



# Towards a cognitive forensic stylistics: An intercoder reliability test for replicable feature finding in the Operation Heron corpus

Language and Literature  
2025, Vol. 0(0) 1–22  
© The Author(s) 2025



Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/09639470251337632  
[journals.sagepub.com/home/lal](https://journals.sagepub.com/home/lal)



**Matthew Voice** 

University of Warwick, UK

**Chloe Harrison** , **Tim Grant** and **Marcello Giovanelli** 

Aston University, UK

## Abstract

This paper reports an initial application of contemporary cognitive stylistics to forensic linguistic contexts. In both areas, a need has been identified for robust analyses. An intercoder reliability study was developed using data from a historic authorship analysis case involving single-authored hate mail. Exploring the applicability of Cognitive Grammar's notion of construal as a reliable framework for describing salient features of the author's style, this test examined the accuracy and consistency of descriptions of schematicity and specificity within the corpus, as applied by independent coders. Iterative coding and testing demonstrated that reliability was achievable, but depended upon a protocol developed through considerable definitional work, refining the concepts of specificity and elaboration as taken from Cognitive Grammar. Our findings support the idea that the identification of stylistic features can be rigorous, retrievable, and replicable, but also that a fuller system of coding will require a substantial research programme. Such an approach, bringing together contemporary stylistics and forensic authorship analysis, would be a productive collaboration between both disciplines and a valuable research method for verifiability in stylistics more generally. **Content:** Readers are advised that the letters analysed for this study contain offensive language, and that short quotes within this paper include racist and hateful language directed at particular groups.

---

## Corresponding author:

Matthew Voice, Department of Applied Linguistics, University of Warwick, Social Sciences Building, Coventry CV4 7AL, Coventry, UK.

Email: [matthew.voice@warwick.ac.uk](mailto:matthew.voice@warwick.ac.uk)

## Keywords

Cognitive stylistics, forensic linguistics, cognitive grammar, intercoder reliability, authorship analysis, construal, qualitative coding

## 1. Introduction: The context of forensic authorship

This paper seeks to bring together forensic linguistics and contemporary literary stylistics to develop methods for assessing analytical reliability applicable to both fields. Reports of authorship analyses being used in forensic casework are increasing in both their quantity and their spread across jurisdictions. [Grant \(2022\)](#) discusses the breadth of activity which includes involvement in civil and criminal cases across a variety of legal questions including accusations of murder and sexual assault, through to civil copyright hearings and critiques of legal procedure in Courts of Arbitration. Within these diverse applications, forensic authorship analyses include (i) comparative authorship analyses, which tend to be those taken as evidence to Court; (ii) profiling an author's sociolinguistic background or history; and (iii) searching for an author across a large dataset of potential authors. Such analyses are also diverse in terms of genre and register with reports of analysis of suicide notes, terrorist conspiracy documents, threatening communications and short SMS texts or communications from messaging apps ([Grant, 2013](#); [Grant and McLeod, 2020](#)).

Forensic authorship analysis has been described as depending on two main approaches: stylometric and stylistic ([Ainsworth and Juola, 2018](#); [Grant, 2008, 2022](#); [Solan, 2013](#)). Stylometric methods are often heavily computational and do not have to be rooted in linguistic theory or linguistically informed features ([Argamon, 2018](#)). Indeed, one of the most powerful approaches in stylometric analysis has been character n-gram analysis. N-grams are low-level strings of characters including spaces and punctuation marks, and are often subject to statistical analyses or machine learning approaches (e.g., [Grieve et al., 2019](#)), which can depend on non-word character strings. Critiques of these approaches include that they are typically seen as resistant to explanation and are widely considered theory-free analyses ([Wright, 2014](#)). Other features commonly explored in stylometry include word and sentence length measures and distributions of function words, and whilst such features might be linked to ideas of individual textual production in terms of linguistic competence and complexity of grammatical structures, such links are backgrounded in favour of demonstrated utility ([Holmes, 1998](#)). For stylometry, the particular features used are less important than a high rate of successful attribution using a particular combination of feature and algorithmic method. Perhaps because of this lack of explanation, such methods are generally not admitted as evidence in many jurisdictions (including the UK). These methods may be used, however, in other non-evidential contexts, such as investigative intelligence gathering.

In contrast to stylometry, where studied features are defined in advance of any specific text, stylistic approaches in a forensic linguistic context can be characterised as involving manual analysis of features arising out of a text, and covering a wide range of lexical, grammatical, and discursive choices. Stylistic approaches depend on within-case validation (rather than proving that features and methods will work universally) and a number

of protocols have been published for feature elicitation (Grant, 2013, 2022), which can be used in qualitative or statistical decision making (e.g., Grant and Grieve, 2022). As is further described below, the essential critique of such analyses is that they are too dependent on the specific linguist involved in the analysis. That is to say, forensic stylistic authorship analyses are seen by some as being subjective, intuitive, and non-replicable and thus unreliable (Howard, 2008).

### 1.1 Critiques of stylistic authorship analysis

Stylistic authorship analyses have attracted a long history of criticism in forensic linguistic literature. Finegan (1990) describes a case, *Brisco v. VFE Corp.*, initially carried out by McMenemy in which he and four other linguists provided expert testimony. McMenemy (with another linguist supporting him) is described as carrying out a stylistic analysis, and Finegan ‘judged the methodology employed by the first two experts to be faulty’ (336). He critiques the broad analytic method that looked first for features in the disputed material and then in the known, and also critiqued the features elicited as being ‘far from being idiosyncratic or personally stylistic, [rather, they] are functional or situational features typical of other situationally similar samples’ (336). In other words, his criticism was that these features may have reflected register or genre differences rather than authorial style (cf. Chaski, 2001).

Howard (2008) examines two American case examples (both involving the same analyst) and describes two ‘forensic stylistic’ analyses. Although he is careful not to generalise to all stylistic methods, he does observe that ‘[t]here is a general trend in the Forensic Stylistic tradition to favor prescriptive variables such as punctuation, spelling, abbreviations, and other non-descriptive grammatical conventions’ (232), later saying that ‘those techniques that rely on these variables [...] tend to have lower reliability’ (234). He summarises these as ‘inherently unscientific in subjective/intuition-based analyses’ (239), while further suggesting that such an approach ‘conflicts with notions of idiolect in theoretical linguistics’ (237).

Although Howard’s critique is tightly focused on two American cases, his critiques are supported and generalised by others. Chaski (2013), for instance, comments that ‘[o]n the one hand, it is essentially impossible to replicate a forensic stylistics analysis, while on the other hand, it is always possible to find an alternative analysis and opposing conclusion’ (363). Based on her own studies, Chaski concludes that ‘[f]orensic stylistics has very poor accuracy on ground truth data’ (365). Such critiques of stylistic approaches can have a quality of a strawman attack in their characterisation of forensic stylistic analyses, but issues around the subjectivity and replicability of analyses bear further examination. Published alongside Chaski (2013) and Solan (2013) remarks that, ‘it will be incumbent upon those whose work is more intuitively stylistic to demonstrate its scientific underpinnings’ (574).

One response to these critiques is provided in Grant (2013), which sets out a protocol and method for a stylistic analysis aimed at addressing some of these issues. In the forensic linguistic literature, however, these debates have set up an unhelpful opposition between the stylistic and stylometric approaches, where the former can be characterised as subjective, intuitive and unreliable in contrast to the latter. However, none of these

discussions have taken account of more recent developments in stylistic theory and method.

## 1.2 Objectivity in stylistics

Criticisms of subjectivity stand against the foundational ethos of stylistics as a discipline, since traditional stylistics has always emphasised and defended its methodological rigour and anti-impressionist approach. Indeed, this foundation is considered to be what sets stylistics apart from other literary studies (see [Carter and Stockwell, 2008](#): 291–302), with three guiding principles ([Simpson, 2014](#): 4) for research:

- Stylistic analysis should be rigorous
- Stylistic analysis should be retrievable
- Stylistic analysis should be replicable

These three prevailing ‘R’s, Simpson argues, *should* predicate any effective stylistic analysis. For an analysis to be rigorous, it should avoid impressionism and be ‘underpinned by structured models of language and discourse that explain how we process and understand various patterns in language’ ([2014](#): 4). Stylistic analysis should also be retrievable; that is, produced via a set of agreed upon terminology, concepts and processes. Finally, stylistic analysis should be replicable by other researchers. Another scholar should be able to follow the same steps to access the same data and observe the same linguistic patterns. Taken together, and when properly followed, these principles are constitutive of stylistics best practice.

In response to the criticisms raised in the forensic linguistic literature, this exploratory study investigates the extent to which these principles hold in practical, forensic application, in reference to contemporary stylistics scholarship. The analytical precision advocated for in recent approaches to language and style might contribute to the view of contemporary stylistics as a nuanced and practical framework for forensic linguistic investigation ([Canning, 2023](#)).

## 2. Data description

As the context for this paper is the application to forensic casework problems, real forensic texts were used, drawing from an abusive letter series known as Operation Heron. A stylometric and corpus analysis of the Operation Heron data set can be found in [Busso et al. \(2022\)](#) and the data are held in the FoLD repository, the Aston University Forensic Linguistic Databank (<https://fold.aston.ac.uk/>).

The Operation Heron data set comprises approximately 60 racially and sexually abusive letters sent between 2007 and 2009 with some of the later letters in the series sent to Gordon Brown, then Prime Minister. This attracted greater police effort, including a *BBC Crimewatch* appeal. The letters contain cartoon images as well as text, and [Figure 1](#) provides an example letter released as part of the appeal.



**Figure 1.** Example letter from the Heron series.

## 2.1 Data preparation

To create a manageably-sized representative dataset, 21 letters from the Operation Heron corpus were randomly selected and prepared for initial analysis. This involved removal of images and provision of a plain text transcription. As can be seen in [Figure 1](#), such transcription removes multimodal information beyond the exclusion of drawings.

## 2.2 Distinctiveness and initial feature-finding

This study involved a multi-stage analysis, beginning with a bottom-up feature-finding exercise which sought to describe the stylistic distinctiveness of the author's language. With an interest in exploring the applicability of Cognitive Grammar ([Langacker, 1987, 1991, 2008](#)) as a descriptive tool in forensic contexts, this was approached via an emphasis on construal. This refers to the ways in which language choices frame the

representation of events and actors within discourse, with an emphasis on specificity (high or low detail descriptions), focusing (foregrounding and backgrounding), conceptual prominence, and the perspective within the viewing arrangement constructed by tense, person, and similar linguistic features (cf. [Langacker, 2008: 55–90](#)). The analysis presented below explores how each of these dimensions may be used to help characterise the Operation Heron author's style.

Overall, the discursive structure of these letters is characterised by their inconsistency, in distinct contrast to [Langacker's \(2008\)](#) definition of 'discourse' as a series of forms which 'develops and builds on the previous one, so as the discourse proceeds an integrated conceptual structure of progressively greater complexity is being constructed' (486). The example below gives a typical example, showing how the letters consist of a series of fragmented statements, often using non-standard grammatical structures:

'Pay for your own vile wogs in your towns, you will be swarming with the flies soon, like England is now, the English demand vote S.N.P kick your evil anti-English coward racist unelected lying patsy Muslim lover boy moral compass shit Gordon Muslim Brown far north of our English border. S.N.P democracy at last S.N.P. Own own parliament. Repatriation exit from Europe' (Letter 13)

As well as demonstrating high frequencies of topic shifting, many of the phrases used in this example are repeated elsewhere in other letters. It is unusual to find full clauses which only appear in a single letter: much of the content is repeated across multiple texts, though exact structures vary between instances.

Analysis of prominence and perspective within these letters gives indications of the social position from which the author writes, and therefore indications of their identity. The nouns 'hospitals' and 'English elderly' are consistently preceded (94% and 100% frequency respectively) by the first person plural possessive 'our'. While this determiner profiles a relationship between the author and the subject being described, the precise nature of this relationship is unclear. In general, the author's identity is backgrounded against the more prominent profile of subject and the relationship itself.

This conceptualisation poses potential problems for author identification. Rather than construing themselves as a member of any particular age or social group (besides 'English'), group identities associated with the author are represented in relation to the author. For example, it does not necessarily follow from this language that the author is female or elderly. There is the single use in Letter 1 of 'us English women', where the possessive determiner construes the author as a member of the group, but this isolated example is insufficient on its own to confirm identity categorisation definitively. Nonetheless, the fact that similar constructions are produced across the texts offers a possible basis for linguistic comparison with further material.

As with phrase-level construction, the author is consistently rendered absent through clause-level language choices. The viewpoint arrangement shows how the use of vague group identities ('killing us') and third person viewpoints ('the English') make the speaker's identity ambiguous. Given the consistency with which the direct association of the author with a group more specific than 'us English' is avoided across the letters, such ambiguity could itself be considered a hallmark of the writing style.

Regular use of imperatives, for example ‘Vote S.N.P’, further foregrounds the addressee rather than the author. ‘Vote’ is almost exclusively imperative, and is the second most frequent word in the corpus (after ‘English’), representing 2.94% of the word count.

‘Clean up our filthy English hospitals’ (Letters 2 and 3)

‘Make our hospitals fit to go in again’ (Letter 2)

‘Kick the filthy foreigners out of our ancestral English tax payers hospitals’ (Letter 5)

While some imperatives are preceded by speech acts (e.g., ‘The English demand...’), most are not. In terms of prominence and focus, the instructor is consistently backgrounded in order to profile the actions being commanded. These imperatives generally instruct the reader to perform specific actions in the world, presumably after reading the letter. However, there are also imperatives which instruct the reader to engage with the discourse:

‘Look at the ruins of England now’ (Letter 3)

‘Open the ward doors look inside’ (Letters 3, 4, 5, 6, 19, and 22)

These imperative forms may also be seen as characteristic of the author’s style. Not only do they appear across the discourse, but their backgrounding of the author mirrors the conceptual structure of other features already observed at the phrasal level.

Other patterns in the viewing arrangement of the Operation Heron letters often lead to the conceptualisation of the author as a victim: perspective is shared with the patients of actions undertaken or caused by immigrants and other aggressors. For example:

‘Protect our spiritual meadows and woodlands from the foreign scum hell bent to destroy our English culture’ (Letter 3)

‘The wog is back in our system again milking our tax and killing us’ (Letter 3)

‘85% living off English tax bleeding us dry’ (Letter 13)

Frequently positioned as grammatical object, the author is construed, alongside the broader group identities to which they belong (‘us’, ‘our English culture’) as being acted upon, diminishing their own agency while highlighting negatively evaluated groups as deliberate actors.

Finally, variation in patterns of specificity correlate to the ideological assessment of the group or individuals being described. Negatively evaluated groups are construed nominally in terms of their ethnicity, nationality, or religion, with frequent adjectival pre-modification. For example:

‘Scrounging Pakis, grabbing Jews, demanding chinkies. Complaining Indians. Muslim scum. Gun-running niggers are in our English tax ancestral hospital beds’ (Letter 6)

‘You voted the foreign leeching-posturing scrounging Asian Black Muslim and minorities into our cities’ (Letters 13, 14, and 15)

The author consistently selects deverbal adjectives as pre-modifiers, which suggest force or motion enacted by the described group, and aligning with their frequent positioning as grammatical agents, foregrounding their perceived behaviour. The first example above shows a typical one-to-one correspondence between adjectives and nouns in these descriptions, though the second example demonstrates how these descriptions are also presented by the author such that any one adjective could be attributed to all groups.

A similar but distinct pattern occurs in descriptions of political figures negatively evaluated by the author, most frequently Gordon Brown. Rather than remaining constant across the letters, the list of adjectives and postmodifiers which describe Brown evolve across the series. For example:

‘Brown paranoid schizophrenic racist Scot’ (Letter 3)

‘Gordon Brown paranoid schizophrenic anti-English racist Scot’ (Letter 4)

‘Gordon Brown paranoid schizophrenic racist ant-English parasite Scot’ (Letter 5)

‘The evil lying cowardly demented anti-English racist Muslim patsy Gordon Mohammed Brown’ (Letter 12)

‘Evil anti-English coward racist unelected lying patsy Muslim lover boy moral compass shit Gordon Muslim Brown’ (Letter 13)

The same adjectives appear repeatedly, although their ordering and inclusion in any one iteration is variable. Other negatively evaluated political figures are similarly construed with high-frequency, repetitive modification:

‘Your dirty freeloading “Martin” crooks’ (Letters 12, 14, 15)

‘Jack Strow [sic] anti-English racist Jew import parasite’ (Letter 5)

‘Jack Straw Jew-import extreme-left parasite’ (Letter 6)

By contrast, individuals and groups with whom the author is more ideologically aligned receive substantially less specific and elaborated description. Of 53 references to ‘[us/the] English’, only four are further specified, and only then through the inclusion of a percentage, for example ‘the English (95% of us)’ (Letter 14). The only political figure referenced in an unequivocally positive light is construed as:

‘Mr Heinz Christian Strache in Austria’ (Letter 15)

Although the author disagrees with David Cameron politically, it is notable that he is not given the same highly-specified description as Gordon Brown. Five of 6 references simply describe him as ‘Cameron’ (Letters 9, 14, 15, 16, and 20), while the last specifies:



‘Is Mr Cameron (Scot) not what you want’ (Letter 12)

In other words, there is a clear contrast in the specificity and elaboration of the author’s language when describing left- and right-wing political figures, indicative of ideological and linguistic trends in the language of the letters. While all dimensions of construal can be used to identify and describe the stylistic features characteristic of the Operation Heron author’s style, the degree of elaboration through modification is highly distinctive, and became the focus of further analysis as detailed below.

2.3 Classifying specificity

Having identified these features as stylistically salient within the corpus, the question then becomes whether or not Cognitive Grammar’s system of describing specificity and elaboration within construal can meet the criteria for rigour, retrievability, and reliability. The discussion below outlines the potential challenges and ambiguities in developing existing theory for this purpose, evidencing the need for a protocol through which to classify examples more precisely, both in forensic and more general stylistic applications.

Specificity is often described within Cognitive Grammar as analogous to visual resolution: ‘A highly specific expression describes a situation in fine-grained detail, with high resolution. With expressions of lesser specificity, we are limited to coarse-grained descriptions whose low resolution reveals only gross features and global organization’ (Langacker 2008: 55). Greater or lesser degrees of specificity can be conferred by a range of linguistic features. Example 1a below (Langacker 2008: 56) shows how pre- and post-modification of a head noun affects the level of detail in the construal of the subject. 1b, on conversely, demonstrates how choices between different lexical options (e.g., choosing the adverb ‘about’ or ‘exactly’) can further influence the precision of a given construal.

- (1)
- (a)

rodent -> rat -> large brown rat -> large brown rat with halitosis
- (b)

hot -> in the 90s -> about 95° -> exactly 95.2°

In any given noun phrase, the head noun and the selection (or absence) of pre- and post-modifiers work together to affect the specificity of a given phrase. In this paper, ‘head

Table 1. Examples of low, mid, and high specificity in head nouns.

	Location	Object	Group	Individual
Low specificity	‘Somewhere’	‘Something’	‘People’	‘Someone’
	‘Place’	‘Item’	‘Crowd’	‘Person’
	‘Outside’	‘Device’	‘Foreigners’	‘Adult’
‘Mid’ specificity	‘City’	‘Clothing’	‘Citizens’	‘Girl’
	‘University’	‘Footwear’	‘Fans’	‘Student’
	‘Building’	‘Trainers’	‘Professors’	‘Researcher’
High specificity	‘London’	‘Nikes’	‘Parliament’	‘Gordon’
	‘Royal Albert Hall’		‘Alex, Jamie, and Charlie’	‘Gordon Brown’

noun specificity' refers to the specificity of a head noun considered in isolation (see Table 1 for examples), while 'phrasal specificity' describes the overall specificity of a noun phrase. Likewise, 'elaboration' is used in this paper to describe the additive effect of pre- and post-modification, which contributes to phrasal specificity.

Though this accounts for the ways in which language might affect specificity and elaboration, the precise classification of different degrees of specificity is problematic in its potential subjectivity. Throughout his work, Langacker describes specificity as a 'scale' (2017: 8) comprising different 'levels' (2009: 2) which produce 'hierarchies' (2010: 7). While the extreme ends of such a scale are relatively straightforward to identify, as in 1a and 1b above, it is somewhat unclear how individual nouns, phrases, and clauses are to be classed within the binary distinctions of 'fine-grained' versus 'coarse-grained' (2008: 55) sometimes used to describe this variation. Highly schematic nouns which connote little detail (e.g., 'someone' or 'something') are clearly low in specificity, while proper nouns (e.g., 'Gordon Brown') afford the conceptualisation of a specific and often unique referent. Unlike other dimensions of construal discussed in the analysis above, whose classification can be largely resolved via the identification of particular linguistic structures within the discourse (e.g., pronouns, imperatives, and transitivity), assessing degrees of specificity requires further decision-making, creating greater risks of inconsistency between analysts describing the same data.

### 3. Methods

Bottom-up coding of sample texts led to the identification of notable features of authorial style in relation to Cognitive Grammar's notion of construal (Langacker, 2008). In line with the view of linguistic style as emergent from a combination of foregrounding, prominence, and deviance from expectations (Leech and Short, 2007: 41), the analysis outlined in §2.2 identified the specificity of noun phrases as a highly distinctive stylistic feature within the corpus. Given this combination of salience within the corpus, and the potential for subjectivity and ambiguity in coders' classifications, specificity and elaboration were selected as the stylistic features to become the focus of an intercoder reliability test. In doing so, the study aimed to investigate the extent to which stylistic analysis informed by features from Cognitive Grammar can conform to Simpson's (2014) principles of rigour, retrievability, and reliability.

#### 3.1 Coding protocols

To produce consistent coding of the primary letters, it was important to develop clear and replicable procedures for identifying and coding specificity and elaboration within the chosen text. As Grant (2022) argues, 'the reliability of stylistic approaches to authorship analysis requires the publication of protocols, which can be critiqued and then subjected to validation studies' (54). The aim of this process was twofold, as it allowed for the study of whether (a) expert stylistic analysts can reliably categorise the degree of specificity and elaboration in the coding process, and (b) the correlation between specificity, elaboration, and the letter writer's personal evaluation remains consistent within new texts, as was identified in the initial bottom-up analysis. The verbatim publication of these protocols therefore seeks to ensure

transparency in the reporting of our method, as well as demonstrating the continued need for refinement in cognitive stylistic descriptions of conceptual features of discourse.

In order to explore and overcome the difficulties of arranging a scalar phenomenon into two categories, we instead determined ‘high’, ‘mid’, and ‘low’ codes through which to organise judgements of head noun specificity, phrasal specificity, and elaboration. The use of a ‘mid’ code sought to reflect that understanding the specificity of reference for many head nouns relies on context and interaction with other elements of the noun phrase. For example, relative to its low specificity alternatives (e.g., ‘place’), ‘building’ describes a more bounded immediate scope with clearly definable characteristics, without referring to a single unique institution (e.g., ‘Royal Albert Hall’). The examples in [Table 1](#) demonstrate the application of these descriptors in isolation from any pre- or post-modification which might affect the specificity of the construal.

To avoid the intermediary ‘mid’ specificity code becoming a catch-all for every construal between total schematisation in the low specificity code and unique reference in high specificity, protocols were established to agree criteria for coding, as well as boundaries between each code. For instance, nouns which connote singular qualities about their referents were defined as belonging within the ‘low’ category. While ‘adult’ is more specific than ‘person’ it remains in the low specificity bracket, as it only connotes age. Comparatively, ‘girl’ appears towards what might be thought of as the lower end of the ‘mid’ specificity scale, as it prototypically connotes both age and gender, and therefore provides more information about the referent in comparison to ‘adult’.

When these head nouns appear within noun phrases, the determiner which accompanies the head noun will significantly affect the perceived specificity, hence our distinction between phrasal specificity and the specificity connoted by the head noun alone. For instance, ‘the building’ connotes the conceptualisation of a particular university, and bounds the scope to an actual location within the world, while ‘a building’ describes an abstracted concept more generally.

While determiner choice is certainly a contributing factor to the overall specificity of a noun phrase, the examples of highly specific sentence structures from [Langacker \(2008: 56\)](#) include indefinite determiners: ‘An alert little girl wearing glasses caught a brief glimpse of a ferocious porcupine with sharp quills’. Here, two ‘mid’ specificity head nouns, which in different contexts could make either a specific or schematic reference, are highly specific in their construal, given the detail afforded by the elaborative adjectives, prepositional phrases, and the non-finite participle clause which provide additional detail. The categorisation of noun phrase specificity, then, is not simply the quantification of linguistic features and their presence or absence, but a subjective appraisal of their effect on conceptualisation, hence the need to test the consistency of such assessments across coders.

The following protocols were therefore designed to guide the coding process for this study:

- (1) Assessing Head Noun Specificity
  - a. Identify the head noun within the phrase:
    - i. Is there a discrete referent (high) or is the noun more schematic (low and mid)?

- ii. If schematic, how many characteristics does the noun connote (null or one = low specificity, >2 = mid specificity)?
  - iii. What is the relationship between the maximal and immediate scope?
  - iv. What other nouns were available in order to describe the subject, and would these have been more or less specific?
- (2) Assessing Phrasal Elaboration
  - a. Identify pre- and post-modifiers within the phrase:
    - i. What is their frequency? How many occur?
    - ii. As a base rule: 0 = low, 1–2 = mid, 3+ = high
    - iii. However, what is their relative degree of elaboration? What other modifiers could have been used, and would these have been more or less precise?
- (3) Assessing Phrasal Specificity
  - a. Consider the relationship between head noun and any modifiers
    - i. Do these affect the scope of the head noun's reference?
    - ii. If so, mid-specificity head nouns likely lead to high phrasal specificity
    - iii. Low specificity head nouns may require additional context for phrasal code
  - b. Additional contextual factors (especially where head noun has mid specificity)
    - i. What choice of determiner appears in the noun phrase?
    - ii. Does the determiner or wider context make the noun schematic or specific (i.e., an abstracted or discrete referent)?
  - c. Where possible, resolve to a 'high' or 'low' coding. See Worked Example (below) for an instance where such a code might not be possible.

As part of the process of becoming familiar with the coding exercise, worked examples were produced in order to demonstrate the application of each code to a different noun phrase. One example, referenced in the Phrasal Specificity protocol, is replicated below:

- (a) Noun Phrase = 'Broken plates'
  - (1) The head noun 'plates' conforms to the definition of **mid-specificity** above – as connoting some qualities of the referent without clear indication of whether or not the subject is specific or schematic.
  - (2) This single adjective affects the scope of reference in relation to the head noun significantly, and is therefore **highly elaborative**
  - (3) However, judgement of the specificity of the phrase as a whole requires **further context** – without which it remains unclear whether the phrase is schematic (i.e., the abstract concepts of broken plates *in general*), or specific (making reference to a collection of actual broken plates within the world)

As shown here, coding involved a combination of quantitative and qualitative feature finding procedures. While elaboration largely focused on the frequency of modifiers, noun-level and phrasal specificity codes required coders to assess the conceptualisation afforded in each instance, and compare this to the range of alternative construals and contextual factors which form the basis of a stylistic approach to textual meaning.

### 3.2 Intercoder reliability test

Having developed clear protocols for the coding of specificity and elaboration within the data, we conducted an intercoder reliability test in order to investigate how consistently two stylisticians could apply the codes independently of one another.

A letter from the Operation Heron corpus was chosen at random as the subject of the test. The text was divided up into clauses and entered into cells in an Excel spreadsheet. Where one phrase was embedded inside another, the embedded clause was presented both as part of the larger phrase, and independently below. For example, the noun phrase and embedded relative clause below were presented as:

- (a) our English tax which is for English doctors
- (b) [English doctors]

Alongside each entry, the researchers were tasked with providing a code for each of the three features outlined in the protocol above: head noun specificity, phrasal elaboration, and phrasal specificity. Entries consisted of four options: 'high', 'mid', and 'low' each correspond to the conceptual categories defined above, while 'uncoded' denoted an entry for which no assessment was applicable.

The letter was divided into 110 clauses for coding, requiring 330 codes from each coder overall. After completing the task independently of one another, both researchers sent their codes to a third member of the project, who compared the consistency of their entries independently using Cohen's kappa to assess agreement beyond chance. Following this statistical comparison, the researchers met to discuss any points of disagreement in coding. The results of this discussion were coded thematically to better understand and further develop the protocols and tasks developed during this study, and are also reported in the results section below.

## 4. Results

Results of the coding test were divided into four types of response. Firstly, if both coders provided a 'high', 'mid', or 'low' code within the same cell on their corresponding spreadsheets, this was checked for consistency, and marked as identical or in disagreement accordingly. Secondly, it was important to know whether the coders had agreed that a cell required a code for nominal specificity and elaboration at all. A category of 'agreed uncoded' cells therefore marked instances in the task where coders had identified that there were no instances of noun phrase specificity to code, for example for the imperative 'go'. Where there was disagreement at this level of the exercise, cells where one coder had overlooked a potential coding were marked as 'missing'. Given the complex mix of coding types, intercoder reliability was measured as simple percentage agreement.

Following this procedure, the intercoder reliability test elicited a generally consistent response between the two coders of 82.1% complete agreement. [Tables 2–4](#) below provides a breakdown of these results across each dimension of specificity assessed within the task:

**Table 2.** Confusion matrix of coder responses to head noun specificity classification task.

	Low	Mid	High	Uncoded	Sum
Low	11	0	1	1	13
Mid	4	35	17	1	57
High	0	1	4	0	5
Uncoded	0	0	2	33	35
Sum	15	36	24	35	110
Relative agreement	73.3%	97.2%	16.6%	94.2%	75.4% ( $k = 0.65$ )

**Table 3.** Confusion matrix of coder responses to elaboration classification task.

	Low	Mid	High	Uncoded	Sum
Low	57	1	0	2	60
Mid	4	5	1	1	11
High	2	0	4	0	6
Uncoded	1	0	0	32	33
Sum	64	6	5	35	110
Relative agreement	89%	83.3%	80%	91.4%	89% ( $k = 0.81$ )

**Table 4.** Confusion matrix of coder responses to phrasal specificity classification task.

	Low	Mid	High	Uncoded	Sum
Low	5	5	3	1	14
Mid	1	7	4	0	12
High	0	1	47	1	49
Uncoded	0	0	2	33	35
Sum	6	13	56	35	110
Relative agreement	83.3%	53.8%	83.9%	94.2%	83.6% ( $k = 0.75$ )

Calculating intercoder reliability on data of this sort is complex, as the coding scheme of ‘high’, ‘mid’ and ‘low’ is essentially ordinal; disagreement between ‘high’ and ‘mid’ codes or ‘mid’ and ‘low’ codes is less than that between ‘high’ and ‘low’ judgements. A review of the literature on reliability measures for this kind of coding confirmed that there is no preferred or standard measure to use. Percentage agreement does not take into account chance agreement (and so may inflate perceived reliability) and Cohen’s kappa accounts for chance, but presumes equal likelihood of confusion between any two categories of judgement, which is not the case here. As the most likely confusion will be between adjacent categories on the ordinal scale, and that kappa treats these as equal degrees of disagreement as that between ‘high’ and ‘low’ judgements, this means that the use of kappa will theoretically at least, deflate the perceived reliability in this situation. Given these difficulties we choose to present both percentage agreement and Cohen’s kappa.

Acceptable intercoder reliability levels are task specific, with no uniform standards for intercoder agreement (Neuendorf, 2016), rates in the 80% range are likely to be considered acceptable across contexts, even when using percentage agreement. However, for Cohen’s kappa there are some conventional thresholds. According to Landis and Koch (1977: 163), a kappa statistic of between 0.61 and 0.80 represents ‘substantial’ agreement, while  $k > 0.81$  is ‘almost perfect’ agreement.

That the results of phrasal elaboration coding met this standard is not surprising, as the coding protocol was largely quantitative, with errors reflecting manual coder mistakes, rather than disagreement. Coding for head noun and phrasal specificity required more subjective decision-making with regard to classification, and their overall proximity to intercoder agreement for elaboration suggests substantial agreement beyond chance in each case. However, the results do reveal variation between each task, as the identification of specificity within head nouns in isolation was coded with lower, though still substantial, reliability ( $k = 0.65$ ) than phrase-level specificity ( $k = 0.75$ ) despite the comparability of the two tasks.

Following the collation of these results, the coders met to discuss their disagreements. These could be of either one place (i.e., a ‘low’ and ‘mid’ code, or a ‘mid’ and ‘high’ code), or two (a ‘low’ and ‘high’ code). As shown in Table 5, only 4 of the 47 disagreements (8.5%) were of two degrees, indicating a consistently close grouping of codes as opposed to strong disagreements in the vast majority of cases.

Of the 47 disagreements, 14 (30%) were identified as human error, where the coder was unable to account for a decision which justified the given code.. Similarly, a further 19 disagreements (40%) were resolved between the coders during the discussion through clarification of protocol procedure, indicating that refinements to coding protocols would have rectified these discrepancies. It is therefore anticipated that with further practice in comparable coding exercises, more than two thirds of the observed disagreements could be avoided. This emphasises the potential value of protocol-driven intercoder reliability testing as a means of ensuring consistency of understanding and application across stylistic analyses, both forensic and more broadly.

The remaining 14 (30%) of unresolved disagreements were identified as resulting from challenges the coders faced in the application of the ‘mid’ specificity protocol, particularly in coding head nouns in isolation. For example, the noun ‘tax’ was coded four times across the test, with one coder consistently marking this as an instance of mid-specificity, and the other coding for high specificity.

Intercoder discussion revealed differing rationales for these separate classifications. The high specificity coder suggested that a single more specific lexical item was not

**Table 5.** The degrees and kinds of disagreements from the intercoder reliability test.

	Disagreements of one degree	Disagreements of two degrees
Coder error	13	1
Adjusted after discussion	16	3
‘Mid’ protocol query	14	0
Overall	43	4

available for selection in lieu of ‘tax’ (beyond less common terminology, such as ‘tithe’), while the mid specificity coder noted that elaborative modification could be used to yield greater specificity of reference (e.g., ‘income tax’). This opened up further discussion about the intuitive understanding of specificity and the extent to which some lexical items may have a conceptual limit on their degree of specificity, such that high specificity can be achieved only through additional elaborative constructions beyond the meaning conferred through the head noun. The discussion also suggests that the final consideration of the head-noun coding protocol (*iv. What other nouns were available in order to describe the subject, and would these have been more or less specific?*) may counter the assessment of the coding criteria raised in i–iii. These considerations are explored in cognitive linguistics and cognitive stylistic more widely. Stockwell’s (2020: 20) overview of ‘basic-level categories’ of schemas that we perceive as more familiar or appropriate in particular contexts, for example, describes how certain conceptual patterns may be preferentially selected over others. While this discussion raises interesting points about potential limits on the specificity of such lexical items, a detailed exploration of this issue is beyond the scope of this paper. For future coding protocols, a more nuanced system which delineates the scale of specificity elicited through phrases and head nouns would better represent the distinct conceptual parameters up the boundary of ‘high’ specificity which has arisen from this initial investigation.

## 5. Discussion

Results of the intercoder reliability test provide empirical support to the assertion that intercoder identification of cognitive stylistic features of language can be shown to be a reliable and replicable process. While only a first step in the development of this style of testing, its approach and findings have implications for both stylistic and forensic linguistic research.

### 5.1 Replicability in stylistic methods

Departures between coder responses reveal areas of learning for the development of future forensic stylistic coding, and points of discussion which feed back into Cognitive Grammar’s model of language and meaning. For example, while the construal of specificity can clearly be described in a scalar fashion from low to high, the process of categorising individual construals along such a scale was initially problematic. The test reported in this paper has demonstrated that this process can be done consistently and accurately by independent coders, although considerable reworking of CG’s definitions of specificity and elaboration was required to define each category, across several iterations of the task. This process was invaluable not only for producing parameters for the reliability test, but also for prompting reflection on the fine-grained distinction between points on the scale of specificity, and identifying areas of disagreement or ambiguity between researchers’ understanding and application of these terms.

Despite disciplinary emphasis on replicability (Carter and Stockwell 2008; Simpson 2014), the development of *coding protocols* for stylistic analysis is a relatively scarce practice. There are a few exceptions within the context of textbooks: Simpson’s (2014)



*Stylistics: A Resource Book for Students*, for example, outlines 6 questions for carrying out a general analysis of a poem which relate to the identification of poetic features, including: general patterns of grammar, foregrounded patterns of grammar, sound and rhythm, graphology, vocabulary and word-structure and stylistic alternatives (113–114). Coding protocols from other branches of applied linguistics have also been implemented: for example, the Metaphor Identification Procedure (Pragglejaz Group, 2007) has been drawn on for stylistic application (e.g., Dorst, 2015; Nuttall and Harrison, 2020), which suggests that such frameworks are relevant and beneficial for producing analytical rigour within the context of particular studies. Jeffries and McIntyre (2010: 21) argue that:

[s]ome fields of linguistics have developed a very clear and agreed set of standard practices, based on a consensus about current theories and models; the field of conversation analysis would be one such example. This has not so far happened with stylistics, perhaps mainly because of the enormous range of practices which seem to shelter under the label.

In addition to the breadth of practices mentioned by Jeffries and McIntyre here, there are other logical reasons for the absence of coding protocols in stylistics scholarship, such as the emphasis in traditional stylistics scholarship on literary texts. Coding for interpretive implications once the stylistic features have been identified is another, and even more complex, process. Generally, stylisticians have argued that certain features and patterns can be equated with broader interpretive effects: Simpson's modal grammar argues that a preponderance of epistemic modality in an extract or text can be classed as a negatively shaded text, suggesting an uncertain or bewildered character-narrator (2004: 126–127); higher degrees of foregrounding are associated with perceptions of literariness (Miall and Kuiken, 1994). While exploring interpretive commonalities can be a helpful point of departure for some analyses, particularly when examining deviations from standardised text constructions, this is not always desirable and can constrain the more deductive consideration of contextual factors and reader- or reception-centred meaning. This may be where stylistics and forensic linguistics diverge and where the research aims of the study in question will establish different preferences.

Intercoder reliability testing therefore supports the view of stylistic analysis as reliable and replicable, whilst at the same time highlighting the potential challenges posed by subjectivity of assessing the middle-ground within scalar models of construal. The agreement on codes lends support to the idea that protocol-driven stylistic categorisations of features, arising from what might be described as 'stylistic intuition' (Leech and Short, 2007: 38), can form the basis of a consistent descriptive protocols that can be used to achieve intercoder reliability. This is a promising initial finding which itself requires replication on a broader scale, across a greater range of features and tested with a larger number of coders. In the case of Cognitive Grammar's construal model, such an exercise has allowed us to identify areas of discrepancy in coder identification of features. The development of precise protocols has provided a vehicle for discussion and reflection on the way coders define and apply aspects of the construal model in their descriptions of text. Moreover, this study's focus on the development of the reliability test itself meant that the features under investigation were pre-agreed, and it remains to be seen whether coders would agree that the same features constitute salient markers of authorial style.

## 5.2 Implications for forensic stylistic approaches to authorship analysis

Our findings both uphold and undermine the concerns about forensic stylistics approaches detailed above. On the one hand, applying a categorisation scheme from cognitive stylistics points to the arbitrary nature of simply eliciting features from a text without a framework in place. In [McMenamin's \(1993\)](#) approach and even in that described in [Grant \(2013\)](#), feature finding is a 'free form' process. As suggested above, such free-form approaches have some advantages in being able to make observations on the basis of truly idiosyncratic features that do occur in all texts, including texts that occur in forensic casework. Freeform or 'stream' forensic stylistics approaches to feature finding, however, have been critiqued as being less rigorous, less replicable and less retrievable than stylometric approaches.

On the other hand, by applying cognitive stylistic frameworks to feature finding we have shown here that some stylistic approaches can avoid these criticisms. This might provide a middle way between freeform stylistics and overly rigid stylometric approaches by way of a guided or structured approach to stylistic feature finding. By applying an approach derived from cognitive stylistics we have shown that we can reliably elicit features from a text - and we hope that such features might be indicative of *idiolectal style*, although this has not been explored explicitly in this paper. However, we have also discovered and demonstrated that achieving such an approach is onerous: it took considerable effort to derive a useable set of protocols and achieve reasonable levels of reliability for a small fragment of the cognitive stylistic system. Furthermore, the selection of specificity as our focus arose from our perception of stylistically salient features within this corpus through bottom-up observation. If a similar analysis were to be undertaken for forensic case work, a practical constraint might be the number of hours that can be given to a commissioned analysis.

One problem for authorship analysis more generally is that of language disguise and identity assumption. [Grant and McLeod \(2020\)](#) observe that low level stylometric features are the easiest to disguise and, in this area, a cognitive stylistic analysis may come into its own as such analyses of authorial style may focus on individuating markers of authorship which are less readily observed than surface-level stylometric features, and therefore less subject to imitation. [Hart \(2014\)](#) describes the selection of construal choices as 'ideologically and rhetorically charged' (123), and the effects of these lexicogrammatical choices might be used to evidence ideological patterns across the discourse. In the initial stylistic analysis above (§2.2), for example, it was noted that highly elaborative phrases were more frequently employed in this dataset to describe groups and individuals who the author evaluates negatively, while those viewed positively are construed with minimal elaboration. Employing cognitive stylistics to describe these patterns may therefore be useful to forensic linguistics as a means of identifying new markers of authorial style at the level of conceptualisation. In cases where disguise is suspected, the extra efforts of these approaches may justify the effort of carrying them out.

## 6. Conclusions

This paper reports an initial application of contemporary cognitive stylistic concepts to forensic linguistic contexts in a tentative first step towards addressing the concerns expressed by some about stylistic approaches to forensic authorship analysis. These concerns can be characterised as aligning with parallel assertions in stylistics that analyses are, could or should be rigorous, retrievable and replicable. In exploring these issues, we have yet to address whether the features studied demonstrate the individual variability required for authorship analysis, but we hope that this will be the start of a longer engagement between contemporary stylistics and forensic authorship analysis.

Primarily, the intercoder reliability test reported in this study provides a framework for demonstrating replicability of stylistic analysis, cognitively-oriented or otherwise. Though isolated to the perception of specificity, the high degree of agreement across the coding tasks demonstrates that conceptual patterns as categorised through Cognitive Grammar's notion of construal can be systematically and consistently coded by different stylisticians, working independently from the same coding protocol. That said, the challenges faced in producing such explicit protocols prompted valuable reflection on the subjectivity of stylistic analysis and description, here in relation to the potentially ambiguous middle ground in Cognitive Grammar's scales of construal features. This exercise provided valuable insight into the potential sites of disagreement between coders' understanding and application of stylistic tools within the same context, with the aim of ensuring that Cognitive Grammar-based descriptions and analyses meet stylistics' standards for rigorous, retrievable, and replicable feature-finding. Future stylistic research could benefit from the development and reporting of similar protocols, as well as testing with a larger number of coders to investigate consistency in the understanding and interpretation of other stylistic features. Further consideration of the other construal dimensions will be a logical next step in expanding this work. Likewise, future studies could address the issue of initial feature selection discussed in §5.1 by comparing the linguistic elements different researchers identify in the same text as stylistically significant.

The exploratory nature of the current study necessitated considerable intercoder discussion of both the theoretical intricacies of the Cognitive Grammar model, as well as of the processes and decision-making involved in the practical application and road-testing of the framework. Consequently, the scope of this project has been limited to an idiolectal study driven by foregrounded style choices of the dataset under consideration, and more work is necessary for to apply this approach in forensic authorship analysis. In identifying, developing, and testing the first steps of this process, we hope that the protocols outlined in this paper may be beneficial for stylisticians interested in the qualitative analysis of texts comprising distinctive patterns of schematicity or specificity. More broadly, we suggest that our approach to publishing the protocols through which intercoder reliability is tested might be adopted in future stylistic research to further advance the replicability, reliability, and retrievability of the research across the discipline.

### Declaration of conflicting interests

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was partly supported by Research England E3 funding to the Aston Institute for Forensic Linguistics.

## ORCID iDs

Matthew Voice  <https://orcid.org/0000-0001-6193-9554>

Chloe Harrison  <https://orcid.org/0000-0002-9963-3208>

Marcello Giovanelli  <https://orcid.org/0000-0001-8470-3800>

## References

- Ainsworth J and Juola P (2018) Who wrote this: modern forensic authorship analysis as a model for valid forensic science. *Washington University Law Review* 96: 1161–1189.
- Argamon S (2018) Computational forensic authorship analysis: promises and pitfalls. *Language and Law* 5(2): 7–27.
- Busso L, Petyko M, Atkins S, et al. (2022) Operation Heron: latent topic changes in an abusive letter series. *Corpora* 17(2): 225–258.
- Canning P (2023) Forensic stylistics. In: Burke M (ed) *The Routledge Handbook of Stylistics*. 2nd edition. London: Routledge, pp. 521–541.
- Carter R and Stockwell P (2008) *The Language and Literature Reader*. London: Routledge.
- Chaski CE (2001) Empirical evaluations of language-based author identification techniques. *International Journal of Speech Language and the Law* 8(1): 1–65.
- Chaski CE (2013) Best practices and admissibility of forensic author identification. *Journal of Law and Policy* 21(2): 333–376.
- Dorst AG (2015) More or different metaphors in fiction? A quantitative cross-register comparison. *Language and Literature: International Journal of Stylistics* 24(1): 3–22.
- Finegan E (1990) Variation in linguists' analyses of author identification. *American Speech* 65(4): 334–340.
- Grant T (2008) Approaching questions in forensic authorship analysis. In: Gibbons J and Turell MT (eds) *Dimensions of Forensic Linguistics*. Amsterdam: John Benjamins, pp. 215–229.
- Grant T (2013) Txt 4n6: describing and measuring consistency and distinctiveness in the analysis of SMS text messages. *Journal of Law and Policy* 22(2): 467–494.
- Grant T (2022) *The Idea of Progress in Forensic Authorship Analysis*. Cambridge: Cambridge University Press.
- Grant T and Grieve J (2022) The Starbuck case: methods for addressing confirmation bias in forensic authorship analysis. In: Picornell I, Perkins R and Coulthard M (eds) *Methodologies and Challenges in Forensic Linguistic Casework*. New Jersey: Wiley Blackwell, pp. 13–28.
- Grant T and McLeod N (2020) *Language and Online Identities: The Undercover Policing of Internet Sexual Crime*. Cambridge: Cambridge University Press.
- Grieve J, Clarke I, Chiang E, et al. (2019) Attributing the Bixby Letter using n-gram tracing. *Digital Scholarship in the Humanities* 34(3): 493–512.
- Hart C (2014) *Discourse, Grammar and Ideology: Functional and Cognitive Perspectives*. London: Bloomsbury.
- Holmes D (1998) The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing* 13(3): 111–117.

- Howard BS (2008) Authorship attribution under the rules of evidence: empirical approaches – a layperson’s legal system. *International Journal of Speech Language and the Law* 15(2): 219.
- Jeffries L and McIntyre D (2010) *Stylistics*. Cambridge: Cambridge University Press.
- Landis JR and Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1): 159–174.
- Langacker RW (1987) *Foundations of Cognitive Grammar, Volume 1, Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker RW (1991) *Foundations of Cognitive Grammar, Volume 2, Descriptive Application*. Stanford: Stanford University Press.
- Langacker RW (2008) *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Langacker RW (2009) *Investigations in Cognitive Grammar*. Berlin: Mouton de Gruyter.
- Langacker RW (2010) *Concept, Image, and Symbol: The Cognitive Basis for Grammar*. 2nd edition. Berlin: Mouton de Gruyter.
- Langacker RW (2017) *Ten Lectures on the Elaboration of Cognitive Grammar*. Leiden: Brill.
- Leech G and Short M (2007) *Style in Fiction: An Introduction to English Fictional Prose*. 2nd edition. London: Routledge.
- McMenamin GR (1993) *Forensic Stylistics*. Amsterdam: Elsevier.
- Miall D and Kuiken D (1994) Foregrounding, defamiliarization, and affect: response to literary stories. *Poetics* 22(5): 389–407.
- Neuendorf K (2016) *The Content Analysis Guidebook*. London: Sage Publications.
- Nuttall L and Harrison C (2020) Wolfing down the Twilight series: metaphors for reading in online reviews. In: Ringrow H and Pihlaja S (eds) *Contemporary Media Stylistics*. London: Bloomsbury, pp. 35–59.
- Pragglejaz Group (2007) MIP: a method for identifying metaphorically used words in discourse. *Metaphor and Symbol* 22(1): 1–39.
- Simpson P (2014) *Stylistics: A Resource Book for Students*. London: Routledge.
- Solan LM (2013) Intuition versus algorithm: the case of forensic authorship attribution. *Brooklyn Journal of Law and Policy* 21: 551–576.
- Stockwell P (2020) *Cognitive Poetics: An Introduction*. 2nd edition. London: Routledge.
- Wright D (2014) *Stylistics versus Statistics: A Corpus Linguistic Approach to Combining Techniques in Forensic Authorship Analysis Using Enron Emails*. PhD Thesis, University of Leeds, UK.

## Author biographies

**Matthew Voice** is an Assistant Professor in the Department of Applied Linguistics at the University of Warwick, UK. His research explores critical and literary applications of cognitive linguistics, and has appeared in *Language and Literature*, *Discourse Studies*, and the *Journal of War and Culture Studies* on topics ranging from pop song lyrics and multiverse fiction to the discourses of drone warfare. Alongside Naomi Adam, Jessica Norledge, and Peter Stockwell, he is co-editor of a forthcoming book on literary immersion.

**Chloe Harrison** is Senior Lecturer in English Language and Literature at Aston University, UK. Her research interests include cognitive stylistics, re-reading, and reader engagement with contemporary fiction. Her publications in these areas include *Cognitive Grammar in Contemporary Fiction* (Benjamins, 2017), *Cognitive Grammar in Stylistics: A Practical Guide* (with Marcello Giovanelli, Bloomsbury, 2018) and *New Directions in*

Cognitive Grammar and Style (edited with Louise Nuttall and Marcello Giovanelli, Bloomsbury, 2021).

**Tim Grant** is Professor of Forensic Linguistics and Director of the Aston Institute for Forensic Linguistics at Aston University, UK. His research interests are in forensic authorship analysis, most recently using linguistics to assist in investigations into the abuse and exploitation of children and adults that occurs on the dark web. He has extensive experience providing expert investigative assistance to UK and overseas agencies, and providing evidence in Court in criminal and civil cases.

**Marcello Giovanelli** is Reader in Literary Linguistics at Aston University, UK. He has research interests in cognitive stylistics, the language of poetry, and in English education. He is the author of ten books and over forty journal articles and book chapters in literary and applied linguistics. Two of his monographs, *Text World Theory* and *Keats' Poetry* (Bloomsbury, 2013) and *The Language of Siegfried Sassoon* (Palgrave, 2022) draw on Cognitive Grammar for stylistic analysis.