Language and Law

Linguagem e Direito

Volume 7

ISSN 2183-3745

Language and Law
Linguagem e Direito

ISSN: 2183-3745 (online)
Volume 7, Issues 1-2, 2020

Editors / Diretores

Malcolm Coulthard & Rui Sousa-Silva

Universidade Federal de Santa Catarina, Brasil & Universidade do Porto, Portugal

Book Reviews Editors / Editores de Recensões

David Wright (English)

Nottingham Trent University, UK

Rita Faria (Português - EMEA)

Universidade do Porto, Portugal

Luciane Fröhlich (Português, Brasil)

Universidade Federal de Santa Catarina, Brasil

Cover / Capa

Rui Effe

Publisher / Editora

Faculdade de Letras da Universidade do Porto

International Editorial Board / Conselho Editorial Internacional

Janet Ainsworth, *University of Washington, USA*

Ron Butters, *Duke University, USA*

Carmen Rosa Caldas-Coulthard, *University of Birmingham, UK*

Le Cheng, *Zhejiang University, China*

Virginia Colares, *Universidade Católica de Pernambuco, Brasil*

Diana Eades, *University of New England, Australia*

Debora Figueiredo, *Universidade Federal de Santa Catarina, Brasil*

Maribel del Pozo Triviño, *Universidad de Vigo, Spain*

Ed Finegan, *University of Southern California, USA*

Tammy Gales, *Hofstra University, USA*

Núria Gavaldà, *Universitat Autónoma de Barcelona, Spain*

Maria Lúcia Gomes, *Universidade Tecnológica Federal do Paraná, Brasil*

Tim Grant, *Aston University, UK*

Patrick Juola, *Duquesne University, USA and Juola Associates*

Krzysztof Kredens, *Aston University, UK*

Iman Laversuch, *University of Cologne, Germany*

Janny Leung, *University of Hong Kong, Hong Kong*

Belinda Maia*, Universidade do Porto, Portugal*

Fernando Martins*, Universidade de Lisboa, Portugal*

Karen McAuliffe, *University of Birmingham, UK*

Frances Rock, *Cardiff University, UK*

Paolo Rosso, *Polytechnic University of Valencia, Spain*

Susan Sarcevic, *University of Rijeka, Croatia*

Roger Shuy, *Georgetown University Washington, USA*

Larry Solan, *Brooklyn Law School, USA*

Editorial Assistants / Assistentes Editoriais

Viviane Maia, *Universidade do Porto, Portugal*

Copyright / Direitos de autor

The articles published in this volume are covered by the Creative Commons "Attribution-NonCommercial" (CC-BY-NC) license (see http://creativecommons.org). They may be reproduced in its entirety as long as Language and Law / Linguagem e Direito is credited, a link to the journal's web page is provided, and no charge is imposed. The articles may not be reproduced in part or altered in form, or if a fee is changed, without the journal's permission. Copyright remains solely with individual authors. The authors should let the journal Language and Law / Linguagem e Direito know if they wish to republish.

Os artigos publicados neste volume estão cobertos pela licença Creative Commons "Attribution-NonCommercial" (CC-BY-NC) (consultar http://creativecommons.org) e podem ser reproduzidos na íntegra desde que seja feita a devida atribuição à Language and Law / Linguagem e Direito, com indicação do link para a página da revista e desde que não sejam cobradas quaisquer taxas. Os artigos não podem ser parcialmente reproduzidos, o seu formato não pode ser alterado, e não podem ser cobradas taxas sem a autorização da revista. Os direitos de autor dos trabalhos publicados nesta revista pertencem exclusivamente aos seus respetivos autores. Os autores devem informar a revista Language and Law / Linguagem e Direito se pretenderem submeter o artigo noutro a outra publicação.

Language and Law / Linguagem e Direito

Language and Law / Linguagem e Direito is a free, exclusively online peer-reviewed journal published twice a year. It is available on the website of the Faculty of Arts of the University of Porto, at http://ojs.letras.up.pt/index.php/LLLD/ .

All articles should be submitted by email to the journal email address (llldjournal@gmail.com) or via the system. See the guidelines for submission at the end of this issue.

Requests for book reviews should be sent to llldjournal@gmail.com.

Language and Law / Linguagem e Direito é uma revista gratuita publicada exclusivamente online, sujeita a revisão por pares, publicada semestralmente e disponível no website da Faculdade de Letras da Universidade do Porto, em http://ojs.letras.up.pt/index.php/LLLD/ .

Os materiais para publicação deverão ser enviados por email para o endereço da revista (llldjournal@gmail.com) ou através do sistema, e devem seguir as instruções disponíveis no final deste volume.

As propostas de recensão de livros devem ser enviadas para llldjournal@gmail.com.

Indexing and abstracting / Indexação e bases de dados bibliográficas

Language and Law / Linguagem e Direito is covered by the following abstracting and indexing services:

A *Language and Law / Linguagem e Direito* encontra-se indexada e catalogada nas seguintes bases de dados:

ERIH PLUS: European Reference Index for the Humanities and the Social Sciences
Google Scholar
Sherpa Romeo
Jornals for Free
CrossRef
Portal RCAAP
QUALIS Periódicos (A2)
JURN

PUBLISHED BIANNUALLY ONLINE / PUBLICAÇÃO SEMESTRAL ONLINE

ISSN: 2183-3745
D.O.I.: https://doi.org/10.21747/21833745/lanlaw
D.O.I. Volume 7: https://doi.org/10.21747/21833745/lanlaw/7_1_2

THE ARTICLES ARE THE SOLE RESPONSIBILITY OF THEIR AUTHORS.
THE ARTICLES WERE DOUBLE BLIND PEER REVIEWED.

OS ARTIGOS SÃO DA EXCLUSIVA RESPONSABILIDADE DOS SEUS AUTORES.
OS ARTIGOS FORAM SUBMETIDOS A ARBITRAGEM CIENTÍFICA DUPLA POR PARES.

Contents / Índice

Editor's Introduction 1
Malcolm Coulthard & Rui Sousa-Silva

Nota Introdutória 4
Rui Sousa-Silva & Malcolm Coulthard

ARTICLES / ARTIGOS

How I Got Started as a Forensic Linguist 8
Roger W. Shuy

How I Got Started in Forensic Linguistics 12
Diana Eades

How I Got Started as a Forensic Linguist 16
Ronald R. Butters

My Thwarted Start as a Forensic Linguist 20
Edward Finegan

How I Got Started as a Forensic Linguist 24
Martha Komter

My First Case 28
John Gibbons

How I Got Started 30
Janet Ainsworth

A Chance Encounter 33
Malcolm Coulthard

Exploring the Distinctiveness of Emoji Use for Digital Authorship Analysis 36
Karoline Marko

"You Have the Right to Keep Quiet": Translation Inadequacies in Nevada's Spanish Miranda Warnings 56
Ashley Rose Keaton

Os Imageboards e os seus Recursos Compartilhados: Um Estudo de Caso em Síntese de Autoria Forense 77
Luiz H. Valle-Nunes

BOOK REVIEWS / RECENSÕES

Análise Fonético-Forense em Tarefa de Comparação de Locutor 94
Recensão de Maria Lúcia de Castro Gomes

Editors' Introduction

Malcolm Coulthard & Rui Sousa-Silva

Federal University of Santa Catarina, Brazil & University of Porto, Portugal

https://doi.org/10.21747/21833745/lanlaw/7_1_2int

2020 has been an extraordinary year in many different ways. One year ago, the world was on the verge of being turned upside down by an as yet largely unknown virus. At the turn of the decade, all the hopes for the New Year were suddenly put on hold as a result of the COVID-19 pandemic; travel was virtually banned, visas were suspended, whole countries were on lockdown. We have witnessed an era of world history like no other. COVID-19 had an impact on life, education, and research, and for citizens around the world life changed in unprecedented ways as they had to adjust to new practices, which included moving their otherwise regular face-to-face activities to online, virtual environments. Academics and researchers were not immune to such changes, as most – if not all – teaching activities moved online and conferences were either cancelled or transformed into web conferences and webinars. All these changes demanded a great deal of time and effort; as teaching had to be adjusted, course modules suddenly redesigned, and teachers had to make themselves more readily available to students, indeed many had to re-learn to do their job or rather learn to do a different job.

For researchers, too, the pandemic brought new and further challenges. While, on the one hand, some live conferences have been cancelled and others moved online, on the other researchers have been 'invited' to more meetings, all of them taking place online, often with brief or no intervals in between, and have received more-than-usual invitations to present at other institutions. Of course, virtual environments are not prejudicial by themselves, rather the contrary; they have allowed people in different parts of the world to attend research events that would otherwise have been out of their reach. However, these circumstances also had a strong negative impact on publication because researchers have been left with comparatively little time to conduct research, to publish it and to review their peers' submissions. Additionally, even those who have taken the time to submit manuscripts for publication have found themselves with less free time to respond to reviewer's comments.

Inevitably, therefore, 2020 has also been a difficult year for journals, many of which have struggled to publish a sufficient number of articles, while at the same time meeting quality standards and deadlines. Paradoxically, this is perhaps the moment in time when more articles, of a higher quality, are needed. The fact that conferences and seminars

moved to virtual environments allowed more people across the world to host, attend and present at online events, thus attracting even more attention from researchers and the general public worldwide. However, a significant proportion of online talks and events have not been subject to any type of peer-review, or offered a guarantee of scientific rigour; instead, they contributed to the already identified counterproductive 'CSI effect'. Forensic Linguistics has indeed witnessed an increase in the number of researchers who are interested in the area and who choose to call themselves 'forensic linguists' even when their qualifications in the field are dubious. Such lack of qualifications is evident, for instance, in the scant attention paid to methodological rigour and accuracy.

Unsurprisingly, 2020 has also been a challenging year for *Language and Law/Linguagem e Direito.* The journal has recently been highly rated by Brazilian indexes in the fields of 'language' and 'law', so the journal attracted the interest especially of Brazilian researchers, who would otherwise submit their manuscripts elsewhere. We have therefore received an extraordinarily high volume of manuscripts, and although a significant proportion was sadly not suited to *Language and Law/Linguagem e Direito*, all manuscripts submitted were pre-edited and peer-reviewed. Busy as they are, especially under the current circumstances, peer-reviewers – whom we take this opportunity to thank for their excellent contribution – have not always been able to respond as quickly as we would have wished.

Book reviews have not been immune to the effects of COVID-19. The previous issue (v.6 n.2) included as many as four book reviews. We had hoped to continue this momentum, with many review copies received from publishers, many reviewers identified and many titles sent out for review. For example, reviews of Vogel's (2019) edited collection *Legal Linguistics Beyond Borders: Language and Law in a World of Media, Globalisation and Social Conflicts* and Murphy's (2019) *The Discursive Construction of Blame: The Language of Public Inquiries* are both in progress, nearing completion, but the pandemic has stalled progress. For other titles, reviewers have been identified but the workload of our English reviews editor has created a bottle-neck for Patrick, Schmid and Zwaan's (2019) edited collection *Language Analysis for the Determination of Origin: Current Perspectives and New Directions* and Heydon's (2019) *Researching Forensic Linguistics.* Meanwhile, there are almost a dozen titles literally quarantined in our English reviews editor's office – which he has not been allowed to visit since March – and ready to be sent out. This pile includes Scott's (2019) *Legal Translation Outsourced*, Leung's (2019) *Shallow Equality and Symbolic Jurisprudence in Multilingual Legal Orders* and Kaplan's (2020) *Linguistics and Law.* It is hoped that our books review process and schedule, both in English and in Portuguese, can return to some sort of normality, along with much else, in 2021.

Thus, we, the Editors, decided to publish this volume as a double issue. In addition to regular articles reporting ongoing research in the field and a solitary book review in Portuguese, this volume also features a collection of 'how I got started' articles. In Spring-2020 Roger Shuy wrote to the Editors of *Language and Law/Linguagem e Direito* suggesting that we commissioned a series of short pieces on the topic of 'How I got started as a forensic linguist', which he thought would be interesting to and useful for young intending forensic linguists. We approached a series of leading scholars, including all of the Presidents of the IAFL – sadly, of course, Peter Tiersma and Maite Turell are no longer with us – and we were delighted with the enthusiastic response. The pieces we have received so far are published in this volume, starting, obviously, with Roger Shuy's.

These first-person accounts will doubtless make exciting reading from a narrative point of view, but they also – and mostly – demonstrate how methodological rigour is crucial to the plethora of applications of linguistic analysis in forensic contexts. Newcomers to the field will learn at least two lessons from these accounts: (1) in order to be a good forensic linguist, one first needs to be an excellent linguist; and (2) contrary to some recent accounts stating that forensic linguistics is a great source of income, one can hardly make a living by working exclusively as a forensic linguist. Notwithstanding, as the eight accounts demonstrate, forensic linguistics remains an extraordinary field of research that is worth exploring further, and more and better research into forensic linguistics is encouraged.

We hope that you enjoy reading this double, though sadly not bumper issue as much as we enjoyed preparing it. We look forward to editing and distributing two exciting journal issues in 2021!

Malcolm Coulthard
Federal University of Santa Catarina
Brazil

Rui Sousa-Silva
University of Porto
Portugal

References

Heydon, G. (2019). *Researching Forensic Linguistics.* London & New York: Routledge.

Kaplan, J. P. (2020). *Linguistics and Law.* Abingdon and New York: Routledge.

Murphy, J. (2019). *The Discursive Construction of Blame.* London: Palgrave Macmillan UK.

P. L. Patrick, M. S. Schmid and K. Zwaan, Eds. (2019). *Language Analysis for the Determination of Origin*, volume 16 of *Language Policy.* Cham: Springer International Publishing.

Scott, J. R. (2019). *Legal Translation Outsourced.* Oxford: Oxford University Press.

F. Vogel, Ed. (2019). *Legal Linguistics Beyond Borders: Language and Law in a World of Media, Globalisation and Social Conflicts – Relaunching the International Language and Law Association (ILLA).* Berlin: Duncker and Humblot.

Nota Introdutória

Rui Sousa-Silva & Malcolm Coulthard

Universidade do Porto, Portugal & Universidade Federal de Santa Catarina, Brasil

https://doi.org/10.21747/21833745/lanlaw/7_1_2int

2020 foi um ano extraordinário de várias formas diferentes. Há um ano, o mundo estava prestes a ser virado do avesso por um vírus então ainda francamente desconhecido. Na viragem da década, todas as esperanças para o novo ano foram, de repente, colocadas em suspenso devido à pamdemia de COVID-19; as viagens foram praticamente proibidas sempre que não urgentes, a atribuição de vistos foi restringida e países inteiros foram colocados em quarentena. Em todo o mundo, presenciámos uma era da história da Humanidade como nenhuma outra. A COVID-19 teve um impacto na nossa vida, nas atividades educativas e na investigação, tendo a vida de cidadãos em todo o mundo mudado de forma jamais vista à medida que se viram forçados a ajustar-se a novas práticas, que incluíram, entre outras, a adaptação das suas atividades – que, noutras circunstâncias, decorreriam presencialmente – a ambientes virtuais, online. Investigadores/pesquisadores e pessoal académico não passou incólume a estas mudanças, uma vez que a maior parte das atividades letivas (senão todas) passaram para plataformas online, enquanto as conferências foram canceladas ou transformadas em conferências virtuais e webinars. Todas estas mudanças foram extremamente exigentes, em tempo e em esforço: com as atividades letivas a terem que ser ajustadas, as unidades curriculares a terem que ser rapidamente redesenhadas e os professores a terem que passar a ter mais disponibilidade para acompanhar os seus estudantes, muitos tiveram, de facto, que reaprender a fazer o seu trabalho ou, inclusivamente, a aprender a fazer um trabalho diferente.

Também para os investigadores/pesquisadores a pandemia trouxe desafios novos e porventura mais profundos. Se, por um lado, algumas conferências presenciais foram canceladas e outras organizadas virtualmente, por outro os investigadores/pesquisadores passaram a receber "convites" para mais compromissos, todos eles online, muitas vezes com intervalos muito curtos – ou mesmo sem intervalos – entre eles, bem como mais convites do que habitualmente para fazer apresentações organizadas por outras instituições. Naturalmente, os ambientes virtuais, por si só, não são prejudiciais, antes pelo contrário: nestas circunstâncias, permitiram a pessoas em diferentes partes do mundo participar em eventos científicos que, noutras circunstâncias, lhes seriam vedados. Contudo, estas circunstâncias também tiveram um forte impacto negativo nas atividades de

publicação, decorrente do facto de os investigadores/pesquisadores ficaram com pouco tempo, comparativamente a outros anos, para desempenhar as suas atividades de investigação/pesquisa, para publicar e para rever as submissões dos seus pares. Além disso, mesmo aqueles que tiveram oportunidade de submeter trabalhos para publicação tiveram que se confrontar com menor disponibilidade de tempo para responder aos comentários dos revisores.

Por isso, 2020 também foi, inevitavelmente, um ano difícil para as revistas científicas, muitas das quais se confrontaram com um número insuficiente de artigos, com padrões de qualidade adequados ou dentro dos prazos. Paradoxalmente, este é, talvez, o período em que á necessário mais artigos, de qualidade mais elevada. O facto de congressos e seminários terem passado para ambientes virtuais permitiu a mais pessoas em todo o mundo organizar, participar e apresentar comunicações em eventos online, atraindo assim ainda mais a atenção de investigadores/pesquisadores e do público em geral. Contudo, uma parte significativa dos eventos e palestras online não foi sujeita a qualquer tipo de revisão por pares, nem ofereceu qualquer garantia de rigor científico; antes, muitos desses eventos contribuíram para o já identificado e contraproducente "efeito CSI". A Linguística Forense foi alvo, efetivamente, de um aumento significativo do número de investigadores/pesquisadores interessados no tema que se auto-intitulam "linguistas forenses", mesmo que as suas qualificações na área sejam, muitas vezes, de natureza duvidisa. Essa falta de qualificações é evidenciada, por exemplo, pela parca atenção prestada à precisão e ao rigor metodológico.

Sem surpresas, 2020 também foi um ano de desafios para a *Language and Law/Linguagem e Direito.* A revista científica foi recentemente muito bem cotada pelo índice brasileiro QUALIS nas áreas da "linguística" e do "direito", pelo que a revista atraiu o interesse sobretudo por investigadores/pesquisadores brasileiros, que, noutras circunstâncias submeteriam os seus manuscritos a outras revistas. Por conseguinte, recebemos um volume extraordinariamente elevado de trabalhos, e, embora um número significativo de submissões, infelizmente, não fosse adequado à *Language and Law/Linguagem e Direito*, todos os manuscritos foram pré-editados e revistos por pares. Os revisores (a quem, aproveitando esta oportunidade, agradecemos publicamente pelo seu contributo), estando extrememente ocupados, sobretudo dadas as circunstâncias atuais, nem sempre conseguiram responder tão rapidamente como desejaríamos e como é apanágio da revista.

As resenhas de livros não ficaram imunes aos egeitos da COVID-19. O número anterior (v.6 n.2) incluiu quatro resenhas de livros e era nossa esperança dar continuidade a este crescendo, tendo já recebido cópias das editoras para revisão, identificado revisores e enviado vários livros para revisão. Por exemplo, as recensões de *Legal Linguistics Beyond Borders: Language and Law in a World of Media, Globalisation and Social Conflicts*, editado por Vogel (2019), e *The Discursive Construction of Blame: The Language of Public Inquiries*, de Murphy (2019), estão em curso e praticamente concluídas, mas a pandemia interrompeu a sua prossecução. No caso de outras obras, os revirores já foram identificados, mas o volume de trabalho do nosso editor de recensões em Inglês provocou uma acumulação do livro editado por Patrick, Schmid e Zwaan (2019) *Language Analysis for the Determination of Origin: Current Perspectives and New Directions* e de *Researching Forensic Linguistics* de Heydon (2019). Entretanto, temos quase uma dúzia de obras literalmente em quarentena no gabinete do nosso editor de recensões em Inglês

– cujo acesso lhe está vedado desde março – e prontos para enviar. Nesta pilha está, também, *Legal Translation Outsourced*, de Scott (2019), *Shallow Equality and Symbolic Jurisprudence in Multilingual Legal Orders*, de Leung (2019), e *Linguistics and Law*, de Kaplan (2020). Esperamos retomar o processo e o calendário de recensão de livros, quer em inglês, quer em português, com alguma normalidade, juntamente com as restantes atividades, em 2021.

Por isso, nós, os Editores da revista, decidimos publicar este volume como um número duplo. Para além de artigos regulares sobre investigação em curso na área da linguística forense e de uma recensão de um livro em português, este volume também inclui uma coletânea de artigos sobre a temática "como comecei". No início de 2020, Roger Shuy escreveu aos Editores da *Language and Law/Linguagem e Direito* a sugerir a publicação de uma série de artigos curtos sob a temático "Como comecei como linguista forense", que ele julgava ser interessante e útil para jovens com interesse na área da linguística forense. Assim, contactámos uma série de investigadores de renome na área, incluindo todos os Presidentes da IAFL – lamentavelmente, como é óbvio, Peter Tiersma e Maite Turell já faleceram – e congratulamo-nos com as respostas entusiásticas que recebemos. Os artigos que recebemos até ao momento são publicados neste volume, começando, obviamente, com o de Roger Shuy. Estes relatos ma primeira pessoa proporcionarão, certamente, uma leitura empolgante numa perspetiva narrativa, mas demonstram também – e sobretudo – de que modo o rigor metodológico é essencial para a série de aplicações da análise linguística em contextos forenses. Aqueles e aquelas que estiverem agora a iniciar o seu trabalho na área irão retirar pelo menos duas lições destes relatos: (1) para se ser um(a) bom(boa) linguista forense, primeiro é necessário ser-se um(a) excelente linguista; e (2) contrariamente a relatos recentes afirmando que a prática da atividade de linguista forense é uma boa fonte de rendimento, é muito difícil sobreviver desempenhando exclusivamente o trabalho de linguista forense. Não obstante, como demonstram os oito relatos aqui publicados, a linguística forense continua a ser uma área de investigação extraordinária, que vale a pena aprofundar, pelo que se incentiva ainda mais e melhor investigação na área.

Esperamos que goste tanto de ler este volume duplo, ainda que não muito volumoso, como nós gostámos de o editar. Esperamos editar e distribuir mais dois excelentes números da revista em 2021!

Rui Sousa-Silva
Universidade do Porto
Brazil

Malcolm Coulthard
Universidade Federal de Santa Catarina
Portugal

Referências

Heydon, G. (2019). *Researching Forensic Linguistics.* London & New York: Routledge.

Kaplan, J. P. (2020). *Linguistics and Law.* Abingdon and New York: Routledge.

Murphy, J. (2019). *The Discursive Construction of Blame.* London: Palgrave Macmillan UK.

P. L. Patrick, M. S. Schmid e K. Zwaan, Orgs. (2019). *Language Analysis for the Determination of Origin*, volume 16 of *Language Policy.* Cham: Springer International Publishing.

Scott, J. R. (2019). *Legal Translation Outsourced.* Oxford: Oxford University Press.

F. Vogel, Org. (2019). *Legal Linguistics Beyond Borders: Language and Law in a World of Media, Globalisation and Social Conflicts – Relaunching the International Language and Law Association (ILLA).* Berlin: Duncker and Humblot.

How I Got Started as a Forensic Linguist

Roger W. Shuy

Emeritus Professor, Georgetown University, USA

https://doi.org/10.21747/21833745/lanlaw/7_1_2a1

It was one of those unusually warm spring afternoons in Fort Worth Texas when three burly policemen approached that city's most flamboyant and wealthiest resident. T. Cullen Davis had just entered a telephone booth to make a phone call when three policemen unceremoniously yanked him out of the booth, arrested, and cuffed him, advising that he was being charged with soliciting murder. He objected, of course, but before he did so he begged the cops to let him get his nickel back from the telephone booth.

Cullen Davis was an unusually frugal man but he had very expensive tastes. He had inherited his father's very successful business of producing oil-drilling equipment that serviced the many oilrigs all over southwestern United States. He built a spacious house in Ft. Worth that everyone referred to as "the mansion." In Davis's social life, however, he did not mingle with the more prominent citizens. He preferred to associate with the lower-classes. He liked to hang around gas stations to chat with mechanics, among others, and he frequented bars where the upper classes would never allow themselves to be seen. At one of these bars he met an attractive blonde woman named Priscilla, married her, and as a wedding present bought her a neckless made of diamonds that spelled out "rich bitch." She wore it unashamedly.

Their marriage didn't last very long. A few years before his arrest at the phone booth, Davis and Pricilla had argued constantly and Davis finally moved out. One night shortly after he left, a masked man entered the mansion and shot and killed Pricilla's new boyfriend and her teen aged daughter by a previous relationship. Pricilla escaped with only a superficial wound to her breast.

Davis was the main suspect because he was thought to have the motivation to kill the woman from whom he was not yet divorced. Although this intruder was masked, Pricilla swore that she recognized him as Davis. Two of Pricilla's friends who happened to be on the street outside the house at the time supported her accusation, claiming that they too recognized the masked man as Davis.

This seemed to be enough evidence for the police to arrest Davis and charge him with murder. He was tried for murder but the trial ended with a hung jury, probably because the evidence against Davis was so weak. Undaunted, the prosecutor retried the case and this time the jury acquitted Davis.

Even though the prosecutor had failed twice to convict Davis, he was still not ready to give up. He had another idea. The new scenario was that if Davis couldn't be convicted of murdering Pricilla's boyfriend and daughter, maybe someone could convince him to hire a hit man to kill Pricilla. The timing for the police to arrange for this scenario became apparent after Davis filed for a divorce from Pricilla. Whether or not it could be verified, a rumor was flying around that Pricilla was in an intimate relationship with the judge in their divorce procedures. Now Davis had even more motivation to kill his wife.

It is not clear how the police originated the plan to get Davis to hire a hit man to kill Pricilla, the judge, and the two alleged witnesses to the earlier murders by the masked intruder, but suddenly one of Davis's midlevel employees, David McCrory, had a conversation with Davis while they sat in the front seat of Davis's Cadillac in a Fort Worth parking lot. McCrory's job was to get Davis to agree that McCrory would find a hit man to kill the four people. McCrory wore a hidden wire to record this conversation while police simultaneously tried to video tape the two men from a van parked directly across from Davis's car.

Although the videotape was virtually inaudible, the prosecutor was delighted with McCrory's audiotape, believing that this time he finally had sufficient evidence to convict Davis at trial. These audio and video tapes could be far better proof against Davis than the previous questionable opinions of witnesses about the identity of the masked man.

This trial for solicitation to murder began in 1979. I heard about the case accidently when I was on an airplane flying to Dallas for a linguistics meeting. The man sitting next to me was reading a manuscript that seemed to look like a sermon. I asked him if he was a minister and he said no, he was a lawyer defending a minister who was charged with slander. My seatmate then asked me what I did for a living and I told him that I was a Georgetown University sociolinguistics professor who analyzed tape-recorded speech. His eyes lit up as he told me that his lawyer-colleague had a case in which the evidence was tape-recorded speech and would I mind if he told his colleague about me. I agreed, not knowing what I might be getting into but always interested in the way people talk. Shortly after I returned home, I got a call from Sam Guiberson, one of the lawyers who was defending Davis in that case. He invited me to fly to his Houston office to get acquainted.

That summer I flew to Houston to meet with Guiberson, who outlined the details of the case. I'd never heard anyone solicit murder before and since I was too insecure to go alone, I talked one of Georgetown's junior faculty members into coming along with me for support. Don Larkin and I were met at the airport by Guiberson, who drove us to his law firm's office.

We had one of those "get-to-know-you" meetings in which Sam didn't mention the tape-recordings that I learned later would be the major evidence in the forthcoming trial. It turned out that Sam was a junior member of the law firm who had told his boss, the famous Richard "Racehorse" Haynes, that believed that the evidence in this case required the expertise of a linguist. He explained that at our first meeting the law firm would assess my physical appearance and demeanor and decide whether I could stand up to the hostile cross-examination that trials inevitably produce. Apparently, I passed their test because Sam then ushered me into the office of the senior partner and introduced me to Haynes, the lead counsel in the case.

As I entered Haynes's office, I quickly buttoned my suit coat before we shook hands. Only much later did Sam tell me that Haynes was impressed that I had performed the politeness gesture of buttoning my coat. The meeting was brief but as I left, Sam told me that Haynes told him that I looked like a "player."

My major test came later, after Guiberson sent me the audiotape and videotape, the evidence upon which the oilman's indictment for solicitation of murder was based. He also advised me that I was now hired to work on the case. Since I no longer needed my young colleague's emotional support, I was now on my own, working on the first law case of my career.

Even though Haynes would do all of the direct questing of me at trial, Guiberson was equally important. After I reviewed and analyzed the taped evidence, Sam had me fly to Houston several times to work intensely with him about how to present my testimony at trial. I remain grateful that he taught me what I needed to know about courtroom procedures and especially what I could expect during the prosecutor's cross-examination. He was a great teacher.

At the trial I was on the witness stand for most of three days. During the first day, Haynes followed the plan that Guiberson and I had created in which my testimony would rely on the homemade charts I had prepared. These charts were important guides for both Haynes and me about the sequence of how my testimony would proceed. But I was disappointed and shocked when Haynes omitted what I considered the most important chart, the keystone of my testimony. When I mentioned this to Sam during one of the breaks, he told me to not worry because Haynes was setting up the prosecutor to ask me about this topic, feeling that it would be more effective if it came through cross-examination rather than direct.

And that's exactly what happened. The prosecutor first asked me the very question that Sam had predicted. Apparently, he thought I had conveniently omitted or over-looked the passage on the tape that Haynes had not questioned me about. When the prosecutor asked me why I didn't deal with this passage, I replied that I had actually analyzed it and I happened to have a chart explaining it if he would like to see it. Of course, he was trapped into saying that he wanted to see it. Expert lawyer that he is, Haynes had successfully trapped the prosecutor into focusing on the most important part of my analysis.

The prosecutor did his best to keep me on the witness stand for most of two days, but he didn't make much progress with me, thanks to the extensive preparation Guiberson had provided. The jury deliberated for a few hours and returned a verdict of not guilty. It was a grueling and exhausting first experience for me as my first time as an expert witness in a trial.

Not all went well for me though. In addition to the problem I had when Haynes didn't ask me the most important question, one of my biggest problems was my nervousness and fear at finding myself in an adversarial situation that reminded me of the final oral exam for my PhD. It was worse than that oral exam because the prosecutor kept trying to trick me into giving answers that I did not want to make. For example, he asked me, "Dr. Shuy, when you did this subjective analysis of these tape recordings, what type of tape recorder did you use?" I paused long enough to realize that he was trying to make me admit that my testimony was not scientifically objective. Therefore, before telling

him what equipment I used, I responded the first clause in his question, telling him that I had done an objective analysis, not subjective. I caught on to what he was doing most of the time but I'm afraid I wasn't always so perceptive. Fortunately, during my direct exam Haynes asked me questions that repaired most of these infelicities.

Another problem for me was how to deal with meaning created in discourse rather than small chunks of data, including any inferences that might be made. I addressed this problem by using topic and response analysis, illustrated with my charts. They enabled me to show that the bad topics were only vaguely introduced by the government's representative, not by Davis, whose responses demonstrated that he did not understand their meaning. As far as I knew, this type of analysis had not been tested before in the courtroom setting and any success I might have with it depended on how effectively I presented it. The prosecutor had never faced such analysis before and he did his best to discredit it. His objections seemed strong at the time, but somehow the jury managed to understand what I was teaching them.

After it was all over, I flew home totally exhausted and vowed to not get involved in law cases ever again. But since this was a highly publicized case, word spread quickly among other lawyers that forensic linguists could be of help to them. Suddenly, my phone began ringing and since that case I have consulted or testified in many other cases that had tape-recorded evidence. Lawyers who specialized in other areas such as disputes about trademarks, contracts, insurance policies, police interrogations and other types of cases also began to call me to help them.

In summary, it was that 1979 conversation with my seatmate on an airplane who helped get me started in this field. After this I was helped greatly by the teaching and preparation I received from attorney Sam Guiberson along with the inspiring courtroom management of attorney Richard "Racehorse" Haynes.

How I Got Started in Forensic Linguistics

Diana Eades

University of New England, Australia

https://doi.org/10.21747/21833745/lanlaw/7_1_2a2

"I'm sorry, I simply don't have time"

In my early 30s I was a recent PhD graduate working in Brisbane, Queensland, at the front-line of an exciting new development in what we now refer to as Indigenous education. The job, at a tertiary college, was to direct a recently established special entry program for Aboriginal and Torres Strait Islander people. Students admitted to this program had not completed secondary education but were admitted into teacher education training via a bridging course and a support centre. I loved working with this great group of lively, committed and diverse students who taught me much about Aboriginal and Torres Strait Islander experiences, expectations, values, attitudes and social life, while I helped them negotiate what was for many of them at times a daunting and alien institution. In those early days of special entry programs, helping college staff learn how to communicate with Indigenous students was one of the most challenging parts of my role.

Despite my energy and youthfulness, I found the job emotionally and socially draining, after what in hindsight seemed like the luxurious life of a PhD student (with two young children). The fact that my office was a small glass-lined inset in the common room of the students I was working with certainly intensified my immersion in the Aboriginal and Torres Strait Islander Group at this college.

Thus, when a young lawyer called me and asked if I might be able to assist with an expert report about the answers attributed to an Aboriginal man in a police interview, my response was clear: "I'm sorry, I simply don't have time". I knew that being able to say "no" to requests is crucial to time management. The lawyer respectfully ignored the propositional content of my answer, and proceeded to explain the seriousness of his imprisoned client's allegation: that he had not said the words typed as his verbatim answers in the police record of interview. This interview typescript had been the major evidence resulting in his murder conviction and imprisonment two years earlier. In the Australian vernacular, his client was arguing that the typescript of his interview was a "verbal" (= fabricated confession).

The lawyer wanted my opinion because he had learned of my linguistics anthropology PhD research into Aboriginal use of English, and was hoping it would be relevant

to the allegedly verbatim typed record of this interview (which had not been audio- or video-recorded).

An expert linguistic opinion about the answers attributed to Mr Condren

It became clear that I should "just look at" the typescript of the interview. As the lawyer had hoped, before I finished reading the 155 questions and answers, it seemed to me that the answers attributed to the suspect, Mr Kelvin Condren, were most unlikely to have verbatim accuracy.

Before long, I had agreed to write an expert report, examining the interview questions and answers in detail, in the light of my knowledge of Aboriginal use of English generally, and Mr Condren's participation in interviews more specifically. An early step was to make a visit to Stuart Prison in the north Queensland city of Townsville where Mr Condren was serving a life sentence for murder. Meeting and interviewing him enabled me to see that his recorded interview with myself and his lawyer in that semi-formal legal setting was consistent with my observations and expectations of Aboriginal English speakers. I had also requested permission for Mr Condren's mother to join this meeting after some time. This enabled me to see that his conversation with his mother was quite different, and was consistent with Aboriginal ways of using English.

My report for the court compared the answers attributed to Mr Condren in the allegedly verbatim typescript with his answers in both the interview in the prison, and his evidence in his trial. My report found that "a significant number of the answers attributed to Condren in the [police record of interview] are not consistent with his speech patterns in the other two interviews, nor are they consistent with Condren's dialect of English." And I concluded that "it is not possible that all of the utterances attributed to Condren in reply to questions in [the police record of interview] are verbatim reports of his actual speech". (I have written about this case in Eades 1993, 2013.)

One of my most terrifying professional experiences

I knew almost nothing about the legal system, and the role of expert witnesses, although I had given evidence the previous year in the same court (Queensland Supreme Court). In that case my evidence considered the police interview of a teenage Aboriginal boy, in which I concluded that it was likely that at least some of his answers of "I don't know" indicated discomfort and possible confusion in the interview, rather than necessarily as a statement about his knowledge of the propositions being questioned.

The solicitor who had engaged me in the Condren case patiently explained my role and the pitfalls of undertaking the work of expert. My pre-trial meeting with the senior barrister was also very enlightening, despite being initially somewhat daunting. This barrister, Tony Fitzgerald, had not only read my report very carefully, but also my entire PhD thesis and my few publications to date, and asked excellent and exacting questions. He did this in the gentlest and most respectful manner, and I decided from that first meeting that he was just the kind of lawyer that Queensland needed. Indeed, within twelve months, he was appointed chair of the ground-breaking Commission of Inquiry into Official Corruption in Queensland ("The Fitzgerald Inquiry"), and later served as a judge of the Supreme Court of Queensland, then first President of that Appeals Division, then judge of the Appeals Division of the Supreme Court of New South Wales Supreme Court, and then judge of the Federal Court of Australia.

Although my experience of giving evidence in Condren's appeal was not my first appearance as an expert in court, it remains one of my most terrifying professional experiences. I shudder to think of some of my answers in cross-examination. On the day after I gave evidence, I came back to court to hear other witnesses. The Crown's barrister (who later went on to become a District Court judge) greeted me outside the court and said something like "I hope I wasn't too hard on you yesterday". I will never forget my reply: "Well, I'll never be scared of a student again!"

What I learnt from the judicial response to my evidence

Although my evidence was heard in full, it was later ruled as legally inadmissible, which of course disappointed me. It meant that my analysis and conclusions, which supported Mr Condren's allegation that he had been convicted on the basis of a verbal, had no influence on the court's decision. But some of the reasons for this ruling were such eye-openers to me that this appeal decision convinced me to pursue my research, about Aboriginal communication in English, in legal contexts specifically. For example, one of the judges ruled that

> …evidence of what are said to be normal characteristics of Aboriginal speech and behaviour is no more admissible than evidence of any other aspect of normal human behaviour would be, or the normal behaviour of persons of Anglo-Saxon descent or the Australian community in general and is not a proper subject for expert testimony.

What alarmed me most about the appeal judges' reasons for ruling my evidence inadmissible involved their discourse of race. While it was no longer at that time accepted in the social sciences, the judges used the then still popular 'pathology of ethnicity' to discount the relevance of my research on Aboriginal English. For example, one of the judges referred to the '…absence of any clear evidence as to the genealogy of the appellant and to the fact that neither of his parents were full-blooded Aboriginals'. He also wrote

> upon my assessment of [Condren's mother's] appearance and manner, I certainly formed the impression that she was of only partly Aboriginal extraction, and indeed that was not predominant.

Terms such as 'half-blooded', 'full-blooded' and 'of partly Aboriginal extraction' are used in the judgments in determining the question of applying evidence about Aboriginal people to Condren specifically. Of course, this is quite different from the findings of the social sciences that it is socialisation and cultural factors, rather than genealogy, that are most important in accounting for behaviour – including speech behaviour.

These judgments brought home to me the legal consequences of ignorance about the social and cultural dimensions of language and communication. At the same time, a few social workers and lawyers were telling me about the direct relevance to legal contexts of my work on Aboriginal ways of seeking and giving information. These two different, but related, dimensions of the participation of Aboriginal people in the legal system needed examination, and became a major focus of my research since then.

So what happened?

So what happened to Mr Condren? While his 1987 appeal, in which I gave evidence, was unsuccessful, the fight for justice in his case did not stop there. One of the country's top

investigative journalists, Chris Masters, undertook the "leg work" that had not been done by either the police or the overworked Aboriginal Legal Service that had represented him at trial. This revealed incontrovertible evidence that the murder for which Mr Condren was convicted and imprisoned had been committed in a time period during which he was in the police cells because of public drunkenness. After six years, his conviction was quashed, he was released from prison, and he received compensation for his wrongful imprisonment.

And what did I learn about saying "no" to requests when you are already overloaded? The seriousness of the allegation of fabricated confession, combined with the realisation that linguistic analysis could certainly shed light on this claim, led me to think "outside the box" and specifically outside my fishbowl office. Serendipitously, the college was starting to talk about Aboriginalisation of my position, which meant that it would not be appropriate for me to request extension of my two-year contract. My departure a few months before the end of my contract gave me time to write my expert report, as well as make up some of the time I had been unable to spend with my young children. Luckily, I was then able to take up a new academic position after a few months, in a time and place where such positions were not as scarce as they now are.

"Being a forensic linguist"

Most of my time, energy and output as a forensic linguist over more than 3 decades has focused on the broad area of forensic linguistics: namely the linguistic study of language in the legal process. Using a predominantly critical interactional sociolinguistics approach, I have focused on the use of varieties of English by, to and about Aboriginal people in the criminal justice system. I have also had many opportunities to provide training and workshops to lawyers and judicial officers about communication issues impacting Aboriginal participants in the legal system.

Much less of my work has been in terms of the narrow meaning of forensic linguistics: namely presenting expert linguistic evidence in courts and tribunals. In 34 years, I have written only 24 expert reports, and given oral evidence in only 10 cases. In my view, anyone starting out with the hope of "being a forensic linguist" needs to first become a good linguist. By specialising in an area of linguistics that is likely to be relevant to legal cases (such as cultural and linguistic aspects of Aboriginal uses of English, especially in legal contexts, in my case), you may have opportunities to respond to requests from lawyers for expert reports. And then some of these cases may result in you being required to give expert evidence in court. But these activities very rarely, if ever, comprise a full-time career.

References

Eades, D. (1993). The case for Condren: Aboriginal English, pragmatics and the law. *Journal of Pragmatics*, 20(2), 141–162.

Eades, D. (2013). *Aboriginal Ways of Using English.* Canberra: Aboriginal Studies Press.

How I Got Started as a Forensic Linguist

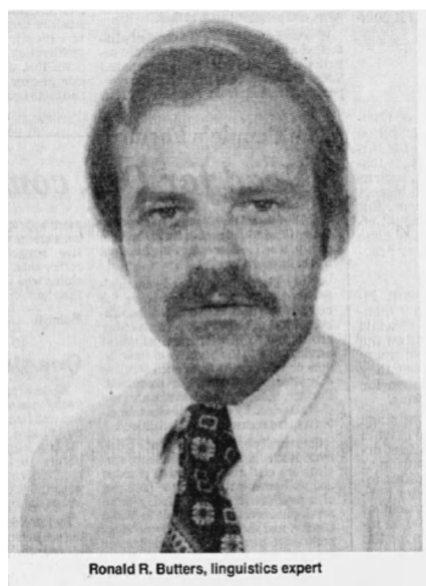
Ronald R. Butters

Professor Emeritus, Duke University, USA
President of the International Association of Forensic Linguists 2009-2011

https://doi.org/10.21747/21833745/lanlaw/7_1_2a3

My academic interests in language and law as well as my practical career in linguistic legal consulting have two different origins and began over a decade after I received my doctorate in English Literature and Linguistics. My first (and only) academic appointment was in the Duke University English Department as a specialist in English linguistics (1967–2007), where I taught courses in the structure and history of the English language, expository writing, and the understanding and appreciation of poetry and fiction. My earliest research began in linguistic theory as a follower of Noam Chomsky, but my interests very soon shifted to the American sociolinguistic school that was inspired by the early publications of William Labov, coupled with an increased interest in the work of the earlier dialectologists such as Raven McDavid. Throughout the 1970s and early 1980s, sociolinguistic approaches to American English were the focus of my scholarly interests, and as editor of the American Dialect Society's journal, *American Speech*, I grew increasingly interested in lexicology and lexicosemantic change.

Speech variation in Southern States English generated a good deal of interest in the local press, and my outreach to the North Carolina community led me to speak to educational and civic groups and to give numerous interviews with the local press. One such interview led to a feature publication in the *Raleigh News and Observer* (October 2, 1983) of a column (by reporter Guy Munger) naming me as "Tarheel of the Week" for my research on speech variation. The headline was, "Studying the language that makes man unique," and the article included my photo, with bushy 1980s moustache and a full head of hair, looking confidently at the camera:


Ronald R. Butters, linguistics expert

The article, with the earnest photograph, caught the attention of a Raleigh criminal lawyer, Howard F. Twiggs, one of the attorneys whose firm had been retained to defend the North Carolina lieutenant governor, James C. "Jimmy" Green, who had recently been arraigned in state criminal court for allegedly accepting a bribe. Mr. Green was then beginning a political campaign for state governor and had established a campaign fund that legally solicited contributions. However, an FBI undercover agent (one Robert Drdak, alias Tom "Doc" Ryan), was assigned to pose as a gangster who attempted to influence Mr. Green to illicitly use his government office to permit "the Detroit mob" to open a legal gambling casino in eastern North Carolina. A series of surreptitiously recorded conversations between the two made up the crucial evidence supporting the state prosecutor's allegations that certain passages (and, indeed, Mr. Green's willingness to continue meeting with Agent Drdak) constituted unambiguous evidence of a criminally culpable quid pro quo—campaign "contributions" in exchange for political favours.

Although Mr. Twiggs and Mr. Green disputed the charges, the defence had no way of contradicting Agent Drdak's sworn assertions as to Mr. Green's allegedly intended meanings short of putting the lieutenant governor on the witness stand. However, when Mr. Twiggs read the October story about my work in linguistics, it occurred to him that a linguist could give expert witness testimony contradicting Agent Drdak's interpretation without subjecting the accused to relinquishing his right not to testify against himself.

And so, Mr. Twiggs called me up. He described some of the passages that were at issue. I described what a linguist could reasonably say about the passages, based upon elementary pragmatic linguistic concepts and conversation analysis. For example, the following passage records how Green responded to the request of "Doc" for the Post Office Box number for campaign contributions:

> GREEN: Well, as I told you the last time I saw you, uh, if something comes up where I can be of assistance, uh, without, uh, jeopardizing myself, well, I'll always be glad to help.
> "DOC": Okay. Ah, the last time we—we got together you, uh, you didn't —you didn't have your box number for the, the uh—you remember you mentioned the campaign, ah, your campaign box number, . . . the uh

> [GREEN gives mailing address]
> "DOC": Okay, ah, well that, that's good. Ah, we, ah, we were, we were kinda wonderin' if, ah, you know, if since, you know we haven't talked in a while that if, ah, we, you know, if our minds are still workin' along the same, same avenue as before, if, if we're . . .
> GREEN: Yes.
> "DOC": we're in the right ballpark.
> GREEN: Yes.
> "DOC": Are we?
> GREEN: Yes.

The prosecution averred that when Mr. Green furnished, as requested, his legal campaign fund address, he was thereby agreeing to use his influence to assist "DOC" in an illegal scheme. They (and of course Agent Ryan) interpreted "without jeopardizing myself" to mean 'without getting caught making an illegal bargain' rather than (as Mr. Green claimed he intended) 'without doing anything illegal'. No clearly culpable "quid" is established: Agent Drdak's vague post facto "our minds still workin' along the same, same avenue as before – in the right ballpark" does not explicitly constitute an agreement to do anything illegal. I noted that Green's first "yes" is simply a conversational back-filler meaning no more than 'I hear you, I am with you in this conversation, fill me in on what you mean', and that Green's other "yes" responses do not explicitly indicate more than an understanding of "same avenue" as he claimed–a perfectly legal promise to fulfil his legitimate role as a government official–"to be of assistance" to the public, if he can do so without "jeopardizing myself" – 'putting myself in jeopardy by being involved in an illicit scheme'. The prosecution, of course, alleged that it could only mean 'put myself in jeopardy of being caught in an illicit scheme'.

In the end, I did not testify in this case, sparing my being subject to the prosecution's cross-examination of a witness who had never testified before (however appealing the earnestness of my newspaper photograph may have made me seem). My forensic contribution thus consisted of pointing out to Mr. Green's attorneys the inconclusiveness of the evidence at each juncture that the prosecution put forth as supporting allegations of corruption. Mr. Twiggs reasoned that the linguistic arguments in support of their position were so powerful that they merely used the details of what I told them in cross-examination of Agent Drdak and in their summary remarks.

The jury stayed out a little more than two hours and found Mr. Green not guilty on all four charges. I like to think that my contribution carried the day, although there was a well-known back story here that may also have influenced the jury. While the investigation was carried out by federal agents, the federal prosecutor refused to prosecute; the case went to state court instead, ultimately under the jurisdiction of the North Carolina Attorney General, who was Mr. Green's political rival in the contest for governor. In the end, neither Mr. Green nor the Attorney General ever became governor.

When Mr. Twiggs contacted me, I was at most only dimly aware that Roger Shuy had already done some innovative and exciting analysis in tape cases and had even begun to write about them in 1981 and 1982.[1] Once my contribution to the Jimmy Green case was finished, I wrote to Roger about my own fledgling attempt, and he responded generously. His fame among criminal lawyers was already bringing him more cases than he had time to take on himself, and he began to recommend me for the overflow.

Since the Jimmy Green case I have testified as a forensic linguistic expert in over 70 cases and consulted in perhaps 250. Roger recommended me for a number of different kinds of cases, and as time went on attorneys recommended me to other attorneys. It became increasingly clear to me that I was much less interested in the forensic analysis of surreptitious conversation than in cases that focused on the lexicographical and pragmatic meanings of individual words and phrases in statutes, contracts, and especially trademarks. In 1995 I testified in my first trademark case, *Circuit City Stores, Inc. v. Speedy Car-X, Inc.*; at issue was the putative likelihood of confusion between the marks *Car-X* and *CarMax*.[2] Since that case, the majority of my consulting work and forensic academic interest has been on trademark issues (chiefly genericness and likelihood of confusion). In my most recent courtroom appearance,[3] I gave evidence establishing that the term *fire cider*, a herbal remedy and condiment based on apple cider vinegar, is a term that has been in generic use for decades among those most likely to have purchased the product; in 2019, the judge ruled in favor of the clients for whom I testified. (See the judge's ruling at https://tinyurl.com/ya7hdjmc; and my paper, "Fire Cider," Dictionary Society of North America, Indiana University, Bloomington, Indiana, May 10, 2019, a copy of which I will be happy to send as a pdf file to anyone who requests it, ronbutters@mac.com).

Notes

[1]See "Can Linguistic Evidence Build on Defense Theory in a Criminal Case?" *Studia Linguistica*, 35(1-2) (1981): 33-49; "Topic as the Unit of Analysis in a Criminal Law Case." In D. Tannen (ed.), *Analyzing Discourse: Text and Talk.* Washington, DC: Georgetown Univ. Press (1982): 113-126; "What Did the Abscam Tapes Really Say?" *Linguistic Reporter* (May 1982): 3–4; "Entrapment and the Linguistic Analysis of Tapes." *Studies in Language*, 8.2 (1982): 215-34.

[2]The details of my contribution to this case are discussed in "Trademarks: Language that one owns," *The Routledge Handbook of Forensic Linguistics*, ed. by Malcolm Coulthard, Alison May, and Rui Sousa-Silva (Routledge, 2021), 364–79.

[3]*Shire City Herbals, Inc. v. Mary Blue, d/b/a Farmacy Herbs, Nicole Telkes d/b/a Wildflower School of Botanical Medicine, and/or Wild Spirit Herbs, and Katheryn Langelier d/b/a Herbal Revolution*, United States District Court District of Massachusetts, Civil No. 15-30069-MGM).

My Thwarted Start as a Forensic Linguist

Edward Finegan

Emeritus Professor, University of Southern California, USA
President of the International Association of Forensic Linguists 2013-2015

https://doi.org/10.21747/21833745/lanlaw/7_1_2a4

My early academic years were spent mostly teaching and in administration. After teaching for a couple of years at the merging (and then merged) Case Western Reserve University, I accepted an assistant professorship in the English Department at the University of Southern California in 1968. A year or so later, I became acting chair and then chair of the university's interdepartmental Ph.D. program in linguistics and went on to establish an independent department. As chair of a department with ties to several others, I served on or chaired an increasing number of college and university committees, a pattern I wasn't keen to continue! Then, in April of 1975, an opportunity arose to change the pattern, and I agreed to direct an English-language teaching program in Tehran under a contract the university had with National Iranian Radio and Television. In Tehran, I considered my future and even weighed the pros and cons of becoming a lawyer. I realized, though, that the likely path forward after law school would be to clerk for a judge and then take a post as an untenured assistant professor of law – a reboot at the bottom of the academic ladder. Besides, at the top of the list of courses I'd want to teach in law school would be "the language of the law," a topic in which I'd developed a serious interest, especially in the relationships between form and function.

Returning to Los Angeles and campus in December of 1976, I realized I could teach the language of the law in the Linguistics Department, and I filed paperwork to institute a course I dubbed "Language and Law." The university's curriculum committee objected to the Linguistics Department teaching law, as the proposed title suggested, but approved the course after a title change to "Linguistic Interpretation of the Law."

In January of 1977, a month after I'd returned from Tehran, the department secretary received a phone call asking whether any faculty member might be able to consult with a team of lawyers. Attorneys at the law firm of Irell and Manella were seeking a linguist to serve as an expert in a pending case. Why, that would be Professor Finegan, the secretary responded, given her familiarity with the paperwork for the new course.

Rose v. Home Savings & Loan Ass'n was a class-action suit filed in California Superior Court, and Irell and Manella represented the plaintiffs. The linguistic issue I was asked to address was whether an English speaker of ordinary intelligence could understand that a

"due on sale" clause – then common in promissory notes accompanying mortgage loans – would require a borrower to pay a pre-payment penalty upon sale of a property secured by the note. As I'd become acquainted with the linguistics of reading and readability while supervising an EFL curriculum and thirty teachers and staff in Tehran, the task seemed reasonably within my lane. When I analyzed a selection of the documents at issue, they fell at the bottom of the charts of standard readability measures and in some cases clear off the charts. It was apparent to me as a linguist that, in addition to the legal jargon in the note and the related documents, the complex syntax of the extremely lengthy sentences was extraordinarily difficult to parse and therefore to understand. It was apparent as well that the discourse structure of the documents seemed designed to disguise the steep penalty borrowers would be obliged to pay upon sale of their home, as an analysis of the discordance between topic sentences and a paragraph's contents could help demonstrate.

Prepped by attorneys – and having tutored them with a reasonably full understanding of my analysis and conclusions – I was ready for a court appearance. On the appointed morning – it was Valentine's Day – I combed my hair, buffed my shoes, and waited outside a Los Angeles courtroom eager to testify. I was keen to explain to a jury not only how difficult it would be for anyone reading the promissory notes to understand the financial consequences of selling a property secured by a note embedding a pre-payment penalty clause, but just what – beyond the jargon – made key points of the promissory note virtually impossible to grasp and just how the kinds of language ordinary people usually read differed in structure from that of the note. The point wasn't to tell jurors how difficult it was to understand the documents – they could judge that themselves. My task was to explain just what made the documents difficult to understand and how much and in what ways they differed from other kinds of texts. The objective was to help jurors understand how a reader-friendly document could have been structured if drafters had intended borrowers to recognize the pre-payment penalty's steep cost: when borrowers sold a home whose promissory note included pre-payment penalty language, they would have to pay six months' interest on the initial amount of the loan, irrespective of how much of the loan had been repaid. Thus, while smaller amounts related to the loan were spelled out in dollars and cents – $16.67, for example – the amount of the pre-payment penalty (often thousands of dollars whose exact amount was known at the outset, given that the penalty was based on the initial loan amount) was disguised in the relatively abstract language of percentage rates.

In 1977, corpus linguistics wasn't far along, but I was familiar with the Brown Corpus and drew on Kučera and Francis's *Computational Analysis of Present-day American English* (1969) for many of my proposed illustrations of word length and frequency, as well as average sentence lengths for various genres, and I could relate them to textual comprehension – to readability. Beyond that, my testimony would address the complexity of sentences, some running to scores of words in length, with clauses embedded in clauses within clauses, and a muddled use of pronouns – first person, for example, sometimes referring to Home and at other times to the borrower and Home together, thereby undercutting accurate understanding. Many of the complexities I recognized as a linguist would not have been apparent to jurors and weren't initially apparent to Irell's lawyers.

Pacing outside the courtroom on that Valentine's Day, I waited, with a tyro's eagerness, to be called. After far longer than I'd anticipated, Dennis Arnold, recently graduated from Yale Law School and the attorney I'd worked most closely with, came out of the courtroom to fetch me – or so I thought. To my utter bafflement, he reported that Home Savings & Loan had objected to allowing a linguistics expert to testify. Counsel for Home argued that, as ordinary speakers of English, jury members could determine for themselves how readable the loan documents were and that testimony by a linguist would unduly bias them. Crestfallen, I left the courthouse.

Little did I know then what transpired at trial about my intended testimony, but afterwards I read the transcripts. While I'd waited outside the courtroom, the judge heard arguments from attorneys on both sides as to the admissibility of a linguist and, in the end, asked the parties to submit briefs on the legality of a linguist's offering expert testimony in this case. During the following two weeks, the court received memoranda of points and authorities in opposition to my offering testimony and in support of it, and also heard oral arguments pro and con in chambers and in open court (though with jurors absent). At the end, in open court but again with jurors absent, Irell's attorneys made an "offer of proof." That is, an Irell senior attorney reported in detail what I would say if allowed to testify.

The offer of proof served two principal aims: it gave the court an opportunity to reconsider permitting my testimony, and, failing that, it created a record of the precluded testimony in the event of an appeal. "[T]he basis of Home's objection," Irell's attorney argued, "[was] that the proffered testimony is irrelevant to any contested issue of fact and [...] is unnecessary, improper, and patently prejudicial" because "the reading of documents is well within the experience of ordinary jurors." But, he went on, "the sophisticated and complex analyses, tests and theories that have been developed with the fields of linguistics and reading to measure the comprehensibility and readability of a particular document are sufficiently beyond the common experience of jurors to justify resort to expert testimony." Despite Irell's offer of proof (comprising nearly forty pages in the court's transcription), the court sustained Home's objection, citing two reasons for the decision. First, in the court's view, what I was to testify to was "not a proper area for expert testimony" because the jurors would "have an opportunity to examine the documents, and they [would] be assisted by the arguments of counsel." As a second reason, the judge said expert testimony giving quantitative measures could lead jurors to decide that "the documents just can't possibly be conspicuous, plain or clear." So, two weeks after I showed up at court to testify, the judge decided against allowing my testimony to the jurors.

That evening Dennis Arnold (today a partner at Gibson Dunn) phoned to let me know the final outcome. Naturally, I felt let down by the court's decision, but it was a valuable and important case for a novice forensic linguist. I learned a great deal, some of it linguistic and much of it legal. On the positive side as well, and this may be a third, though informal, reason for offers of proof, the litigants settled the case soon after my testimony was read into the record. It is tempting to think that the offer of proof containing my analysis helped persuade Home to settle the case.

As the promissory notes widely used at the time by S&Ls and other lending institutions included due-on-sale clauses and pre-payment penalties – the law journals in the years surrounding the case are filled with discussions of the subject – the Irell attorneys

and others brought class-action suits against one lender after another, and reached settlement after settlement. Eventually, with no linguist involved (to my knowledge), the underlying legal issue reached the California Supreme Court, where *Wellenkamp v. Bank of America* banned such due-on-sale clauses under most circumstances. Following that, the issue was litigated across the United States with similar results.

In 1977, I didn't know that I was part of the nascent field of forensic linguistics, as Jan Svartvik had named the field nine years earlier. My involvement came about because, as a linguist, I was particularly interested in the language of the law and had proposed teaching a course on the subject. Following the Home S&L matter, I went on to serve as a linguistics expert in a couple hundred cases, most of them civil matters, not criminal ones. Since that first case in 1977, I've served as an expert linguist chiefly in three arenas of the law. In trademark disputes, I've given expert advice or testimony for such firms as Bayer, Delta Airlines, DuPont, Victoria's Secret, and the United States Postal Service. In matters of contract dispute, I've interpreted insurance policies and assessed their linguistic comprehensibility for thousands of policy holders in dozens of cases brought against insurers. In defamation cases, I've been retained on behalf of Martha Stewart, Aretha Franklin, and Tom Cruise, among others, analyzing how discourse structure can convey propositions not expressly stated anywhere in a newspaper or other source.

Earlier in this decade I was honored to serve as president of the International Association of Forensic Linguists, and that was quite a surprise, but the most surprising development in my career came in 1996 when – without a law degree – I was appointed Professor of Linguistics and Law at the University of Southern California and taught in the law school for more than two decades.

References

Kučera, H. and Francis, W. N. (1969). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

How I Got Started as a Forensic Linguist

Martha Komter

Netherlands Institute for the Study of Crime and Law Enforcement (NSCR)
The Netherlands

https://doi.org/10.21747/21833745/lanlaw/7_1_2a5

I don't really know if I can call myself a regular forensic linguist, at least not in the sense that my research is aimed at applying my knowledge for solving problems of legal practice. It is more accurate to describe myself as someone who just happens to have drifted into the forensic linguistic world.

Let me explain. I studied English and Sociology, and then I worked at the departments of English, General Linguistics, Criminology, Law, Language and Communication, and the NSCR research centre. These jobs were a result of chance encounters, temporary contracts, reorganisations and unexpected opportunities. It is my research that took care of the consistency in my academic life. In the course of my studies I became interested in the interrelations between language and society and, more specifically, in the details of talk and interaction in institutional settings. My PhD research dealt with job interviews.

I became intrigued by the operation of this kind of 'people-sorting institution', because the applicants were treated at the same time as the best and the worst source of information. The best because they obviously had first-hand knowledge about their activities and motives, the worst because they had a stake in the outcome. So, after my PhD research was completed, I wanted to study the criminal trial. In the criminal process the stakes are higher and the power differences greater. And where else do interviewers start the interview by telling the interviewees that they don't have to answer questions?

After a long and tedious period of entry negotiations, I collected audio recordings of criminal trials, and transcribed and analysed them. Let me first give you some background. In the inquisitorial criminal law system of the Netherlands, trial judges question the suspect (who is called 'suspect' throughout the criminal process) about the case, with the help of the reports in the case file. The reports of police interrogations are required by law to be written "as much as possible in the suspect's own words". In the course of their questioning judges routinely confront suspects with their earlier statements to the police. They would say to the suspect: "you said to the police ..." and then read to them the weirdest sentences from the police report. Let me give an example from my courtroom recordings:

> Judge: You said to the police "You tell me that I am guilty of factual indecent assault. That is correct. There, in that place."

I was confused when I heard this. What was going on? Didn't the police officer who wrote this down know that people do not talk like this, and that it is obvious that these are not "the suspect's own words"? And why is the suspect presented as giving voice to both sides of the interaction with the police officer? And does it make any sense to transform the monologue into a dialogue? Then the interaction between the police officer (P) and the suspect (S) would be like this:

> P: You are guilty of factual indecent assault.
> S: That is correct. There, in that place.

It is hard to believe that this is what was actually said in the interrogation. A little later, I read an article in the newspaper about a rape trial. A man and a woman had met in a cafe and went home separately. Later in the evening the man knocked on the woman's door and had a bottle of wine in his hands. She let him in, they drank the wine and, according to the woman the man raped her, and according to the man they had consensual sex. In court, the judge read aloud how the woman, according to the police report, described the man:

> He was tall and well built, and had a tan as if he had been sunbathing a lot.

The suspect's lawyer commented:

> It is inconceivable that a woman who has just been raped and lies on the bed crying, would look at her rapist in this way.

I do not know whether this played a part in the judges' considerations, but the suspect was acquitted. Of course, the lawyer was right. No woman would describe her rapist like this. But was he also right in taking the text of the police report at face value? And, as in that period the interrogations were reported in a monologue form, did the words of the police officers not count?

These episodes made me curious to find out what actually happens in police interrogating rooms. So, after my courtroom research was completed, I went to the police officer on the beat in the area where I live in Amsterdam, and asked him if I might observe his interrogations. He and some of his colleagues were very cooperative and, after a while, they also let me record the interrogations. Then I learned that there were problems at the police station, and that the police detectives' superiors did not know what I was doing there. So, I decided to take the official route and, after a long and tedious period of entry negotiations, I got permission from the top of the police and prosecutor organisations to record police interrogations and collect the corresponding reports.

In the course of this project I discovered to what extent and in what manner police interrogators dominate the talk and the text of the written report. The following sentences from a police report are a simple example:

> Then I walked out of the store without paying.
> Outside I was stopped by two security men.

Contrary to what this monologue would suggest, these sentences were written down in the course of a number of question-answer exchanges. The police officer took care of most of the talk, while the suspect's answers amounted to: Yes... Yes... Outside... Yes...

Two... Two men. The monologue form made it impossible to see that most of what seemed to be "the suspect's own words" were in fact those of the police detective.

Now let's get back to the example from the rape trial. I came to realise the obvious: the woman did not think about what the man looked like when she was crying on the bed, but when she was interviewed by the police. And she must have told the details in answer to questions by the police officer. So, it might have gone something like this[1]:

> P: Can you describe what the man looked like, was he short, or tall,
> W: Tall
> *P types*: **He was tall**
> P: Okay. And was he thin or fat?
> W: Uh neither.
> P: Well built?
> W: Hm.
> *P types*: **and well built**
> P: Okay. And what was his skin colour?
> W: Brown.
> P: Do you mean that he was from Surinam?
> W: No, from the sun.
> P: You mean as if he had been sunbathing a lot?
> W: Yes.
> P: Okay.
> *P types*: **and had a tan as if he had been sunbathing a lot.**

This is guesswork of course, because I don't know what really went on in this police interrogation. Neither did the suspect's lawyer or the judge, but they did not take into account that the monologue that was written down was elicited in interaction with the police detective, and that it was probable that at least some of the words in these sentences came from the police officer, not from the witness.

When I thought I had studied police interrogations sufficiently to understand what was going on there, I still felt something was missing. I had first analysed trials and then I had analysed police interrogations, but that was the wrong order. I should have analysed police interrogations first and then the trials of the same cases, in order to see not only how police reports of suspect interrogations were drawn up, but also how these reports were read, understood and quoted in court. So, after a long and tedious period of entry negotiations, new materials were collected: audio recordings of police interrogations and video recordings of some of the trials of the same cases.

In spite of the cooperative attitude of many of the police officers that participated in the research, they were reluctant to let us record interrogations for serious crimes. The materials that were collected covered 'ordinary' street crimes, such as robbery, theft or drugs dealing. So, it was not surprising that I did not come across the kind of miscarriages of justice that are publicised in the media. I did not mind, because that was not my primary research motivation. What I wanted was to study what happens with the suspect's statement in the routine, run of the mill dispatch of criminal cases. This allowed me to discover the vulnerabilities of a criminal law system that relies so heavily on the construction and use of written documents (Komter, 2019).

At the moment steps are taken to digitise criminal case files in the Netherlands. Police interrogations for the more serious crimes are audio or video recorded, and they

are developing speech recognition software for automising transcriptions of police interrogations. This opens up a complete new field of possible problems. How accurate are these transcriptions? What must be done to transform these mountains of data into manageable portions? And what will be the status of these transcriptions in relation to the original recordings? These might be the challenges for future generations of forensic linguists.

Notes

[1]This is the typical question-answer-typing format of these interrogations.

References

Komter, M. (2019). *The suspect's statement. Talk and text in the criminal process.* Cambridge: Cambridge University Press.

My First Case

John Gibbons

Adjunct Professor, Monash University, Australia
President of the International Association of Forensic Linguists 2003-2005

https://doi.org/10.21747/21833745/lanlaw/7_1_2a6

My first case happened in the early 80s. Sadly the full details were lost in multiple changes of home, work and computer. It all began with a phone call from a solicitor out of the blue, saying that he noticed I taught a course on bilingualism, and wondered if I could help. The case involved an elderly man who had migrated from Italy to Australia in adulthood, and had acquired only limited English. In question was a police Record of Interview, which purported to be a verbatim account. (This was before major scandals involving confected police evidence based on interviews led to the imposition of audio and later video recording of any interview to be used as evidence.) The solicitor said that the Record of Interview did not seem to represent the man's English, and asked me to look at the data. It was intriguing, and a justice issue, so I agreed.

The lawyer sent me the Record of Interview, where the language attributed to the suspect looked like a lay person's idea of pidgin English. I interviewed the suspect, and found that his English proficiency consistently fitted a classic second language acquisition stage, which did not include passives, had limited control of tense, and only sentence coordination (not subordination). The Record of Interview included passives, tenses outside his range, and subordination. I therefore informed the lawyer that the Record of Interview was probably not verbatim, even if it represented the semantic content of the interview. As we prepared for court the lawyer warned me that the Prosecution would probably ask whether the suspect had artificially depressed his English performance when I interviewed him in order to invalidate the Record of Interview. My answer was that it was highly unlikely that over a lengthy interview with me he could consistently reduce his performance to an earlier acquisition stage. The problem of course was that it is impossible to fully impart second language acquisition research during courtroom examination. I therefore had to convince the court of the very existence of this field, and its relevance to the case. Fortunately, the court accepted my evidence.

One learning experience was the process of producing a report that would be acceptable in court. It turned out that the report had to follow a strict genre structure and (to use Heffer's terms) had to be in a paradigmatic legal framework. There was consid-

erable to and fro with the solicitor to achieve this. After the case concluded, I received a letter from the solicitor in the case, which included the following:

> ... Counsel closely cross-examined both the Arresting Officer (Detective Sergeant X) who conducted the Record of Interview in question and Police Constable Y who was the typist. The latter in particular was most confused and decidedly uncomfortable when confronted with your report ... It is our belief that he did not stand up well under cross-examination and we would deduce from this that much of what our client had verbally stated in this interview was conveniently transcribed into better English. The Constable admitted as much...

The main impression that I took away from my courtroom experience was amazement at the language and communication in court. It seemed to be at an opposite extreme to the linguistic yardstick of everyday casual conversation. This became a fascination with legal communication in the broadest sense that stayed with me for the remainder of my career.

Although I worked as an expert on more than thirty cases during my career, it was the more general field that became my focus.

All this out of the blue, following that day when I received a phone call from a solicitor.

References

Heffer, C. (2005). *The Language of Jury Trial: A Corpus-Aided Analysis of Legal-Lay Discourse.* Basingstoke & New York: Palgrave.

How I Got Started

Janet Ainsworth

John D. Eshelman Professor of Law Seattle University School of Law, USA
Current President of the International Association of Forensic Linguists

https://doi.org/10.21747/21833745/lanlaw/7_1_2a7

Although this might be surprising, given that I am a law professor, my interest in language and linguistics long predates my interest in the law. While I was an undergraduate student, I enrolled in an introductory linguistics course taught by Ray Jackendoff, and I found the subject fascinating. I went on to take a second course from him, but at the time, the university had no linguistics department and no major available for students to take to concentrate on the field, so I moved on to study other things, but continued to read linguistics books in my spare time. Language is the faculty more than any other than makes us human, and it seemed to me that understanding language and how it works was the best practical window into the problem of what it means to be human.

After finishing my undergraduate degree, I went on to graduate school in Chinese history, and in preparation for an academic career in that field, studied both the Chinese and Japanese languages. Both languages challenged me to think about the ways in which languages differ–and believe me, both of these languages differed dramatically from English, and from each other as well. But studying these Asian languages also impressed me with the ways in which unrelated languages nevertheless share a great many common features. I assumed that further language study would be in my future as a scholar in Asian history, but life often takes unexpected turns, and I ended up losing my dissertation advisor to an untimely death, and therefore needed to think about what a career Plan B might look like. Plan B ended up with an application to law school – not because I wanted to be a lawyer, but because I could get a joint degree in history and law, and finish my dissertation on Tang Dynasty taxation practices, while getting generous law school financial aid for my entire graduate education.

This was the plan, then. But again, my academic life took an unanticipated turn when, during my first year in law school, I happened to volunteer for an extra-curricular project in which law students served as legal representatives to prisoners charged with violating prison rules for which they could have their sentences increased if they lost their disciplinary cases. I had no intention to do anything more than take on a case or two, but once I took on my first such case, I was hooked. Before long, I began to wonder if finishing a dissertation on medieval Chinese law was really the best use of my

time, and started thinking more seriously about practicing law – specifically serving as a public defender representing people charged with crimes who could not afford to hire lawyers to represent them. And, in the end, that is exactly what I did – working at the Seattle public defenders' office for what I thought might be a couple of years, but turned out to be eight years of serving as both a trial lawyer in the felonies division and as an appellate lawyer handling appeals of convictions. After several years in the trenches of the criminal justice system, I was asked to become the training coordinator for the public defender and work both with young lawyers as they gained early experience as trial lawyers and also with more seasoned lawyers to deepen their expertise in legal analysis and practice.

Through another unexpected turn of events, I was asked to consider taking a temporary position teaching at a law school for a year—sort of a sabbatical from practice. This struck me as a great opportunity to recharge my professional batteries and perhaps to inspire young law students to consider public interest work as a career. I also assumed that I would have more time to pursue interests that had necessarily taken a back seat to my hectic schedule as 'counsel for the damned.' Although I continued to read books about language and linguistics in my spare time while a public defender, spare time was a rare commodity in that job. Now that I was going to be a professor–if only for a year–I would have the time to catch up with my pleasure reading in linguistics.

As it turned out, my life as a newly appointed law professor was nearly as time consuming as my position as a public defender had been. Figuring out how to teach law students—at a time when the only accepted method of instruction in American law schools was a professor-centric use of Socratic-style questioning of students in class—turned out to be a lot more challenging than I had imagined. And, as the only female professor in the law school at that time, my office hours were more occupied with mentoring and advising students – particularly women students in need of role models – than with leisurely contemplation of scholarly books. I reluctantly put aside any thoughts I had entertained about catching up with the world of language and linguistics during my one year 'sabbatical from practice.'

All this changed partway through that supposed 'sabbatical from practice' when the dean of the law school asked me if I would consider applying for a tenure-track permanent job. This was a difficult decision; I loved my work as a public defender, but I also found teaching law students to be deeply satisfying. In the end, the siren call of the academy won out; if the 'teaching gig' didn't work out, I would happily have returned to the public defender. But, as you might assume from the fact that you are reading this, in the end the 'teaching gig' did work out, and I have remained at the law school for more than thirty years and counting.

One of the privileges and duties of being a professor is that you have the opportunity and the obligation to engage in scholarship. Frankly, I was unclear as to exactly what legal scholarship might look like and what topics I might have something to say about. My earliest law review articles came out of my experiences in practice. I had seen many, many instances in which the criminal justice system had failed to serve the needs of the community and failed to provide justice for those caught up in it. In my second year as a professor, my dean took me aside to ask me what area of scholarly endeavor I was thinking of pursuing to justify an eventual grant of tenure. "Well," I began, tentatively, "I was thinking about exploring the intersection of law, language, and culture". "Hmm...

are you sure there is enough to say there for you to establish yourself as a tenure-worthy scholar?" was his reply.

In the end, of course, there was. Indeed, there is enough to explore in that intersection to keep an untold number of scholars busy for a lifetime. For my first scholarly project located at that intersection, I chose to look at the law regarding a suspect's invocation of the famous Miranda rights. I knew from my criminal trial practice that police officers and judges often failed to respect attempts at invoking Miranda rights by suspects undergoing police interrogation, and I believed that part of the reason was that judges and police officers had ideas about how people express themselves that linguistic science could show were simply empirically wrong. I decided to write an article to that effect, to be published in a law review where I hoped lawyers and judges would find it. To my surprise, it did receive a lot of attention in the legal scholarly community, and that encouraged me to keep writing about language issues in the law.

However, writing about language and the law as a law professor was a lonely endeavor at the start. My law school colleagues knew little if anything about linguistics, so it was a challenge to get feedback on my ideas from peers. Luckily, in the course of researching my articles, I became aware of the fertile scholarly ground being plowed by scholars identified with the linguistic subfield of forensic linguistics. I read their works voraciously – they insightfully applied linguistic science to dozens of areas that were highly relevant to me, including analyses of language in the legal process, language issues relevant to legal causes of action, and language evidence admitted, or more unfortunately, denied admission in legal disputes. I realized that there was indeed a scholarly community out there, consisting of both highly accomplished 'giants in the field' as well as more junior scholars taking the field into new and exciting areas. I began to send in paper proposals to conferences where forensic linguistic work was likely to be featured, and in the course of attending those conferences, large and small, I was getting a post-graduate education in the field of forensic linguistics as well as refining my own scholarly work through presentation and critique.

I have sometimes been asked whether there is a role for legally trained scholars in this field of study, to which my answer is a resounding "Of course!" Legal training and law practice experience can contribute an insider's perspective on legal doctrines, policies, and practices that can illuminate why law so often misfires. If lawyers and judges better understood how language works, this awareness could contribute to making the law fairer and more responsive to society's needs. By its very nature, forensic linguistic scholarship is highly inter-disciplinary. That means that all of us come to depend in our work upon knowledge and perspectives situated outside of our own disciplinary homes. Through my formal and often informal collaboration with researchers grounded in linguistics, I have found a community where the norm is that we build on the work of others in the field, adding to it our own contributions, in that way pushing the boundaries of forensic linguistics farther than before. The forensic linguistics community also happens to be the friendliest, most welcoming scholarly community that I have been privileged to become a part of. While I often joke that I am not licensed to practice linguistics in any jurisdiction, I am nevertheless deeply honored to be a member of the forensic linguistics community.

A Chance Encounter

Malcolm Coulthard

Professor Emeritus Aston University, UK
Foundation President of the International Association of Forensic Linguists, 1993-1995

https://doi.org/10.21747/21833745/lanlaw/7_1_2a8

Roger Shuy has long insisted that to be a good forensic linguist one must first be a good linguist. Although now there are growing numbers of academically qualified forensic linguists, when my generation started out the discipline of forensic linguistics did not even exist and most of us began working on cases by pure chance.

Roger reports that his entry to forensic linguistics consultancy was through a chance encounter in a plane, mine was a chance encounter in a university corridor. Roger's involved intensive coaching by skilled lawyers, followed by three days of examination and cross-examination in the witness box; I was never even called to give evidence... But I get ahead of myself.

I spent the first 37 years of my academic life working in a university English Language and Literature Department. For my first 20 years, the majority of my students had little initial interest in studying English language having entered the department to study literature and then discovered there were some compulsory language classes, so I did my best to link the teaching of language analysis to the study of literature – linking phonetics to identifying sound patterning in poetry, lexis and grammar to features of prose style and the analysis of spoken discourse to dramatic effects. One of the topics seen to be more 'useful' was the identifying difference between real spoken interaction and pseudo-interaction on the stage which I exemplified by showing how the conversations in plays written by the popular contemporary playwright Harold Pinter, which audiences thought sounded like 'real people speaking', were also artificial.

I had a colleague whose academic specialty was also seen by many students to be peripheral. The Department was a leading centre for the study of Shakespeare and his contemporaries and the task of one of my colleagues was to teach students to read Elizabethan handwriting. One evening he was at a party and another guest asked him what he did. "Pretty boring really", he said "I teach students to read Elizabethan handwriting". "Fascinating", said the man, "I wonder if you can help me, I am a solicitor and I have a client accused of falsifying signatures on cheques – could you analyse the signatures and show they are not his?" So, my colleague became a Forensic Document Analyst, specialising initially in handwriting analysis and later in using a newly developed tool,

labelled ESDA for short, (**E**lectro**S**tatic **D**ocument **A**nalyser), which allowed one to read the indentations on a given piece of paper, created by someone writing on a sheet resting above it – the significance of this will become clear shortly.

One day, in the mid-80s, my colleague passed me in the corridor and thrust into my hands what turned out to be an incriminating interview with a suspect, which had been recorded in handwritten form, as was the custom at the time, by a police officer and later typed up. Apparently, the accused, about to be tried for armed robbery, claimed the police had made up some of the utterances. "If you teach about the difference between invented and real interaction you should be able to say something about this" said my colleague. The accused's solicitors commissioned a report and I struggled with the analysis. In the first five pages I could find nothing to suggest manipulation, but the final page was different. It read as follows:

> Interviewer: I take it from your earlier reply that you are admitting been [sic] involved in the robbery at the M.E.B.
> Suspect: You're good. Thursday, Friday, Saturday, Sunday and you've caught me. Now you've got to prove it.

I could find no preceding incriminating "earlier reply" in the interview, to which the Interviewer could have been referring and the challenge to the Police by the suspect, "Now you've got to prove it", was not only odd in itself, but different in kind from the co-operative, non-aggressive tone of all his previous replies. In my report I was able to draw attention to the discoursal oddity of the two utterances which suggested that at least there was some text missing and thus undermining the evidential validity of the confession.

When the case went to court the Defence mounted an attack on the police. At the time there was an elite police group called the West Midlands Serious Crime Squad, two of whose members had interviewed the accused. The Defence produced an anonymous letter from someone claiming to be the wife of one of the officers saying "They write confessions in my living room at weekends"; there was evidence that this accused was not the only one who, in his interview, happened to praise the skills of the group, but the crucial evidence presented came from my colleague. I sat in court – as an expert witness I was allowed to do so – while he presented his analysis of the same interview record on which I had written my report.

At that time the British police conducted interviews in pairs, one asking the questions, the other making a written record of the interview, not just notes of the main points, as still happens in many jurisdictions world-wide, but a verbatim record of both the questions asked and the answers given. British Judges had indicated at the beginning of the 20th century that they wanted full access to the actual locutions, so they could decide the illocutions and perlocutions they conveyed. Physically the policeman/scribe would start with a pile of blank record sheets in front to him and write on the top one, when completed he would place it to one side and start on the next sheet. In so doing, he was, unwittingly, creating a multiple record of the interview, because apart from the top sheet, all subsequent sheets had not only a visible handwritten record but also an invisible dentation record of the page above – indeed, if the policeman pressed hard with his pen there could be a two-page indentation record. My colleague demonstrated to the court how, when processed by his ESDA machine, page 2 had an almost perfectly legible record of page 1 and so on until the final page, which, as I had previously noted,

was discoursally odd. This final page did not have, as expected, a record of the previous page, but instead a record of what appeared to be an earlier non-incriminating version of the final page – in other words the final page had been re-written. This evidence of malpractice was unchallengeable; the judge stopped the trial immediately and dismissed the case. Within hours the Chief Constable disbanded the Serious Crime Squad, 52 officers were suspended and later disciplinary action was taken against seven of the officers, although eventually, despite significant opposition and protests from those who claimed they had been wrongfully convicted on other falsified evidence, none were prosecuted.

Although I was not actually called to give evidence, my name appeared in the press and soon letters began to arrive from prisoners claiming they too had been 'verballed' by police officers, and these cases gave me insight into some of the strategies used by police officers when acting as amateur dramatists and falsifying interview records and statements. It also meant that for quite a long time my colleague and I felt nervous whenever we noticed a police car in the rear-view mirror.

I also began to be contacted by solicitors with clients claiming wrongful conviction and journalists investigating historic cases of miscarriage of justice. Fortunately, although coincidentally, the major facilitator of these wrongful convictions, the handwritten police record, was phased out by the Police and Criminal Evidence Act of 1984, which required that all significant interviews with suspects be audio-recorded. Even so, the major cases I worked on in the 1990's were all concerned with police falsification of written versions of oral evidence: the Birmingham Six and the Bridgewater Four appeals, both involving officers who were or had been in the West Midlands Serious Crime Squad and the Derek Bentley Appeal against a 1950's murder conviction (see Coulthard *et al.* 2017, chapters 6, 8, 9).

So, in conclusion, to reprise Roger, I would say to any intending forensic linguist, first become a recognised expert in one area of forensic linguistic analysis, preferably computer-assisted and you won't need to look for work; eventually the world will beat a path to your door. Although, one caveat, if you want to become a **full-time** forensic linguist, specialise in phonetics, so far there are very few people who can support themselves analysing texts.

References

Coulthard, M., Johnson, A. and Wright, D. (2017). *An Introduction to Forensic Linguistics: Language in Evidence.* London and New York: Routledge.

Exploring the Distinctiveness of Emoji Use for Digital Authorship Analysis

Karoline Marko

University of Graz, Austria

https://doi.org/10.21747/21833745/lanlaw/7_1_2a9

Abstract. *This exploratory study sets out to investigate the potential distinctiveness of emoji use on the social media platform Instagram. The use of emoji has become popular in digital media, as they provide additional information about a given message; they are said to serve similar purposes to non-verbal cues in face-to-face interactions (e.g Gawne and McCulloch 2019). Several studies have investigated emoji use related to geographical origin, personality traits, age and gender, but their distinctiveness and use for authorship analysis has remained relatively unexplored. Based on a sample of 60 individuals, this study researches not only group-specific characteristics of emoji use on Instagram, but also explores whether or not the use of emoji is distinctive enough to identify individuals simply based on their use of emoji. For this purpose, this study mainly draws on, but also expands, Evan's (2017) emoji function classification framework. The results suggest that individuals do indeed exhibit emoji usage patterns that can be valuable for authorship analysis.*

Keywords: *Emoji, uniqueness, authorship analysis, sociolinguistic profiling.*

Resumo. *Este estudo exploratório tem como objetivo investigar a possível singularidade da utilização de emojis na rede social Instagram. A utilização de emojis popularizou-se nas plataformas digitais por proporcionarem informações adicionais acerca de uma determinada mensagem; eles servirão fins semelhantes às indicações não-verbais em interações face-a-face (e.g Gawne e McCulloch 2019). Diversos estudos investigaram a utilização de emojis relativamente à origem geográfica, a traços de personalidade, idade e género, mas a sua singularidade e utilização para efeitos de análise de autoria continua por explorar. Com base numa amostra composta por 60 pessoas, este trabalho investiga, não só características da utilização de emojis específicas de determinados grupos no Instagram, como também se a utilização de emojis é suficientemente distintiva para identificar indivíduos simplesmente com base na sua utilização de emojis. Para o efeito, este estudo, sobretudo, baseia-se – mas também aprofunda – a grelha de classificação*

das funções dos emojis, de Evans (2017). Os resultados indicam que os indivíduos apresentam realmente padrões de utilização potencialmente preciosos para análise de autoria.

Palavras-chave: *Emoji, singularidade, análise de autoria, perfis sociolinguísticos.*

Introduction

Situated in the field of authorship analysis, this study addresses prevailing new challenges posed by digital media. Authorship analysis, which aims at comparing two or more different texts with the goal of investigating whether the authors of the texts are the same or different, or of creating sociolinguistic profiles based on the language used in order to classify the authors into some category (e.g. gender, age, education, etc.), has been facing many challenges. For instance, it has long been an issue that sample texts used in authorship attribution are too short to yield valid results through the use of statistical methods (e.g. Coulthard 2006; see also Eder 2015, Brennan and Greenstadt 2009). This particular problem has been intensified by technological developments, as newer forms of communication such as text messages are often even shorter than 100 words, and yet have frequently been implicated in crimes (e.g. Coulthard *et al.* 2017, Grant 2010). Sousa Silva (2018) and, even more recently, Heydon (2019) have pointed out that technology impacts on authorship analysis also due to developments like the inclusion of multimodal resources such as gifs, videos, and pictures (including emoji). The focus of this article will be on the use of emoji.

Reflective of its place of invention, emoji is Japanese for "picture character" (Goldman, 2018: 1231). Emoji are particularly common on social media platforms, which are sites "that promote social interaction between participants" (Page *et al.*, 2014: 5). Since emojis are often used to substitute for non-verbal cues and gestures (e.g. McCulloch 2019; Miller *et al.* 2016), it is hardly surprising that they appear so commonly in digital media. Evans (2017: 22) estimates that more than six billion emoji are exchanged on social media every day, and Goldman (2018) reports that "2-3 trillion mobile messages incorporate emojis in a single year" (p. 1229). Further, and particularly relevant for the context of this article, Dimson (2015) argues that approximately 50% of the texts on Instagram contain at least one emoji. These figures are indeed intriguing and give rise to important questions relating to authorship analysis, such as whether it is possible that the use of emoji differs so much among different people that it is possible to identify the author of a post simply based on their use of emoji. This is the question I will investigate in this paper. First, however, it is necessary to establish what previous research has found out about the use of emoji in relation to different groups of people and to provide an overview of research in the field of authorship analysis in relation to digital communication.

Emoji, which can represent emotions, objects, ideas, and even actions (Donato and Paggio, 2017), were introduced in 2011 and due to the sharp rise in use in 2015 (Evans, 2017: 10), many researchers have taken on the task of investigating how emoji use relates to individual users. Theoretically, as emoji (as opposed to emoticons, see e.g. Ai *et al.* 2017), are unified by the Unicode Consortium, all emoji should look the same on each platform. Practically, however, each platform adapts the original Unicode Code so that in reality, emoji appear differently on each platform or device. This has led to many misunderstandings, as Goldman (2018) points out. A much-discussed example in

the context of misunderstandings is the "Astonished face" (U+1F632[1]). As Figure 1 illustrates, the "Astonished face" is rendered quite differently across platforms. The black and white emoji on the left-hand corner is the original emoji provided by Unicode. As Goldman (2018) discusses, the renderings of this particular emoji on some platforms can be interpreted as death threats, while others are more clearly associated with feelings of anger, astonishment, shock, or annoyance. Similarly, the "Grinning Face with Smiling Eyes" (U+1F604) has been associated with both happiness and anger, depending on the platform on which they are depicted (see Miller *et al.* 2016).

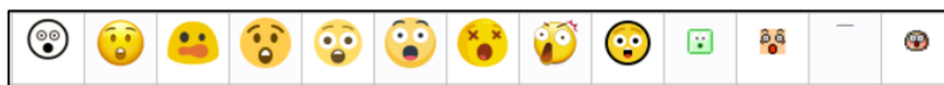


Figure 1. The "Astonished Face" across different platforms (Goldman, 2018: 1258).

Further, Chen *et al.* (2018), for instance, have researched emoji use in relation to gender. They have not only found that women use significantly more emoji than men, but they have also found that women and men prefer different kinds of emoji and use different emoji for expressing similar sentiments (p. 1). Even though many emoji are co-used by males and females, it appears that these emoji are used for different purposes and to convey different emotions or sentiments (p. 2).

Other researchers such as Li *et al.* (2018) have focused on the connection between emoji use and personality. They have correlated emoji use patterns with the Big Five Personality Traits[2] and reached the following conclusions: the trait of openness does not seem to be related to emoji use, while people with higher scores on conscientiousness tend to use fewer emoji. Further, people who scored low on extraversion used the most emoji; people scoring high on agreeableness also tend to use more emoji than those scoring low on this trait. Additionally, people who scored high on neuroticism are the ones who prefer the use of exaggerated expressions (pp. 649-650).

Another interesting strand of research in this respect has investigated links between emoji use and the living conditions of the respective users. Concentrating on a Twitter corpus, Ljubesic and Fiser (2016), for example, have focused on first, second, third and fourth world clusters[3] and obtained the following results: while people in the first world cluster used almost no face emoji, people in the second world cluster used many emoji conveying positive emotions. People in the third world cluster exhibited a high use of unhappy face emoji and the praying hands emoji, while people in the fourth world cluster used many hand gesture emoji. Thus, they conclude, it is possible to track the user's living conditions in different parts of the world simply by taking a close look at which emoji people use most commonly (pp. 87-89).

As outlined above, particular attention has been paid to how emoji use and patterns of emoji use are related to different groups of individuals. Other researchers have taken different approaches and have investigated emoji semantics (Barbieri and Camacho-Collados, 2018), redundancy (Donato and Paggio, 2017), the role of emoji in the law (Goldman, 2018), emoji as gestures (Gawne and McCulloch, 2019), and emoji ambiguity (Miller *et al.*, 2016, 2017). Even though these studies have investigated how people differ in their use and interpretation of emoji, none of them has looked at patterns of emoji use from an authorship perspective, even though Na'aman *et al.* (2017) have hinted at, but

do not elaborate on, idiolectal differences in the use of emoji, by stating that "there can be little question that individuals use emoji differently" (p. 141). The present study is not simply interested in the use of individual emoji, however. Rather, it is the aim to look at how people use emoji, as it is possible that individuals develop certain preferences for emoji use, comparable to "preferred co-selections" (Coulthard, 2008: 146) in language.

It is the aim of this paper to answer the following research questions: (1) Can the use of emoji on Instagram be related to demographic categories such as gender, age, and social group so that findings can be useful for sociolinguistic profiling? (2) Is the use of emoji distinctive enough to identify the author of posts only based on emoji usage patterns in a simple authorship comparison task?

Research in Digital Authorship Analysis

Research into short texts produced online has, for example, garnered the interest of researchers focusing on computational approaches to authorship analysis: Orebaugh and Allnutt (2009) have focused on the classification of instant messaging communications, and Layton *et al.* (2010) have concentrated on authorship attribution in a Twitter corpus. Other examples of automated authorship analysis are Sousa-Silva *et al.* (2011); Rocha *et al.* (2016), and Ishihara (2017). Interestingly, Sousa-Silva *et al.* (2011) have found that in their study, the use of emoticons "outperforms all other feature groups tested" (p. 167), demonstrating the usefulness of non-language features in authorship analysis.

MacLeod and Grant (2012), and Johnson and Wright (2014) have also investigated short online communications (Twitter and e-mails, respectively). Since the analysis of short texts usually defies the use of statistical methods and because "traditional [authorship analysis] methods do not easily translate into computer-mediated communication" (MacLeod and Grant, 2012: 210), these researchers have adapted the use of Jaccard's coefficient and Delta-s calculations to authorship analysis problems. This approach will also be taken in the present article.

Sociolinguistic profiling, in contrast to authorship attribution, is non-comparative (Ehrhardt, 2018). Rather, it categorizes individuals in terms of pre-defined social and demographic categories. As Coulthard *et al.* (2011: 538) state, "profiling involves taking a single example and, by matching it to a well-founded generalization, drawing a conclusion about that instance." Even though sociolinguistic profiling has been improved in recent years (see Nini 2018b), the potential of emoji has not yet been addressed in this connection.

Emoji Classification Systems

A variety of emoji classifications and categorizations exist. Many of these classifications have separated emoji into categories, such as 'traveling/commuting', 'events', 'places', 'feelings', 'people', 'eating and drinking', etc. (Donato and Paggio, 2017). Other such classifications are provided by Emojipedia (online), Lin *et al.* (2014), Barbieri *et al.* (2016), and Vidal *et al.* (2016). Although these classifications are interesting and important (as shown in experiment 2), they only classify individual emoji into different groups but disregard how these emoji function in context.

One of the most important classifications of emoji which takes their function into account comes from Evans (2017: 130-135). He differentiates between six different functions, which are similar to the functions non-verbal cues serve in face-to-face conversation, such as adding emphasis, repeating what is said, or referring to objects and lo-

cations, among others (see also Gawne and McCulloch 2019). The first of Evans' (2017) categories is substitution, which refers to the actual replacement of a word with an emoji. Secondly, emoji can serve the function of reinforcement. This, for instance, means that the emoji conveys the same meaning as the words do, which simultaneously emphasizes the meaning of the words. Further, emoji can be used in a contradictory way, which usually happens in cases when the writer intends to be ironic. Emoji can also serve a complementary function, which refers to something similar as a meta-comment to the words. This can also be regarded as a politeness strategy which has the potential of mitigating possible face-threatening acts. The fifth function is an emphasizing function, meaning that emoji are used to highlight an idea. Lastly, the discourse management function focuses on emoji in initial and final positions. For instance, utterance initial emoji are often used to respond to a previous message, while utterance final emoji can be seen as similar to transition relevance points in conversation, which signal that an idea is complete (see, e.g. Clift 2016).

From a semiotic and communication studies perspective, Danesi (2016) has additionally found that emoji are often used to replace punctuation marks at the end of sentences or salutation formulae at the beginning of messages – an idea that is compatible with Evans' (2017) discourse management function. Also similar to Evans (2017), Danesi (2016) differentiates between an adjunctive and a substitutive use of emoji, but he additionally discusses phatic and emotive functions. The former refers to ways of using emoji for small talk, such as utterance openers, utterance enders, and silence avoiders – thus, they simultaneously serve a pragmatic function; the latter refers to emoji being used to substitute for facial expressions or for emphasizing an idea visually, which is comparable to what Evans (2017) calls substitution or reinforcement.

Data & Methodology

The data for the present study was collected manually from the social media platform Instagram, which is a photo sharing "platform best known for selfies and self-representation" (Leaver *et al.*, 2020). Instagram was chosen for several reasons: first of all, little linguistic research has focused on Instagram. It allows the addition of yet another layer of multimodality to the study of emoji in that emoji can also be used in relation to the image, not just to the text (see below). The second advantage is that Instagram does not have a function that allows users to direct a post to someone in particular. Public posts are visible to potentially anyone who has access to the internet and it is rarely the case that someone is directly addressed through the use of tagging, which is a common feature on Twitter (Zappavigna, 2013). The lack of this affordance on Instagram diminishes the effect of the addressee on the post (Bell, 1984), as all individuals have potentially the same audience – any person with access to the internet.

The data sample in this study is composed of 60 individuals: 30 males and 30 females. The age range is 14 to 69, with a mean age of 26.4 years, a median of 24 years and a mode of 15 years and is thereby reminiscent of the general age distribution of Instagram (see Statista 2020).

8873 posts were included in this analysis. Posts were excluded for the following reasons: if the emoji were only used in hashtags, and if they were used in reposts with the original post present and thus in direct response to someone else's post. This was done for the following reasons: first of all, emoji in hashtags seemingly portray a different

function (see, e.g. Zappavigna 2013) which was not focused on in this analysis. Secondly, if posts are re-posts and include the original post, the language and emoji use of the individual is likely to be influenced by the audience (Bell, 1984). In order to prevent the skewing of findings, such posts were entirely excluded from the analysis.

Ethical Considerations

Great care needs to be taken to avoid harm to the people whose data is included in an analysis based on social media data (Townsend and Wallace, 2016). For this study, only public profiles that can be accessed and viewed by anyone with or without an Instagram account have been included. According to the Instagram policy (Instagram, 2018), data that is publicly available can be accessed and used by third parties. Importantly, since the minimum age for Instagram use is 13, no individuals younger than 13 are included in this study. The individuals' names have been anonymized and no inferences about their identities can be made. Since the posts are publicly available, the wordings of the posts used in this study have been altered to prevent any detection of the individuals through google searches by replacing content words, as previously practiced by Gawne and McCulloch (2019), who refer to Ayers *et al.* (2018).

Emoji Classifications

Emoji Functions

For this analysis, Evans' 2017 classification of emoji functions outlined above was adapted and linked in parts to Danesi's (2016) functions for practical reasons. For instance, the discourse management functions described by Evans (2017) could not be identified in the Instagram data on which this study is based, as there are no ongoing conversations to be analyzed. Thus, although emoji can appear in initial and final positions, they have rather different functions than responding to previous messages or serving as transition-relevance points. Further, based on the data, it was impossible to distinguish clearly enough between emoji that serve as reinforcements and those that serve as emphasizers. The remaining adapted categories are shown in Table 1.

Emoji classifications have concentrated on how emoji can be classified according to the function they serve in relation to the language they accompany. However, on Instagram, emoji use might not only be related to the language in the post; it might as well reflect the picture itself, complement, emphasize, or contradict it. This dimension has to be accounted for by the classification system. Thus, the classification system outlined in Table 1 was further adapted to suit this particular need (see column 4). In the present context this additional dimension is crucial, as individuals may have (un)intentional preferences for their emoji use either reflecting the picture and/or the text, which can have further important implications for authorship analysis.

Prior to the experiments, the applied categories of emoji functions were tested for inter-rater reliability in several steps. Following from these initial tests in which each trial was rated by two researchers and which revealed an inter-rater reliability of 68%, the definitions of the categories of emoji functions were slightly adapted to provide clearer boundaries between the categories. The adapted definitions can be seen in Table 2 below.

Function	*Definition*	*Example*	*Functions related to pictures*
Substitution	Replacing a word with an emoji	I❤️this dress!	n/a
Reinforcement, emphasis	Repeating meaning of words; emphasizing the idea the words convey	I love this dress❤️	Emoji can either reinforce or emphasize the picture they accompany
Contradiction	Emoji expresses the opposite of what the words convey; often used for irony	Thank you very much 🙄	Emoji can contradict the picture
Complementation	Meta-comment to the accompanying words; politeness strategy, frequently encountered in conversation openings (therefore, non-reactive)	Hi, how are you? 🙂	Emoji can complement the words or the image, either in semantic terms, or, for instance in terms of color
Discourse management; placement[4]	Initial position: response to previous turn; Final position: replacing punctuation marks, thereby	🙂 I agree Fantastic 🙂	n/a

Table 1. Emoji Functions (adapted from Danesi 2016 & Evans 2017)[4]

Function	*Definition*	*Example*
Substitution	Replacing a word with an emoji in a sentence; not applicable to emoji outside sentence/clause boundaries	I❤️this dress!
Reinforcement, emphasis	Repeating meaning of words actually present in the sentence; emphasizing the idea the words convey; illustrative purpose	I love this dress❤️
Contradiction	Emoji expresses the opposite of what the words convey; often used for irony	Thank you very much 🙄
Complementation	Meta-comment conveying an additional sentiment/idea; politeness strategy to save face; meaning of the emoji is not present as a word within the sentence it complements	Hope you're ok 🙂
Discourse management; placement	*Excluded as a function*	

Table 2. Adapted Emoji Functions.

Table 2 shows the changes made to the original definitions (Table 1): The category of Substitution is only used when a word inside a sentence is replaced by an emoji. In order to avoid confusion between Reinforcement/emphasis and Complementation, the former will only be used in cases where an emoji conveys the same meaning as the words within the sentence. That is, for an emoji to count as reinforcement or emphasis, a word or phrase needs to be present that has the same meaning as the emoji itself. If an emoji conveys an additional thought or concept that is not directly encoded in the

words used, the emoji will be coded as serving a complementary function. The Discourse Management function was excluded from the present analysis for the reason that any emoji serving any of the other functions can also simultaneously serve a discourse management function. The placement of emoji was thus looked at separately from the emoji functions, since it is still expected to be of analytical value (see Table 4).

After these adaptations were made, another researcher was asked to classify the emoji into their respective categories. This time, the researcher received more prior input and more detailed instructions in addition to the adapted definitions. With these changes in definitions and preparations, inter-rater reliability increased to 85%.

Emoji Taxonomy

The second experiment conducted for this study is not based on the functions the emoji serve but rather on the use of emoji types. Therefore, the following taxonomy of emoji based on Apple's iOS version, as shown in Table 3, is used. The original version of the taxonomy was adapted to the specific needs of this study. Thus, some categories were split into further descriptive categories in order to make follow-up calculations more accurate.

Level 5	Level 4	Level 3	Level 2	Level 1
Emoji	Smileys & People	Smileys	Positive	
			Neutral	
			Negative	
		People	Body parts	
			Real people	Male
				Female
			Mythical people	
		Families		
		Hand gestures		
		Clothing & accessories	Male	
			Female	
		Other	Positive	
			Neutral	
			Negative	
	Animals & Nature	Animals		
		Nature	Astronomy	
			Weather	
			Plants	
			Other	
	Food & Drink	Fruit		
		Vegetables		
		Meals		
		Beverages		
		Utensils		
	Activity	Sports		
		Music		
		The Arts		
		Hobbies		
		Other		
	Travel & Places	Scenes		
		Locations		
		Buildings		
		Modes of transport		
	Objects	Household items		
		Celebrations		
		Stationary		
		Miscellaneous objects		
	Symbols	Hearts		
		Clocks		
		Arrows		
		Signs		
		Shapes		
	Flags	Country		
		Other		

Table 3. Emoji Taxonomy and Corresponding Levels.

Jaccard's Coefficient & Delta-S

As mentioned above, many computational and statistical methods tend to be unreliable with short texts. The use of Jaccard's Coefficient, however, provides a solution to this problem. Jaccard's Coefficient, as outlined in detail in MacLeod and Grant (2012), Grant (2013) and also in Johnson and Wright (2014) and Nini (2018a), was used in the present study to evaluate distances between texts based on emoji. Jaccard's coefficient "establish[es] degrees of similarity between cases" (MacLeod and Grant, 2012: 2013) and Grant (2013: 482) further outlines that Jaccard's coefficient "is a correlation for binary values"; features identified in texts are assigned either 1 (presence) or 0 (absence). Results close to 0 indicate that the investigated texts are completely different; results close to 1 indicate that the texts are the same. An important advantage of Jaccard's Coefficient outlined by

Grant (2013) is that “it does not inflate similarity on the basis of two absences [and thus] does not risk overstating the explanatory power of a single text” (p. 482).

In order to make Jaccard calculations possible, the following features were identified as variables in the analysis: emoji functions (see Table 2), specifics of use, and the placement or position of the emoji (see Table 4). These variables have emerged as important features of emoji use patterns in previous research and in the present corpus and were thus chosen for the analysis.

Variable	*Definition*	*Examples*
Emoji functions	See Table 1	See Table 1
Specifics of use • Strings • Compositions	Strings: consecutive use of the same emoji Composition: use of several different emoji	String: 🌴🌴🌴🌴🌴🌴🌴🌴🌴🌴 Composition: ✨☄️💫
Position	Beginning or end of the post; around a word or phrase for emphasis; “naked” (Provine et al., 2007) or stand-alone	Initial position: 🤩 This! Final position: Happy to be home 💗 Emphasis: 🤩 Fantastic🤩 Stand-alone: 💫

Table 4. Variables used in the Jaccard calculation.

Delta-S (Δs), an extension of Jaccard, “allows the recognition of similar but not identical stylistic choices” (MacLeod and Grant, 2012: 213). Originally used in marine biology and adapted to forensic psychology (Woodhams *et al.*, 2007), MacLeod and Grant (2012) have successfully applied these measures to attribute authorship of Twitter messages. In contrast to the first experiment which uses Jaccard measures, Δs will be used in order to attribute messages to authors based on the emoji types used in the posts rather than on the emoji functions.

In order to simulate a simple authorship attribution analysis, twelve individuals were randomly chosen as authors. These individuals were then randomly grouped into pairs of two, resulting in six pairs. From each individual in each pair, ten posts from the respective data collection period were randomly chosen (‘known’ writings) and treated as a collective in the analysis; further five posts from one individual in each pair from outside the data collection period were chosen to represent the ‘unknown’ writings. The reason for choosing only a minimum of posts from each individual is that it is rarely the case that plenty of material for comparison is available in real world authorship analysis cases, and thus a method has to deliver useful results under extreme conditions (e.g. Johnson and Wright 2014). Future studies will take into account other scenarios, such as authorship attribution with more than two candidate authors, but for this exploratory study, it was decided to focus on a simple authorship attribution problem.

Analysis & Results

In total, 10573 emoji (690 different ones) were used by the individuals in the sample. Only the 25 most common ones occur more than 100 times, and 199 occur only once. Among the 20 most common emoji are seven different types of hearts, with the red heart (U+2764) as the most frequent one. The second most common emoji are the camera (U+1F4F7) and the camera with flash (U+1F4F8), which is likely explained by the fact that Instagram is a photo sharing platform and the camera emoji is used to substitute for the phrase "picture credits".

Emoji Use for Sociolinguistic Profiling

As a first step, the 60 individuals' emoji use related to gender was investigated in more detail. This analysis has revealed the following patterns: females use 56.3% of all emoji in the data set, while males use only 43%. Interestingly, a female dominance in emoji usage also emerges when the variety of emoji use was investigated: the females in the sample use a mean of 63.9 different emoji in their posts, while the males use a mean of 46.3 different emoji.

A further aspect that can be important for sociolinguistic profiling is the estimation of age of an individual. Figure 2 shows that the age group of 14-19 exhibits the highest emoji use with a combined total of 43% of all emoji in the data. It is clearly visible that emoji use steadily declines with increasing age.

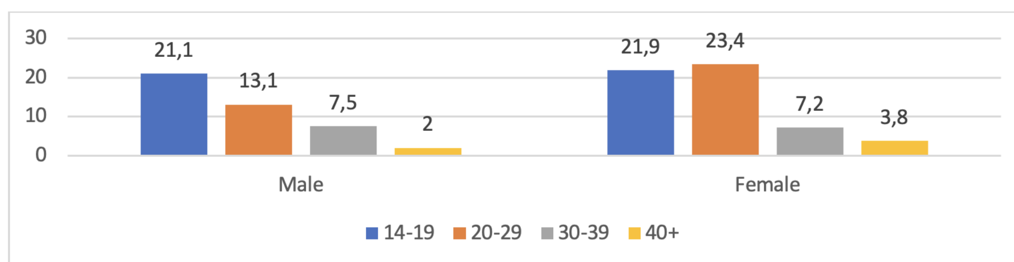


Figure 2. Emoji use according to age and gender (percentages).

Figure 3 further shows that the females in the age group of 20-29 use slightly more different emoji than the younger ones, which is also visible in Figure 2 with regard to overall emoji use. The high variety of emoji use in the category male/40+ is due to one outlier. If this outlier is disregarded, the figure is at 18.5. Thus, the findings support the results of previous studies and, for the first time, demonstrate that emoji use patterns are similar on Instagram with regard to sociodemographic features as it is on other social media platforms or instant messaging, thereby suggesting potential generalizability of results.

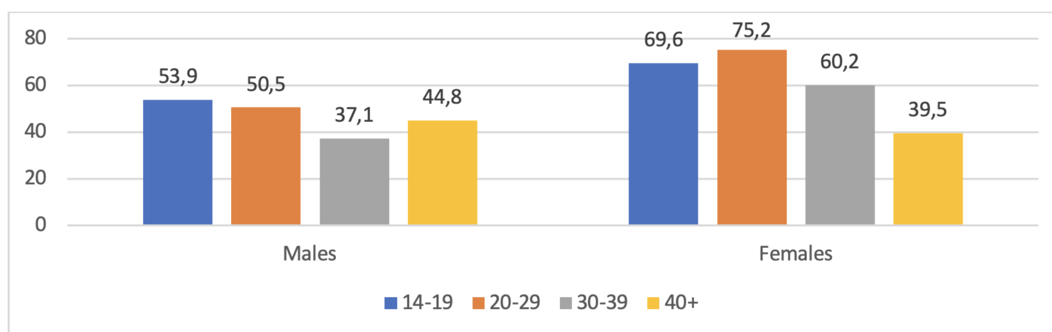


Figure 3. Comparison of mean variety of emoji according to age and gender.

Further highly promising results were obtained when investigating actual emoji use by the different age groups. Firstly, based on the 20 most common emoji in each group, the youngest age group uses hearts most frequently, followed by faces and gestures. In the age group of 30-39-year-olds, gestures are the most common, followed by objects and faces, while hearts are the least common. The age group of 20-29-year-olds employs faces and hearts the most, while gestures are used more sparsely. The oldest age group makes common use of hearts and almost no use of gestures (see Figure 4). Additional research is required in order to allow generalizations of these findings on any level, but these initial results from this exploratory study seem promising for authorship profiling.

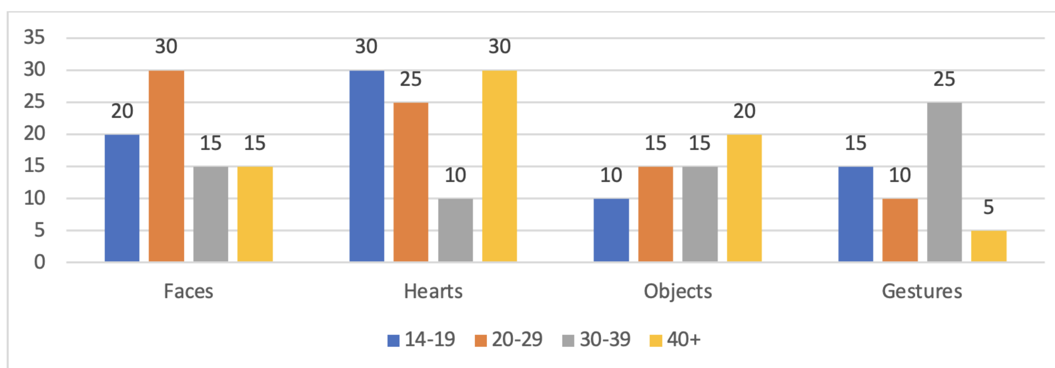


Figure 4. Emoji use according to age group (percentages).

Differentiating Individuals Based on their Use of Emoji

Subsequently, the findings of the Jaccard and Δs calculations will be outlined. The first experiment employs Jaccard's coefficient to attribute 'unknown' posts to one individual in a pair based on the functions of the emoji used (see Table 2). For the second experiment employing Δs, it was not the emoji functions but emoji types that were focused on, since each emoji can potentially serve any of the functions outlined in Table 2.

Experiment 1: Jaccard's Coefficient

As briefly outlined above, 12 individuals were randomly chosen from the data set and then grouped in pairs, resulting in six pairs (i.e. six trials). Each trial consists of 25 posts in total. The analysis proceeded as outlined in section 'Jaccard's Coefficient & Delta-S' above.

Results

In order to illustrate the most important findings from this experiment one trial will serve as an example, but results will be provided for all six trials.

The following examples are taken from Trial 5. Examples (1) to (3) are selected from the 'anonymous' posts of one of the authors in this trial; examples (4) to (6) are taken from the ten randomly selected posts by M.S. and examples (7) to (9) are randomly selected from the ten posts by S.A. The respective categories are indicated in square brackets.

Examples (1) – (3): "Anonymous" posts

(1) [name] is getting married, sorry guys 💍 [complementary] 👰🏼 [reinforcement]

(2) We are now officially open – [name] and I have been working hard, we are really excited! Check it out!!! 💗 [complementary]

(3) My sweet [name], you are the best boyfriend on this planet and I love you! Happy birthday! 💕[reinforcement]

Examples (4) – (6): M.S.'s posts

(4) Americans 🇺🇸 [reinforcement]

(5) This is✈️ boring [substitution]

(6)⚠️ [reinforcement of the picture showing a similar sign]

S.A.'s posts

(7) Only a few more weeks with this guy 🌊 [reinforcement of the picture showing the ocean]

(8) First day in Venice ⭐️ [complementary]

(9) smiling all day 🌼 [reinforcement of the picture showing yellow flowers]

Applying Jaccard's Coefficient to Trial 5 revealed the following results: S.A.'s posts are slightly closer to the anonymous posts in terms of emoji functions compared to M.S.'s posts. Both of them use emoji in a similar way, but M.S. uses emoji in more different positions than S.A. does. In contrast, S.A. uses emoji more directly in conjunction with the respective picture. Table 4 below shows the results of all six trials.

Trial	Jaccard (A + anonymous)	Jaccard (B + anonymous)	Results & Solution
1	0.4 (B.Y.)	0.5 (Co.C.)	Correct
2	0.6 (C.S.)	0.25 (C.P.)	Correct
3	0.6 (S.Cu.)	0.7 (Ca.S.)	Correct
4	0.3 (M.Y.)	0.4 (R.J.)	Incorrect
5	0.5 (M.S.)	0.6 (S.A.)	Correct
6	0.6 (K.C.)	0.6 (J.B.)	Undecided

Table 5. Overview of Results of Experiment 1 (Jaccard).

As indicated in Table 5, four trials can be considered successful, while one trial lead to a wrong result and one is undecided. It can be seen that the outcome of most trials is very close, with the exception of Trial 2, which shows a very clear result. This can be attributed to the test itself not being sensitive enough. That is, the test only differentiates whether specific emoji functions are present or absent but disregards the extent to which a specific function is employed.

A further interesting finding related to the meanings of emoji has emerged. Examples (10) and (11) below show that although both H.B. and J.P. use the rainbow emoji, they use them to denote different things: H.B.'s use of the rainbow emoji is related to her sport: nature, the ocean, and (actual) rainbows. In contrast, J.P.'s use of the rainbow in example (11) is connected to gay pride. Although these are only two examples, the same use of the rainbow by these individuals can be found throughout their posts. Chen et al.'s (2018) study has yielded similar results in relation to gender differences, and these differences might arise from the different contexts these individuals find themselves in. Thus, in this respect, even though the rainbows are potentially used with the same function (even though this is not the case in Examples 10 and 11), they are used to refer to different concepts and ideas, which indicates a qualitative difference the purely quantitative analysis cannot account for.

Example (10) H.B.

> Surfing 🌈 📸: @name

Example (11) J.P.

> Proud to have been a part of this… thank you @name and happy pride 🌈 💗

Further interesting individual differences emerge when looking at the use of emoji by individuals in relation to how emoji are officially defined. Two common emoji in the dataset will serve as examples, namely the

(U+1F919) and the

(U+26A1). The first one is illustrated in Example (12). K.S. uses the emoji with the meaning 'hang loose' and thereby signifies her identity as either Hawaiian and/or member of a particular local community. However, the official name of this emoji is "call me" (emojipedia, online) – a meaning which is not intended in this context.

Example (12) K.S.

Mahalo NY for all the good memories!! 🙏🤙🏻

A further case in point is the following post by J.H. (Example 13), who uses the

emoji to represent movement. Officially, it is described as a symbol for high voltage, or a representation of lightning. Thus, J.H. clearly adapts the latter meaning and uses it in a metaphorical sense.

Example (13) J.H.

Another wave⚡

These examples illustrate the individual use of emoji in the following ways: individuals use emoji with a specific meaning in mind. This specific meaning might be shared by their community of practice (see Eckert 2006), which could lead to other meanings of the same emoji being lost. As Goldman (2018) has pointed out, this has already happened for a small number of emoji, resulting in what he calls 'emoji dialects'. Taking these individual meanings into account might be of high value in authorship analysis, even though further research is required.

Experiment 2: Δs

A detailed description of the procedure of calculating Δs can be found in Woodhams *et al.* (2007) and in Izsak and Price (2001). As mentioned above, Δs allows for the recognition of similarities much more so than Jaccard's coefficient does. Therefore, this measure was chosen to investigate whether or not it is possible to differentiate between authors in the trial data simply based on the emoji types the individuals use. The emoji taken from examples (14) to (22) are presented below for illustration of how they were classified according to Table 3 above.

Examples (14) – (16): "Anonymous" posts

(14) 💍 [Smileys & People/Clothing & Accessories/Female] 👱🏼‍♀️ [Smileys & People/People/Real/Female]

(15) 💗 [Symbols/Hearts]

(16) 💕[Symbols/Hearts]

Examples (17) – (19): M.S.'s posts

(17) 🇺🇸 [Flags/Country]

(18)✈️ [Travel & Places/Modes of Transport]

(19)⚠️ [Symbol/Sign]

Examples (20) – (22): S.A.'s posts

(20) 🌊 [Animals & Nature/Nature/Other]

(21) ⭐️ [Animals & Nature/Nature/Astronomy]

(22) 🌼 [Animals & Nature/Nature/Plants]

All emoji in the selected posts were classified according to this taxonomy. Table 6 shows the results of all trials.

Trial	∆s (A + anonymous)	∆s (B + anonymous)	Results & Solution
1	0.4 (B.Y.)	0.5 (Co.C.)	correct
2	0.33 (C.S.)	0.16 (C.P.)	correct
3	0.32 (S.Cu.)	0.33 (Ca.S.)	incorrect
4	0.05 (M.Y.)	0 (R.J.)	correct
5	0.08 (M.S.)	0.2 (S.A.)	correct
6	0.43 (T.W.)	0.45 (J.B.)	incorrect

Table 6. Overview of Results of Experiment 2 (Delta-s).

Four of the anonymous posts were attributed to their actual authors (trials 1, 2, 4, 5), while two trials resulted in incorrect attributions (trials 3, 6). Since it was not expected that posts can be attributed to individuals based only on the actual emoji used, these results are quite surprising. These results indicate that it is indeed worth looking at emoji types in addition to emoji functions, particularly when emoji meanings are taken into account as well. Moreover, the taxonomy of emoji has much potential for development, which will also very likely improve results.

Discussion & Conclusions

The overview of existing literature has revealed that emoji are used quite differently by different groups of people; and yet, emoji have been largely neglected in authorship analysis. For the purpose of filling this research gap, two analyses were carried out: first of all, the data was used to reveal whether or not patterns identified in previous studies, particularly those relating to gender and age, would also be found in posts on the social media platform Instagram. The findings support the results of previous studies and they indicate similar trends: females and younger people use both more and a larger variety

of emoji. Another promising avenue for future research is the investigation of emoji functions in relation to age and gender in order to investigate whether differences exist in this respect as well. Additionally, it was shown that even though people use the same emoji, they might use them in different contexts, thereby denoting different meanings. An important issue for authorship analysis that needs to be investigated further is how far individuals might be aware of different meanings of emoji and whether or not they consciously switch between these meanings, or whether they consistently use the same emoji with the same meaning. In the latter case, several of the more ambiguous emoji could prove useful in authorship analysis. If these findings are tested and developed further, they could make a valuable contribution to sociolinguistic profiling tasks in digital media.

Secondly, in the context of a simple mock authorship attribution task, an experiment was conducted in order to find out about individual's use of emoji with regard to the emoji functions. The calculation of Jaccard's coefficient demonstrates that an investigation of emoji functions can indeed be valuable for authorship analysis, even though the classification system itself still needs to be improved. The main limitation of this study remains the inter-rater reliability. Importantly, however, this pilot study reveals the potential of an analysis of an individual's emoji use in addition to a purely linguistic analysis. Regarding emoji functions, this paper demonstrates the importance of complementary qualitative analyses in conjunction with quantitative analyses. Neglecting a qualitative analysis in this context could result in a loss of valuable information and might even mislead the analysis. Additionally, using complementary qualitative analyses that can easily be explained to a judge or jury is likely to be viewed favorably in actual forensic cases (e.g. Grant and Baker 2001; Solan 2013; Grant and MacLeod 2020).

Thirdly, Δs was used to investigate an individual's use of emoji regardless of the emoji function. The relative success of the second experiment might be due to emoji being used as identity markers, as previous research has indicated (e.g. Robertson *et al.* 2018; Ge 2019). Further, this portrayal of identity might be particularly strong on Instagram (Leaver *et al.*, 2020), which could explain the relatively good results for this platform.

Overall, it was the aim to investigate the potential of emoji as authorship markers. Even though any other linguistic markers were neglected in this paper, the results are very promising. Nevertheless, the classification systems require further refinement to be used in forensic cases. As this paper demonstrates, even though emoji use should not be relied upon as an individual marker of authorship, it should not be neglected either and can serve as a valuable addition to authorship analysis methods.

Notes

[1]Codes in brackets refer to the original Unicode Codepoints.

[2]Agreeableness, openness, extroversion, conscientiousness, and neuroticism (see, e.g., Roccas *et al.* 2002).

[3]The first world cluster includes North America, Western Europe, the Russian Federation, and Australia; the second world cluster covers most of South America, India and China, Eastern Europe, Morocco, Algeria, and Tunisia; the third world cluster includes Angola, Nigeria, Sudan, Jordan, Saudi Arabia, Yemen, Pakistan, Nepal, and the Philippines; the remaining African states are subsumed under the fourth world cluster (Ljubesic and Fiser, 2016: 86)

[4]The description of the 'Discourse Management Placement' category is based on research carried out by Evans (2017) and Danesi (2016). Due to this category overlapping to a large degree with some of the other categories, the discourse management function was later abolished for the purposes of this study (see Table 2) and later reintroduced and adapted an additional feature of emoji use (see Table 4).

References

Ai, W., Lu, X., Liu, X., Wang, N., Huang, G. and Mei, Q. (2017). Untangling Emoji Popularity through Semantic Embeddings. In *Proceedings of the 11th Int'l AAAI Conference on Web and Social Media*, 2–11.

Ayers, J. W., Caputi, T. L., Nebeker, C. and Dredze, M. (2018). Don't quote me: reverse identification of research participants in social media studies. *Digital Medicine*, 1(1), 30.

Barbieri, F. and Camacho-Collados, J. (2018). How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter. In *Proceedings of the 7th Joint Conference on Lexical and Computational Semantics*, 101–106.

Barbieri, F., Ronzano, F. and Saggion, H. (2016). What does this emoji mean? In *Language Resources and Evaluation Conference, LREC.*, Portoroz, Slovenia.

Bell, A. (1984). Language Style as Audience Design. *Language in Society*, 13(2), 145–204.

Brennan, M. and Greenstadt, R. (2009). Practical attacks against authorship recognition techniques. In *Proceedings of the 21st Innovative Applications of Artificial Intelligence Conference*, 60–66.

Chen, Z., Lu, X., Ai, W., Li, H., Mei, Q. and Liu, X. (2018). Through a Gender Lens. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 763–772, New York, New York, USA: ACM Press.

Clift, R. (2016). *Conversation Analysis.* Cambridge: Cambridge University Press.

Coulthard, M. (2006). …and then… Language Description and Author Attribution. Inaugural Lecture at Aston University.

Coulthard, M. (2008). By their words shall ye know them: On linguistic identity. In C. R. Caldas-Coulthard and R. Iedema, Eds., *Identity Trouble.* New York, NY: Palgrave Macmillan, 143–170.

Coulthard, M., Grant, T. and Kredens, K. (2011). Forensic linguistics. In R. Wodak, B. Johnstone and P. Kerswill, Eds., *The SAGE handbook of sociolinguistics.* London: SAGE, 531–544.

Coulthard, M., Johnson, A. and Wright, D. (2017). *Introduction to Forensic Linguistics.* London England & New York, NY: Routledge.

Danesi, M. (2016). *The Semiotics of Emoji.* Bloomsbury.

Dimson, T. (2015). Emojineering Part 1: Machine Learning for Emoji Trends. *Instagram Blog, Engineering.*

Donato, G. and Paggio, P. (2017). Investigating Redundancy in Emoji Use: Study on a Twitter Based Corpus. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 118–126.

Eckert, P. (2006). Communities of Practice. In K. Brown, Ed., *Encyclopedia of Language & Linguistics.* Amsterdam, The Netherlands: Elsevier Science, 683–685.

Eder, M. (2015). Does size matter? Authorship attribution, small samples, big problem. *Digital Scholarship in the Humanities*, 30(2), 167–182.

Ehrhardt, S. (2018). Authorship attribution analysis. In J. Visconti and M. Rathert, Eds., *Handbook of communication in the legal sphere, 14.* Berlin, Germany: de Gruyter, 169–200.

Evans, V. (2017). *The Emoji Code.* New York: Picador.

Gawne, L. and McCulloch, G. (2019). Emoji as Digital Gestures. *Language@Internet*, 17.

Ge, J. (2019). Emoji Sequence Use in Enacting Personal Identity. In *Companion Proceedings of The 2019 World Wide Web Conference*, 426–438, New York, NY, USA: ACM.

Goldman, E. (2018). Emojis and the law. *Washington Law Review*, 92, 1227–1291.

Grant, T. (2010). Txt 4n6: Idiolect free authorship analysis. In M. Coulthard and A. Johnson, Eds., *Routledge Handbook of Forensic Linguistics.* Abingdon and New York: Routledge.

Grant, T. (2013). Txt 4n6: Method, consistency, and distinctiveness in the analysis of SMS text messages. *Journal of Law and Policy*, 21(2), 467–494.

Grant, T. and Baker, K. (2001). Identifying reliable, valid markers of authorship: A response to Chaski. *Forensic Linguistics*, 8(1), 66–79.

Grant, T. and MacLeod, N. (2020). *Language and Online Identities.* Cambridge: Cambridge University Press.

Heydon, G. (2019). *Researching Forensic Linguistics.* London & New York: Routledge.

Instagram, (2018). Privacy and Security.

Ishihara, S. (2017). Strength of linguistic text evidence: A fused forensic text comparison system. *Forensic Science International*, 278, 184–197.

Izsak, C. and Price, A. R. G. (2001). Measuring β-diversity using a taxonomic similarity index and its relation to spatial scale. *Marine Ecology Progress Series*, 215, 69–77.

Johnson, A. and Wright, D. (2014). Identifying idiolect in forensic authorship attribution: an n-gram textbite approach. *Language and Law / Linguagem e Direito*, 1(1), 37–69.

Layton, R., Watters, P. and Dazeley, R. (2010). Authorship attribution for Twitter in 140 characters or less. *IEEEI*, 1–8.

Leaver, T., Highfield, T. and Abidin, C. (2020). *Instagram.* Cambridge: Polity.

Li, W., Chen, Y., Hu, T. and Luo, J. (2018). Mining the Relationship between Emoji Usage Patterns and Personality. In *Proceedings of the 12th International AAAI Conference on Web and Social Media*, 648–651.

Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, 740–755.

Ljubesic, N. and Fiser, D. (2016). A Global Analysis of Emoji Usage. In *Proceedings of the 10th Web as Corpus Workshop and the Empirist Shared Task*, 82–89.

MacLeod, N. and Grant, T. (2012). Whose Tweet? Authorship analysis of micro-blogs and other short-form messages. In *Proceedings of the International Association of Forensic Linguists' 10th Biennial Conference*, 210–224, Birmingham, UK: Aston University.

McCulloch, G. (2019). *Because Internet.* New York: Riverhead Books.

Miller, H., Kluver, D., Thebault-Spieker, J., Terveen, L. and Hecht, B. (2017). Understanding Emoji Ambiguity in Context: The Role of Text in Emoji-Related Miscommunication. In *Proceedings of the 11th International AAAI Conerence on Web and Social Media*, 152–161.

Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L. and Hecht, B. (2016). "Blissfully Happy" or "Ready to Fight": Varying Interpretations of Emoji. In *Proceedings of the 10th AAAI Conference on Web and Social Media*, 259–268.

Na'aman, N., Provenza, H. and Montoya, O. (2017). MojiSem: Varying linguistic purposes of emoji in (Twitter) context. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics- Student Research Workshop*, 136–141.

Nini, A. (2018a). An authorship analysis of the Jack the Ripper letters. *Digital Scholarship in the Humanities*, 33(3), 621–636.

Nini, A. (2018b). Developing forensic authorship profiling. *Language and Law / Linguagem e Direito*, 5(2), 38–58.

Orebaugh, A. and Allnutt, J. (2009). Classification of instant messaging communications for forensic analysis. *The International Journal of Forensic Computer Science*, 1, 22–28.

Page, R., Barton, D., Unger, J. W. and Zappavigna, M. (2014). *Researching Language and Social Media.* London & New York: Routledge.

Robertson, A., Magdy, W. and Golwater, S. (2018). Self-representation on Twitter using emoji skin color modifiers. In *Proceedings of the 12th International AAAI Conference on Web and Social Media*, 680–683.

Roccas, S., Sagiv, L., Schwartz, S. H. and Knafo, A. (2002). The Big Five personality factors and personal values. *PSPB*, 28(6), 789–801.

Rocha, A., Scheirer, W., Forstall, C., Cavalcante, T., Theophilo, A., Shen, B., Carvalho, A. R. B. and Stamatatos, E. (2016). Authorship attribution for social media forensics. *IEEE Transactions on Information Forensics and Security*, 12(1), 1–30.

Solan, L. (2013). Intuition versus algorithm: The case of forensic authorship attribution. *Journal of Law and Policy*, 21, 551–576.

Sousa Silva, R. (2018). Computational Forensic Linguistics: An Overview. *Language and Law / Linguagem e Direito*, 5(1), 80–96.

Sousa-Silva, R., Laboreiro, G., Sarmento, L., Grant, T., Oliveira, E. and Maia, B. (2011). 'twazn me!!! ;(' Automatic Authorship Analysis of Micro-Blogging Messages. In R. Muñoz, A. Montoyo and E. Métais, Eds., *Lecture Notes in Computer Science 6716 Springer 2011*, volume Natural La, 161–168, Berlin and Heidelberg: Springer – Verlag.

Statista, (2020). Distribution of Instagram users worldwide as of April 2020, by age and gender.

Townsend, L. and Wallace, C. (2016). *Social media research: A guide to ethics.* University of Aberdeen & ESRC.

Vidal, L., Ares, G. and Jaeger, S. R. (2016). Use of emoticon and emoji in tweets for food-related emotional expression. *Food Quality and Preference*, 49, 119–128.

Woodhams, J., Grant, T. and Price, A. R. G. (2007). From marine ecology to crime analysis: improving the detection of serial sexual offenders using a taxonomic similarity measure. *The Journal of Investigative Psychology and Offender Profiling*, 4, 17–27.

Zappavigna, M. (2013). *Discourse of Twitter and Social Media. How We Use Language to Create Affiliation on the Web.* London: Bloomsbury.

"You Have the Right to Keep Quiet": Translation Inadequacies in Nevada's Spanish Miranda Warnings

Ashley Rose Keaton

University of Nevada, Reno, USA

https://doi.org/10.21747/21833745/lanlaw/7_1_2a10

Abstract. *In the landmark case Miranda v. Arizona (1966), the United States Supreme Court required law enforcement agencies to advise all suspects of their "Miranda warnings," or Constitutional protections, prior to interrogation. Previous research demonstrates that the Miranda warnings in the United States are largely unregulated and highlights how inadequate translations can impact comprehensibility. The present study evaluates the translation problems found in the Spanish Miranda warnings in Nevada, including complex grammar, formal lexicon, and the assumption by law enforcement agencies that detainees will have a baseline familiarity with their rights. In some instances, these errors are significant enough that they might preclude a listener from understanding their Constitutional rights. This study suggests specific areas where the Spanish Miranda warnings require specific revision in order to conform to case law and best practices based on research.*

Keywords: *Miranda warning, reading of the rights, Miranda rights, Spanish translation, caution.*

Resumo. *No caso paradigmático Miranda v. Arizona (1966), o Supremo Tribunal dos Estados Unidos determinou que as forças policiais teriam de passar a informar todos os suspeitos das suas "advertências de Miranda," ou proteções Constitucionais, antes de qualquer interrogatório. Estudos anteriores demonstraram que as advertências de Miranda nos Estados Unidos não são, em grande parte, regulamentadas e realçam que as traduções inadequadas podem influenciar a compreensibilidade. O presente estudo avalia os problemas de tradução existentes nas advertências de Miranda em espanhol no Estado de Nevada, incluindo gramática complexa, léxico formal e o pressuposto, por parte das forças policiais, de que os detidos possuem conhecimento de base dos seus direitos. Nalguns casos, estes erros são suficientemente sérios para impedir o interlocutor de compreender os seus direitos Constitucionais. Este estudo sugere algumas áreas específicas das advertências de Miranda em espanhol que necessitam de revisão específica para, com base na*

investigação, cumprir a lei e as boas práticas.

Palavras-chave: *Advertência de Miranda, leitura dos direitos, Direitos de Miranda, tradução para espanhol, advertência.*

Introduction

This study examines linguistic elements of the Spanish-language *Miranda* warnings in Nevada with a focus on two specific issues: the translational relationship between the respective English and Spanish *Miranda* warnings from a given law enforcement agency, and the grammatical and lexical quality of an agency's Spanish-language warning. A comparison of the translations between the English and Spanish versions from each agency reveals substantive errors as well as grammatical and lexical problems caused by mistranslation. The present study provides evidence that in some instances, the Spanish *Miranda* warnings used by Nevada law enforcement agencies do not satisfy the legal requirements and best practices for the reading of the rights in the United States.

Relevant Case Law

The Supreme Court decision in *Miranda v. Arizona* (1966) set the precedent that law enforcement agencies must advise all suspects of their Constitutional rights prior to interrogation. As a result, law enforcement agencies in the United States were left to determine how they would fulfill these requirements, now known as the "*Miranda* rights." Various scripts, collectively referred to as the "*Miranda* warnings," were produced in order to satisfy the guidelines provided by the decision in *Miranda v. Arizona* (1966):

> The foremost requirement, upon which later admissibility of a confession depends, is that a four-fold warning be given to a person in custody before he is questioned, namely, that he has a right to remain silent, that anything he says may be used against him, that he has a right to have present an attorney during the questioning, and that, if indigent he has a right to a lawyer without charge. To forgo these rights, some affirmative statement of rejection is seemingly required. (p. 33)

Following this decision, several more cases were brought to the Supreme Court that further codified the process of the reading of the rights to suspects. In 1989, the Supreme Court heard the case *Duckworth v. Eagan*, in which the petitioner alleged that he was not properly advised of his rights because the officer used an unusual phrasing in the *Miranda* warnings. The Court decided against the petitioner, thereby broadening the already imprecise requirements proposed in *Miranda v. Arizona* (1966).

Nine years later, the Ninth Circuit Court of Appeals decided law enforcement officers must verify that the suspect understands English well enough to "knowingly and intelligently" waive their *Miranda* rights (*United States v. Garibay*, 1998). This decision created the precedent that if non-native English speakers cannot understand their rights in English, they must be communicated in the suspect's native language.

In 2013, the Ninth Circuit Court of Appeals heard the case *United States v. Botello Rosales*, where a Spanish-speaking respondent alleged that he had not been advised of his rights because the officer had explained them to him in near-incomprehensible Spanish. The Court decided all foreign-language readings of the *Miranda* warnings must "reasonably convey" the meaning of the rights. To date, this decision is the most specific standard for reading the rights in Spanish.

These cases represent only a small slice of the decisions based on the case law in *Miranda v. Arizona* (1966), and for this reason, discussion continues about the meaning of the *Miranda* rights and what the *Miranda* warnings should convey (see *Commonwealth v. Ochoa*, 2014; *State v. Carrasco-Calderon*, 2008; *Riviera-Reyes v. Commonwealth*, 2006; *State v. Ortez*, 2006; *State v. Teran*, 1993; *United States v. Castro-Higuero*, 2007; *United States v. Higareda-Santa Cruz*, 1993; and *Wisconsin v. Santiago*, 2017).

Linguistic Issues Surrounding the Miranda Warnings

Research concerning systemic problems with the *Miranda* warnings is divided here into three subsections: problems with the length and order of the warnings, problems with the language used in the warnings, and problems with translations of the rights into other languages. These issues are compounded by the significant variation in quality and accuracy of different *Miranda* warning texts. While it is possible that many different, high-quality versions of the *Miranda* warnings exist, research demonstrates that variation in the *Miranda* warnings can have a somewhat unpredictable effect on a listener’s ability to understand their rights.

Before discussing the various problems with the warnings texts themselves, it is best to highlight the importance of understanding the *Miranda* warnings prior to interrogation. While police officers might interpret many different responses by a suspect to signal that they understand their rights, officers are far less likely to recognize that a suspect is signaling that they want to invoke their rights during interrogation (Ainsworth, 2008). This is supported by Mason’s (2013) work, where suspects who made indirect requests to exercise their rights were frequently ignored or challenged by interrogating officers. Pavlenko’s (2008) case study produced similar findings. Even if suspects understand their rights perfectly, it is unlikely that they will succeed in invoking them; therefore, if suspects have an incomplete understanding of their rights, the likelihood that they will utilize them is further diminished.

Some scholars argue that the *Miranda* warnings are not structured in a way that enables suspects to easily understand them. Shuy (1997) outlines how the current order of the *Miranda* warnings is not conversationally logical, and therefore decreases listener comprehension. Kurzon (2000) demonstrates that a dilemma exists between brevity and completeness; that is, the shorter the warning is, the more difficult it is to unambiguously interpret, but a longer warning may cause the listener to lose focus and ignore important information. Eades and Pavlenko (2016) suggest that suspects would better understand the importance and meaning of the *Miranda* warnings if they were delivered in a question-and-answer format, rather than as a rote recitation.

Moreover, the *Miranda* rights themselves use language that is inaccessible to the average listener. In an attempt to resolve these comprehension barriers, *The Communication of Rights Group* (CoRG) treatise (2016) provides guidelines for communicating legal rights to non-native speakers of English. In order to accommodate the widest possible range of language proficiency, the rights should use the simplest grammar constructions available. Conditional statements and prepositional phrases should be avoided because they require the listener to parse multiple levels of embedding (Eades and Pavlenko, 2016; Gibbons, 2001). Shuy (1997) demonstrates that prepositional phrases are frequently used in the *Miranda* warnings and are likely to be recognized but not understood by the listener. Further, Pavlenko (2008) and Shuy (1997) argue that suspects who do not com-

pletely understand the meaning of the *Miranda* warnings will assume that they got "the gist" since they recognized words within the sentence (see also Rock 2007.

Additionally, the *Miranda* warnings typically use formal lexicon characteristic of the legal register. Suspects who are unfamiliar with legal language are more likely to not understand it, which puts first-time offenders and recent immigrants at a disadvantage (Pavlenko, 2008; Rock, 2007; Rogers *et al.*, 2007). To avoid an uneven understanding of the warnings by detainees, the CoRG proposals suggest that the *Miranda* warnings should use plain language wherever possible, as well as high-frequency lexical items and phrases that are not culture-specific (Eades and Pavlenko 2016; see also Rock 2007). This position is supported by research from Rogers, Hazelwood, Sewell, Blackwood, and Rogstad (2009b), which demonstrates a relationship between higher grade-level difficulty of *Miranda* vocabulary and worse average comprehension by detainees.

Further, Rogers *et al.* (2007) indicate that some words in the *Miranda* warnings might have significantly different reading levels depending on the context. For instance, the word "right" (*derecho* in Spanish), is a polysemic word that can evoke both a sense of direction and legal right. When used to convey direction, the word right has an elementary level of difficulty, whereas it has an eighth-grade level difficulty when used to mean legal claim. Rogers *et al.* (2007) demonstrate that the grade-level difficulty of the *Miranda* warnings can vary from a third-grade reading level to a collegiate reading level, depending on the jurisdiction. Thus, a detainee's ability to comprehend the *Miranda* rights might depend on the vocabulary used in that particular jurisdiction, and their understanding of the correct use of the word in context.

Rogers *et al.* (2007) and Rogers *et al.* (2008) demonstrate that the *Miranda* warnings in the United States exhibit a great degree of variance in substantive content, length, and quality. While these variations already present challenges to English-speaking detainees, these problems are compounded when the warnings texts must be translated into another language. The American Bar Association (ABA) identified this issue in ABA Resolution 110 (2016) and suggests a standard Spanish translation that could be used by all law enforcement agencies. However, the Spanish *Miranda* warning proposed in the Resolution includes translation errors, such as omission of the requisite subject pronouns, and is written in only one dialect of Spanish; therefore, it fails to provide an adequate template for law enforcement agencies to implement.

If anything, the imperfect translation by the ABA suggests that insufficient attention has been given to the Spanish *Miranda* warnings in the United States. Rogers *et al.* (2009a) analyzed the Spanish *Miranda* warnings from 121 jurisdictions and indicated a number of significant translation errors. However, Rogers *et al.* (2009a), do not offer an in-depth textual analysis in their study. The current study aims to build upon the existing evidence of translation inadequacies in the Spanish Miranda warnings, and identify persistent problems within these translations.

Given the requirements of *Miranda v. Arizona (1966)* and subsequent case law combined with the linguistic issues seen in previous research, there is significant potential for the Spanish *Miranda* warnings to be incomprehensible to the listener. This is especially true in jurisdictions where a linguist or court-certified interpreter was not involved in the construction of a standard Spanish *Miranda* warning. The present study, therefore, seeks to answer the following questions: 1) Is there a standard *Miranda* warning in

Nevada, 2) Is there a Spanish standard translation of the *Miranda* warnings in Nevada, and 3) Are the Spanish translations of the *Miranda* warnings adequate to ensure listener comprehension?

Methodology

Data Collection

Nine of the twenty-two law enforcement agencies in Nevada were initially contacted for participation in this study. These law enforcement agencies are located in the most populous regions of Nevada (that is, the northwest and south). From these nine, those that were able to provide both a Spanish and English version of their Miranda warning were included in this study. The resulting sample includes five law enforcement agencies: two rural sheriff's offices, two urban police departments, and the state highway patrol. These agencies are anonymized in this report by letters A through E.

The data in this study was collected through visits to law enforcement agency headquarters; each law enforcement agency was asked for a physical copy of its *Miranda* warnings in English and Spanish. Some law enforcement agencies were able to provide physical copies of the warnings during that visit, while others required email correspondence in order to obtain the English and Spanish versions of the warnings.

The scope of this study is limited to written translations only and assumes, for the purposes of analysis (unlikely as it may be), that the listener would hear the *Miranda* warnings exactly as they appear on the card. This study does not examine recordings of the *Miranda* warnings as read by law enforcement officers, nor does it attempt to comment on the ability of law enforcement officers to speak Spanish with sufficient proficiency to guarantee the warnings were read correctly.

Data Analysis

The English and Spanish versions from each agency were reviewed to ensure each of the five *Miranda* rights was present. Next, the samples were reviewed for lexical or grammatical errors caused by translation inadequacies. Many small translation errors were present in the data collected; the only translation errors described in the current study are those which, in the view of the analyst, could substantially preclude a listener's ability to understand the intended meaning of the rights.

Results

Absence of Standard Warning

Based on the data collected, Nevada does not have a standard English *Miranda* warning. Of the nine law enforcement agencies initially surveyed, no two law enforcement agencies had an identical English Miranda warning. While it appears that two out of five law enforcement agencies in this study have an official, agency-specific version, the other three do not. Further, data suggests that even at agencies with standard *Miranda* warnings, some officers use other versions that do not match the official agency warning (for the complete set of data, see Appendix A). As one example, an officer at Agency D showed me that he carried the agency's official English language *Miranda* warning as well as a different version he had obtained elsewhere, because he believed the unofficial version was more descriptive.

Further, law enforcement officers at Agencies A, B, and C did not have the Spanish *Miranda* warnings readily available, nor did they seem familiar with where to find their agency’s version of the warning. An officer at Agency A was able to find a version of the *Spanish* Miranda warnings, but mentioned that he frequently would do an ad-hoc interpretation of the English *Miranda* warning into Spanish. The officer explained, because he was a heritage Spanish speaker, he felt “more than confident” in his ability to accurately interpret the *Miranda* warnings. As a measure to ensure that his interpretation was accurate, the officer would type the warnings into Google Translate and ask the Spanish speaker to read the translated text. At Agency B, an officer stated that he was not sure whether police officers were issued a copy of the *Miranda* warnings by his law enforcement agency or whether they were supposed to obtain them by their own accord, but he did not indicate where an officer might obtain a *Miranda* card.

More generally, law enforcement officers seemed unaware that different versions of the *Miranda* warnings existed. At Agency E, when I requested a copy of the *Miranda* warning, the police officer looked at me skeptically and asked, “you know they’re all the same, right? Like the decision in *United States v. Miranda*, it makes them all the same.” A second officer at the same agency firmly challenged the idea that there were different *Miranda* warnings until he was shown evidence to the contrary; when shown the warnings from Agency C, he remarked that Agency C was giving “too much information.” These interactions suggest that law enforcement officers are not aware that multiple versions of the warnings exist.

There is also no standard version for the Spanish translation. Most law enforcement agencies in this study use Spanish translations that closely mirror their English versions. However, Agency C’s warnings have no apparent similarity between the English and Spanish versions, nor do they closely resemble the Spanish warnings from another law enforcement agency in the study.

Quality of Translated Miranda Warnings

To examine whether the Spanish translations adequately meet the requirements in *Miranda v. Arizona* (1966), the results have been divided into the four sections described in the decision of the case: 1) you have the right to remain silent, 2) anything you say may be used against you, 3) you have a right to have present an attorney during the questioning, and 4) if indigent, you have a right to a lawyer without charge. These sections are based on the precise language from the decision in *Miranda v. Arizona* (1966). The Court’s decision also requires law enforcement agencies to “inform accused persons of their right of silence and assure a continuous opportunity to exercise it,” which has been noted in a fifth section (5).

1. *You have the right to remain silent.*

Agency D translates *remain silent* as *mantenerse callado*, which would translate as *keep quiet. Mantenerse* is not conversationally equivalent to *remain*, and would have questionable meaning to a Spanish speaker; in the federal court case *United States v. Higareda-Santa Cruz* (1993), the court decided *mantenerse* was not an acceptable translation of *remain.* Additionally, *callado* is more often used to mean *reserved* or *quiet* than silent, and is only a medium-frequency Spanish word, whereas *silencio* is more common (Collins, 2020). The *Miranda* Vocabulary Scale (MVS) evaluates the word silent as one of the highest-import words in the *Miranda* warnings, earning a five-point score on a scale of

one to five; because of its importance in understanding this right, it is critical that the word silence be translated accurately (Rogers *et al.*, 2009b).

2. *Anything you say may be used against you.*

Agencies C and E use the phrasing *can and will* in their English version of the *Miranda* warnings, but translate this as *puede ser*, which in Spanish is only equivalent to *can*. *United States v. Botello-Rosales* (2013) supports that *puede ser* is not equivalent to *will*.

Further, Agency E translated *court of law* from the original English version as *corte de ley*, which is not a meaningful translation according to *State v. Ortez* (2006).

3. *You have the right to have an attorney present during the questioning.*

Agency C has a substantive error in the Spanish version of the right to counsel; that is, it does not specify when the listener can have an attorney or what the attorney is for. The court in *Riveria-Reyes v. Commonwealth* (2006) determined that the English and Spanish *Miranda* warnings at the same agency must have identical content, but Agency C’s warnings do not meet this standard.

English version	Spanish version	Back-translated into English
You have the right to speak with an attorney before answering any questions,		
and to have an attorney present with you while you answer any questions.	Usted tiene el derecho de tener un abagado presente.	You have the right to have a lawyer present.

Table 1. Agency C *Miranda* warning, third prong.

Agency A omits the second-person singular pronoun *usted*, or *you*, when it describes the listener’s right to an attorney. Critically, Agency A’s two-part warning lacks a subject pronoun, and does not include any anaphoric references in either sentence that would provide necessary clarity. It is possible that a suspect would hear this warning and interpret it to mean that someone else has the right to an attorney rather than themselves.

Agency B also omits the subject pronoun *usted*. Agency B does make passing reference to the subject; however, the subject is only found in a cataphoric reference within a subordinate clause. This sentence also features three levels of embedding following the main clause.

Agency B translates the second subordinate clause of the warning as *mientras se le interroga*. In Spanish, *se* and *le* can serve as second or third person object pronouns. *Se* can also function as a syntactic element to reference the impersonal *se*. The function of these pronouns in this clause is ambiguous and could be clarified by including another referent to the subject, such as *mientras se le interroga a usted*. The current representation of this statement requires the listener to hypothesize about who this right applies to.

English version	Spanish version	Back-translated into English
You have the right to talk to a lawyer	Tiene el derecho de hablar con un abogado	[You] have the right to talk with a lawyer
and have him present with you	y que esté presente con usted	and that [s/he] is present with you
while you are being questioned.	mientras se le interroga.	while [ambiguous] interrogate.

Table 2. Agency B *Miranda* warning excerpt, third prong.

4. *If indigent, you have the right to an attorney without charge.*

As is the case with the previous warning, many law enforcement agencies in this study omit the subject of the main clause or use ambiguous pronouns instead of explicitly stating the agent of the action. Agency A uses the phrasing *se le asignará* to mean *[an attorney] will be appointed [for you]*, but this does not provide sufficient context for the listener to derive precisely that the pronoun *le* refers to *you*.

English version	Spanish version	Back-translated into English
If you cannot afford an attorney,	Si usted no puede pagar un abogado	If you cannot pay a lawyer
one will be appointed before questioning.	uno se le asignará antes del interrogatorio.	One will be assigned [ambiguous] before the interrogation.

Table 3. Agency A *Miranda* warning excerpt, prong four.

Agencies C and E omit the necessary diacritic marks over the words *asignará* and *adjudicará* (*will assign* and *will award* or *will determine*), respectively, and thereby change the verb mood and tense from indicative to subjunctive and from future to past. Instead of the intended meaning *one will be assigned to you*, this clause may be taken to mean *one might be assigned to you*, and implies a significant degree of uncertainty or doubt that the event will take place. Additionally, *adjudicara* is almost exclusively used in the legal register in Spanish. For many Spanish speakers without a college education, *adjudicara* might be a completely unfamiliar word. The use of a rare translation of *assign*, which earns a score of 4.33 on the *Miranda* Vocabulary Scale, could cause a suspect to misunderstand their right to an attorney.

On the whole, the fourth *Miranda* warning is the most likely to be presented in complex grammar constructions. Although all law enforcement agencies in this study use at least one conditional statement in their translations of the fourth *Miranda* warning, Agency E has a particularly complex statement that contains four subordinate clauses within five levels of embedding:

This sentence requires the listener to parse the agent and the action of a main clause, which is buried in a haystack of subordination. Then the listener must interpret how

Part 4 of Agency E's English-language Miranda Warning				
1-Main	2	3	4	5
	If you do not have the means			
		to pay a lawyer		
the court will appoint one for you without cost				
			before starting the interrogation	
				if you want.
Part 4 of Agency E's Spanish-language Miranda Warning				
1-Main	2	3	4	5
	Si usted no tiene medios			
		para emplear un abogado		
el juzgado de la corte [sic] adjudicara [sic] uno para Usted				
			antes de iniciar el interrogatorio	
				si Usted quiere.
A back-translation of Part 4 of Agency E's Spanish-language Miranda Warning				
1-Main	2	3	4	5
	If you do not have means			
		to hire a lawyer		
the tribunal of the Court [sic] will determine [sic] one for you				
			before starting the interrogation	
				if you want.

Table 4. Agency E *Miranda* excerpt, prong four.

each of the four surrounding clauses relate to the court will appoint one for you without cost. Rimmer (2009) offers subordination as one method to demonstrate sentence complexity; here, this measure illustrates the unusually great complexity of this warning.

Agency B's warning is very similar to Agency E's, but without any of the appropriate subject pronouns to render a comprehensible sentence. The subject of the sentence, *usted*, is omitted without the necessary anaphoric references to define it. The object noun *abogado* is only present in the first of the three clauses in the sentence. Though it is possible to connect the pronouns *le*, *le*, and *uno* to the referent *abogado*, it presents a significant challenge to the listener.

The clause *if you wish one* in Agency B's warning is especially ambiguous in the Spanish translation because it relies on embedded anaphoric references to define the subject noun and object noun, and because the verb desea could refer to the second person singular *usted*, or the third person singular *él* or *ella* (*he* or *she*). Though the listener may recognize all the words used in this sentence and therefore believe they comprehended it, it is highly unlikely that a listener could understand the meaning of this sentence (Shuy, 1997).

Lastly, for the clause *if you cannot afford*, most law enforcement agencies in this study use low-frequency or legalistic translations. Agency B's translation uses *costear* to mean *to pay*. *Costear* has multiple meanings in the Spanish language and is not equivalent to *afford* in the English version. Agency E uses the complex phrase *tiene medios*

English version	Spanish version	Back-translated into English
If you cannot afford a lawyer	Si no puede costear los gastos de un abogado,	If cannot finance the expenses of a lawyer,
one will be appointed to represent you before any questioning	Se le asignará uno para que le represente ante cualquier interrogatorio	[ambiguous pronoun] will assign one so that [ambiguous pronoun] represent before any interrogation
if you wish one.	si desea uno.	if want one.

Table 5. Agency B *Miranda* warning excerpt, prong four.

para emplear (*have the means to employ*) when *puede pagar* is considerably simpler and uses higher-frequency Spanish words. Agency C uses the formal *si no tiene usted recursos* (*if you do not have the resources*), when *dinero* would work as well. Additionally, Agency D uses *fondos* as the translation for *funds. Fondos* has a variety of distinct meanings in Spanish, and *money* is listed as the tenth out of twelve possible definitions in the COBUILD (2020) dictionary. Each of these translations is unsatisfactory for the high-import word afford, which earns a score of 4.67 on the *Miranda* Vocabulary Scale.

5. *Continuous opportunity to exercise your rights.*

Agency E is the only agency in this study that includes any language to suggest to the listener that their rights are ongoing. Many of the Spanish warnings in this study contain conditional clauses and subjunctive verbs that imply uncertainty that the rights mentioned therein are absolute, much less that a detainee can invoke them at any time. Meanwhile, most agencies use language that implies that an interrogation is imminent or inevitable, like *while we interrogate you* and *during the interrogation.* This creates the implicit message that an interrogation by police is compulsory, when in fact the accused person is not obligated to participate.

Though Agency E does partially explain that the accused person can ask for a lawyer at any time, a more accurate representation would inform the listener that they have the right to an attorney at any time. This warning also fails to mention that the right to silence is ongoing, as is required by *Miranda v. Arizona* (1966). Interestingly, in the English version from Agency E, there is no warning that advises the listener that any of their rights are ongoing.

Analysis

The results of this study indicate that no version of the Nevada Spanish *Miranda* warnings is entirely accurate and that the Spanish versions vary substantially in quality and completeness among agencies. Each problem described above can itself create confusion for the listener; when combined with other translation errors, these problems can compound to create significantly more confusing translations. The consistent problems with the *Miranda* warnings in this study fall into three categories: errors caused by unskilled translation, complex grammar constructions, or the false assumption that suspects have a baseline familiarity with their rights.

English version	Spanish version	Back-translated into English
	Si usted decide contestar las preguntas sin tener un abogado presente,	If you decide to answer the questions without having a lawyer present,
	usted puede,	you can,
	cuando quiera,	when you want,
	pedir un abogado durante el interrogatorio.	ask for a lawyer during the interrogation.

Table 6. Agency E *Miranda* warning excerpt, prong five.

Errors Caused by Unskilled Translation

All Spanish *Miranda* warnings in this study contain at least one translation that is not a direct equivalent to its English counterpart. For instance, law enforcement agencies in this study tend to use imprecise or incorrect verb translations especially in modal verb phrases. In the warning *anything you say can and will be used against you*, Agencies C and E translated *can and will* as *puede*. The modal *puede*, or *can* in English, suggests a low degree of probability that an event will occur, whereas *será*, or *will*, marks a significantly higher degree of probability. This creates an uneven standard for communicating the rights between the two languages, which was deemed unconstitutional in *Riviera-Reyes v. Commonwealth* (2006).

Other law enforcement agencies in this study do not include the requisite diacritic marks in certain verb translations. In the Spanish language, future tense verbs which lack the appropriate diacritic marks will appear to be the past subjunctive form; therefore, the omission of written accents can affect a reader’s ability to identify the intended mood and tense. In the fourth part of the *Miranda* warning where the suspect is advised, *if you do not have means to hire a lawyer, the court will appoint one for you*, the verb phrase *will appoint* is incorrectly translated as *might appoint* by Agencies D and E. These same agencies require Spanish speakers to sign a form that contains the Spanish *Miranda* warnings prior to interrogation. On these forms, none of the future tense verbs have the appropriate accents. Since the past subjunctive mood and tense is consistent throughout the statement rather than the future indicative, the reader is likely to interpret that to be the correct conjugation and therefore interpret the fourth warning incorrectly.

Alongside the persistent verb translation issues, the data contains several examples of the omission of necessary subject pronouns, usually the second person singular *usted*, or *you*. In Spanish, anaphoric references sometimes allow the referent to be inferred by the listener. However, in many of the *Miranda* warnings in this study, not only is the subject pronoun absent from the clause but there is no compensating anaphoric reference to provide context. Under these conditions, it is difficult for the listener to unequivocally determine the subject pronouns.

This problem is especially common in the clauses which advise the listener of their right to an attorney during questioning, and their right to have an attorney provided by the court without cost. Without the second person singular pronoun *usted*, it is unclear whether the listener has this right or some other person. Additionally, Rock (2007) suggests that detainees are more likely to exercise their rights when the rights texts ex-

plicitly contain *you*. If the subject *usted*, or *you*, is consistently absent from the warnings, then suspects are significantly less likely to understand the relevance of these statements to their own situation.

As with anaphoric references to subject pronouns, cataphoric referents for direct and indirect object pronouns are at times left to be inferred by the listener in Spanish. For example, *le vendió el coche* may be a truncated form of *ella le vendió el coche a Juan*. Here, the subject pronoun *ella* and the object noun *a Juan* may or may not appear. However, the *Miranda* warnings do not provide the necessary lexical or syntactic context to easily make these inferences. For instance, Agency B's version contains the clause *mientras se le interroga*, which translates as *while* + [*se*: impersonal *se*] *interrogate* [*le*: indirect-object pronoun]. This clause does not provide sufficient context for the listener to reasonably infer *se* has a syntactic function as opposed to a pronominal one. Likewise, it is not sufficiently clear that *le* refers to *usted*, since the subject pronoun does not appear in this clause. In any case, the *Miranda* warnings should use language that explicitly defines the subject and object pronouns in order to eliminate any ambiguity.

More generally, the Spanish *Miranda* warnings in this study frequently contain low-frequency words where more commonplace equivalents are available. In some cases, these translations have been deemed unacceptable in federal court (see *State v. Ortez*, 2006; *United States v. Higareda-Santa Cruz*, 1993). *Afford*, a word of "high importance" on the *Miranda* Vocabulary Scale, is especially likely to be translated into complex verb phrases rather than more straightforward translations.

The formal language used in Spanish *Miranda* warnings is likely derived from the culture-specific legal register that exists in the English versions as well. As such, the average detainee is unlikely to have the vocabulary necessary to understand the *Miranda* warnings (Rock, 2007; Rogers *et al.*, 2009b). Rogers *et al.* (2009b) demonstrate that a failure to understand just one or two key words in the *Miranda* warnings can cause detainees to misunderstand their rights. To ensure comprehension, the warnings should incorporate higher-frequency words and plain language standards, which would increase the likelihood that listeners understand their rights Eades and Pavlenko (2016).

Complex Grammar Constructions

Every Spanish *Miranda* warning in this study contains several complex grammar constructions which could be revised into much simpler equivalents. At Agencies B, D, and E, all but one of the sentences in their respective warnings have two or more clauses and feature several levels of embedding. All three agencies have at least one conditional statement in their warning; Agencies B and E each have two conditional statements in one sentence. Agency E's fourth warning has four subordinate clauses spread across five levels of embedding.

In general, the third and fourth warnings are the most likely to have complex grammar, conditional statements, and multiple prepositional phrases, all of which should be avoided in rights texts (Eades and Pavlenko, 2016; Gibbons, 2001; Rock, 2007; Shuy, 1997). While it would be impractical to attempt to remove every dependent clause in every agency's *Miranda* warnings, there is potential for the warnings to be broken into simpler clauses or more sentences.

Rimmer (2009) indicates that with each additional level of embedding, a sentence becomes more difficult to understand, and places a greater cognitive load on the listener.

Complex constructions such as those seen in the third and fourth warnings demand that suspects have the memory and literacy skills necessary to recall antecedent glosses and connect them to later clauses. As Rock (2007) demonstrates, detainees are frequently unable to connect glosses to clauses that appear later in the warning, and as a result do not understand their rights. Thus, it is essential that the *Miranda* warnings are easily understood and accessible to people of various literacy levels. The warnings in this study overall do not meet such a requirement.

Assumption of Familiarity with the Miranda Rights

Moreover, the most significant issue with the *Miranda* warnings in this study is the assumption by law enforcement agencies that detainees will have a baseline familiarity with their Constitutional rights and will be able to fill in linguistic gaps through assumptions and inferences. Of all the shortcomings in the *Miranda* warnings investigated in this study, the lack of a statement that the *Miranda* rights are continuing has the most impact on listener comprehension and is the most frequent problem across agencies. Four out of five agencies in this study do not advise suspects that their rights are continuing through the entire interrogation, and nothing in the language of the warnings would imply that legal fact to the listener. Rogers *et al.* (2007) demonstrated that in the Western United States, inclusion of a statement of continuation of rights is rare, but the authors do not provide an explanation for this. In any case, the lack of a statement of continuation of rights significantly compromises detainees' understanding of their rights and is in direct violation of the decision in *Miranda v. Arizona* (1966).

Furthermore, many agencies summarized essential parts of the *Miranda* rights with excessively simple phrases in Spanish, such as Agency C's Spanish warning which merely states *usted tiene el derecho de tener un abagado* [sic] *presente* (*you have the right to have an attorney present*), while the English version contains the much more complete *you have the right to speak with an attorney before answering any questions, and to have an attorney present with you while you answer any questions.* While a native to the United States might understand the importance of the presence of an attorney, the Spanish version of this warning does not explain it in explicit terms. Some agencies in this study use the English word *appointed* to imply that a lawyer will be provided free of charge, but the chosen Spanish equivalents like *nombrar* or *adjudicar* do not have the same implication.

Moreover, none of the *Miranda* warnings in this study explain exactly how the detainee can invoke their rights, even though it is evident that police officers expect detainees to use specific language to invoke them (Ainsworth, 2008, 2010; Mason, 2013; Pavlenko, 2008; Shuy, 1997). While a citizen of a Common Law country might be able to understand their rights and invoke them without issue, it is mistakenly taken for granted that all Spanish speakers in the United States would be able to do the same. In fact, research by Rock (2007) and Pavlenko (2008) indicates that most citizens of Common Law countries are familiar with their constitutional protections because of television reenactments of the legal process. Recent immigrants to the United States will lack this cultural exposure to the United States criminal justice system, and therefore may not realize the paramount importance of their constitutional right to counsel and protection against self-incrimination. The *Miranda* warnings in this study do not have sufficient contextual or conversational markers to indicate their importance to a listener.

Discussion

In Nevada, where 21% of the population speaks Spanish and 40.9% of Spanish speakers report they speak English less than “very well” (American Community Survey, 2016), law enforcement agencies have a responsibility to explain the *Miranda* warnings completely and accurately in Spanish. To meet this need, the translations in this study require revision before they can wholly reflect the meaning of the decision in *Miranda v. Arizona* (1966).

The results of this study indicate a need for improvement to the Spanish *Miranda* warnings in Nevada and suggest key areas that require specific revisions in order to conform to case law and best practices based on linguistic research. When writing the *Miranda* warnings, law enforcement agencies should take into consideration the sociopolitical background and native dialect of the demographic population of Spanish speakers under the agency’s jurisdiction. It is likely necessary that law enforcement agencies increase the length of their warning in order to provide better context for the detainee’s legal rights, and to decrease the level of complexity of the warnings. In order to best capture the meaning of the *Miranda* rights, any revisions should be made by a cooperative of linguists, legal experts, law enforcement officers and native Spanish speakers, as is recommended by Eades and Pavlenko (2016) and Rock (2007).

One limitation of this study is that not all *Miranda* warnings were provided directly from *Miranda* cards; since some were obtained via email, they may not accurately reflect the actual practices of law enforcement officers when they read the *Miranda* warnings to suspects. Likewise, this study does not reflect how the *Miranda* warnings are actually communicated to suspects, because audio or video footage was not examined; the written translations of the *Miranda* warnings serve only as a guideline for law enforcement officers and may or may not be used in actual practice.

In future studies, audio and video footage could also be used to evaluate the role of pragmatics (e.g., interrogation strategies used by police), proxemics, and articulation in listener comprehension when a suspect is read their rights. In addition, the efficacy of standard translations of the *Miranda* warnings as compared to the efficacy of sight interpretations could be tested. Lastly, further research could investigate to what extent officer’s competence in the Spanish language, their beliefs, and biases affect the quality of the Miranda warnings read to Spanish-speaking suspects.

It is important to note that the present study represents a fraction of law enforcement agencies in Nevada. Future research with a larger sample size would provide a better picture of the overall quality of the Spanish Miranda warnings. The current study serves as a pilot for future studies as it provides an in-depth textual analysis of the shortcomings of the translated *Miranda* warnings. In doing so, this study expands upon research on linguistic issues related to the *Miranda* rights and indicates systemic problems with translation of the *Miranda* warnings into Spanish.

Acknowledgements

I would like to thank Dr. Jerry McMenamin for his enthusiastic support both professionally and personally. I appreciate his academic mentorship and willingness to remind me to “lighten up” when need be. I also thank Dr. Cassie Isabelli and Dr. Erin Edgington for their guidance throughout the data collection and writing process. Thank you as well

to Catherine Bain for her frank advice and encouragement not only during this process, but in all things.

References

Ainsworth, J. (2008). “You Have the Right to Remain Silent. . . But Only If You Ask for It Just So”: The Role of Linguistic Ideology in American Police Interrogation Law. *International Journal of Speech Language and the Law*, 15(1).

Ainsworth, J. (2010). Miranda Rights. In M. Coulthard and A. Johnson, Eds., *The Routledge Handbook of Forensic Linguistics*. London & New York: Routledge, 111–125.

American Community Survey, (2016). American factfinder: Percent of specified language speakers. U.S. Census Bureau.

Collins, (2020). *Collins Spanish Dictionary (COBUILD).* COBUILD2020.

Eades, D. and Pavlenko, A. (2016). Translating research into policy: New guidelines for communicating rights to non-native speakers. *Language and Law / Linguagem e Direito*, 3(2), 45–64.

Gibbons, J. (2001). Revising the language of New South Wales police procedures: applied linguistics in action. *Applied Linguistics*, 22(4), 439–469.

Kurzon, D. (2000). The right to understand the right of silence: a few comments. *International Journal of Speech Language and the Law*, 7(2), 244–248.

Mason, M. (2013). Can I get a lawyer? A suspect’s use of indirect requests in a custodial setting. *International Journal of Speech Language and the Law*, 20(2).

Pavlenko, A. (2008). “I’m very not about the law part”: Nonnative speakers of English and the Miranda warnings. *TESOL Quarterly*, 42(1).

Rimmer, W. (2009). Can what counts in complexity be counted? *Language Studies Working Papers*, 1(1), 25–34.

Rock, F. (2007). *Communicating rights: The language of arrest and detention.* Hampshire: Palgrave MacMillan.

Rogers, R., Correa, A. A., Hazelwood, L. L., Shuman, D. W., Hoersting, R. C. and Blackwood, H. L. (2009a). Spanish translations of Miranda warnings and the totality of the circumstances. *Law and Human Behavior*, 33(1), 61–69.

Rogers, R., Harrison, K. S., Shuman, D. W., Sewell, K. W. and Hazelwood, L. L. (2007). An analysis of Miranda warnings and waivers: Comprehension and coverage. *Law and Human Behavior*, 31(2), 177–192.

Rogers, R., Hazelwood, L. L., Sewell, K. W., Blackwood, H. L., Rogstad, J. E. and Harrison, K. S. (2009b). Development and initial validation of the Miranda Vocabulary Scale. *Law and Human Behavior*, 33(5), 381–392.

Rogers, R., Hazelwood, L. L., Sewell, K. W., Harrison, K. S. and Shuman, D. W. (2008). The language of Miranda warnings in American jurisdictions: A replication and vocabulary analysis. *Law and Human Behavior*, 32(2), 124–136.

Shuy, R. W. (1997). Ten unanswered language questions about Miranda. *International Journal of Speech Language and the Law*, 4(2), 175–196.

Legal References

Commonwealth v. Ochoa, 32 Mass. L. Rptr. 153 (Mass. Super. 2014)
Duckworth v. Eagan, 492 U.S. 195 (1989).
Miranda v. Arizona, 384 U.S. 436 (1966)
Rivera-Reyes v. Commonwealth, 2006 WL 2986495 (Ky. 2006)
State v. Carrasco-Calderon, 1-CA-CR 07-1082. (A.Z. Ct. App. 2008)
State v. Ortez, 631 S.E.2d 188, 192 (N.C. Ct. App. 2006)
State v. Teran, 862 P.2d 138, 138 (Wash. Ct. App. 1993)
United States v. Botello Rosales, 728 F.3d 865 (9th Cir., 2013)
United States v. Castro-Higuero, 473 F.3d 880, 883 (8th Cir. 2007)
United States v. Garibay, 143 F. 3d 534 (9th Cir., 1998)
United States v. Higareda-Santa Cruz, 826 F.Supp. 355, 357 (D. Or. 1993)
Wisconsin v. Santiago, 900 N.W.2d 344 (Wisc. Ct. App. 2017)

Appendix A

Agency A

ADVISEMENT FOR CUSTODIAL INTERROGATION (Adults)

You have the right to remain silent.

Anything you say can be used against you in a court of law.

You have the right to consult with an attorney before questioning.

You have the right to the presence of an attorney during questioning.

If you cannot afford an attorney, one will be appointed before questioning.

Do you understand these rights?

INFORMACIÓN SOBRE EL PROCEDIMIENTO DE INTERROGACIÓN BAJO CUSTODIA

Adultos

Tiene el derecho de guardar silencio.

Cualquier cosa que usted diga puede ser usado en su contra en un tribunal de justicia.

Tiene el derecho de consultar con un abogado antes del interrogatorio.

Tiene el derecho a la presencia de un abogado durante el interrogatorio.

Si usted no puede pagar un abogado, uno se le asignará antes del interrogatorio.

¿Entiende estos derechos?

Juveniles

Tiene el derecho de guardar silencio.

Cualquier cosa que usted diga puede ser usado en su contra en un tribunal de justicia.

Tiene el derecho de consultar con un abogado antes del interrogatorio.

Tiene el derecho a la presencia de un abogado durante el interrogatorio.

Si usted no puede pagar un abogado, uno se le asignará antes del interrogatorio.

¿Desea que alguno de sus padres o tutores esté presente?

¿Entiende estos derechos?

Agency B

> I included what is available to our Officers regarding Miranda Rights in both Spanish and English. Its important to note that Officers usually wont attempt to interview a suspect in a language that they are not fluent in. If needed we will utilize an Officer or Detective fluent in a given language to ensure accuracy of Miranda Rights.
>
> ADMONITION
>
> 1. You have the right to remain silent.
> 2. Anything you say can be used against you in a court of law.
> 3. You have the right to talk to a lawyer and have him present with you while you are being questioned.
> 4. If you cannot afford a lawyer, one will be appointed to represent you before any questioning, if you wish one.
>
> WAIVER OF RIGHTS
>
> 1. Do you understand each of these rights I have explained to you?
> 2. Having these rights in mind, do you wish to talk to us now?
>
> AMONESTACIÓN
>
> 1. Usted tiene el derecho a permanecer en silencio
> 2. Todo lo que diga puede ser usado en su contra en un juzgado.
> 3. Tiene el derecho de hablar con un abogado y que esté presente con usted mientras se le interroga.
> 4. Si no puede costear los gastos de un abogado, se le asignará uno para que le represente ante cualquier interrogatorio, si desea uno. _____
>
> RENUNCIACIÓN DE DERECHOS:
> 1. ¿Entiende cada uno de estos derechos que le he explicado?
> 2. ¿Considerando estos derechos, desea hablar con nosotros ahora?

Agency C

You have the right to remain silent if you give up that right to remain silent anything you say can and will be used against you in a court of law. You have the right to speak to an attorney before answering any questions, and to have an attorney present with you while you answer any questions. If you cannot afford an attorney, an attorney will be appointed for you at no cost to you, and you need not answer any questions until that attorney has been appointed for you. If you decide to answer questions now, you may stop at any time and ask to talk to an attorney before any questioning continues. If you decide to stop answering questions once you have begun, all questioning will stop. Do you understand each of these rights I have explained to you? Understanding these rights, would you like to speak to me now?

I hope this email finds you well. Your information was passed along to me with the message that you were looking for a Spanish version of our Miranda Rights. Here is the version I use:

Usted tiene el derecho de permanencer en silencio. Caulquier cosa que usted declare puede ser usada contra usted en la corte. Usted tiene el derecho de tener un abagado presente. Si no tiene usted recursos para pagar un abogado, la corte le asignara un abogado antes de serinterrogado. Entiende usted estos derechos?

For juveniles (under 18 years of age), there's an additional line: Desea un padre o tutor este presente?

Please let me know if I may be of further assistance.

Best regards,

Agency D

MIRANDA WARNING
(Your Rights)

You have the right to remain silent.
Anything you say can and will be used against you in a court of law.
You have the right to talk to an attorney and have him present during any questioning.
If you desire and cannot afford one, an attorney will be appointed to represent you prior to any questioning.
Do you understand each of these rights I have explained to you?____________.
Having these rights in mind, do you wish to talk to me now?______________.

MIRANDA WARNING

... have the right to remain silent

...ything you say can and may be used against you in a court of law.

...u have the right to talk to an attorney and have him present during any questioning.

If you so desire and cannot afford one, an attorney will be appointed to represent you prior to any questioning.

Do you understand each of these rights I have explained to you?

Having these rights in mind, do you wish to talk to me now?

MIRANDA WARNING

- You have the right to remain silent.
- Anything you say can and will be used against you in a court of law.
- You have the right to talk to a lawyer and have him present with you whi... you are being questioned.
- If you cannot afford to hire a lawyer, one will be appointed to represent you before any questioning, if you wish.
- You can decide at any time to exercise these rights and not answer any questions or make any stateme...s.
- Do you understand each of these rights I have expla...ed to you?
- Having these rights in mind, do you wish to talk to us now?

LA ADVERTENCIA MIRANDA
(Sus Derechos)

Usted tiene el derecho de mantenerse callado.
Cualquier cosa que usted diga sera usada en su contra en un jusgado.
Usted tiene el derecho de hablar con un abogado y tenerlo presente durante cualquier interrogatorio
Si usted desea un abogado pero no tiene fondos, un abogado sera nombrado para que lo represente antes de empezar cualquier interrogatorio.
Entiende usted cada de estos derechos que yo le he explicado?
Teniendo estos derechos en mente, desea usted hablar conmigo ahora?

Agency E

Miranda Rights

-You have the right to remain silent. _______________

-Anything you say can and will be used against you in a court of law. _______________

-You have the right to talk to a lawyer and have him/her present while you are being questioned. _______________

-If you cannot afford to hire a lawyer, one will be appointed to represent you before any questioning, if you wish. _______________

Juvenile Admonishment

-You have the right to have your parent or guardian present during questioning. _______________

-Anything you say can and will be used against you in Juvenile Court. _______________

-(If 14 years or older and accused of a felony)
You may be certified as an adult and tried in Adult Criminal Court.
Anything you say can and will be used against you in Adult Court. _______________

Waiver

-Do you understand each of these rights I have explained to you? _______________

-Having these rights in mind, do you wish to talk to us now? _______________

SUS DERECHOS DE MIRANDA

- Usted tiene el derecho de guardar silencio.
  ¿Entiende Usted? ________
- Cualquier cosa que Usted diga puede ser usada contra Usted en una corte de ley.
  ¿Entiende Usted? ________
- Usted tiene el derecho de hablar con un abogado, y tener un abogado presente durante nuestras preguntas ahora y en el futuro.
  ¿Entiende Usted? ________
- Si Usted no tiene medios para emplear un abogado, el juzgado de la corte adjudicara uno para Usted, sin costo, antes de iniciar el interrogatorio, si Usted quiere.
  ¿Entiende Usted? ________
- Si Usted decide contestar las preguntas sin tener un abogado presente, Usted puede, cuando quiera, pedir un abogado durante el interrogatorio.
  ¿Entiende Usted? ________
- ¿Entiende Usted sus derechos como yo se los expliqué? ________
- ¿Quiere Usted contestar mis preguntas sin tener un abogado presente? ________

Os Imageboards e os seus Recursos Compartilhados: Um Estudo de Caso em Síntese de Autoria Forense

Luiz H. Valle-Nunes

Universidade do Porto, Portugal

https://doi.org/10.21747/21833745/lanlaw/7_1_2a11

Abstract. *The development of the internet and new information and communication technologies has led to new enunciative practices which occur via emerging textual genres, such as forum threads. These new communicative habits, especially mob behaviours, combined with an anonymity that can be real or just felt may result in different cybercriminal activities, including those related to language. Forensic Linguistics has focused on such practices through studies in forensic authorship analysis with the aim of authorship synthesis, either based on individuals or communities, such as that of imageboards, collaborating in tasks associated with undercover investigations. The current analysis is a case study in forensic authorship synthesis applied to imageboards in Brazilian Portuguese, focusing on the category of resources shared by the community of practice, which is part of the resources and restrictions model presented by Grant and MacLeod (2018, 2020a). The corpus consists of 20 forum threads collected between October and November 2019, with a total of 306 posts and 24,587 tokens. The approach results in a holistic description of linguistic markers found at the microstructural level. It is based on a study of the n-grams obtained through the 'Corpógrafo' corpora management tool and a manual analysis of a sub corpus comprised of 10 threads, which point out to phenomena such as code-switching and othering manifestations.*

Keywords: *Undercover investigations, forum threads, code-switching.*

Resumo. *O avanço da internet e de novas tecnologias de informação e comunicação tem ocasionado o surgimento de novas práticas enunciativas, realizadas por meio de gêneros textuais emergentes, como os fios de discussão. Estes novos comportamentos comunicativos, dentre eles os mobilizadores, aliados a um anonimato que pode ser real ou apenas sentido, resultam em diferentes crimes cibernéticos, dentre eles os relacionados à linguagem. A Linguística Forense tem se debruçado sobre tais práticas por meio de estudos em análise de autoria forense com o intuito de síntese de autoria, seja de indivíduos ou seja de comunidades, tais como a dos imageboards, colaborando em tarefas periciais associadas a investigações*

infiltradas. A presente análise trata-se de um estudo de caso em síntese de autoria forense acerca de imageboards em português brasileiro, utilizando a categoria de recursos compartilhados em comunidades de prática, presente no modelo de recursos e restrições proposto por Grant and MacLeod (2018, 2020a). Foram coletados 20 fios de discussão entre outubro e novembro de 2019, compondo um corpus de 306 publicações e 24.587 átomos. A abordagem resulta numa descrição holística de marcadores a nível microestrutural, localizados a partir de um estudo de n-gramas obtidos por meio da ferramenta de gestão de corpora "Corpógrafo" e de uma análise manual de um subcorpus de 10 fios, de modo a evidenciar fenômenos como a alternância de códigos e a alterização.

Palavras-chave: *Investigações infiltradas, fios de discussão, alternância de códigos.*

Introdução

O desenvolvimento de novas tecnologias digitais, aliado à democratização do acesso à internet, possibilitou a criação de plataformas digitais cada vez mais multimodais e focadas na coletividade, permitindo o surgimento de práticas enunciativas (Fiorin, 2008: 4) e gêneros textuais emergentes altamente dinâmicos (Marcuschi, 2005: 13), tais como os tweets, os reels, os fios (*threads*), entre outros. Esta evolução trouxe grandes mudanças para a Comunicação Mediada por Computador (CMC) por contribuir para uma nova relação interativa entre os usuários e as redes. Isto se reflete, num sentido amplo, em diferentes ferramentas e estratégias adotadas pelos usuários para interagirem entre si e em suas percepções de pertença a comunidades de prática (Bucholtz and Hall, 2005), o que influencia as suas projeções enquanto *personas* digitais (Rashid *et al.*, 2013). Associado a tais mudanças, o advento de um efeito de despersonificação (KhosraviNik and Esposito, 2018: 49) teve como resultado a materialização de discursos que potencialmente não se realizariam em interações pessoais (Sousa-Silva, 2018: 119). Com efeito, ocorreu uma escalada na disseminação dos crimes cibernéticos relacionados à linguagem nos mais diferentes espaços da rede, dentre eles os discursos de ódio contra minorias, bem como comportamentos mobilizadores (Bernstein *et al.*, 2011; Fontanella, 2010) de usuários que compartilham de ideologias radicais.

Espaços que começaram a receber grande atenção a partir da década de 2010 são os fóruns de discussão online, que não só utilizam a sensação de anonimato proporcionada pelas plataformas, por conta da possibilidade de criação de contas sem a necessidade da inserção de informações reais dos utilizadores, como também possibilitam o recurso a ferramentas de criação e edição de conteúdo (Lisecki, 2013), que podem aumentar a efemeridade da comunicação materializada (Bernstein *et al.*, 2011), o que dificulta as investigações acerca das publicações dos utilizadores. Para além deste anonimato sentido, outro desafio está relacionado à criação de ferramentas de encriptação da rede, que propiciam um anonimato real por proteger os dados do usuário, de modo a permitir uma comunicação não rastreável em diferentes níveis, fazendo com que a linguagem se torne uma peça-chave para se identificar autores de crimes virtuais e se obter informações sobre as atividades de grupos organizados.

No âmbito dos fóruns de discussão que cultivam o anonimato destacam-se os *imageboards*, que são espaços inerentemente hipertextuais (Xavier, 2002: 29-33) e colaborativos (Koch, 2007: 35) a ponto de constituírem uma cultura online própria (Thibault, 2015). Popularmente conhecidos como *chans*, trata-se de ambientes nos quais o anoni-

mato é requerido e em que diferentes estratégias para a sua manutenção são empregadas (Bergman, 2001: 1), desde a possibilidade do uso de redes privadas virtuais (VPNs) e do roteamento cebola (Okazaki *et al.*, 2015: 371), por meio de navegadores anônimos como o Tor, até a codificação da própria linguagem, que opera por meio de informações e códigos compartilhados, estabelecidos por meio do exercício do poder ideológico (Van Dijk, 2006) dos membros mais integrados à comunidade. Este tipo de fórum já foi alvo de diferentes investigações, como em trabalhos que discutem o compartilhamento de informações (Manivannan, 2012; Richoux, 2016) e a articulação de grupos radicalizados, seja no que diz respeito aos discursos de ódio (Manivannan, 2013; Ruocco, 2020), ou seja na prática do ciberterrorismo (Weimann, 2016) e de ataques online orquestrados (Fontanella, 2010), entre outros. No entanto, trata-se ainda de um objeto de estudo novo que não recebeu grande atenção de trabalhos em português, principalmente numa perspectiva da Linguística Forense, que pode contribuir para que se tenha uma visão mais global acerca da linguagem em fóruns, de modo a identificar, a partir de tarefas periciais, estratégias de codificação importantes para o auxílio a outros profissionais das Ciências Forenses na solução de crimes cibernéticos, além de se possibilitar eventuais atividades de monitoramento.

No quadro da linguagem como prova, a Linguística Forense tem se dedicado a problematizar as questões relativas à identidade linguística (Coulthard and Johnson, 2010), discutindo-a a partir de noções importantes como a do idioleto (Coulthard, 2004), a do estilo idioletal (Turell, 2010) e a da performance linguística (Grant and MacLeod, 2018), com o intuito de se aprimorar o entendimento acerca das recorrências linguísticas e assim avançar na eficácia das tarefas periciais, sejam manuais ou seja também na implementação de softwares para as análises automatizadas sobre grandes quantidades de dados. Desde os anos 2000, não só as discussões se estenderam como também novas tarefas periciais relacionadas aos gêneros textuais emergentes começaram a ser requisitadas aos linguistas forenses, como é o caso da síntese de autoria forense (Grant and MacLeod, 2020b), um desdobramento da análise de autoria forense.

A análise de autoria forense prevê a identificação de marcadores linguísticos recorrentes materializados na comunicação de um ou mais determinados autores (Grant, 2008; Sousa-Silva and Coulthard, 2016). Já a síntese de autoria forense ocorre quando um linguista perito, a partir das ocorrências consistentemente localizadas, constitui um modelo de *persona* linguística a ser assumido por um investigador infiltrado, que poderá simular a autoria de uma vítima ou de um membro de um grupo organizado para que seja possível se obter informações acerca do caso investigado e possibilitar o seu avanço (Grant and MacLeod, 2020a). Para isto, o eventual infiltrado não só precisará compreender com profundidade a performance de uma *persona* linguística (Grant and MacLeod, 2020b: 82), como evitar ao máximo o vazamento de informações que revelem a sua própria identidade por meio da materialidade linguística, o que pode comprometer a sua investigação.

As primeiras experiências de aplicação da síntese da autoria forense foram testadas num programa de treinamento para investigadores infiltrados intitulado *Pilgrim* no Reino Unido, ao longo da década de 2010 (MacLeod and Grant, 2017: 159). Com a participação de linguistas forenses, as tarefas periciais tiveram como intuito adquirir inteligência relativamente as atividades de aliciamento de menores para fins sexuais, a partir de interações entre investigadores infiltrados e potenciais pedófilos em salas de bate-papo online (Grant and MacLeod, 2020b: 90). Partindo desta experiência, Grant and MacLeod

(2018, 2020a) elaboraram uma abordagem para as *personas* linguísticas que chamaram de *modelo de recursos e restrições*, que parte de domínios propostos pela Análise do Discurso Mediado por Computador (ADMC), como o estrutural, o do significado, o interacional e o sociocomportamental (Herring, 2004: 360), utilizando-os como base para uma taxonomia de marcadores linguísticos relevantes para a síntese de autoria (MacLeod and Grant, 2012).

O modelo de recursos e restrições leva em conta o fato de as performances linguísticas estarem inerentemente associadas a performances de identidades, dotadas de processos de extração de recursos para a interação que podem conter traços mais estáveis ou mais dinâmicos (Grant and MacLeod, 2018: 92). Segundo os autores, os recursos para a interação estão disponíveis em diferentes níveis, nomeadamente o do histórico sociolinguístico do autor, de sua fisicalidade, do contexto interacional e da relação do falante com a audiência (Grant and MacLeod, 2018: 93-94). Estes níveis podem atuar como categorias de marcadores linguísticos com maior ou menor produtividade a depender dos contextos, dos gêneros textuais e da persona linguística que o autor pretende assumir.

O presente estudo leva em consideração a quarta categoria proposta no modelo de recursos e restrições, nomeadamente a das informações compartilhadas pelos membros da comunidade de prática, com o intuito de se identificar marcadores persistentes e relevantes para a síntese de autoria forense de *imageboards* em português brasileiro. Com isto, visa-se obter uma visão geral acerca das características compartilhadas de autoria neste ciberespaço, alargando o entendimento sobre as variedades linguísticas da rede em português, principalmente no que diz respeito as suas relações com a integração e o sentimento de pertença em comunidades de prática associadas ao cibercrime. Ao se localizarem as características partilhadas pelo grupo torna-se mais eficaz a prevenção de potenciais vazamentos identitários dos investigadores infiltrados, principalmente os ocasionados por conta de marcadores a nível estrutural, que são tipicamente os mais detectáveis por parte dos membros do grupo, por revelarem o nível de integração do locutor. Para além disto, a identificação de recorrências na materialidade linguística das publicações possibilita a criação de bancos de dados acerca deste tipo de comunidade online para que, por meio de ferramentas de *software*, seja possível implementar diferentes tipos de sistemas para o auxílio de investigações infiltradas, como glossários, bem como maneiras de se gerenciar diferentes perfis sintetizados dos fóruns e assim realizar comparações entre eles.

Com estes objetivos em consideração, nas secções seguintes serão apresentados os dados e a metodologia de análise escolhida para este estudo de caso em síntese de autoria forense.

Dados

Foram coletados 20 fios de discussão (*threads*) compostos por 306 postagens, com um total de 24.587 átomos contabilizados pela ferramenta de gestão de corpora "Corpógrafo", provenientes de categorias do *imageboard* brasileiro 55chan. As categorias escolhidas para a coleta foram /b/, /pol/, /escoria/, /lit/, /an/ e /esp/, que correspondem respectivamente às temáticas: aleatoriedade, política, questões politicamente incorretas, literatura, cultura japonesa e esportes. As amostras foram coletadas entre outubro e novembro de 2019, levando em consideração as respostas síncronas e assíncronas relativamente a publicação iniciadora.

Método

Em primeiro lugar, vale ressaltar que a abordagem adotada para a obtenção, o armazenamento e o subsequente tratamento das amostras coletadas, teve como premissa o princípio da confidencialidade, de modo a salvaguardar a privacidade dos dados de usuário e o anonimato das publicações, sem que para isto fosse perdida a integridade do material linguístico. Para este efeito, uma vez realizadas as capturas de tela de cada publicação, foram removidos os seus elementos paratextuais, como a identificação (ID), os números de hiperligação e a localização temporal. No entanto, o corpo das publicações foi mantido de maneira integral, com o espaçamento, a tabulação, as travas maiúsculas/minúsculas e a acentuação originais. Uma vez que isto pode se traduzir em questões como a falta de espaço entre os caracteres e até mesmo entre frases, a medição escolhida para o volume de dados se deu em átomos e não em palavras. Responsável pela contagem dos átomos, a ferramenta selecionada para a criação e a gestão do corpus foi o "Corpógrafo" (Sarmento *et al.*, 2004).

Após a coleta, dividiu-se os 20 fios em dois subcorpora a partir de suas temáticas, com um recorte de 216 publicações compostas por 12.945 átomos para o subcorpus intitulado "A" e 102 publicações com 11.642 átomos para o subcorpus "B". Ambas as partes foram inseridas no Corpógrafo para a componente quantitativa da análise, que se baseou num estudo de uni, bi e trigramas, com o intuito de se localizar escolhas linguísticas recorrentes a nível microestrutural, como alterações morfossintáticas ou da ordem das frases, além de se quantificar ocorrências com nomes e sintagmas nominais. Além da ferramenta automatizada, para uma segunda análise, manual e de natureza qualitativa, optou-se pelo subcorpus A, no qual se inserem as categorias /b/, /pol/ e /escoria/ e localizam-se os textos cujas questões iniciadoras e as respostas dos usuários estão relacionados aos discursos de ódio, que são tipicamente o foco das investigações infiltradas.

Realizada a análise de autoria forense do grupo em questão, optou-se por alocar as ocorrências mais relevantes para a tarefa pericial de síntese nas quatro categorias de recursos e restrições propostas por Grant and MacLeod (2018: 87-88), nomeadamente a do histórico sociolínguistico, a da fisicalidade, a do contexto da interação e a das informações compartilhadas para a integração da comunidade de prática. Uma vez que o escopo deste estudo diz respeito às informações partilhadas entre os membros da comunidade, o que corresponde unicamente à quarta categoria, recorreu-se maioritariamente a trabalhos que abordam o Discurso Mediado por Computador como o de Androutsopoulos (2007) e de Danet and Herring (2007), entre outros; que se debruçam sobre a comunicação multilíngue como o de Dor (2004), Mozzillo (2009), Seargeant and Tagg (2011), entre outros; que tratam dos discursos de ódio e fenômenos de alterização como o de Schneider (2004), Van Dijk (1984, 2006, 2018) e Wodak (2001); que discutem hegemonias culturais em ciberculturas como Ging (2019) e Salter (2018); além de trabalhos sobre o sistema linguístico do português como é o caso dos trabalhos de Alves (2007) e Villalva (1994).

Assim, nas secções a seguir, serão apresentados os resultados obtidos para a quarta categoria do modelo de recursos e restrições, de modo a apontar para as hegemonias ideológicas presentes nos *imageboards* em português brasileiro.

Resultados

A alternância de códigos

Fatores socioculturais e políticos têm contribuído para a hegemonia de algumas línguas para a comunicação online em detrimento de outras (Danet and Herring, 2007: 17). Este é o caso do inglês, que se estabeleceu como língua franca e passou a influenciar as escolhas linguísticas de usuários nativos de línguas como o português, fazendo parte das características das *personas* digitais criadas nas redes sociais e servindo como pilar para a identificação e a integração de comunidades de práticas de natureza multilíngue (Seargeant *et al.*, 2012: 528-529). Como resultado do processo de anglicização (Dor, 2004), os usuários passaram a alternar segmentos de sua língua materna com segmentos da segunda língua (L2), tanto em comunicações síncronas quanto assíncronas (Androutsopoulos, 2007: 357-359), num fenômeno apontado como alternância de códigos (Seargeant and Tagg, 2011). O processo de alternância influencia diretamente o domínio da estrutura, que já havia sido apontado por MacLeod and Grant (2017: 167-168) como o fator que mais causa o vazamento identitário no caso da síntese de autoria forense. Tendo isto em consideração, daremos atenção à maneira como este fenômeno materializa-se no corpus analisado em diferentes níveis, com base na proposta de Mozzillo (2009), que destaca uma distinção entre ocorrências de natureza intra e intersegmental, além das materializações por meio de acrônimos.

Em relação às ocorrências de natureza intrassegmental, cabe destacar que elas tipicamente se materializam quando elementos internos de um enunciado são influenciados a nível morfossintático de modo a gerar mudanças que têm como base uma segunda língua. Como aponta Mozzillo (2009: 189), este tipo de interferência pode afetar desde a raiz lexical de nomes até o nível da ordem da frase, sendo os tipos mais comuns a inserção e estratégias de sufixação e composição, bem como o recurso a traduções literais. Outros trabalhos como o de Danet and Herring (2007) apontaram o fenômeno da alternância intrassegmental como prototípico da formação identitária de comunidades, uma vez que pressupõem conhecimentos compartilhados das circunstâncias por meio das quais as trocas podem ocorrer de modo a serem aceitas e interpretadas pela audiência.

Ao se analisar o corpus, verificou-se a presença dos três tipos de materializações apontadas por Mozzillo (2009), sendo que as inserções e as traduções literais foram integradas a categoria das realizações intrassegmentais. Com isto, para que fosse possível se obter uma visão geral acerca desta categoria, considerou-se o seguinte esquema das ocorrências encontradas de acordo com a sua tipologia:

Com base nos fios de discussão coletados, como mostra a figura 1, o que se localizou foi uma preferência por realizações intrassegmentais, com um total absoluto de 71 ocorrências, face a apenas 2 intersegmentais. Os acrônimos, com 14 ocorrências localizadas, foram identificados sobretudo na referenciação entre os usuários integrados a comunidade da prática, com materializações como "OP" (*Original Poster*) para o usuário iniciador do fio e NEET (*Not in Education, Employment or Training*) para os indivíduos com poucas relações sociais fora das redes.

Dada a sua presença maioritária, procedeu-se a uma análise mais pormenorizada das ocorrências intrassegmentais, levando em consideração três categorias, nomeadamente a das inserções diretas, das criações lexicais ou neologismos, e das traduções literais do inglês. A distribuição destas categorias foi organizada em forma de gráfico para a sua melhor visualização:

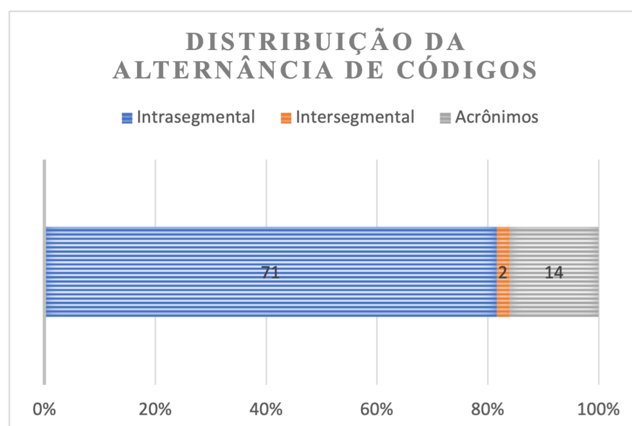


Figura 1. Distribuição da alternância de códigos por meio de anglicização.

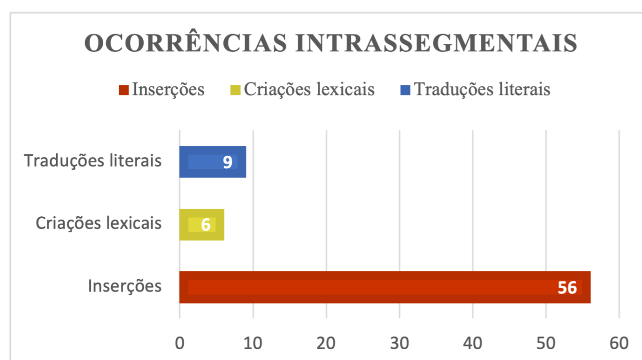


Figura 2. Distribuição das ocorrências intrassegmentais por tipo.

No caso do *imageboard* analisado, as amostras apontam para uma preferência pela utilização de inserções diretas de palavras de língua inglesa, com um valor absoluto de 56 ocorrências, em contraste às traduções literais e as criações lexicais, cujas ocorrências materializaram-se, respectivamente, em apenas 9 e 6 casos. Para uma melhor compreensão de cada tipo de realização da alternância intrassegmental, pode-se recorrer a exemplos de entradas do corpus como:

(1) “Mas os desenhos do Liefield têm tantos problemas técnicos que se for citar todos a **thread** entra em **bump limit** [...]” [TB3_R3].

(2) “[...] logo **razoa por** termos puramente profanos, pois de fato, é a única coisa que você tem, sendo um endemoniado que é; ou talvez não, talvez você seja um **literal demônio**, nunca se sabe” [T9_R4].

(3) “Não sou o **quotado**, mas ele está certo” [T11_R49].

(4) “Nem estou nessa discussão mas tive vergonha alheia de você sendo **samefag, criançafag**” [T11_R48].

No exemplo (1), a alternância intrassegmental ocorre por meio de inserções diretas de sintagmas nominais provenientes do inglês, os quais ocorrem de maneira integral, sem interferência na ordem da frase em português brasileiro. Já no exemplo (2), o que se verifica no segmento em destaque é uma tradução literal que influencia diretamente a ordem da frase, mais especificamente no argumento interno do sintagma verbal "ser literalmente um demônio". Também no exemplo (2), há um outro tipo de anglicização por meio de tradução literal a alterar o sintagma verbal, sendo o foco da alternância o verbo "pensar", que é realizado como "razoar por", numa alusão a "*reason through*". O exemplo seguinte (3) demonstra como a inserção pode sofrer derivações morfológicas do português (Alves, 2007), sendo que em "quotado" o processo de nominalização incide sobre o empréstimo "quote" no lugar de "citar". Por fim, os processos de alteração morfológica podem ocorrer por meio da composição (Villalva, 1994) inglês+inglês, português+inglês ou ainda inglês+português, tal como no último exemplo (4), em que "*same*" e "criança" são adicionados a "*fag*", um modo dos usuários se referirem uns aos outros como integrantes dos *boards*.

A questão das traduções literais foi apontada anteriormente por autores como Paredes *et al.* (2016: 194), os quais ao se debruçarem sobre culturas da rede como a comunidade dos *gamers*, identificaram um nível de anglicização rapidamente assimilado e interpretado no âmbito do processamento da leitura pelos demais participantes da interação. A influência na ordem da frase é um tipo de alternância de códigos que aparece no corpus analisado principalmente nos sintagmas nominais, sendo o movimento sintático tipicamente realizado pelos complementos e não pelo núcleo, tal como nos exemplos:

(5) "GAME DEV / **DESENVOLVIMENTO DE JOGOS FIO** [...]" [TB1_OP].

(6) "[...] Apesar da história ser meio bobinha, é realmente uma obra de arte que deve ser apreciada por todos os **animesamigos** [...]" [TB4_R4].

Nos exemplos em destaque, os núcleos sintagmáticos "fio" e "amigos" mantêm-se a direita dos seus complementos, sob a influência da ordem em inglês para "*game dev thread*" e "*anime friends*". Uma característica presente nos dois exemplos, (5) e (6), diz respeito a ausência da preposição "de" entre os núcleos e os complementos, embora o exemplo (5) ainda a preserve dentro de seu complemento, em "desenvolvimento [de] jogos". Já no caso do exemplo (6), a ausência da preposição "de" pode relacionar-se ao fato de se tratar de uma composição, sendo que o movimento sintático escolhido relativamente a escolha da posição do complemento também apareceu anteriormente no exemplo (4) com "criançafag" como escolha linguística em detrimento de "fagcriança". Finalmente, cabe dizer que a alteração da ordem de palavras, embora ocorra tipicamente na estrutura interna dos sintagmas nominais, não se materializa a nível dos sintagmas verbais e no caso de expressões idiomáticas:

(7) "[...] mas como eu já **havia dumpado** todos os meus jogos de Wii, não teve jeito. [...]" [TB2_R7].

(8) "[...] >literatura brasileira contemporânea **numa casca de noz**" [TB10_R4].

Ambos os exemplos (7) e (8) mantêm a ordem do português, sendo que no caso de (8) a expressão "numa casca de noz" como tradução de "in a nutshell" não apresentou no corpus nenhum tipo de ocorrência invertida. Uma hipótese pode estar relacionada a natureza semântica do sintagma "casca de noz" em relação, por exemplo, a uma ocorrência como "desenvolvimento de jogos fio", uma vez que a primeira aponta para uma tipologia e segunda para um fim.

A alternância de códigos com ocorrências intrassegmentais, para além das traduções literais e de inserções no âmbito do predicado, ocorre também nos nomes escolhidos para a referenciação. As escolhas linguísticas localizadas dizem respeito tanto ao grupo de pertença quanto às pessoas que não participam do *imageboard*. Nas próximas secções, apresentam-se estas escolhas de modo a apontar para a hegemonia ideológica da comunidade.

A referenciação entre membros

As escolhas feitas pelos usuários para se referirem uns aos outros e a si próprios configura um aspecto importante relativamente à representação dos atores sociais envolvidos e aos níveis hierárquicos da comunidade. Os utilizadores atribuem diferentes valores àqueles que fazem parte do grupo de pertença face a quem não participa de suas práticas. Há uma relação direta entre os valores considerados positivos e o grupo de pertença, além de uma relação entre valores negativos e os Outros, tal como tipicamente ocorre quando há uma associação ideológica forte prevista para a integração (Wodak, 2001). Dar-se-á início pelos valores associados aos autores do grupo de pertença, de modo a explorar a lexicalização e a significação das formas materializadas.

Em primeiro lugar, verificou-se que as formas utilizadas para referenciação aos atores sociais, para além de apresentarem inserções do inglês e se referirem sempre aos homens, remetem para relações sociais que podem ser divididas entre aquelas associadas ao universo do trabalho, as interpessoais e a fisicalidade dos integrantes:

(i) **Fisicalidade**: "*chad*";

(ii) **Interpessoais**: "mago", "falho", "escravoceta";

(iii) **Laborais**: "*ex-NEET*", "*wageslave*".

Relativamente à fisicalidade, o uso de "*Chad*" para se referir aos homens considerados dentro do padrão de beleza da sociedade já havia sido foco de atenção de autores como Ging (2019) ao estudar o espectro de misoginia – a que chama de *Manosphere* – nos fóruns online. Ao analisar as relações entre a cultura *incel* e a performance da masculinidade nas redes, a autora concluiu que há um chamado "dilema da masculinidade alfa-beta", em que, por um lado, os usuários enaltecem a figura do macho alfa, isto é, do homem dentro dos padrões normativos da sociedade, mantedor da hegemonia pa-

triarcal e dotado de valores tradicionais e conservadores; e, por outro lado, abraçam os aspectos autodepreciativos da figura do macho beta, que não recebe a mesma atenção e prestígio social, sendo marginalizado pelos seus valores e *hobbies* (Ging, 2019: 648-650). Esta masculinidade híbrida recorre a diferentes formas de referenciação para as características alfa, dentre elas "*chads*", "*normies*" e "*fratboys*" (Ging, 2019: 650). Embora não tenham sido encontrados no corpus exemplos destes últimos dois nomes, as ocorrências com *chad* se encaixam no mesmo paradigma apresentado pela autora, apontando assim para a presença do dilema alfa-beta nesta comunidade:

(9) "O pior e **o que mais me irrita** nem é **o fato das depósitos escolherem chad** e chads, blá-blá-blá" [T10_R5].

(10) "**O chad que atenta contra a integridade física e emocional de um falho** é o mesmo que namora a depósito descolada que posta estórias no instagado disseminando correntes de amor ao próximo" [T10_R6].

(11) "**O estereótipo de chad valentão que agride o mais fraco** é mesmo só **um estereótipo falho e deturpado**" [T10_R6].

Há uma dualidade de valores associados a *Chad*, que é aceito pela sociedade e é bem sucedido com as mulheres (exemplo 9), facto que é visto como positivo pelos usuários por conta da manutenção da dominância masculina, mas que também pode receber valores negativos uma vez que este indivíduo "agride o mais fraco" (11), referido também como "falho" (10 e 11), evidenciando que a forma híbrida alfa-beta, um meio-termo, é de facto a mais almejada. É importante destacar que embora *Chad* pertença ao espectro relacionado a fisicalidade, por estar associado a uma cultura popularmente conhecida por *jock*, i.e., de indivíduos que frequentam o ginásio e dão atenção redobrada à sua aparência, "falho" não se encaixa na mesma categoria pelo facto que a fraqueza mencionada não está somente relacionada ao físico, mas às relações que os chamados machos betas têm no seu dia a dia com a sociedade.

Ao se verificar as ocorrências com "falho", foi possível relacionar a própria noção de falha ao modo como estes indivíduos percebem a sua aceitação pelas pessoas de fora do grupo. Isto é evidenciado no fio identificado no corpus como [TB7], no qual um indivíduo expõe que quer deixar de ser falho, pedindo "ajuda para o seu desenvolvimento" [TB7_OP]. Aqui é utilizado o adjetivo "falho", associado ao valor negativo de "largado":

(12) "[...] quero dizer que essa minha versão **largada** e **falha** vai deixar de existir. Irei excluir minhas redes sociais, só ficarei entrando no chan para relatar o progresso e contribuir com coisas que aprendo nessa jornada [...]" [TB7_OP].

Inicialmente, embora o fio em questão seja dedicado a falha no sentido da fisicalidade, outros usuários começam a expor diferentes valores associados ao nome "falho" que dizem respeito ao intelecto e às interações sociais. As ocorrências a seguir expan-

dem o conceito de "falho" adotado pelos membros da comunidade de prática, além de apontar para uma vontade de se encontrar uma solução:

(13) "[...] Estudarei interações pessoais e oratória para melhorar essa forma ruim que tenho para me comunicar com as pessoas. Achei uns exercícios para os músculos da face na Internet, vou tentar fazer [...]" [TB7_OP].
Valor associado: dificuldades articulatórias (oratória);

(14) "[...] estou tentando me tornar uma pessoa mais interessante, por isso estou lendo mais [...]" [TB7_R1].
Valor: falta de arcabouço de conhecimentos;

(15) "[...] eu digo que vale a pena, estou me sentindo mais confiante para falar com as pessoas e estou só no começo" [TB7_R6].
Valor: dificuldade na comunicação.

Os exemplos apontam para discursos em que as "interações pessoais" a que os usuários se referem têm uma relação direta com "ser mais interessante" (14) e "ser mais confiante" (15), ou seja, são exemplos em que a falha está associada à capacidade de se relacionar com o mundo exterior e ser compreendido por ele. É pautado neste tipo de discurso que o fórum possui uma categoria intitulada "falha e aleatoriedade", na qual são postados fios relacionados aos problemas da vida cotidiana, tipicamente relacionados ao facto destes indivíduos não se encaixarem em padrões sociais. Ao conhecimento de saber transitar pelas situações do mundo real, os usuários deram o nome de "magia", o que explica o facto de se referirem uns aos outros como "magos". O uso deste nome no corpus evidencia uma outra faceta da relação entre os usuários e o mundo exterior, nomeadamente a das relações laborais, como no exemplo:

(16) "O Brasil consegue ser o pior lugar para ser um **mago** , pois na maioria das vezes , se você é um inepto social , nem os empregos mais inóspitos sobram pra você" [T10_R5].
Valor: dificuldades nas interações laborais.

No exemplo (16), o que está em causa é o despreparo daquele que ainda não detém as habilidades interpessoais para o trabalho. É preciso lembrar, no entanto, da dualidade apontada por Ging (2019), já que este despreparo é ao mesmo tempo posto em causa e celebrado. Os usuários escrevem sobre o problema do desemprego, mas também o enaltecem por estarem longe da sociedade tóxica que acreditam existir. Para falarem sobre o mal do trabalho, por exemplo, os usuários recorrem a uma composição de valor negativo que foi importada dos fóruns em inglês, nomeadamente "*wageslave*", como no exemplo:

(17) "Atualmente sou um *ex-NEET*, *wageslave* sem amigos que mora com os pais , mais alguns meses e eu planejo sair daqui e cortar completamente os laços com a minha família [...]" [T4_R7].

A temática dos utilizadores como incompreendidos pela sociedade reaparece quando o usuário se refere a "cortar os laços" (17) com a família. Esta dificuldade de integração, tanto do ponto de vista interpessoal quanto econômico, foi discutida por autores como Uchida and Norasakkunkit (2015: 2) ao estudarem o fenômeno dos "*Hikikomori*", que se popularizou no ocidente sob o acrônimo NEET para "*Not in Education, Employment or Training*" (Smith and Wright, 2015: 402-403) e que pode ser associado à cultura *incel* como um culto ao isolamento. Ruocco (2020: 27) já havia afirmado que parece existir uma chamada "cultura NEET" em *imageboards* internacionais como o 4chan, em que a reclusão é vista como uma característica a ser reforçada. Tal cultura parece existir no *imageboard* brasileiro analisado, o que explica o porquê de ex-*NEET* estar associado ao uso de *wageslave*, como alguém que é explorado somente pelo dinheiro, pois odeia o trabalho. Outra hipótese levantada por Ruocco (2020: 27) é a de que esta cultura seja apenas de fachada, parte de uma performance que condiz apenas a *persona* digital destes indivíduos. Mesmo que este seja o caso, deve-se atentar ao fato que uma síntese de autoria forense é baseada na performance linguística materializada, já que o intuito é o de realizar uma performance específica ao contexto do fórum.

Obtidos os resultados acerca da referenciação entre membros, a próxima secção tem por enfoque a referenciação às pessoas que não fazem parte da comunidade de prática, ou seja, os Outros (Wodak, 2001). Apresentam-se, portanto, os apontamentos relativamente às estratégias de alterização (Schneider, 2004), sem deixar de lado a presença da alternância de códigos, persistente característica compartilhada entre as *personas* linguísticas.

A representação dos Outros

Uma das estratégias utilizadas por grupos de pertença para a manutenção de sua hegemonia e para o exercício de poder ideológico é a representação positiva dos membros e a negativa dos Outros, em discursos de persuasão que apontam para comportamentos potencialmente agressivos àqueles que são considerados antagonistas, tipicamente os grupos minoritários (Van Dijk, 1984: 4). A representação de si a partir de características consideradas negativas no Outro constitui um processo de criação identitária chamado de alterização (Schneider, 2004), no qual as práticas discursivas, materializadas por meio de uma gramática pautada no contraste e no conflito (Baumann and Gingrich, 2004), podem realizar marcadores linguísticos em diferentes níveis do texto, como na escolha linguística dos nomes e na manifestação de tópicos prototípicos (Wodak, 2001: 73), que podem compreender características relevantes à síntese de autoria, uma vez que focalizam as ideologias (Van Dijk, 2018) mais pungentes do grupo de pertença.

No que diz respeito às estratégias de alterização, destacam-se no corpus as representações das mulheres e dos negros, sobretudo nas escolhas de nomes pejorativos, que, junto aos adjetivos axiologicamente negativos e os deíticos pessoais, posicionam estes grupos como antagônicos aos membros da comunidade. Exemplos que evidenciam este antagonismo podem ser localizados em sintagmas como:

a) **Mulheres**: "uma mãe solteira", "as mães solteiras", "msol", "vadia maluca", "pitanga", "depósitos", "a depósito descolada", "fêmeas", "diabolher", "merdalher";
b) **Negros**: "pedaços de carne ambulantes", "precto", "esses macacos", "jovens negros suburbanos", "animais humanos".

Efetivamente, os sintagmas apontam para a materialização da identidade masculina excludente e opressora descrita por autores como Salter (2018: 256) e Ging (2019: 639) para os fóruns anônimos, na medida em que as ocorrências coisificam a mulher de forma misógina, reduzindo-a a um papel de total submissão e objetificação sexual. Para a manutenção deste tópico recorrem, por exemplo, a traduções literais do inglês, como é o caso de "a vadia maluca" (*crazy bitch*) e "a depósito" (*dumpster*). Há também casos como "as mães solteiras", também materializado como "msol", que adaptam do inglês o acrônimo "MILF" (*Mom I'd Like To Fuck*). Estas escolhas revelam uma interdiscursividade presente na organização ideológica do grupo, fundamental para a sedimentação de suas práticas sociais (Van Dijk, 2006: 117), resgatando discursos de ódio produzidos em outros *imageboards* internacionais, como o 4chan. No entanto, embora a alternância de códigos esteja presente aqui novamente, ela não impede a emergência de novas materializações específicas do português. Como exemplo, a representação do grupo "mulher" materializa-se também em composições português+português com valores negativos, em nomes como "diabolher" e "merdalher".

O único contexto que aponta para uma diferenciação no que tange a referenciação às mulheres ocorreu com o nome "pitanga", que se refere às companheiras dos utilizadores. Neste caso, há uma tentativa de colocá-las num patamar menos pejorativo que outras mulheres, mas ainda assim objetificado:

(18) "Quero muito arrumar **uma pitanga** , tentar ter uma vida normal [...]" [T7_R14].

No exemplo, o nome "pitanga" aparece associado a valores positivos, tais como o de "ter uma vida normal", isto é, estar num relacionamento. Não foram encontradas no corpus ocorrências com os nomes "namorada", "esposa", "amante", "ficante", "companheira" ou "mina". A questão da objetificação, no entanto, não se restringe às mulheres, mas afeta ainda mais pejorativamente os negros.

Os negros são referenciados de forma coisificada e animalizada, por meio de tópicos de sobrecarga (Wodak, 2001: 73), em ocorrências como "pedaços de carne" e "animais humanos". Este uso, entretanto, não parece estar restrito apenas a locutores que se identificam como brancos, como pode ser visto no exemplo:

(19) "Eu me sinto trancado com esses macacos , estou perdendo minha sanidade" [T4_R9].

Em (19), o locutor se refere à própria família com o nome "macacos", distanciando-se por meio do uso do deítico. Este tipo de ocorrência já havia sido apontado por Ruocco (2020: 27), que salientou a utilização não esperada de determinados nomes por parte de locutores que são parte do próprio grupo alvo, dada a cultura violenta cultivada nos *imageboards*. Assim, ao se fazer suposições acerca do perfil sociolinguístico destes usuários,

é preciso se ter uma atenção redobrada à reprodução das ideologias da cibercultura, que pode se materializar até mesmo por parte da própria minoria a frequentar estes espaços.

A partir destas considerações, é possível dizer que no âmbito da integração da comunidade de prática, a alternância de códigos incide sobre a alterização, embora não seja o único tipo de recurso a que os usuários recorrem. Verifica-se também o uso de nomes que remetem a valores ideológicos possivelmente compartilhados nos *imageboards* em inglês, constituindo assim uma ligação interdiscursiva entre esses espaços, o que pode ser objetivo de futuros estudos comparativos. Com isto em mente, na próxima secção serão apresentadas algumas considerações finais, bem como outras aberturas de pesquisa numa perspectiva da Linguística Forense.

Considerações finais

O estudo analisou fios de discussão de fóruns do tipo *imageboard* em português brasileiro, com o intuito de se identificar marcadores recorrentes na performance linguística da comunidade de prática, mais especificamente no âmbito dos recursos compartilhados para a sua formação e integração, categoria presente no modelo de recursos e restrições (Grant and MacLeod, 2018). A análise visa chamar a atenção às características linguísticas compartilhadas, de modo a auxiliar na criação de uma inteligência para a mitigação do vazamento identitário a nível estrutural de policiais infiltrados durante tarefas periciais de síntese de autoria forense.

A partir da análise foi possível detectar a presença acentuada da alternância de códigos por meio de processos de anglicização, que ocorrem tipicamente de maneira intrassegmental. A alternância atua sobre o nível morfossintático, sendo os sintagmas nominais os mais afetados, embora tenha-se verificado alterações da ordem da frase em português, principalmente por conta de traduções literais. A anglicização também foi encontrada nas estratégias de referenciação intragrupo e relativamente aos Outros, nos processos de alterização, em que os valores associados aos nomes escolhidos são negativos e em que a hegemonia ideológica da comunidade se materializa em discursos de ódio.

Dada a perspectiva de uma visão geral adotada nesta pesquisa relativamente aos recursos a nível da integração da comunidade, trabalhos futuros podem debruçar-se sobre uma tipologia das ocorrências de anglicização, bem como apontar para outras línguas potencialmente utilizadas como base para a alternância de códigos no contexto dos *imageboards* brasileiros. Este tipo de identificação dá abertura a potenciais aplicações forenses computacionais tais como a criação de bancos de dados de referência em português para que investigadores infiltrados possam ter acesso mais facilmente a marcadores compartilhados por este tipo de grupo. Possibilita também a elaboração de glossários no que diz respeito aos neologismos e outras inovações de natureza morfossintática como as composições e os truncamentos mais comuns, para que a decodificação dos textos escritos seja mais eficaz. Investigações mais exaustivas acerca deste fenômeno podem resultar, por exemplo, num glossário de valores associados aos nomes, a ser implementado em *softwares*, de modo a permitir que outros profissionais das Ciências Forenses possam aceder mais rapidamente a buscas específicas acerca do léxico do texto em causa durante as investigações infiltradas. A partir da criação de um banco de dados, investigações sobre outros fóruns nesta perspectiva possibilitariam também, por um lado, a realização de análises comparativas relativamente aos marcadores compartilhados nas redes e, por

outro lado, a realização de análises computacionais supervisionadas em aprendizagem automática. Todas estas possibilidades poderão ser utilizadas não só para a resolução de crimes cibernéticos específicos, mas também para o potencial recurso a ferramentas de monitoramento destes ciberespaços.

Nota

Este artigo baseia-se na dissertação de Mestrado em Linguística, aprovada pela Faculdade de Letras da Universidade do Porto e publicada no Repositório Aberto: Valle-Nunes, L.H. (2020). *O cibercrime e as estratégias linguísticas do anonimato: A síntese de autoria forense aplicada aos imageboards em português brasileiro.* Dissertação de Mestrado. Porto: Faculdade de Letras da Universidade do Porto. Disponível em: https://hdl.handle.net/10216/129313. Acesso: 16/01/2021.

Referências

Alves, I. (2007). *Neologismo: criação lexical.* Ática: São Paulo, 3 ed.

Androutsopoulos, J. (2007). Language Choice and Code Switching in German-Based Diasporic Web Forums. In B. Danet and S. Herring, Eds., *The Multilingual Internet: Language, Culture, and Communication Online.* New York: Oxford University Press.

Baumann, G. and Gingrich, A. (2004). *Grammars of Identity/Alterity - A Structural Approach.* New York: Bergham Books.

Bergman, M. K. (2001). White Paper: The Deep Web: Surfacing Hidden Value. *The Journal of Electronic Publishing*, 7(1).

Bernstein, M., Monroy-Hernández, A., Harry, D., André, P., Panovich, K. and Vargas, G. (2011). 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *Proceedings of the Fifth International AAAI Conference on Web and Social Media, North America.*

Bucholtz, M. and Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, 7(4-5), 585–614.

Coulthard, M. (2004). Author Identification, Idiolect and Linguistic Uniqueness. *Applied Linguistics*, 25(4), 431–447.

Coulthard, M. and Johnson, A. (2010). The Routledge Handbook of Forensic Linguistics.

Danet, B. and Herring, S. (2007). *The Multilingual Internet: Language, Culture, and Communication Online.* New York: Oxford University Press.

Dor, D. (2004). From Englishization to Imposed Multilingualism: Globalization, the Internet, and the Political Economy of the Linguistic Code. *Public Culture*, 16(1), 97–118.

Fiorin, J. (2008). A Internet vai acabar com a língua portuguesa? *Texto Livre: Linguagem e Tecnologia*, 1(1), 2–9.

Fontanella, F. (2010). Nós somos Anonymous: anonimato, trolls e a subcultura dos imageboards. In *XXXIII Congresso Brasileiro de Ciências da Comunicação.*

Ging, D. (2019). Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere. *Men and Masculinities*, 22(4), 638–657.

Grant, T. (2008). Approaching questions in forensic authorship analysis. In J. Gibbons and M. Turell, Eds., *Dimensions of Forensic Linguistics.* Amsterdam-Philadelphia: John Benjamins.

Grant, T. and MacLeod, N. (2018). Resources and constraints in linguistic identity performance: a theory of authorship. *Language and Law/Linguagem e Direito*, 5(2), 80–96.

Grant, T. and MacLeod, N. (2020a). *Language and online identities: The Undercover Policing of Internet Sexual Crime.* Cambridge: Cambridge University Press.

Grant, T. and MacLeod, N. (2020b). Recursos e restrições na Manutenção de Identidades Linguísticas: uma Teoria de Autoria. In D. Almeida, M. Coulthard and R. Sousa-Silva, Eds., *Perspectivas em Linguística Forense.* Campinas: IEL-UNICAMP, 76–94.

Herring, S. C. (2004). Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. In S. A. Barab, R. Kling and J. H. Gray, Eds., *Designing for Virtual Communities in the Service of Learning.* Cambridge: Cambridge University Press, 338–376.

KhosraviNik, M. and Esposito, E. (2018). Online hate, digital discourse and critique: Exploring digitally mediated discursive practices of gender-based hostility. *Lodz Papers in Pragmatics*, 14(1), 45–68.

Koch, I. (2007). Hipertexto e construção do sentido. *Alfa*, 51(1), 23–38.

Lisecki, M. (2013). Discourse features of Internet message board on the example of 4chan. *Media i Społeczeństwo*, 3, 21–40.

MacLeod, N. and Grant, T. (2012). Whose Tweet? Authorship analysis of micro-blogs and other short-form messages. In *Proceedings of the International Association of Forensic Linguists' 10th Biennial Conference*, 210–224, Birmingham, UK: Aston University.

MacLeod, N. and Grant, T. (2017). "go on cam but dnt be dirty": linguistic levels of identity assumption in undercover online operations against child sex abusers. *Language and Law/Linguagem e Direito*, 4(2), 157–175.

Manivannan, V. (2012). Attaining the Ninth Square: Cybertextuality, Gamification, and Institutional Memory on 4chan. *Enculturation*, Online.

Manivannan, V. (2013). Tits or GTFO - The logics of misogyny on 4chan's Random – /b/. *The Fibreculture Journal*, 22(FCJ-158), 109–132.

Marcuschi, L. (2005). Gêneros textuais emergentes no contexto da tecnologia digital. In L. Marcuschi and A. Xavier, Eds., *Hipertexto e gêneros digitais: novas formas de construção de sentido.* Rio de Janeiro: Editora Lucerna, 2 ed., 13–67.

Mozzillo, I. (2009). O Code-switching: fenômeno inerente ao falante bilíngue. *PAPIA: Revista Brasileira de Estudos do Contato Linguístico*, 19, 185–200.

Okazaki, N., Toyoda, K., Yokoyama, E., So, H., Katayama, T. and Park, M. (2015). Countermeasure against fingerprinting attack in Tor by separated contents retrieval. *IEICE Communications Express*, 4(12), 370–375.

Paredes, L., Ferreira, G. and Gomes, N. (2016). A linguagem dos jogos de vídeo game e sua influência na formação de neologismos: reflexões preliminares sob a perspectiva da Linguística. *Revista Philologus*, 22(66), 191–200.

Rashid, A., Baron, A., Rayson, P., May-Chahal, C., Greenwood, P. and Walkerdine, J. (2013). Who Am I? Analyzing Digital Personas in Cybercrime Investigations. *Computer*, 46(4), 54–61.

Richoux, N. (2016). *Computer Mediated Communication: Enregisterment of Gamerspeak and Intertextual Borrowings by 4chan Users.* , Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Ruocco, J. (2020). Cómo la extrema derecha se apoderó de 4chan. *Nueva Sociedad (NUSO)*, 286, 25–34.

Salter, M. (2018). From geek masculinity to Gamergate: the technological rationality of online abuse. *Crime, Media, Culture: An International Journal*, 14(2), 247–264.

Sarmento, L., Maia, B. and Santos, D. (2004). *The Corp{ó}grafo - a Web-based environment for corpora research.*

Schneider, J. (2004). Discursos simbólicos e símbolos discursivos: considerações sobre a etnografia da identidade nacional. *Mana*, 10(1), 97–129.

Seargeant, P. and Tagg, C. (2011). English on the internet and a 'post-varieties' approach to language. *World Englishes*, 30(4), 496–514.

Seargeant, P., Tagg, C. and Ngampramuan, W. (2012). Language choice and addressivity strategies in Thai-English social network interactions. *Journal of Sociolinguistics*, 16(4), 510–531.

Smith, R. and Wright, V. (2015). The possibilities of re-engagement: cultures of literacy education and so-called NEETs. *Research in Post-Compulsory Education*, 20(4), 400–418.

Sousa-Silva, R. (2018). Computational Forensic Linguistics: An Overview of Computational Applications in Forensic Contexts. *Language and Law / Linguagem e Direito*, 5(2), 118–143.

Sousa-Silva, R. and Coulthard, M. (2016). Linguística Forense. In R. J. Dinis-Oliveira and T. Magalhães, Eds., *O que são as Ciências Forenses? – Conceitos, Abrangência e Perspetivas Futuras*. Lisboa: Pactor, chapter Forensic L.

Thibault, M. (2015). "Do not talk about Anonymous" Censura, autocensura e anonimato nelle periferie del web. *Lexia - Rivista di semiotica*, 21(22 (dezembro)), 237–254.

Turell, M. T. (2010). The use of textual, grammatical and sociolinguistic evidence in forensic text comparison. *The International Journal of Speech, Language and the Law*, 17(2), 211–250.

Uchida, Y. and Norasakkunkit, V. (2015). The NEET and Hikikomori spectrum: Assessing the risks and consequences of becoming culturally marginalized. *Frontiers in Psychology*, 6.

Van Dijk, T. (1984). *Prejudice in discourse*. Amsterdam: John Benjamins.

Van Dijk, T. (2018). Discourse and Migration. In R. Zapata-Barrero and E. Yalaz, Eds., *Qualitative Research in European Migration Studies. IMISCOE Research Series*. Cham: Springer, 227–245.

Van Dijk, T. A. (2006). Ideology and discourse analysis. *Journal of Political Ideologies*, 11(2), 115–140.

Villalva, A. (1994). *Estruturas morfológicas, unidades e hierarquias nas palavras do português.* , Faculdade de Letras da Universidade de Lisboa.

Weimann, G. (2016). Terrorist migration to the Dark Web. *Perspectives on Terrorism*, 10(3).

Wodak, R. (2001). The discourse-historical approach. In R. Wodak and M. Meyer, Eds., *Methods of Critical Discourse Analysis*. London: Sage, 63–94.

Xavier, A. (2002). *O hipertexto na sociedade da informação: a constituição do modo de enunciação digital.* , Universidade Estadual de Campinas.

Análise Fonético-Forense em Tarefa de Comparação de Locutor

Recensão de Maria Lúcia de Castro Gomes

Universidade Tecnológica Federal do Paraná, Brasil

https://doi.org/10.21747/21833745/lanlaw/7_1_2r1

Análise Fonético-Forense em tarefa de comparação de locutor
Plínio Almeida Barbosa (Coord.) (2020)
Campinas-SP: Millenium Editora

Introdução

A Fonética Forense (FF) é uma área de estudos recente e interdisciplinar que, principalmente no Brasil, carece de pesquisas e de textos em língua portuguesa. Por isso, a obra aqui resenhada é de extrema importância. Além de contribuir com relevantes trabalhos de pesquisa em Comparação de Locutor (CL), também apresenta, na forma de um protocolo, uma proposta bastante detalhada de técnicas e procedimentos para o trabalho de perícia. Esse protocolo é o resultado do projeto intitulado "Análise fonético-acústica e elaboração de protocolo para comparação de locutor em casos forenses", financiado pela Fapesp em parceria com a Escola Superior do Ministério Público do Estado de São Paulo, e desenvolvido pelo Grupo de Estudos de Fonética Forense (GEFF) do Instituto de Estudos da Linguagem da Unicamp.

O livro, em duas partes, apresenta nos primeiros cinco capítulos o Protocolo de análise fonético-forense para comparação de locutor e fecha a primeira parte com o capítulo "Por uma formação em Fonética Forense". A Parte 2, que se intitula "Questões atuais em análise fonético-forense para comparação de locutor", compõe-se de mais seis capítulos escritos por pesquisadores envolvidos na pesquisa e/ou na prática da CL.

Neste texto, apresento uma descrição dos capítulos que compõem cada uma das partes do livro, ao tempo em que emito meu parecer sobre os textos que, já adianto, devem trazer uma grande contribuição para a área da Fonética Forense.

Parte 1 – Protocolo de análise fonético-forense para comparação de locutor

O Capítulo 1, que tem como título “Protocolo de análise fonético-forense”, faz uma apresentação dos objetivos desse protocolo, divulga o *link* para um repositório do Protocolo[1], apresenta um roteiro para a análise acústico-auditiva e detalha os primeiros procedimentos para a análise fonético-forense: a investigação da qualidade acústica do material de fala. Com o objetivo principal de contribuir para o trabalho de peritos, a disponibilização do Protocolo pode vir a ser um divisor de águas no trabalho da perícia que, por ser um trabalho essencialmente interdisciplinar, precisa dos conhecimentos de diversas áreas, principalmente, como se afirma no Capítulo 6, da Linguística e da Fonoaudiologia. Não são apenas conhecimentos dessas duas áreas, no entanto, que o Protocolo abarca. Conceitos de áreas como a Física, a Matemática, a Estatística, a Genética também estão inseridos ao longo dos capítulos.

Continuando a exposição do Protocolo, o Capítulo 2, intitulado “Segmentação e transcrição da fala para fins forenses”, trata da preparação do material gravado para as análises, ou seja, detalha a segmentação e transcrição dos áudios no *software* PRAAT; o Capítulo 3, com o título “Análise linguística”, faz um foco mais específico em uma análise de oitiva para destacar elementos idiossincráticos e sociolinguísticos do locutor; o Capítulo 4, intitulado “Análise fonético-acústica”, traz questões sobre a análise fonético-acústica, discorrendo sobre os parâmetros acústicos mais robustos para a confiabilidade da CL, a avaliação do efeito de ruído para a extração desses parâmetros, o uso de *script* para extração automática de valores e possibilidades para tratamento estatístico; e, finalmente, o Capítulo 5, com o título “Apresentação de resultados no laudo/parecer técnico”, orienta o leitor para a apresentação dos resultados na confecção do documento final do trabalho, o laudo pericial/parecer técnico.

A equipe que organiza o livro (o GEFF), coordenada por Plínio Almeida Barbosa, assina os dois primeiros capítulos. Os Capítulos 3 e 5 são de autoria de Lucilene Aparecida Forcin Cozumbá e Ana Paula Sanches, que também fazem parte do GEFF. O Capítulo 4 é de autoria do Prof. Plínio Barbosa, que também assina o capítulo que fecha a primeira parte do livro. Este último capítulo, com o título “Por uma formação em Fonética-Forense”, faz uma importante defesa de parcerias na formação do profissional da FF. Apresenta as contribuições da Linguística, especialmente nas disciplinas de Fonética Acústica e de Sociolinguística, e da Fonoaudiologia na análise de voz. Finaliza com a nomeação dos principais centros de formação no Brasil e no exterior e com a defesa ao estabelecimento de parcerias entre órgãos governamentais e a universidade para a formação de peritos.

Nessa primeira parte da obra, os capítulos estão bem organizados, os conteúdos estão adequadamente distribuídos e o aspecto visual do livro como um todo, as imagens e tabelas, são de alta qualidade. No entanto, há alguns problemas no texto e nas figuras, possivelmente por uma revisão apressada, que não chegam a comprometer a qualidade da obra e que podem ser corrigidos em uma próxima edição. Um exemplo está na repetição das Figuras 3 e 4 nas páginas 19 e 20, que tratam de camadas diferentes do processo de etiquetagem. Os quadros e tabelas seguramente serão muito úteis à prática da perícia. No entanto, fica uma dúvida em relação ao Quadro 3 da página 26. Embora bem explicado o propósito do uso do código ASCII, por utilizar apenas fontes comuns dos teclados e possibilitar o uso de *scripts*, não se explica o porquê da correspondência com os códi-

gos do IPA ter sido feita pela Fonologia, com o uso de arquifonemas, e não com Fonética, como é comum nas transcrições de fala. A tabela também contém algumas falhas como, por exemplo, a falta do /s/ em ataque silábico.

Parte 2 – Questões atuais em análise fonético-forense para comparação de locutor

Conforme apresentação dos organizadores, a segunda parte do livro se compõe de trabalhos práticos ou acadêmicos de pesquisadores comprometidos (seja na área da Linguística, seja da Fonoaudiologia), com técnicas para a CL, alguns deles envolvidos com a formação de peritos para análise de fala. Com o título “O peso da evidência sociofonética na perícia de Comparação de Locutor”, o Capítulo 7 assinado por Cláudia Regina Brescancini e Cíntia Schivinscki Gonçalves, traz uma metodologia que, utilizando a escala verbal qualitativa de Eriksson (2012) de razão de verossimilhança, propõe um sistema para mensurar as evidências sociofonéticas com potencial distintivo, discutindo os conceitos de similaridade e tipicidade. Infelizmente, quando propõem a escala de Eriksson e discutem sobre o seu uso em laudos/pareceres periciais, as autoras não mencionam o Capítulo 5, que discorre sobre laudos/pareceres e também recomenda a escala de Eriksson. Um ponto negativo do capítulo, e do livro como um todo, é a falta de uma maior articulação entre os textos. Nesse caso especificamente, o mais grave é que há divergência no ano da referência, certamente pelo problema de revisão já mencionado. Apesar disso, a proposta de um valor numérico para o confronto de amostras nos processos sociofonéticos, aplicado a uma escala verbal qualitativa, parece ser uma ideia bastante interessante para a perícia.

Sandra Madureira e Zuleica Camargo assinam o Capítulo 8, “O Protocolo de análise perceptiva de voz VPA e seus usos para a área forense”, que se dedica a descrever o protocolo de análise de voz criado por Laver nos anos 1980 como ferramenta de análise em contexto forense. Esse protocolo tem demonstrado sua utilidade em análise de voz em fonoaudiologia clínica (Camargo e Madureira, 2008) e pode ser também útil na área forense.

O Capítulo 9, “Os efeitos da transmissão telefônica e do estilo de fala telefônico no sinal de fala”, escrito por Renata Regina Passetti, faz uma ótima discussão sobre os efeitos técnicos e os efeitos de locutor no sinal de fala em transmissão telefônica. A autora expõe questões metodológicas e resultados de dois trabalhos que conduziu sobre esses dois efeitos, oferecendo um rico material para pesquisadores e profissionais envolvidos em CL em dados do português brasileiro, uma vez que muito do material de análise se origina de conversa telefônica.

Se os três capítulos já mencionados têm objetivos bem definidos e atendem às expectativas do leitor a partir do título, o mesmo não acontece com o Capítulo 10. Com o título “Os efeitos individuais e a variação regional”, assinado por Ana Carolina Constantini e Aline de Paula Machado, o capítulo não é claro em seus objetivos. As autoras afirmam que vão tratar do comportamento de parâmetros prosódico-acústicos com foco em variedades de fala, mas tratam também de efeito do ruído e passam mais da metade do capítulo descrevendo bancos de dados. Embora os três temas tratados sejam pertinentes ao que o capítulo anuncia no título, a forma como foram apresentados deveria ter uma melhor articulação.

O capítulo seguinte também gera no leitor uma sensação de objetivo não atingido, principalmente em relação ao seu título, “Aspectos metodológicos e ferramentas para análise forense”. Na verdade, o autor Pablo Arantes trata de um tema relevante para a CL, que é a duração de uma amostra de fala, mas o título gera diferentes expectativas. Os métodos de que o capítulo trata são para definir tamanho ideal de amostra ou quantidade de unidades linguísticas para a extração de parâmetros confiáveis na comparação de vozes. Como o próprio autor menciona, é uma questão que deve ser colocada em um momento anterior à análise forense em si.

O último capítulo, traz uma boa revisão de trabalhos sobre o papel da genética e do ambiente, considerando locutores geneticamente relacionados. Com o título “Análise fonético-acústica em gêmeos idênticos: os limites da variação entre locutores”, de autoria de Julio Cesar Cavalcanti, o capítulo discute sobre questões relacionadas às abordagens articulatória, perceptiva e acústica e, nesta última, discorre sobre diversos parâmetros de análise. O autor traz também alguma discussão sobre a velha dicotomia genética versus ambiente.

Finalizando esta resenha, reafirmo com veemência que se trata de uma obra de grande valor para a área da Fonética Forense no Brasil. Vai indubitavelmente ajudar estudantes, pesquisadores e, especialmente, profissionais da área em tarefas de comparação de locutor. Sejam linguistas, fonoaudiólogos, engenheiros, ou outros profissionais e pesquisadores, todos se beneficiarão dos conhecimentos compartilhados pelos autores que assinam os textos que compõem o livro. As falhas apontadas não prejudicam a qualidade da obra, mas devem ser corrigidas em uma próxima edição.

Notes

[1]Nesse repositório estão vários textos que detalham os diversos procedimentos no trabalho de CL, desde a coleta de áudio e condução de entrevista até a confecção de laudo pericial/parecer técnico. Lá se encontram também *scripts* para o software PRAAT, tabelas de Excel e material de áudio. Muitos dos textos de orientação são bem similares a capítulos do livro, mas com maior detalhamento, principalmente para o uso do PRAAT.

Referências

Camargo, Z. e Madureira, S. (2008). Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). In *Proceedings Of The Fourth Conference On Speech Prosody*, Campinas, SP.

Eriksson, A. (2012). Aural/acoustic vs. automatic methods in forensic phonetic case work. In A. Neustein e H. Patil, Orgs., *A Forensic Speaker Recognition: Law Enforcement and Counter-terrorism.* New York: Springer-Verlag, 41–69.