology*, 27, 76–105.

Sporer, S. L. and Schwandt, B. (2006a). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, 20, 421–446.

Sporer, S. L. and Schwandt, B. (2006b). Moderators of nonverbal indicators of deception: A meta–analytic synthesis. *Psychology, Public Policy, and Law,* 13, 1–34.

Stech, F., Heckman, K., Hilliard, P. and Ballo, J. R. (2011). Scientometrics of deception, counter-deception, and deception detection in cyber-space. *PsychNology Journal*, 9(2), 79–122.

Steller, M. and Köhnken, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), Psychological methods in criminal investigation and evidence (pp. 217-245). New York: Springer-Verlag.

Stiff, J. B. and Miller, G. R. (1986). “Come to think of it...”: Interrogative probes, deceptive communication, and deception detection. *Human Communication Research*, 12, 339–357.

Stiles, W.B. (1992). *Describing talk: A taxonomy of verbal response modes*. Newbury Park, CA: Sage.

Stirman, S. W. and Pennebaker, J. W. (2001). Word use in the poetry of suicidal and non–suicidal poets. *Psychosomatic Medicine*, 63, 517–522.

Strader, J. K. (2002). *Understanding White Collar Crime*. Newaark, NJ: LexisNexis.

Strömwall, L. and Granhag, A. (2005). Children's repeated lies and truths: effects on adults' judgements and reality monitoring scores. *Psychiatry, Psychology and Law*, 12(2), 345–356.

Suckle-Nelson, J. A., Colwell, K., Hiscock-Anisman, C., Florence, S., Youschak, K. E. and Duarte, A. (2010). Assessment Criteria Indicative of Deception (ACID): Replication and Gender Differences. *The Open Criminology Journal*, 3, 23–30.

Suengas, A. G. and Johnson, M. K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General*, 117(4), 377–389.

Talbot, S. D. (2011). *Evaluating the Ability of Law Enforcement and Mental Health Professionals to Detect Deception.* Unpublished PhD dissertation, Northcentral University, Arizona, Tucson, AZ.

Tausczik, Y. and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24–54.

Taylor, R. and Gozna, L.F. (2010). *Deception: A Young Person's Life Skill?* Psychology Press.

Toma, C. L. and Hancock, J. T. (2010a). Looks and lies: the role of physical attractiveness in online dating self-presentation and deception. *Communication Research*, 37, 335–351.

Toma, C. L. and Hancock, J. T. (2010b). Reading between the lines: Linguistic cues to deception in online dating profiles. In *Proceedings of the ACM conference on Computer–Supported Cooperative Work, CSCW 2010* (pp. 5–8).

Toma, C. L. and Hancock, J. T. (2012). What lies beneath: The linguistic traces of deception in online dating profiles. *Journal of Communication*, 62, 78–97.

Trankell, A. (1972). *Reliability of evidence*. Stockholm, Sweden: Beckmans.

Twitchell, D. P. (2005). *Automated Analysis Techniques for Online Conversations with Application in Deception Detection.* Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Twitchell, D. P., Adkins, M., Nunamaker, J. F. and Burgoon, J. K. (2004). Using Speech Act Theory to Model Conversations for Automated Classification and Retrieval. Paper presented at the *Ninth International Working Conference on the Language Action Perspective on Communication Modeling*, New Brunswick, NJ.

Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen [Veracity assessment of statements]. In U. Undeutsch (Ed.), *Handbuch der Psychologie*: *Forensische Psychologie*, 11 (pp. 26–181). Gottingen: Hogrefe.

Undeutsch, U. (1989). The Development of Statement Reality Analysis. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 101–119). The Netherlands: Kluwer Academic Publishers, NATO ASI Series.

Üstün, B., Melssen, W. J., Buydens, L. M. C. (2006). Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. *Chemometrics and Intelligent Laboratory Systems*, 81, 29–40.

Van Swol, L. M. and Braun, M. (2012). Evidence for the Pinocchio Effect: Linguistic Differences between Lies, Deception by Omissions, and Truths. *Discourse Processes*, 49, 79–106.

Van Swol, L. M., Malhotra, D. and Braun, M. (2011). Deception and its detection: Effects of monetary incentives and personal relationship history. *Communication Research*, 39, 217–238.

Veiga Díaz, M. T. (2008). El inglés como vehículo de la ciencia: influencia sobre la redacción y traducción de textos científicos. In L. Pegenaute, J. Decesaris, M. Tricás, and E. Bernal (Eds.), *Actas del III Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación. La traducción del futuro: mediación lingüística y cultural en el siglo XXI* (pp. 1–10). Barcelona: PPU.

Vilariño, M. (2010). *¿Es posible discriminar declaraciones reales de imaginadas y huella psíquica real de simulada en casos de violencia de género? [Is it possible to discriminate between authentic and false statements and between real and feigned psychic signs in cases of spousal abuse?].* Santiago de Compostela: Servizo de Publicacións e Intercambio Científico de la Universidade de Santiago de Compostela.

Vrij, A. (2010). *Detecting lies and deceit: Pitfalls and opportunities* (2^{nd} ed.) Chischester: John Wiley and Sons.

Vrij, A. and Semin, G. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behaviour*, 20, 65–80.

Vrij, A., Edward, K., Roberts, K.P. and Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behaviour*, 24, 239–264.

Vrij, A. and Mann, S. (2001). Telling and detecting lies in a high–stake situation: The case of a convicted murderer. *Applied Cognitive Psychology*, 15, 187–203.

Vrij, A. and Mann, S. (2004). Detecting deception: The benefit of looking at a combination of behavioral, auditory and speech content related cues in a systematic manner. *Group Decision and Negotiation*, 13(1), 61–79.

Vrij, A., Mann, S., Fisher, R., Leal, S., Milne, R., and Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, 32(3), 253–265.

Vrij, A., Akehurst, L., Brown, L. and Mann, S. (2006). Detecting lies in young children, adolescents and adults. *Applied Cognitive Psychology*, 20, 1225–1237.

Vrij, A., Akehurst, L., Soukara, S. and Bull, R. (2004a). Detecting deceit via analyses of verbal and nonverbal behaviour in children and adults. *Human communication research*, 30, 8–41.

Vrij, A., Fisher, R., Mann, S. and Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, 10(4), 141–142.

Vrij, A., Mann, S., Kristen, S. and Fisher, R.P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, 31(5), 499–518.

Vrij, A., Granhag, P. A. and Porter, S. (2011). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, 11, 89–121.

Walters, S. B. (2000). *The Truth about Lying: How to Spot a Lie and Protect Yourself from Deception*. Naperville: Sourcebooks, Inc.

Wang, G., Chen, H. and Atabakhsh, H. (2004). Criminal identity deception and deception detection in law enforcement. *Group Decision and Negotiation*, 13(2), 111–127.

Watson, K. W. (1981). *Oral and Written Linguistic Indices of Deception during Employment Interviews.* Unpublished doctoral dissertation, Louisiana State University. Dissertations Abstracts International, 42, 06A 2367.

Weinstein, N. and Hodgins, H. S. (2009). The moderating role of autonomy and control on the benefits of written emotion expression. *Personality and Social Psychology Bulletin*, 35(3), 351–364.

Weiner, M. and Mehrabian, A. (1968). *Language within Language: Immediacy, a channel in verbal communication.* New York: Appleton Century Crofts.

Zhou, L. (2005). An Empirical Investigation of Deception Behavior in Instant Messaging. *IEEE Transactions on Professional Communication*. 48(2), 147–160.

Zhou, L., Burgoon, J. and Twitchell, D. (2003a). A longitudinal analysis of language behavior of deception in e-mail. In *Proceedings of the 1st NSFNIJ conference on Intelligence and Security Informatics* (pp. 102-110). Berlin: Springer Verlag.

Zhou, L., Burgoon, J., Nunamaker, J. and Twitchell, D. (2004a). Automating linguistics–based cues for detecting deception in text–based asynchronous computer–mediated communications. *Group Decision and Negotiation*, 13(2), 81–106.

Zhou, L., Burgoon, J. K., Twitchell, D. P., Qin, T. and Nunamaker, Jay F. (2004b). A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems,* 20(4), 139–165.

Zhou, L., Burgoon, J. K., Zhang, D. and Nunamaker, J. F. (2004c). Language dominance in interpersonal deception in computer-mediated communication. *Computers in Human Behaviour*, 20, 381–402.

Zhou, L., Shi, Y. and Zhang, D. (2004). Catch Deceptive Messages Online. In *Proceedings of the 14th Workshop on Information Technology and Systems, WITS 2004* (pp. 230–235).

Zhou, L., Shi, Y. and Zhang, D. (2008). A Statistical Language Modeling Approach to Online Deception Detection. *IEEE Transactions on Knowledge and Data Engineering*, 20(8), 1077–1081.

Zhou, L., Twitchell, D., Qin, T., Burgoon, J. K. and Nunamaker, J. F. (2003b). An Exploratory Study into Deception Detection in Text-Based Computer-Mediated Communication. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003* (pp. 1–10).

Zhou, L. and Zenebe, A. (2008). Representation and reasoning under uncertainty in deception detection: A neuro–fuzzy approach. *IEEE Transactions on Fuzzy Systems*, 16(2), 442–454.

Zhou, L. and Zhang, D. (2006). A comparison of deception behavior in dyadic and triadic group decision making in synchronous computer–mediated communication. *Small Group Research*, 37(2), 140–164.

Zhou, L. and Zhang, D. (2007). Typing or messaging: Modality effect on deception detection in computer-mediated communication. *Decision Support Systems*, 44, 188–201.

Zipf, G. K. (1935). *The Psycho–Biology of Language*. Cambridge: M.I.T. Press.

Zuckerman, M., DePaulo, B. M. and Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, vol. 14, 1–59. New York, NY: Academic Press.

APPENDIXES[18]

1. *Appendix I:* Data set in Spanish

2. *Appendix II:* Chapter 3

    2.1. *Appendix II.1:* Analysis with LIWC and further stylometric dimensions

    2.2. *Appendix II.2:* ML experiment with dimensions

    2.3. *Appendix II.3:* ML experiment with BoW model

    2.4. *Appendix II.4:* Discriminant function analysis

    2.5. *Appendix II.5:* Kolmogorov-Smirnov test

    2.6. *Appendix II.6:* Binary logistic regression

[18] All Appendixes are contained in the enclosed CD-ROM.