

TXT 4N6: METHOD, CONSISTENCY, AND DISTINCTIVENESS IN THE ANALYSIS OF SMS TEXT MESSAGES

*Tim Grant**

INTRODUCTION

This paper presents a case study in forensic authorship analysis for SMS text messages. The case involves a domestic murder where the husband attempted to disguise the timing and mode of his wife's death through sending a series of SMS text messages from her phone.¹ Late in the evening on Sunday, January 17, 2009, the fire brigade was called to the home of Christopher and Amanda Birks in Stoke-on-Trent, UK. After phoning the emergency services, Christopher Birks entered the house and rescued his sleeping children from the lower floors. On the arrival of the firefighters, he informed them that his wife had gone to bed in the attic bedroom and that she must be trapped there. Placing themselves at considerable risk, firefighters entered the building to attempt a rescue from the top floor bedroom, but were only able to recover Amanda Birks' severely burned body. Apparently Amanda had been in bed

* Centre for Forensic Linguistics, Aston University.

¹ Details of the case as described in this introduction were supplied to me in my role as External Expert Advisor to Staffordshire Constabulary and were provided in the course of a series of police briefings and prosecution case conferences. The case received limited local and national news coverage in the United Kingdom. *See, e.g., Businessman Admits Murdering Wife and Setting Fire to Her Home*, TELEGRAPH (Nov. 3, 2009, 7:00 AM), <http://www.telegraph.co.uk/news/uknews/crime/6488982/Businessman-admits-murdering-wife-and-setting-fire-to-her-home.html> [hereinafter *Businessman Admits Murdering Wife*].

when the fire began, and there was no indication that she had woken during it or made any attempt at escape.

Subsequent forensic examination showed that fibers recovered from Amanda's body were from her daytime clothes, and toxicology reports indicated that Amanda's lungs contained little or no carbon monoxide. Those findings indicate that when Amanda went to bed, she had not changed into her nightclothes and that when she was burnt, she had not been breathing and thus had not inhaled fumes from the rising fire. As can be typical where a body is badly damaged by fire, no precise cause of death could be determined.

Christopher Birks asserted that Amanda had been at home during most of the day on January 17. Midmorning, an employee stopping by to collect wages had seen Christopher and Amanda at the house, and when Christopher was out during the afternoon, a series of SMS text messages were sent from Amanda's phone. These messages, sent to Christopher, employees of their joint business, and other family members, indicated that Amanda had been going about her normal business at home. They also suggested that Amanda had had important discussions with Christopher about the state of their relationship and that she was going to bed early "relaxing with candles" in the attic bedroom. Since Amanda's phone was destroyed in the fire, these messages were collected from the various recipients' phones.

Christopher Birks' account, however, seemed contradictory to the evidence from the domestic burglar alarm. Expert technical examination of the alarm showed that it had been fully set during the afternoon of January 17 but also showed that it had not recorded Amanda's movements about her home.

A circumstantial case was built against Christopher Birks, which included the evidence of the textile fibers on Amanda's body, the lack of carbon monoxide in her lungs, the burglar alarm log, and finally, forensic linguistic evidence concerning the authorship of the text messages. Christopher Birks was charged with the murder of his wife and with the endangerment of the lives of his children and of the firefighters.

This article will describe the linguistic analysis carried out and consider the methodological and theoretical basis for that analysis. The method employed in this specific case was purely

descriptive. This article, however, uses that method as a starting point from which to discuss and develop a general statistical method for forensic analysis of text messages. This article posits that such a method will be a helpful tool in future cases for analyzing text messages and other short form messages.²

I. FORENSIC AUTHORSHIP ANALYSIS

A. *Stylometric Versus Stylistic Approaches to Analysis*

Forensic authorship analysis of written texts is achieving increasing acceptance in the United Kingdom's courts. Academically, there is a significant literature developing around the discussion of the theoretical presuppositions and implications of this work, the necessity and limits of quantification in the field,³ and the law and application of the law concerning admissibility of such evidence.⁴

While much of this discussion is beyond the scope of this article, it is relevant to note that in the UK, admissibility is subject to review by the UK Law Commission.⁵ On the basis of the published Law Commission report, it seems that the UK

² Such short form messages include Twitter feeds, Blackberry Messenger communication, and Facebook status updates.

³ See generally Tim Grant, *Text Messaging Forensics: TXT 4N6: Idiolect Free Authorship Analysis?*, in THE ROUTLEDGE HANDBOOK OF FORENSIC LINGUISTICS 508, 508–22 (Malcolm Coulthard & Alison Johnson eds., 2010) [hereinafter Grant, *TXT 4N6*] (discussing the theoretical assumptions about the nature of an individual's linguistic personae and how such assumptions are expressed through methods adopted in authorship analysis work). See also Tim Grant, *Quantifying Evidence for Forensic Authorship Analysis*, 14 INT'L J. SPEECH LANGUAGE & L. 1, 1–25 (2007) [hereinafter Grant, *Quantifying Evidence*]; Moshe Koppel et al., *Authorship Attribution in the Wild*, 45 LANGUAGE RESOURCES & EVALUATION 83, 83–94 (2011).

⁴ See Lawrence M. Solan & Peter M. Tiersma, *Author Identification in American Courts*, 25 APPLIED LINGUISTICS 448, 448–65 (2004); Blake Stephen Howald, *Authorship Attribution Under the Rules of Evidence: Empirical Approaches in the Layperson Legal System*, 15 INT'L J. SPEECH LANGUAGE & L. 219, 222–24 (2009).

⁵ The UK Law Commission is a statutory body, independent of Parliament, whose function is to monitor and review laws and, where appropriate, make proposals for reform.

courts are likely to require a showing of the scientific validity of techniques not dissimilar to the *Daubert* criteria currently applicable in United States federal courts.⁶ Although scientific validity and reliability do not require quantification, and quantification is by no means sufficient to demonstrate appropriate scientific status, there appears to be a presupposition in some literature that an appropriate quantified method can make it easier to demonstrate that a method is both reliable and valid.⁷

Following previous work,⁸ I here draw a distinction between *stylometric* and *stylistic* approaches to authorship analysis. Stylometric approaches exemplified by scholars across the field seek to find or describe quantifiable markers of authorship, which in the general sense vary more between authors than within authors.⁹ Typical stylometric markers include relative frequencies of different word classes or even nonword letter clusters referred to as *n*-grams. The demonstration of the usefulness of a stylometric marker of authorship requires that, for almost any sampled set of authors, there be significant differences in the occurrence of the marker between authors, regardless of other textual variables such as topic, register, or

⁶ THE LAW COMMISSION, EXPERT EVIDENCE IN CRIMINAL PROCEEDINGS IN ENGLAND AND WALES 65–67 (2011), available at http://lawcommission.justice.gov.uk/docs/lc325_Expert_Evidence_Report.pdf; cf. *Daubert v. Merrell Dow Pharm.*, 509 U.S. 579, 579–95 (1993).

⁷ See Solan & Tiersma, *supra* note 4, at 454; see also Howald, *supra* note 4, at 236.

⁸ Grant, *TXT 4N6*, *supra* note 3, at 510–13.

⁹ See Shlomo Argamon & Moshe Koppel, *A Systemic Functional Approach to Automated Authorship Analysis*, 21 J.L. & POL'Y 299 (2013); John Burrows, *Questions of Authorship: Attribution and Beyond*, 37 COMPUTERS & HUMAN. 5, 5–13 (2003); Carole E. Chaski, *Empirical Evaluations of Language-Based Author Identification Techniques*, 8 INT'L J. SPEECH LANGUAGE & L. 1, 2–8; Carole E. Chaski, *Best Practices and Admissibility of Forensic Author Identification*, 21 J.L. & POL'Y 333 (2013); Grant, *Quantifying Evidence*, *supra* note 3, at 1–5; David I. Holmes et al., *Stephen Crane and the New York Tribune: A Case Study in Traditional and Non-Traditional Authorship Attribution*, 35 COMPUTERS & HUMAN. 315, 315–31 (2001); Patrick Juola, *Stylometry and Immigration: A Case Study*, 21 J.L. & POL'Y 287 (2013); Moshe Koppel et al., *Authorship Attribution: What's Easy and What's Hard?*, 21 J.L. & POL'Y 317 (2013).

genre. This I call *population-level distinctiveness*.¹⁰ With longer texts involving some thousands of words, such approaches can be used to make successful attributions.¹¹ Accordingly, a good stylometric marker should demonstrate that, on that marker, an individual is distinctive against the background population from which it is drawn. The presuppositions of this research are revealed through the language of some of its proponents who refer to linguistic fingerprinting¹² or the discovery of a stylome.¹³ This is the language of a forensic science discipline which can provide individuation—the discrimination of one individual from any other in a population.

In contrast to these stylometric approaches, forensic practitioners working on shorter and sometimes fragmentary texts have tended to use more stylistic approaches.¹⁴ Such approaches do not assume that the discovery of population-level discriminants is necessary to authorship analysis but focus on variation between specific individuals.¹⁵ Furthermore, that variation is understood as being created by habitual choice across a wide and unpredictable range of features.¹⁶ Thus, one author might fall into a habit of using unusual punctuation whereas another author might exhibit a preference for elaborate adjective use. Because, before examining a text, one does not know precisely what sort of feature one is looking for, quantitative methods tend to be less well defined

¹⁰ See Grant, *TXT 4N6*, *supra* note 3, at 515.

¹¹ See, e.g., Holmes et al., *supra* note 9, at 322–28.

¹² See, e.g., Sebastian Bernhardsson et al., *The Meta Book and Size-Dependent Properties of Written Language*, *NEW J. PHYSICS* 6 (Dec. 10, 2009), http://iopscience.iop.org/1367-2630/11/12/123015/pdf/1367-2630_11_12_123015.pdf.

¹³ See Hans van Halteren et al., *New Machine Learning Methods Demonstrate the Existence of a Human Stylome*, *12 J. QUANTITATIVE LINGUISTICS* 65 (2005).

¹⁴ See generally GERALD R. MCMENAMIN, *FORENSIC STYLISTICS* 161 (1993) [hereinafter MCMENAMIN, *FORENSIC STYLISTICS*]; GERALD R. MCMENAMIN, *FORENSIC LINGUISTICS: ADVANCES IN FORENSIC STYLISTICS* (2002) [hereinafter MCMENAMIN, *ADVANCES*].

¹⁵ MCMENAMIN, *FORENSIC STYLISTICS*, *supra* note 14, at 161; MCMENAMIN, *ADVANCES*, *supra* note 14, at 171–72, 174.

¹⁶ MCMENAMIN, *FORENSIC STYLISTICS*, *supra* note 14, at 162–70; MCMENAMIN, *ADVANCES*, *supra* note 14, at 45–65.

within this approach. At a population level, it has been suggested by McMenamin,¹⁷ among others, that a naïve Bayesian approach could be used to quantify stylistic features, and such an approach would seem to be appropriate. However, distributional assumptions and assumptions of independence are hard to demonstrate in the linguistic context, and it is difficult to model how their indeterminacy or violation affects the posterior probability of assigning a text to a particular author.¹⁸ In the UK legal context, a recent judgment criticized a Bayesian approach to footprint analysis, in part because of the lack of certain population data.¹⁹ Linguistic stylistic analysis might also stumble at this hurdle.

One aim of this paper is to demonstrate how it is possible to derive a methodologically rigorous approach to stylistic authorship analysis that can result in statistically described results. The approach is based on the analysis of vocabulary choices and morphological features and is particularly tailored to text messages and allied genres (such as Twitter feeds), though it may be generalized to other text types and other features. Rather than focusing on population-level distinctiveness, this approach gives primacy to *pairwise distinctiveness* between

¹⁷ MCMENAMIN, *ADVANCES*, *supra* note 14, at 171–72.

¹⁸ A full discussion of Bayesian approaches to forensic science problems is beyond the scope of this article, but a good introductory discussion is provided by DAVID LUCY, *INTRODUCTION TO STATISTICS FOR FORENSIC SCIENTISTS* 6, 108–12 (2006). The essence of the Bayesian approach is to ask, “What is the probability of the evidence given two opposing hypotheses?” The ratio of the two probabilities (in this case, the probability that the text was authored by AB over the probability that the text was authored by CB) gives a measure of the weight of evidence in favor of one or the other of the probabilities. In authorship problems this likelihood ratio can be applied to the occurrence of each individual feature and summed to provide an overall weight of evidence for authorship given a specified basket of features.

In discussion at this symposium, and through a useful commentary provided by discussant Professor Ed Cheng, it was advocated that I might use a Bayesian approach to examine the case described in this paper. Pressure of time and space meant that this was not pursued here, but a paper examining this should follow.

¹⁹ *R v. T*, [2010] EWCA (Crim) 2439, [86], [2011] 1 Crim. App. 9 (Eng.).

potential authors of messages, thus avoiding some of the population sampling issues which otherwise arise.

B. Consistency and Distinctiveness

All comparative authorship analysis depends upon two theoretical assumptions.²⁰ The first assumption is that there is a sufficient degree of consistency of style within relevant texts by an individual author. The second assumption is that this consistency of style inherent in an author's writings is sufficiently distinctive to discriminate the one author from other relevant authors. Ultimately, the idea that comparative authorship analysis rests upon a strong theoretical assertion of an idiolect is false. The empirical discovery of consistency and distinctiveness can, however, be a sufficient foundation for such work.²¹

The first assumption, that there is "a sufficient degree of consistency of style within relevant texts," requires further discussion. It is not necessary to identify features of an author's language that are wholly consistent. As shall be seen in the Birks case, a weight of evidence for authorship may be built upon a *degree* of consistency. It must be recognized, however, that the greater the degree of consistency in any comparison corpus, the greater the weight of evidence there will be for an attribution. Identifying consistency within relevant texts also requires the creation of a linguistically *relevant* comparison corpus, which accounts for genre²² as well as other sources of linguistic variation. For example, it must take account of accommodation effects between different recipients of messages and between the possible modes of production—whether the text message was created using a twelve-key alphanumeric system (as is found on more old-fashioned phones), a touch sensitive qwerty keyboard such as is found on an iPhone, or even through a speech-to-text system.

The second assumption raises different considerations. There may be *degrees* of distinctiveness between pairs of individuals or

²⁰ Grant, *TXT 4N6*, *supra* note 3, at 509.

²¹ *Id.* at 521–22.

²² To avoid comparison, for example, of text messages with emails.

within smaller or larger groups. It seems obvious that some pairs of authors will produce writings wherein the authors are easily distinguished, whereas other authors may generally produce texts that are stylistically similar. In the former case, a comparative authorship analysis will be easier, whereas in the latter case, it may be impossible to distinguish between the authors' writings. Therefore, it may not be necessary to show a writer's distinctiveness against all possible authors; it may only be only necessary to compare one author with other relevant authors in the case. A linguistic fingerprint or stylome may be a holy grail for some stylometric researchers; but, should this grail prove as elusive as the Arthurian Holy Grail, comparative authorship analysis can still proceed and provide useful forensic evidence. Generally, investigators or the circumstances of a case will provide the definition of the relevant set of authors, and, as will be demonstrated in the Birks case, in some circumstances it can be sufficient to provide evidence of distinctive style between authors without hazarding to provide evidence of author identification.

C. Linguistic Analysis of SMS Text Messaging in Previous Cases

Increasingly, linguists—interested in describing the nature of text messaging as textual, functional, and social phenomena—are studying the language variety used to communicate with mobile telephones and similar devices.²³ Text messaging is shown to cross age, gender, and cultural boundaries. Stylistically, text messages generally are not full of “texting language”—abbreviations and initialisms. In fact, these tend to comprise less than twenty percent of vocabulary choices in text messages.²⁴ What is characteristic is that there is little or no censure for nontraditional spelling variants or for syntactic ellipses (such as omission of articles, auxiliaries, and other parts of speech), and

²³ See DAVID CRYSTAL, *TXTNG: THE GR8 DB8 37–62* (2008) (identifying six principal distinctive features of text messages); see also David Bamman et al., *Gender in Twitter: Styles, Stances, and Social Networks*, 2–6, 30–31 (2012), <http://www.cc.gatech.edu/~jeisenst/papers/GenderInTwitter923.pdf>.

²⁴ CRYSTAL, *supra* note 23, at 22, 156.

thus in text messaging, creative language play is positively reinforced.²⁵ This creativity results in an area of linguistic production where idiosyncratic use can flourish, and this can be advantageous for the forensic analyst.

There have been several UK criminal cases involving text messaging which have not yet been directly reported in the research literature.²⁶ One example is Professor Malcolm Coulthard's involvement in the Hodgson case.²⁷ Coulthard assisted the successful prosecution of David Hodgson for the murder of Jenny Nicholl, even though her body has never been found.²⁸ Coulthard observed a style shift in the sequence of text messages sent from Nicholl's phone.²⁹ Specifically, the latter messages were not compatible with Nicholl's previous texting style but were compatible with the previous style of Hodgson.³⁰ Coulthard's method involved discriminating between the two potential writers by identifying consistent and distinctive alternate vocabulary choices.³¹ He shows that Jenny Nicholl, in her undisputed messages, tended to write "my" and "myself" whereas the defendant, David Hodgson, tended to follow the North English pronunciation and use "me" and "meself."³² With

²⁵ *Id.* at 74.

²⁶ Examples of cases include one brought against Stuart Campbell for the murder of Danielle Jones, *Text Messages Examined in Danielle Case*, BBC NEWS (Oct. 9, 2002, 7:27 PM), http://news.bbc.co.uk/2/hi/uk_news/england/2314389.stm (describing the case as the first UK case to turn on text messaging evidence, with analysis provided by Professor Malcolm Coulthard), and a case of murder against Margaret James, *Margaret James Fails to Overturn Plot Conviction*, BBC NEWS (May 15, 2012, 1:49 PM), <http://www.bbc.co.uk/news/uk-england-cornwall-18078998> (discussing the background of the case). The prosecution in the James case initially introduced text messaging analysis but later withdrew it on the sight of expert response reports provided by Professor Coulthard and myself.

²⁷ See Owen Amos, *The Text Trap*, N. ECHO (Feb. 27, 2008, 11:38 AM), <http://www.thenorthernecho.co.uk/news/2076811.print/>; see also Grant, *TXT 4N6*, *supra* note 3, at 508–09.

²⁸ Amos, *supra* note 27.

²⁹ *Id.*

³⁰ *Id.*

³¹ Grant, *TXT 4N6*, *supra* note 3, at 515–17.

³² *Id.* at 516.

this and eight additionally selected opposing features, Coulthard demonstrates degrees of distinctive and consistent use in the undisputed messages and then goes on to show that the disputed messages are inconsistent with Nicholl's previous style.³³ In a similar way, he also concludes that the disputed messages are compatible with David Hodgson's known style.³⁴

While Coulthard's work in this case was admitted at trial and upheld at an appeal hearing,³⁵ his methodology can be developed and broadened in two respects. First, notwithstanding Coulthard's considerable linguistic skills and reputation, his method depends upon the expert identification and selection of potential distinctive vocabulary choices. This leaves open the possibility that confirmation bias may lead to the selection of features favorable to an analyst's implicit or explicit expectations about a case.³⁶ Describing a replicable process for feature selection may mitigate against this possibility to some degree. Second, while Coulthard's method is wholly descriptive rather than statistical, development of a quantified method may lead to a better demonstration of the reliability of the conclusions and the validity of the methods. The development of methods for the Birks investigation was intended to safely build on Coulthard's successful methods while simultaneously addressing these issues. The method described here draws on well-established methodological and statistical approaches used in behavioral case linkage as undertaken by forensic psychologists.³⁷

³³ *Id.*

³⁴ *Id.* at 515.

³⁵ See *The Failed Appeal*, JENNY NICHOLL (June 29, 2009), <http://jennynicholl.blogspot.co.uk/search?updated-min=2009-01-01T00:00:00Z&updated-max=2010-01-01T00:00:00Z&max-results=2>.

³⁶ See, e.g., Itiel E. Dror et al., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 FORENSIC SCI. INT'L 74, 76-77 (2006) (finding that fingerprint experts made different judgments when misled with extraneous contextual information).

³⁷ See, e.g., Jessica Woodhams et al., *From Marine Ecology to Crime Analysis: Improving the Detection of Serial Sexual Offences Using a Taxonomic Similarity Measure*, 4 J. INVESTIGATIVE PSYCHOL. & OFFENDER PROFILING 17, 17-27 (2007); Jessica Woodhams et al., *The Psychology of Linking Crimes: A Review of the Evidence*, 12 LEGAL & CRIMINOLOGICAL PSYCHOL. 223, 223-49 (2007).

In conclusion, the aim in the Birks analysis was to provide a sound evidentiary analysis based on a pairwise discrimination of Amanda and Christopher Birks' respective vocabulary choices in their text messages. My method seeks to avoid, as much as possible, my own biases affecting the selection of features, the analysis, or the drawing of conclusions. In addition, in the analysis presented in this article, I shall extend the actual approach taken in the case to demonstrate a statistical approach that can further support future analyses.

II. METHODS

This section follows with a description of the available data and the analytic approach taken in the Birks analysis. Justifications are provided for the decisions taken.

A. Data

The text message data were provided by Staffordshire Police as part of their investigation into the death of Amanda Birks, and I was a paid expert witness for the prosecution.³⁸ Case details and the forensic analysis were released to the UK media.³⁹

Two small corpora of text messages were provided in the form of spreadsheets identifying phone numbers, names of recipients, sending times, and text for each SMS message. The first spreadsheet contained collated text messages sent by Amanda Birks ("AB") to eleven separate recipients over a period of several days prior to January 17, 2009, the date of her death. As AB's phone was destroyed in the house fire, police collected this information from reports produced by forensic telecommunication engineers who downloaded the messages from the recipients' telephones.⁴⁰ After a number of automated

³⁸ Subsequent to the conclusion of the case, permission was obtained from the police to use these materials in teaching and research.

³⁹ See, e.g., *Businessman Admits Murdering Wife*, *supra* note 1.

⁴⁰ I was also provided with the raw telecommunications engineers' forensic reports, but I used these only to check the accuracy of the spreadsheets.

and repeat messages were removed from the spreadsheet, 204 text messages apparently sent by AB remained.

On a preliminary version of this spreadsheet, investigators indicated that a shift in texting style occurred on January 17, 2009 at 12:07 p.m. After that point, investigators believed the messages sent from AB's phone were not crafted by AB. The basis of this hypothesis was essentially a shift in the manner of signing off the messages using "kiss" marks. Prior to this time, the assertion was that AB tended to sign off using "Xxx" or "xxx" with no spacing. From about midday, the style shifts to include spaces in the sign off—"X x." When investigators brief a forensic expert, it is common for them to have already formulated a hypothesis based on such observations. Since confirmation bias may affect expert findings, this is not ideal. In the interest of full disclosure, the investigators' hypothesis and the concerns it raises were reflected in the final forensic report.

In addition to the AB data, a parallel corpus was provided containing undisputed text messages sent by Christopher Birks ("CB") to ten separate recipients. CB had deleted all sent messages from his phone such that they could not be recovered, and so these messages were also collected from recipients. The spreadsheet also contained a small number of unsent messages from CB's draft message folder, which had not been deleted. After a number of automated messages and repeat messages were removed from this spreadsheet, there remained 203 text messages apparently sent by CB.

Where a message had been automatically split into two messages for sending because the original was greater in characters than the permitted SMS length, these were left as two messages.

B. Feature Analysis, Coding, and Preliminary Selection

The method depended upon the identification of vocabulary choices with their associated spelling variants. As the intention was to avoid selectivity, all word forms were listed using corpus linguistics software WordSmith tools⁴¹ and coded as being

⁴¹ See generally Mike Scott, *WordSmith Tools Manual, Version 3.0*,

present or not present in every message. In addition to coding these lexical items, a number of further features of texting abbreviations were also coded. These features included the use or absence of spacing between words in certain contexts; letter and number substitutions (using “c” for “see” or “4” for “for”) for syllables and whole words; accent stylizations (“ad” for “had” or “cuz” for “because”); initialisms (“imho” for “in my honest opinion”); and so on. In total a basket of 154 features were developed, and every text message was scored as containing or not containing each of these features.

As the demonstration of consistency requires a certain amount of repetition of a feature, the next decision made was to reject any feature that had fewer than ten occurrences across the 407 text messages. As might be expected, this resulted in the removal of the majority of features from the analysis, leaving a reduced set of just twenty-eight features that formed the basis of the ongoing analysis.

C. Identifying Undisputed Text Messages

The text messages were then examined more closely to determine which messages could be considered to be, without dispute, of known authorship of either Amanda Birks or Christopher Birks. As described above, the police had indicated that they were suspicious of messages sent after midday on January 17, 2009. Taking a precautionary approach, all texts written after midnight on January 16, 2009 were considered disputable and set aside.⁴² In addition to this, two further texts apparently sent by CB but from AB’s phone were discarded—one was explicitly signed “Chris” and the other from its content appeared to have been sent by CB. Removing these messages left a total of 165 messages, and it was taken to be a reasonable but not infallible assumption that these messages had indeed been sent by AB.

LEXICALLY.NET (1998), <http://www.lexically.net/wordsmith/version3/manual.pdf> (explaining how software analyzes word behavior in texts).

⁴² It was only later learned that AB had been seen alive by independent witnesses at approximately 11:00 on the morning of the January 17, 2009.

Examining CB's messages in a similar way, there was no indication for removal of any messages, and this left the original set of 203 messages intact.

D. Identifying Features with Discriminative Power

The selected undisputed texts and features were further examined to determine whether there were features that discriminated consistently to some degree between the two writers in their known texts. Features were only retained in the analysis if one author used them in at least twice as many messages as the other—that is to say where one author used a feature at a rate of more than sixty-six percent of its total occurrence. This left a set of just eighteen features that are tabulated in Table 1.

This frequency table based on the known messages of AB and CB clearly indicates, in a general sense, that CB and AB write text messages using distinctive styles, and for some features the degree of distinctiveness is absolute. AB never, in this data, writes "with" as "wiv" and CB never writes "had" as "ad." Other features are features of predominately one author; CB tends to use commas, and AB does so very rarely. This contrasts, for example, with AB's tendency to use "t" for "the," which is only rarely used by CB. For each author, lists can thus be created of features characteristic of their text messaging style; ten features for AB and eight features for CB. In contrast to Coulthard's method in the Jenny Nicholl inquiry, these feature lists are not necessarily reciprocal alternates. For two words, "don't" and "with," each writer has as a feature a preferred spelling variant of that word, but for the other features this is not the case.

For neither author can these lists be considered identifying in an absolute sense. The features contained in these lists are not linguistic "fingerprints" identifying individuals against a population. Rather, they demonstrate a relative consistency of habit and a pairwise distinctiveness which thus can be used to stylistically discriminate between messages of the suspect and the victim in this case. As we do not have good knowledge of the distribution of texting features across the population of all

texters, we cannot say with certainty how many other people will share each of these sets of features. I did test these feature sets against a corpus containing ten messages from each of 500 texters, and no other individuals demonstrated the use of either complete set of features. Such information is useful but cannot be employed in statistical calculations, as theoretical linguistic difficulties remain over how any such reference corpus can be considered representative of the population of texters.⁴³

Table 1: Frequency distribution of elicited features

	Feature	# in CB texts	# in AB texts	Total	% in AB texts	% in CB texts
Features characteristic of AB's texting	"ad" for "had"	0	13	13	100%	0%
	"dont" for "don't"	0	9	9	100%	0%
	"t" for "the"	1	15	16	93.8%	6.3%
	"bak" for "back"	1	10	11	90.9%	9.1%
	"av" for "have"	1	9	10	90.0%	10.0%
	"wud" for "would"	2	9	11	81.8%	18.2%
	"w" for "with"	3	10	13	76.9%	23.1%
	"y" for "yes"	2	6	8	75.0%	25.0%
	"wil" for "will"	4	9	13	69.2%	30.8%
	"wen" for "when"	4	9	13	69.2%	30.8%
Features characteristic of CB's texting	"dnt" for "don't"	8	0	8	0%	100%
	"jst" for "just"	12	0	12	0%	100%
	"wiv for "with"	15	0	15	0%	100%
	4 for "for" with no trailing space	35	0	35	0%	100%
	2 for "to" with no trailing space	58	0	58	0%	100%
	Use of comma	87	5	92	5.4%	94.6%
	"4get" for "forget"	15	1	16	6.3%	93.8%
	"thanx" for "thanks"	16	2	18	11.1%	88.9%

⁴³ See, e.g., Grant, *Quantifying Evidence*, *supra* note 3, at 6–9, 7 fig.1 (discussing issues of population sampling for authorship analysis work).

E. Quantification of Distinctiveness

In the analysis that contributed to the investigation, Table 1 was used to demonstrate the distinctiveness between the two authors. In developing the model further, it is possible to follow methods rooted in forensic psychology to demonstrate the distinctiveness statistically. Each feature is marked as present or absent in each text message by using a “one” for a present feature and a “zero” for an absent feature. The resulting array of zeros and ones can then be used to compare messages using a binary correlation known as Jaccard’s coefficient.⁴⁴ Pairs of messages were placed into three categories: the first two categories were within-author pairings, each comprising two texts by AB and two texts by CB. Using the random case selection feature within SPSS statistical analysis software, a sample of 100 within-author pairings was taken for each author.⁴⁵

Table 2: Jaccard values for linked and unlinked pairs of messages

	AB-AB pairs	CB-CB pairs	AB-CB pairs
Mean Jaccard	0.195	0.199	0.09
Standard Deviation	0.24	0.25	0.12

⁴⁴ Jaccard’s coefficient is a correlation for binary values and can be typically read as a distance measure. Results vary between zero and one where one indicates that two binary arrays are identical and zero indicates that they are completely different. Decimals between zero and one indicate variation between these two extremes. One advantage of Jaccard is that it does not inflate similarity on the basis of two absences. Absence of evidence of a stylistic feature in a particular text message is not evidence of its absence from that individual’s stylistic range when texting generally, and thus using Jaccard does not risk overstating the explanatory power of a single text.

⁴⁵ A within-author pairing comprised either two AB texts or two CB texts; a between-author pairing comprised one AB text and one CB text. Given 165 AB texts and 203 CB texts were used, this sample was taken from more than 13,000 potential AB pairings and more than 20,000 potential CB pairings.

In addition, a further random sample of 100 between author pairings was taken. For each of these three sets of pairings, Jaccard's coefficients were calculated.

As Table 2 suggests, the mean Jaccard values for linked pairs of messages show roughly similar levels of within-author consistency; Jaccard values for AB-AB pairs of messages and CB-CB pairs of messages are relatively close. The Jaccard values for unlinked pairs, each containing an AB and a CB message, tend to have considerably lower Jaccard scores. That is to say that linked pairs appear to have a greater degree of stylistic similarity than unlinked pairs. On further examination, however, it can be seen that the Jaccard values fall into nonnormal distributions (Kolmogorov-Smirnov Z values, respectively: AB-AB pair = 2.30, CB-CB pair = 2.19, AB-CB pair = 3.21; $N = 100$ in each group; $p < 0.0005$) indicating a nonparametric approach is required.⁴⁶

Using Mann-Whitney U comparisons, pairs of messages containing only AB texts can be shown to be significantly more similar than pairs of messages each containing a CB and AB text ($U = 3832$; $N = 200$; $p = 0.002$). Additionally, pairs of messages containing only CB texts can be shown to be significantly more similar than pairs of messages each containing a CB and AB text ($U = 3730$; $N = 200$; $p = 0.001$). These findings both demonstrate a significant degree of consistency of style within each author and at the same time distinctiveness between the authors' text messaging styles.

F. Disputed Text Messages

Having determined distinctiveness in style between the two authors by examining the undisputed material, it is possible to analyze the disputed text messages as a group. The disputed messages are hereafter referred to as queried or Q messages.

⁴⁶ Many statistical tests (such as t -tests) require data to approximate a normal distribution and the Kolmogorov-Smirnov Z test can be used to test whether a distribution is normal. Where data is not normally distributed, one must choose an alternative test. The Mann-Whitney U is in essence equivalent to a t -test but can be used on nonnormal, nonparametric data.

We can follow the quantified method described above. Jaccard values were obtained for a random sample of 100 AB-Q pairs of messages and 100 CB-Q pairs. The AB-Q messages were compared with AB only linked pairs. The Mann-Whitney comparison showed a significant lower Jaccard value for the AB-Q pairs ($U = 4119$; $N = 200$; $p = 0.016$). In contrast, no significant difference can be shown between the CB-only linked pairs and the Jaccard values for pairs of messages containing a CB and a Q message ($U = 3572$; $N = 200$; $p = 0.203$). Taken as a set, the stylistic choices made in the disputed messages show significant dissimilarity from the stylistic choices in Amanda Birks's undisputed messages but no equivalent difference can be shown for the undisputed messages of Christopher Birks.

G. Examining the Sequence of Individual Messages

The quantified method can demonstrate the general point of consistent style within authors, which can discriminate between them, but this takes one only so far. Appropriate caution, however, requires that the queried texts not be treated as a homogenous group but rather that the authorship of each queried message be considered on its own merits. Quantified classification of individual data points in a nonnormal distribution would create a considerable statistical challenge, and because of this, a more qualitative descriptive analysis is preferred.

The queried messages fall into a rough time-ordered sequence. Caution is required—since the messages were collected from recipients' phones, the time tagged on each messages may not be an accurate indication of when the message was sent.

To demonstrate the descriptive method, a test case is provided by the data. A message sent on January 13 was removed from the spreadsheet of known texts attributed to AB because, although it was sent from AB's phone, it was signed "Chris." The message was longer than the permitted SMS length, and so sent as two messages, but in full it reads:

I think u were abit hard on me earlier, 1. I wasnt l8 purposely, 2, i nd a car 4assessments. Iv had a luk at myself in mirror n undastand 1, y u dnt fancy me n 2, y u dnt like me. Sory 4all the pain iv caused u, i love u amanda..... Chris. Xxx

Messages sent 01/13/09

Qualitatively, this message can be analyzed for features consistent with AB's texting style and features consistent with CB's texting style. Of the eight features considered characteristic of AB's text messaging style indicated in Table 1, none of them appear in this message. Further, given that "had," "the," and "don't" are used in the message, these might be considered opportunities to display these features. On this descriptive basis, the analysis suggests this message can be considered inconsistent with AB's style.

In contrast, the message contains features considered characteristic of CB's style: six commas, the use of "dnt" for "don't," and two uses of the digit "4" for the word "for" where there is no trailing space, i.e., "4assessments" and "4all." On this basis one can reasonably conclude that the message is consistent with CB's texting style.

In the forensic context the expression of results is important. In this case the language used is that of consistency and distinctiveness; thus, initial opinions in the forensic report for this case are as follows:

- i. Some of the messages sent from Amanda Birks phone . . . on the 17 January 2009 are stylistically distinctive from messages known to have been sent by Amanda Birks before that date.
- ii. Analysis of text messages known to have been written by Christopher Birks has given rise to a description of his habitual style which is described in this report. This described style will be shared by a limited number of people and is distinctive from the habitual style of Amanda Birks.
- iii. Some of the messages sent from Amanda Birks phone . . . on the 17 January 2009 show stylistic consistency

with text messages known to have been sent by Christopher Birks on and before that date.

The absence of identifying language is deliberate and significant. I am able to say that there are “a limited number of people” who may share Christopher Birks’s texting style because I have tested the feature set against my independent corpus of 500 texters, but it is a weakness of my position that I have no substantial basis to say how many individuals in a population will share this feature set. Another way of thinking of this is that the question before the court is, “How likely is it that Christopher Birks wrote the text messages on the 17 January 2009?” I do not answer this question; rather, I would respond that “as the texts are distinctive from AB’s historic style and consistent with CB’s historic style, it is more likely to have been Christopher Birks than Amanda Birks who wrote those texts.” While this information is useful to the court considered in conjunction with other evidence, it is by no means identification evidence standing alone.

H. Messages of January 17, 2009 Sent from AB’s Phone

Turning to the list of messages sent from AB’s phone on January 17, 2009, each message was evaluated in turn to arrive at a qualitative conclusion for stylistic consistency with each of CB’s and AB’s previously described style.

Table 3: Examination of disputed messages

Time	Sent To	Message	Comment
00:40	Friend 1	I love u my gorgeous sexy babe! Xxx	Contains none of the specific features listed for either AB or CB.
10:04	Friend 1	Got go fetch milly. Val cant cope w her x	Contains “w” for “with,” identified as a feature of AB’s style. CB tends to use “wiv” but does use “w” on occasions.

AB was seen alive at 11 a.m.

11:39	Friend 1	Wen r u up 4a repeat performance? X x	Digit “4” used for “for” with no trailing space is used only by CB. “Wen” is used by AB with twice the frequency of CB.
11:49	Friend 1	Wot do think? Cant believe theresa! X x	Contains none of the identified features. The use of “wot” is not discriminating between AB & CB.
12:07	Friend 1	Txt me, talkin with chris. X x	Contains a comma, which is rare in AB’s texts. Slight inconsistency with AB, consistent with CB.
12:10	Friend 1	About your route, spk lata, talkin with chris. X x	Contains commas, which are rare in ABs texts. “Spk” and “lata” are only used in the disputed messages.
12:39	Friend 1	U wen u filled ur application in. X x	Contains “wen” used twice as often by AB than CB.
12:39	Friend 1	Am talkin wiv chris, am confused. Ur 2 young 4me. X x	Contains the use of commas, the use of “wiv” rather than “w” or “with” and the use of “4” without a trailing space. <i>First text to be judged inconsistent with AB and consistent with CB.</i>
12:41	Friend 1	Txt u lata. X x	Contains none of the identified features.
12:54	Friend 1	Chris is sayin the same, giv me space, u know wot i think of u. X	Contains commas, which are rare in AB’s texts.
13:02	Friend 2	Not sure yet, am jst talkin wiv chris so will txt u lata, dont worry.	Contains commas, the use of “wiv” rather than “w” or “with,” and the use of “jst.” Also contains “dont” (with no apostrophe) which otherwise is used only by AB.

14:03	Friend 1	Dnt kno, need think, am goin relax in bath then go sleep, really tired. X	Contains commas and "dnt." Judged inconsistent with AB; consistent with CB.
14:05	Friend 1	Will do when i wake up, so tired. Need talk 2 chris. Still hav feelins 4him. X	Contains commas and the use of "4" without a following space. Judged inconsistent with AB; consistent with CB.
14:12	Friend 1	Just give me time and space please, im tired, confused n wud like runaway, x	Uses "wud" which is and unusual for CB and "just" rather than "jst."
14:51	Friend 3	Iv bin out all week, me n chris not been gettin on so probably stay in 2talk.x	Contains commas and the use of "2" without a following space. Judged inconsistent with AB; consistent with CB.
16:58	Friend 1	Just need get my head straight, maybe 2mro	Contains comma but also "just."
17:00	Friend 1	I can try?	Contains none of the identified features.
17:06	Friend 4	Talkin with chris, tryin c if theres a way 4ward 4us. Love u.	Contains commas and the use of "4" without a following space. Judged inconsistent with AB; consistent with CB.
17:27	CB	Take 1day at at time but u still mean alot 2me n i do love u. X	Contains "2" without a following space. Judged inconsistent with AB; consistent with CB.
17:32	Friend 2	Sorry just want time 2think. Been talkin 2chris so tryin get my head 2getha! R u out with wayne 2nite? X	Contains "2" without a following space. Judged inconsistent with AB; consistent with CB.
17:37	Friend 2	Thanx tim, i just want chris talk 2me right, he needs learn not take his stresses out on me, then we can b happy. X	Contains commas and the use of "2" without a following space. Judged inconsistent with AB; consistent with CB.
18:11	CB	Goin stay in and look	Contains comma and the use

		through my of “2” without a following aromatherapy books, space. <i>Judged inconsistent with AB; consistent with CB.</i> 2tired u hav fun. Xx
19:34	Friend 3	Dont really want talk, am shattered, nan says i shud sort it, we had a talk n probably will, he just needs talk 2me better! Xx
19:48	Friend 3	Weve come so far, had a great xmas, things just ontop of us, think we need a holiday. Xx

In Table 3, each message sent on January 17, 2009 is treated individually, but it is possible to consider the set of messages as a roughly sequentially ordered group. The early text messages in this series demonstrate more consistency with AB’s known texting style than with CB’s known texting style. The later messages demonstrate a reciprocal pattern of more consistency with CB’s style and more inconsistency with AB’s style. This pattern, however, is not absolute.

The first message that is clearly and substantially inconsistent with AB and consistent with CB was received at 12:39 p.m. and reads:

Am talkin wiv chris, am confused. Ur 2 young 4me. X x

Of the seven messages that are timed before this message, two are entirely consistent with AB’s style; two are consistent with CB’s style and are inconsistent with AB’s style; and two demonstrate none of the identified features at all. One message in this earlier set demonstrates consistencies and inconsistencies with both authors.

Of the sixteen messages timed as occurring after this point, eleven messages demonstrate consistent features with CB’s texting style and or features inconsistent with AB’s texting style. No messages are wholly consistent with AB’s texting style. Three messages demonstrate consistencies and inconsistencies with both authors. Finally, two messages demonstrate none of the identified features.

Explanations for this slightly unclear picture are both linguistic and technical. The technical issues concern the fact that the timings of these messages are from the recipient phones. It is possible that delays in the telephone network created a time gap between sending and receipt. In addition to technical issues, the simple issue of linguistic variability has to be taken into account. Although I have attempted to describe levels of consistency within the known texts of AB and CB respectively, this does not preclude the possibility of shifts in texting style by either author for unknown reasons. Even where a consistency of style has been demonstrated over a stretch of two hundred messages, it must be considered that such a pattern could change.

III. DISCUSSION

I have presented here a method for the forensic authorship analysis of SMS text messages. In some ways, the case is straightforward: the police evidence indicates a pair of candidate authors. Although a further author cannot be precluded as a possibility, the presence of a pair of candidate authors makes the analytic task easier.⁴⁷

Framing the task in terms of consistency and distinctiveness allows for a combination of statistical and descriptive methods. Describing the points of consistency in the two corpora of undisputed messages allows one to quantify what is essentially a stylistic description and thereafter conclude statistically that a pairwise discrimination can be obtained between them. Avoiding claims about any population distribution of the identified features limits the conclusions that can be drawn. The *R v. T* case⁴⁸ suggests that quantification of identification requires some approximate knowledge of distributional data, and this is not available or perhaps even not obtainable for language data.⁴⁹ Given these concerns, it is not possible to identify Christopher Birks as the sender of the last messages from his wife's phone,

⁴⁷ In my experience, many comparative, forensic authorship analyses are similar comparisons between small sets of potential authors.

⁴⁸ *R v. T*, [2010] EWCA (Crim) 2439, [86], [2011] 1 Crim. App. 9 (Eng.).

⁴⁹ Grant, *Quantifying Evidence*, *supra* note 3, at 14.

but it is possible to assert that his style is distinctive from his wife's and that the last messages are inconsistent with her previously described style and compatible with his style. As a contribution to a wider criminal case involving other forms of evidence, this is strong information and can be useful enough to contribute to the evidence in the case.

With regard to methodology, one of the perceived weaknesses of stylistic analysis can be an overreliance on subjective expertise and an apparent lack of method in the identification of features. The method employed here attempts to limit that subjectivity. A further remedy would be the explicit statement of a protocol for feature identification and analysis, which could be designed and stated in advance of approaching an individual case. Casework invariably involves working with awkward situations and imperfect data. Consequently, one aspect of practitioner expertise, generally underreported, is the negotiation of this real world difficulty.

*A. Proposed Protocol for Stylistic Analysis in
Classification Problems⁵⁰*

*1. Try to Know as Little as Possible
About the Wider Details of the Case.*

The aim here is to mitigate the well-documented cognitive biases that occur across forensic disciplines.⁵¹

2. Describe the Features of the Known Texts First.

Once it has been established that the known texts are

⁵⁰ I divide cases into classification, inclusion, and exclusion problems. Classification problems take the form, "Which of these set of authors is the most likely to have written the query text?" The definition of the set of potential authors will be defined by nonlinguistic evidence, and it must be explicitly stated that linguistic conclusions presuppose the soundness of this evidence. This protocol is only for such classification problems.

⁵¹ See Itiel E. Dror et al., *Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison*, 208 FORENSIC SCI. INT'L 10 (2011); Dror et al., *supra* note 36, at 74.

linguistically relevant to the queried texts, the queried texts should be put aside. Ideally this preliminary examination might be carried out by a first analyst who then provides just the known texts and very limited information to a second analyst. In classification problems, this contrast provides a detailed but unbiased description of the known texts. The analysis must allow for possible outcomes where no distinction can be drawn between authors' styles in the known texts and for the possible outcome that the query texts are distinctive from all known authors' previously described styles. Ordering the analysis in this way allows for these important possible outcomes.

3. The Contrastive Analysis Should Elicit Consistent and Distinctive Features Within the Known Texts.

a. Within-Author Consistency

This requires several, and sometimes many, texts for each possible author. The number of texts of known authorship that make good comparison documents in terms of genre, recipient effects, and other linguistic variables becomes key in determining whether a case should be taken.

b. Between-Author Distinctiveness

It seems likely that stylistic distinctiveness can only be demonstrated pairwise or for small groups. That we do not have population distributions of stylistic features is not just a question of inadequacy or a lack of effort in carrying out linguistic surveys; linguistic complexity in the sources of language variation may mean that it is not possible to collect representative population samples of stylistic features.

The output of the contrastive analysis becomes a "locked" feature list that cannot be altered hereafter.

4. Carry out an Examination of the Query Texts for the Identified Features

If at this stage further features are found which seem useful

that are not contained in the locked feature list, it is too late as any such features cannot be included in the final analysis.

On the other hand there will be some features in the locked feature list that do not occur in the query texts. The significance of any such features should be considered as possible evidence that none of the included authors wrote the query texts.

*5. Draw Conclusions Based on the Consistency
and Distinctiveness of the Query Texts with
Each Set of Texts of Known Authorship.*

The Birks case was brought to me by Staffordshire police in the absence of any such protocol; as such, there were numerous times when my analysis did not follow this outline. For example, early on the police explained to me their hypothesis, which they without a doubt hoped my analysis would support. This is not a criticism of the police, who will have little knowledge of the needs of forensic linguistic analysis, but it may point to the need for an intermediary between the investigators and the forensic analyst such that the intermediary might control the information and data that reach the analyst.

No claim is made that the features that distinguish Amanda and Christopher's text messages will be useful in distinguishing between other pairs of authors or for the same authors in other genres. They are not population-level stylometric markers of authorship. Furthermore, and in contrast to Coulthard's analysis of the Nicholl-Hodgson case,⁵² the selection of markers did not depend on my individual skill in linguistic observation; rather, the features were elicited from the data according to a set of linguistically and statistically justifiable criteria. The method can be (and has been) developed and tested in other similar cases—and in a more recent case involving email analysis, the use of two analysts has proved invaluable. No claim, however, is made for the reliability of the specific set of markers used, and there are no grounds to generalize their use to other cases.

⁵² See *supra* Part I.C.

IV. CONCLUSION

Ultimately, the theoretical importance of the distinction between population level distinctiveness and pairwise or small group distinctiveness is crucial to the success of this approach and to the statistical method employed. While this limits the opinions of the expert to considerations of consistency and distinctiveness, given the richness of linguistic variation, such a limitation is appropriate to forensic authorship analysis and provides an expert with more certain and more credible evidence to offer the courts.

On November 2, 2009, Christopher Birks was due to be tried. On the morning before trial, he changed his pleas to “guilty” of the murder of his wife, “guilty” of arson, and “guilty” of the endangerment of his children and of the firefighters. He was subsequently handed down a life sentence with a minimum term of incarceration of nineteen years.⁵³

⁵³ *Man Jailed over Wife Fire Murder*, BBC NEWS (Dec. 11, 2009, 7:28 PM), http://news.bbc.co.uk/2/hi/uk_news/england/staffordshire/8408020.stm.