

# Language as Evidence

Victoria Guillén-Nieto • Dieter Stein  
Editors

# Language as Evidence

Doing Forensic Linguistics

palgrave  
macmillan

*Editors*

Victoria Guillén-Nieto  
Departamento de Filología Inglesa  
University of Alicante  
Alicante, Spain

Dieter Stein  
Anglistik III Englische Sprachwissenschaft  
Heinrich Heine University Düsseldorf  
Düsseldorf, Germany

ISBN 978-3-030-84329-8      ISBN 978-3-030-84330-4 (eBook)  
<https://doi.org/10.1007/978-3-030-84330-4>

© The Editor(s) (if applicable) and The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cover illustration: Zoonar GmbH / Alamy Stock Photo

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG. The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

This volume aims to respond to calls for a survey of modern approaches to forensic linguistics, narrowly construed here as using linguistic science in an applied field with an extraordinary number of different types of challenges. It also tries to position forensic linguistics as a subtype of and with the claim to theoretical status and a theory of its own as part of larger forensic science.

Just as other scientific fields of enquiry, forensic linguistics has internal discussions about subjects, approaches, theories, methodologies and ethics. Because of this wide ambit of issues and the limited scope of a manageable volume, a selection had to be made that hopefully represents not a monoculture but a representative variety of approaches. The fact that a subject is not represented does not mean that it is less significant in itself. It only means that the editors have, at this point, made a choice that was deemed necessary.

The volume would not have seen the light of the day had it not been for several peoples' idealistic and unremunerated efforts. We are deeply indebted to Ingo Plag, Andrew Hammel, Donato Mancini and Borja

Navarro, who read and commented in detail on chapters of this volume. We are very grateful to Julika Weber for her thorough copy editing work. Special thanks also go to Anna Stein and Rebeca Ferrero for their assistance in technical translation.

Alicante, Spain  
Düsseldorf, Germany  
June 2021

Victoria Guillén-Nieto  
Dieter Stein

# Contents

- 1 Introduction: Theory and Practice in Forensic Linguistics** 1  
*Victoria Guillén-Nieto and Dieter Stein*
- 2 Serving Science and Serving Justice: Ethical Issues Faced  
by Forensic Linguists in Their Role as Expert Witnesses** 35  
*Janet Ainsworth*
- 3 Linguistic Expert Evidence in the Common Law** 55  
*Andrew Hammel*
- 4 Expert Evidence in Civil Law Systems** 85  
*Mercedes Fernández-López*
- 5 Interacting with the Expert Witness: Courtroom  
Epistemics Under a Discourse Analyst's Lens** 105  
*Magdalena Szczyrbak*
- 6 A Lie or Not a Lie, That Is the Question. Trying to  
Take Arms Against a Sea of Conceptual Troubles:  
Methodological and Theoretical Issues in Linguistic  
Approaches to Lie Detection** 131  
*Martina Nicklaus and Dieter Stein*

<b>7</b>	<b>Authorship Identification</b>	185
	<i>Eilika Fobbe</i>	
<b>8</b>	<b>Automatic Authorship Investigation</b>	219
	<i>Hans van Halteren</i>	
<b>9</b>	<b>Speaker Identification</b>	257
	<i>Gea de Jong-Lendle</i>	
<b>10</b>	<b>Plagiarism Detection: Methodological Approaches</b>	321
	<i>Victoria Guillén-Nieto</i>	
<b>11</b>	<b>The Linguistic Analysis of Suicide Notes</b>	373
	<i>Monika Zaśko-Zielińska</i>	
<b>12</b>	<b>Fighting Cybercrime through Linguistic Analysis</b>	419
	<i>Patrizia Anesa</i>	
<b>13</b>	<b>Linguistic Approaches to the Analysis of Online Terrorist Threats</b>	439
	<i>Julien Longhi</i>	
	<b>Index</b>	<b>461</b>

## Notes on Contributors

**Janet Ainsworth** is Professor of Law at Seattle University in the USA. Her research interests lie at the intersection of law, language and culture. She is President of the International Association of Forensic Linguists.

**Patrizia Anesa** is a researcher in English Language and Linguistics at the University of Bergamo in Italy and a member of the Research Centre on Specialised Languages (CERLIS). Her research interests lie mostly in the area of specialised discourse, with special reference to legal language. She also cooperates with several international organisations as a consultant in discourse, conversation and frame analysis.

**Gea de Jong-Lendle** is Senior Scientist and Lecturer in the Phonetics Arbeitsgruppe at the Philipps-University of Marburg in Germany. She holds a PhD in Linguistics from the University of Florida and an M.Phil. in Computer Speech and Language Processing from the University of Cambridge. Her research interests focus on the area of forensic phonetics and perceptive phonetics. As the Director of Forensic Research Associates, she has undertaken forensic investigations since 1994 for both prosecution and defence.

**Mercedes Fernández-López** is Senior Lecturer at the Commercial and Procedural Law Department at the University of Alicante. Her main line of research is evidence law (in civil and criminal law), on which she has published two monographs and numerous articles. Since 2008, Fernández has been a reserve judge at the Provincial Court of Alicante. She directs the Master's Degree in Law at the same University and the School of Legal Practice of Alicante.

**Eilika Fobbe** is Senior Scientist in Forensic Linguistics at the Bundeskriminalamt in Germany. After studying Indo-European linguistics, German and Sanskrit philology at the University of Göttingen, she received her doctorate in linguistics and worked as a postdoctoral fellow in the universities of Göttingen and Greifswald. In 2012, she published an introductory book on forensic linguistics that has become an important reference in Germany. Fobbe has also worked as an expert linguist for law firms and courts of justice.

**Victoria Guillén-Nieto** is Senior Lecturer in Applied Linguistics at the University of Alicante in Spain. She has published articles on trademarks linguistics, plagiarism detection and language crimes such as defamation, harassment and hate speech. Since 2009, she has served as an expert linguist in Spain, Switzerland, Sweden and the USA. From September 2019 to September 2021, she was President of the International Language and Law Association (ILLA) for Linguistics.

**Andrew Hammel** is a lawyer, defence counsel, writer and translator living in Düsseldorf, Germany. From 2003 to 2016 he taught Anglo-American Common Law and Comparative Law at the Heinrich Heine University Düsseldorf.

**Julien Longhi** is a Full Professor of Linguistics at the CY Cergy Paris Université in Paris and a Junior Member at the Institut Universitaire de France (IUF). He has published books, articles and edited volumes in semantics, pragmatics, discourse analysis and corpus linguistics. Longhi is working on two major projects: one investigating ideology detection in Twitter and the other looking at risk and security discourses in collaboration with security forces in France.

**Martina Nicklaus** is Senior Lecturer in the Department of Romance Languages at the Heinrich Heine University Düsseldorf, where she also received her PhD in Linguistics (Romance Languages). Her research interests lie in veracity assessment.

**Dieter Alfred Stein** is Emeritus Chair of English Linguistics at Heinrich Heine University Düsseldorf. One main focus of his work is language development; others deal with pragmatics, open access publishing, the linguistics of the Internet, language in the legal domain and forensic linguistics. In addition to teaching at his home university, he has taught at several foreign universities, including China, UCLA Los Angeles law school and the linguistic institute of the Linguistic Society of America.

**Magdalena Szczyrbak** is Assistant Professor in the Institute of English Studies at Jagiellonian University in Krakow. Her research interests lie primarily in discourse analysis and corpus-assisted discourse study applied to legal discourse and, particularly, to the study of stance and evaluation. She is the current President of the International Language and Law Association (ILLA) for Linguistics.

**Hans van Halteren** is Assistant Professor at the Centre for Language Studies (CLS), Radboud University Nijmegen. His research interests lie in corpus linguistics, computational linguistics and machine learning. Two of his focus areas are language variation and forensic linguistics, which are two aspects of the same natural language processing task.

**Monika Zaśko-Zielińska** is Associate Professor in the Institute of Polish Philology at the University of Wrocław. -Monika Zaśko-Zielińska received her PhD in genre studies and habilitation on forensic linguistics at the University of Wrocław. She has contributed to Polish Emo WordNet.

# List of Figures

Fig. 1.1	Language and/in the law: disciplines	3
Fig. 7.1	The original extortion letter	197
Fig. 7.2	English Translation of the original extortion letter	197
Fig. 7.3	The first anonymous e-mail	203
Fig. 7.4	The English translation	204
Fig. 8.1	Histograms for the frequency counts of ‘the’ and ‘suddenly’ in various subsets of text samples	227
Fig. 9.1	Voice identification scores for different retention intervals based on the values reported by McGehee (1937)	271
Fig. 9.2	The <i>same</i> (creaky) male speaker reading ‘had today’ in the left recording with a rising F0 (‘uptalk’), in the right with a final fall. The speaker is SSBE-speaker nr. 37 from the DyViS database (Nolan et al., 2009)	277
Fig. 9.3	Two <i>different</i> female speakers, German students at the university of Marburg, with the same accent and a similar voice quality (left, slightly breathier towards the end) reading ‘Nordwind und Sonne’	277
Fig. 9.4	The region defined by REDE, based on the pronunciation of the words ‘stand’, ‘have’ and ‘are’ (Kehrein, 2021)	287
Fig. 9.5	Articulation rate distribution (syll./s) for 35 female German speakers (20–25y.) speaking spontaneous compared with the AR rates found for the two emergency calls and the reference recording. Calculations are based on a minimum of 15	

	Memory Stretches per person (Mean 24,4 MS) using the measuring method described in Jessen (2007). Study carried out at the University of Marburg to provide background data for a forensic case involving a 23-year woman exhibiting an extremely high articulation rate above 7 syll./s.	290
Fig. 9.6	The SSI-4 stutter frequency for 3 stutter patients in 3 different speaking conditions. The calculations were based on the Stuttering Severity Instrument for Adults and Children (SSI-4), see Riley (2009)	293
Fig. 9.7	An example of a transcript with different levels using PRAAT TextGrids	301
Fig. 10.1	Similarity threshold comparisons	350
Fig. 10.2	Shared vocabulary more than once comparisons (1)	351
Fig. 10.3	Shared vocabulary more than once comparisons (2)	352
Fig. 10.4	Hapax legomena comparisons	353
Fig. 10.5	Vocabulary that is only in one translation	353
Fig. 11.1	Scan of the handwritten suicide note to everyone (source: PCSN repository)	393
Fig. 11.2	Scan of the handwritten suicide note to the girlfriend (source: PCSN repository)	395
Fig. 11.3	Scan of the handwritten poem (source: PCSN repository)	397
Fig. 13.1	Descending hierarchical classification (themes) of the corpus	451
Fig. 13.2	Graphic grouping of texts based on their grammatical characteristics	453
Fig. 13.3	Prototype of authorship attribution model	455
Fig. 13.4	Authors connected by the analysis model	456
Fig. 13.5	Descending hierarchical classification (themes) of the second corpus	456

# List of Tables

Table 5.1	The most common clusters with <i>I, you</i> and <i>we</i>	116
Table 7.1	Error distribution	198
Table 7.2	Error distribution	204
Table 7.3	Thematic patterns of the letter's first section	206
Table 7.4	Thematic patterns of the letter's second section	207
Table 7.5	Thematic patterns of the letter's closing section	207
Table 8.1	Statistics for various feature types in BNC measurements	231
Table 8.2	Quality measurements for various systems for verification of Howard within M&B	248
Table 9.1	Description of the main tasks in forensic phonetics	260
Table 9.2	Transcript of the phone call of kidnapper Ferdi Elsas with the receptionist of the Okura Hotel played in a Documentary by Huys and Krabbé in 2019	266
Table 9.3	Speaker identification methods used over time	273
Table 9.4	An overview of the speaker characteristics analysed in the auditory-acoustic method	282
Table 9.5	A phonetic analysis of a German speaker saying the words 'stand', 'have' and 'are'	286
Table 9.6	An example of a transcription coding format	299
Table 9.7	An example of a transcript using the transcription code format described in Table 9.6	300
Table 10.1	Suspicious pair of Spanish translations of Oscar Wilde's <i>The Nightingale and the Rose</i> (1888)	343

Table 10.2	Distractor Spanish translations of Oscar Wilde's <i>The Nightingale and the Rose</i> (1888)	343
Table 10.3	Stylometric analysis	355
Table 10.4	Spanish translations comparison	357
Table 10.5	Spanish translations comparison	360
Table 10.6	Inductive probability scale	362
Table 11.1	Transcript and English translation of the handwritten suicide note to everyone in Fig. 11.1 (slashes indicate end of line in the Polish text)	394
Table 11.2	Transcript and English translation of the handwritten suicide note to the girlfriend in Fig. 11.2 (slashes indicate end of the line in the Polish text)	396
Table 11.3	Transcript and English translation of the handwritten poem in Fig. 11.3 (slashes indicate end of line in the Polish text)	397
Table 11.4	Transcript and English translation of the Polish suicide note	408
Table 12.1	Male profiles: claimed professions	430
Table 12.2	Male profiles: claimed ethnicity	431
Table 12.3	Strategic semantic fields	433
Table 13.1	Names of authors listed in the articles	454