

Hearing voices in the poetry of Brendan Kennelly¹

Carl Vogel* and Sandrine Brisset **

*Trinity College Dublin

**St Patrick's College and Université Sorbonne Nouvelle – Paris 3

Abstract

Text classification methods from corpus linguistics are applied within the poetry of Brendan Kennelly in order to determine how they cluster into categories defined by dominant voices within the poems. Using within-text objective analyses, certain strong patterns emerge: the voices of Ozzie, the Author of the Letters, the Women, etc., prove to be distinct centers of clusters.

I Introduction

A poem may be considered oral by virtue of its semblance in presentation to that within the oral formulaic tradition (Beye 1972), to that of a songster's ballads, to that common to dialogues, to that known from cadences of the poet's speaking voice, and undoubtedly in many other ways. Objective metrics can be appealed to for assessment of orality in textual artifacts (Biber 1988, Sampson 2001, Biber, Conrad and Reppen 1998). Yet a question emerges when one writer acts as a conduit of all of the oral messages contained in a text. The question is: are the voices truly distinct or has the author placed an indelible imprint on them making all her own, and can they be objectively (non-trivially) identified as such. Objectively, they are all her own in a trivial sense if they are known to have all been written by her, but in a non-trivial sense, they are objectively her own if they are independently identifiable as of a single source on the basis of textual artifacts alone. In analogy, the contributions of characters in Yeats' play, *The Land of Heart's Desire*, can more reliably be identified as belonging to that play than the voices of the characters of Eugene O'Neill's *Hairy Ape* (Mencke 2004). The question obtains relevance in terms of psychological archetypes when one chooses to over-interpret the author's words in an introduction as "listening to voices". However, even in split personality disorder diagnosed according to the DSM-IV manual,² the affected individual has just one brain and one tongue: all the voices are channelled through the same composite conduit. In an altogether different domain, ballistic

¹ This paper benefited from presentation of an earlier draft at *Varieties of Voice*, the third international conference of the Belgian Association of Anglicists in Higher Education (BAAHE), at the University of Leuven, December 7-9, 2006. We are grateful for the constructive feedback which emerged from the review process. This research is supported by a grant from Science foundation Ireland (RFP 05/RF/CMS002). Poetry quoted is cited with permission of Brendan Kennelly.

² Diagnostic and Statistical Manual of Mental Disorders (fourth edition). Washington: American Psychiatric Association, 1988.

experts in forensic investigations study the impact of the channel on the bullet, and the bullet on the target. The poet's bullets are amenable, hypothetically, to comparable analysis, especially when a poem is sometimes taken to stand for the poet (Kennelly 1995b). The form of analysis involved here is actually suggested by research in forensic linguistics in the case of authorship attribution (Chaski 1997, Chaski 1999, Chaski 2001).

In this paper we hypothesize a version of Brendan Kennelly which does not correspond to reality, but which nonetheless provides a coherent reading of his corpus of poetry. To discriminate the two, we will pretend that another poet, K, has managed, like the Pierre Menard character in Borges' *Fictions*, to write all of Kennelly's poems separately himself, and not as an act of plagiarism (Borges 1991). K actually does have a split personality, and does hear voices (K is actually attentive to the fluids and vocalizations of all of the body's apertures, as well as the voices inside the head). Some of these are localized voices, local to specific poems, voices that provide one-off commentary. Some of the women and children are in this category. Some are located in dialect regions or regions of time—Ozzie or *Cromwell's* Buffun—craftily contrived. Others are persistent and recurrent voices—Ace or Janey-Mary. Others are simply the names of places (as in the street names around the Liberties or Smithfield in Dublin), personified places.

We argue the coherence of our subjectively determined reading of K. We do so on the basis of close, if superficial, readings of the texts, attending to K's constructed biography. We bolster the claim of coherence in this reading with two forms of reality cross-checking. First, we establish external inter-rater and intra-rater judgements of identity among voices that we isolate across and within poems. Secondly, we apply blind evaluation techniques of the sort identified within corpus linguistics literature (e.g. Biber). The outcome is a—mitigated—success. We do find some support (which we document below) of our claim that the *dramatis personae* of the K corpus is such that there are in fact identifiable voices in the corpus which, unlike the voices of Shakespeare's characters, can be reliably individuated from each other.³ This supports our claim that K has split personality, and by extension supports a claim of uncommon creativity on the part of Brendan Kennelly.

2 Materials and methods

2.1 Materials

The text under analysis derives from four collections of poetry composed by K: *Cromwell* (1987), *The Book of Judas* (1991), *Moloney Up and At It* (1995a) and *Poetry My Arse* (1995c). The present analysis does not include the entirety of any of those volumes, only the poems indicated in Appendix A. It is not novel to raise questions about voices within the poetry of K's counterpart, Brendan Kennelly. The issues are discussed in reviews of the works subsequent to publication. In the course of one review, for example, Sowday (1984) writes, "The poem [*Cromwell*] contains many voices, and the voice upon which I am concentrating here is but one of the voices with which Cromwell speaks." As a comment on the framing of thematic structure, Lucy (1984) wrote, "As in a dream, there is no strict logical progression in the sequence of poems; images cluster; themes weave in and out; figures, voices vanish and reappear." If voice is equated with character, analysis of voice also figures into theoretical studies of the works. For example, Roche (1994) examines the content of the poetry and psychology of the characters delivered by voice. Quinlan (1991) focused on the

³ This claim is made on the basis of recent experimental analysis of the text of *Hamlet* by the first author in collaboration with Myriam Mencke.

conflict among characters as revealed by clashing voices. Persson (2000) also considered issues of conflict among characters, but without addressing the linguistic features of each voice explicitly. In addressing Kennelly's "mastery of different registers and discourses," Villacanas wrote that it "enables him *speak from the mind of Oliver Cromwell*, to inhabit him...." (2003:139). Voice in the work of Kennelly is certainly an established dimension of discussion and classification. So it is also with K.

Below is an example of a poem from *Cromwell*, in Cromwell's voice:

In Oliver's Army

No man shall depart a mile out of the Army, upon pain of death
 No man shall draw his sword without order, upon pain of death
 No man shall hurt a man bringing food, upon pain of death
 A sentinel asleep or drunk or forsaking his place shall die without mercy
 No man shall give a false Alarum, upon pain of death
 He that makes known the Watchword without Order shall die for it
 If a Pike-man throw away his pike, he shall die for that
 No man shall abandon his colours, upon pain of death
 None shall kill an Enemy who yields and throws down his Armes
 Rape, Ravishments, Unnatural Abuses shall meet with death
 Let God be served, Religion be frequented
 Let sellers of meat avoid the unsound, the unwholesome
 Let Heaven be praised with sermon and prayer
 Let all faults be punished by the Laws of War.

This is to be contrasted with the voice of Ozzie in *Book of Judas*.

prades

ozzie is stonemad about prades
 so he say kummon ta Belfast
 for the 12th an we see de Orangemen
 beatin de shit outa de drums
 beltin em as if dey was katliks heads

so we set out from Dublin
 an landed in Belfast for de fun
 it was brill
 dere was colour an music an everyone
 was havin a go at sumtin i dunno

what but ill never forget ozzie in
 de middul of all de excitement
 pickin pockets right left and centre

on de train back to Dublin he was laffin his head
 off, dere shud be more fukken prades he said

Other voices, with personalities and expressive behaviours in between these extremes are also evident, for instance, Janey-Mary and a child. The perception of these voices is evident

to some readers, but a matter of contention to others. The authors of this paper vociferously disagree about whether Janey-Mary and Janet are the same voice, for example.

Of the many collections of poetry that K has published, four are clearly organized around a cast of characters:⁴ *Cromwell*, *The Book of Judas*, *Poetry My Arse* and *Moloney Up and At It*. Within those books there are a total of 1399 poems (254, 587, 548 and 10, respectively). As a starting point for analysis we selected 76 poems on the basis of subjective reading and the perception of characteristic voices represented by the poems in the K repertoire. The two voices, OZZIE and the AUTHOR OF THE LETTERS, selected from *The Book of Judas* seemed the most homogeneous voices in the book, for instance. The selection for this study is limited in size in part because of the practical limitations of not having the works in machine readable form. The poems had to be typed out and proofread against published copy; the non-canonical orthography associated with OZZIE, for example, makes this a particularly onerous task. The main point here is that the selections were based on subjective reading. The question we explore at the heart of this research is whether the individuation of texts on the basis of perception of voice yields a body of texts that, using objective assessments of them, cluster into categories which correspond to subjective perceptions of voice.

The next section describes the methods from text classification tasks corpus linguistics that we applied to these texts to objectively analyze voice within the body of poetry considered. Here we point out the selection of poems (as indicated above) and how the materials are structured. Following considerable discussion about the level of granularity in which voice can be individuated, the second author constructed a set of electronic files labelled by source and a personal opinion on voice. In some cases, a voice spans more than one poem. In other cases, a poem is split into two voices.

Voice	Book
ACE	<i>Poetry My Arse</i>
AUTHOR OF THE LETTERS	<i>The Book of Judas</i>
BUFFUN	<i>Cromwell</i>
CHILD	<i>The Book of Judas</i>
CHILD	<i>Poetry My Arse</i>
CHILD	<i>Cromwell</i>
CROMWELL	<i>Cromwell</i>
JANEY-MARY	<i>Poetry My Arse</i>
MISTAKE	<i>Poetry My Arse</i>
MOLONEY	<i>Moloney Up and At It</i>
NARRATOR 1	<i>The Book of Judas</i>
NARRATOR 2	<i>Poetry My Arse</i>
OZZIE	<i>The Book of Judas</i>
WOMAN	<i>Moloney Up and At It</i>

Table I Voices and books

Table I indicates the voices analyzed and how they span books of poetry. Tables in Appendix A show the selection of files constructed from each book in order to have each

⁴ Thus, we ignore, for the present study, poems that appeared in isolation, or in heterogeneously authored volumes, or those which in an eclectic K collection exhibit constancy of voice from one poem to another.

file be homogeneous with respect to the perceived voice. A convention adopted in this paper is to format the name of a voice in small capital letters (e.g. OZZIE). Two distinct narrators are evident; however, the voice of the CHILD spans three books.

Notice for example, that “She’s there” is split into two files. One part of the poem provides the voice of the NARRATOR, a voice which is evident in others of K’s poems, akin to the CHORUS. By comparison, McDonald (2005) notes that in *Antigone* the role of the Chorus is displaced to individual characters:

My soul is her lower lip but only for
A moment, then it’s the story she becomes
Before my eyes in the coughing street where
I am trying to remember her name.

While another part of the same poem offers the voice of the CHILD.

‘Ginnie, Ginnie Green, sir. I takes this blanket, yes,
An’ I makes through the streets, I’m a beggar,
I looks right into the tourists’ eyes
An’ I takes what I can for me sisters an’ brothers.

An’ why am I tellin’ you this? Once, a bad day
You gave me bread, gave me white bread,
That’s why I’m talkin’ like this.’

In some places, as in “She’s there” a change of voice is signalled explicitly with quotation marks. But this is not universally the case. We individuated NARRATOR voices as sequences typically involving a mixture of first person and third person description. Sometimes quoted speech is of a voice distinct from the NARRATOR’s, as in “She’s there”, and sometimes the quoted text is essentially the voice of the NARRATOR as well (in the sense that as editor, selecting text to quote can be an endorsement of the content or form of the text).

A total of 95 files were thus constructed. The files ranged in length from one line to 194 lines (both of these, interestingly enough, were in the voice of CROMWELL), with an average of 25.4 lines. In many cases, the files coincide with entire poems.

2.2 Methods

The first method of analysis was described above. The authors individually classified the poems, and in some cases agreed on perception of persistent voices echoing from the poems. The second method, and focus of the paper, is to ascertain whether techniques in authorship attribution from forensic linguistics and the text classification literature discriminate the voices in comparable manner (Vogel 2007). The approach considers all pairwise comparisons of the files and ranks them in similarity according to their characterization: as similar to the rest of the book the file appears in (Section 3.1), as similar to the body of files constituting a voice (Section 3.2), or as similar to individual files (Section 3.3). The exact techniques used here have been applied to a number of different sorts of categorization tasks involving correlations with subjective and objective categories (see e.g., Van Gijssel and Vogel 2003, O’Brien and Vogel 2003, Mencke 2004, Medori 2005, Hogan 2005).

The core of the method used here for judging similarity is based on objective criteria. We consider both distributions of words and distributions of individual letters. Given sometimes unique orthography (witness “fukken” and “laffin” above), it is clear that a word-based analysis will fairly uniquely identify a voice. Thus, the problem is made more interesting by examining letter distributions as has been argued in forensic linguistics (Chaski 1999).⁵ Note that normally word choice determines spelling, and in general people do not impose conscious control over their spelling. Thus, if the files cluster by voice under an analysis of letter distributions, then the result is stronger than if it obtains because of particular words. While we have performed the analyses at both the word and letter level, here we report specifically on the letter distribution analysis; the two analyses largely coincide.

The exact method is to compare two files on the basis of the frequencies of occurrences of letters in each, using a chi-square analysis for each letter (which thus relativizes the frequencies to file sizes), calculating the sum of the chi-square values for each comparison, and essentially dividing by the total number of comparisons (Kilgarriff 2001). Chi-square testing is typically done to prove that two frequency distributions are significantly different. When the value for any one comparison is large enough, it exceeds a threshold that is generally accepted for establishing the level of statistical significance. However, we are interested in similarity, so the smaller the chi-square value, the smaller the difference. This cumulative chi-square value is our index of the difference between two files that we can then use in a Mann-Whitney analysis which evaluates the significance of rank ordering according to similarity between two categories. Thus, the variation in the experiments reported is in terms of the categories assigned: by books, by voice, by (part of) poem.

3 Analysis of results

The analysis generates reams of output from which results are extracted and interpreted. Only similarity rankings that achieve significance are reported, and the extracted tables of only the comparison involving four categories is included explicitly here as an appendix (B). The remainder of the results are simply stated and discussed.

3.1 Poems and books

The relevant question here is which book a given poem most resembles. It is easy to imagine that it should always be most like the book it appears in, but this is not the case. This analysis involved four categories, one for each of the books considered. Notice that the files drawn from *Moloney Up and At It* were *a priori* classified either as of the voice of the Woman or of Moloney in their file names (see Table 7 in Appendix B). The results show that only one of the files was classified as more like the rest of the files comprising *Moloney Up and At It* than the other poems. In the main, both voices are like those of *Poetry My Arse*. The files of *Cromwell* (Table 8) are mainly like the poems of that book, but some are actually more like those of *Poetry My Arse*. The files of *The Book of Judas* (Table 9) are most similar to their own book, but many are also attributable to *Poetry My Arse* and *Cromwell*. Finally, files of *Poetry My Arse* split in assignment to *Poetry My Arse* itself, *Cromwell* and *Moloney Up and At It* (Tables 10 and 11). What these results indicate is that the books are not individually homogeneous. Poems in some of the books are more like poems in others of the books than they are like the rest of the poems in their own book.

⁵ The argument in forensic linguistics is not actually about making the problem more interesting, but the conclusions more reliable.

3.2 Poems and voices

Here we inquire into the similarity between a subjectively constructed voice and the file from each voice. To structure the discussion, we proceed via the books that the files are derived from. We do not mention the cases where statistical significance did not obtain. Because of the subjective construction of voice as a category, it is natural to expect noise, and indeed we find it. However, we also find clear signals.

The single file representing the NARRATOR 1 of *The Book of Judas* is most like the AUTHOR OF THE LETTERS poems of the same volume, but also quite similar to MOLONEY and CROMWELL. All manifest a mature formal register. OZZIE is overwhelmingly self-similar, and less-so NARRATOR 2. Similarly, the AUTHOR OF THE LETTERS is also a dominating voice – self-similar for each file comprising the voice, but with additional matches to NARRATOR 2.

The WOMAN of *Moloney Up and At It* is equally like JANEY-MARY of *Poetry My Arse* and herself at the highest levels of significance. There is also similarity to both ACE and the NARRATOR 2 of *Poetry My Arse*. Despite being an eponymous character, files associated with MOLONEY are actually scattered in similarity: some each to NARRATOR 2, JANEY-MARY, ACE, the CHILD, the WOMAN and BUFFUN.

From *Poetry My Arse*, the NARRATOR 2 is most significantly matched to himself, and also to: the AUTHOR OF THE LETTERS, CROMWELL, and ACE. Files of ACE match JANEY-MARY and ACE equally, also NARRATOR 2, and the WOMAN. The MISTAKE is matched with the AUTHOR OF THE LETTERS (*Book of Judas*) most strongly, and also the CHILD, NARRATOR 2 and MOLONEY. The CHILD is another scattered voice, matching NARRATOR 2, BUFFUN, ACE, JANEY-MARY, the WOMAN, the AUTHOR OF THE LETTERS and MOLONEY. In turn, files of JANEY-MARY match her own most closely, followed by ACE, NARRATOR 2 (both of the same book), the WOMAN and MOLONEY of *Moloney Up and At It*, and the AUTHOR OF THE LETTERS.

Voice	Book	Match
ACE	<i>Poetry My Arse</i>	equivocal
AUTHOR OF THE LETTERS	<i>The Book of Judas</i>	univocal
BUFFUN	<i>Cromwell</i>	equivocal
CHILD	<i>The Book of Judas</i>	equivocal
CHILD	<i>Poetry My Arse</i>	equivocal
CHILD	<i>Cromwell</i>	equivocal
CROMWELL	<i>Cromwell</i>	narrators
JANEY-MARY	<i>Poetry My Arse</i>	women
MISTAKE	<i>Poetry My Arse</i>	equivocal
MOLONEY	<i>Moloney Up and At It</i>	equivocal
NARRATOR 1	<i>The Book of Judas</i>	narrators
NARRATOR 2	<i>Poetry My Arse</i>	narrators
OZZIE	<i>The Book of Judas</i>	univocal
WOMAN	<i>Moloney Up and At It</i>	women

Table 2 Results: Voices and books

Analyzing *Cromwell*, files of BUFFUN match most strongly and most frequently with NARRATOR 2 and also with the CHILD, the AUTHOR OF THE LETTERS, the MISTAKE and CROMWELL, matching BUFFUN's own files only twice. Recall that CROMWELL and BUFFUN are from the same book. As for CROMWELL, those files also match with most significance and

frequency with NARRATOR 2. Some files of this voice also match the AUTHOR OF THE LETTERS, BUFFUN, CROMWELL himself, and JANEY-MARY. Similarity between parts of *Cromwell* and the AUTHOR OF THE LETTERS is to be expected from the fact that the associated poems were composed contemporaneously.⁶ Table 2 summarizes the results.

It is evident that *a priori* individuation of voices does not cluster the files uniformly. The AUTHOR OF THE LETTERS and OZZIE are the strongest voices. Other strong voices appear to cluster on types: the narrators, and the women cluster together.

3.3 Poems and poems

Under this analysis, we ask what other file an individual file most resembles. This is different from the preceding analysis in that there the analysis compares a file as a good fit within some voice as opposed to the complement category of all possible voices. This analysis takes into account all of the files within a voice category in making the estimation. The alternative outlined here identifies the most similar individual file and considers how often the two files are within the same *a priori* classification of voice (of course this is not possible for those voices constituted by exactly one file, but it is still interesting to note what files those are most similar to).

Persistently, individual files that we classified as OZZIE are most like other files in the OZZIE category. While ACE, BUFFUN, JANEY-MARY and NARRATOR 2 have files that are absolutely similar to more than one other file of their voice, they also closely resemble a range of other voices.

4 Discussion

While it is possible (and possibly preferable) to engage in direct textual analysis in order to analyze similarity of poems in terms of voices that emerge from them, we have shown it possible to meaningfully apply less subjective techniques from corpus linguistics to the task as well. Because the metrics are objective, they are guaranteed reliability. Anyone who divides the poems into files as we have arranged them and who applies the same statistically driven reasoning will reach exactly the same conclusions. The fact that these similarity measures cluster the files in ways that reflection justifies with thematic (or in some cases, historical) coherence further establishes their validity. While the results reported have only been of the letter unigram analyses, because letter unigram distributions tend to give sharpest discriminations, comparable results emerge when longer strings of letters are considered, such as letter bigrams, and when the tokens counted are individual words instead of letters. The truly interesting results emerge when one considers the results from alternative characterization—poems clustering with their books, the externally classified voices or individual poems, as has been explored above. What remains to be done is to carry out the study with respect to the entire corpus of K's poetry.

It is in the context of the full analysis that one can discriminate whether there is an intervening effect of our having selected the 76 poems for this analysis initially on the basis of subjective reading. We find that the statistical analysis employed here clusters poems into categories that show correlation with the subjective readings. It is a slightly different question to consider whether the same clusters will emerge when there is no initial subjective selection. Works that cluster together here may well fall into different clusters when the full set of poems is analyzed. In that situation, voice, as addressed here may well

⁶ Brendan Kennelly, personal communication to the second author.

not be the dominating feature that separates the texts, and reflection would be necessary to assess the best interpretation of the clustered poems. On the other hand, the larger set of poems may also cluster by voice. Nonetheless, when a substantial number of poems is chosen on the basis of a reading of voice, the corpus based techniques demonstrate substantial agreement with that reading. The value of this kind of study in computational stylistics is interactive. In one direction, it provides construct validity to stylistics—objective support for subjective readings—and it can yield via objective means discovery of clusters of similarity that require subjective interpretation. In the other direction, it provides to corpus linguistics evidence that methods which cluster texts on the basis of even orthographic distributions, and thus have little face validity, actually do deliver judgements of similarity that merit interpretation.

What emerges from our clinical analysis of K is that his poetry is not produced in homogeneous collections or even univocalic poems. Clear, distinct, personalities are evident either by name or archetype: OZZIE and the AUTHOR OF THE LETTERS are crisply audible within the analysis by voice; OZZIE is the strongest voice across analyses, and the WOMEN, the NARRATORS (blending in CROMWELL and the AUTHOR OF THE LETTERS) are also clearly heard. That K does not produce poems with homogeneous form is suggestive of schizophrenia, and forcefully suggests that his counterpart, Kennelly, composes poetry with a very tight control over characterization of voices.

Works Cited

- Beye, C. R. (1972) *The Iliad, the Odyssey and the epic tradition*. Gloucester, Massachusetts: Peter Smith.
- Biber, D. (1988) *Variation across speech and writing*, Cambridge: Cambridge University Press.
- Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus linguistics: Investigating languages Structure and use*. Cambridge: Cambridge University Press.
- Borges, J. L. (1991) Pierre Menard, author of the Quixote, in A. Kerrigan (ed.) *Fictions*. Paris: Calder Publications.
- Chaski, C. (1997) Who wrote it? Steps toward a science of authorship identification. *National Institute of Justice Journal* 233: 15-22.
- Chaski, C. (1999) Linguistic authentication and reliability. *Proceedings of national conference on science and the law*: 97-148.
- Chaski, C. (2001) Empirical evaluations of language-based author identification techniques. *Forensic Linguistics* 8.1: 1-65.
- Hogan, L. (2005) *A corpus linguistic analysis of American, British and Irish political speeches*. Master's thesis. Centre for language and communication studies, Trinity College Dublin.
- Kennelly, B. (1987 [1983]) *Cromwell*. Newcastle upon Tyne: Bloodaxe Books.
- Kennelly, B. (1991) *The Book of Judas*. Newcastle upon Tyne: Bloodaxe Books.
- Kennelly, B. (1995a) *Moloney Up and At It*. Dublin: Mercier.
- Kennelly, B. (1995b) Poems are Cheeky Bastards. In *Poetry My Arse*. Newcastle upon Tyne: Bloodaxe Books.
- Kennelly, B. (1995c) *Poetry My Arse*. Newcastle upon Tyne: Bloodaxe Books.
- Kilgarriff, A. (2001) Comparing corpora. *International Journal of Corpus Linguistics* 6.1: 97-133.
- Lucy, S. (1984) Review. *The Tablet*. 952-953.

- McDonald, M. (2005) Rebel women: Brendan Kennelly's versions of Irish tragedy. *New Hibernia Review* 9.3: 123-36.
- Medori, J. (2005) *Experiments testing the reliability and validity of author identification methods*. Master's thesis. Computational linguistics group, Trinity College Dublin.
- Mencke, M. (2004) *Experiments to validate scientifically reliable author identification techniques*. MPhil in Linguistics, Centre for language and communication studies, Trinity College, University of Dublin.
- O'Brien, C. and Vogel, C. (2003) Spam filters: Bayes vs. Chi-squared; Letters vs. Words. *Proceedings of the international symposium on information and communication technologies*: 298-303.
- Persson, A. (2000) *Betraying the age: social and artistic protest in Brendan Kennelly's Work*. Gothenburg: Universtiy of Gothenburg Press.
- Quinlan, K. (1991) Review. *World Literature Today* 65.4: 711.
- Roche, A. (1994) The Book of Judas: Parody, double cross and betrayal. In R. Pine (ed.) *Dark fathers into light*. Newcastle upon Tyne: Bloodaxe Books.
- Sampson, G. (2001) *Empirical linguistics*. New York: Continuum.
- Sowday, J. (1984) Review. *Poetry Ireland Review* 10.2: 57-60.
- Van Gijssel, S. and Vogel, C. (2003) Inducing a cline from corpora of political manifestos. *Proceedings of the international symposium on information and communication technologies*: 304-310.
- Villacanas, B. (2003) Brendan Kennelly's Cromwell: Black comedy as exorcism. *Estudios Ingeles de la Universidad Complutense* 18: 137-57.
- Vogel, C. (2007), N-gram distributions in texts as proxy for textual fingerprints. In A. Esposito, E. Keller, M. Marinaro and M. Bratanic (eds.) *The fundamentals of verbal and non-verbal communication and the biometrical issue*. IOS Press.

Appendix A. Key to poems used and index of electronic files

Our file name	Voice	Poem
BOJ1991Ozzie1	OZZIE	Words
BOJ1991Ozzie10	OZZIE	skool
BOJ1991Ozzie2	OZZIE	sumtimes ozzie
BOJ1991Ozzie3	OZZIE	madmanalive
BOJ1991Ozzie4	OZZIE	no trubbal
BOJ1991Ozzie5	OZZIE	prades
BOJ1991Ozzie6	OZZIE	Flushed
BOJ1991Ozzie7	OZZIE	ozzie smiled
BOJ1991Ozzie9	OZZIE	skool
BOJ1991Ozzie8	OZZIE	bang
VoiceofChildren3Narr	NARRATOR 1	She's There
VoiceofChildren3Child	CHILD	She's There
VoiceofChildren4Child	CHILD	A Deeper Tyrant
Authorofletters10Author	AUTHOR	Coming Soon
Authorofletters11Author	AUTHOR	Inadvertent
Authorofletters1Author	AUTHOR	A Former Particular Friend
Authorofletters13Author	AUTHOR	The Common People

Authorofletters2Author	AUTHOR	Beyond Suspicion
Authorofletters3Author	AUTHOR	Worth Watching
Authorofletters4Author	AUTHOR	Mr Watson
Authorofletters5Author	AUTHOR	A Country Gentleman
Authorofletters6Author	AUTHOR	An Incoming Tide
Authorofletters7Author	AUTHOR	The Transaction
Authorofletters8Author	AUTHOR	Let me Survive
Authorofletters9Author	AUTHOR	A Man Named Clarke

Table 3 Texts drawn from *The Book of Judas*

Our file name	Voice	Poem
C1983Buffun10Buffun	BUFFUN	Saskatoon
C1983Buffun1Buffun	BUFFUN	I was there
C1983Buffun2Buffun	BUFFUN	The Curse
C1983Buffun3Buffun	BUFFUN	You Would Have Blessed Me
C1983Buffun4Buffun	BUFFUN	The Crowd
C1983Buffun5Buffun	BUFFUN	The Crowd and the Curse
C1983Buffun6Buffun	BUFFUN	Journey to the Golden Man
C1983Buffun7Buffun	BUFFUN	Beyond the Warning Sign
C1983Buffun8Buffun	BUFFUN	Some Tiny Right
C1983Buffun9Buffun	BUFFUN	Silver
C1983Oliver10Oliver	OLIVER	Oliver to His Daughter
C1983Oliver1Oliver	OLIVER	Oliver to His Brother
C1983Oliver2Oliver	OLIVER	In Dublin
C1983Oliver3Oliver	OLIVER	A Friend of the People
C1983Oliver4Oliver	OLIVER	History
C1983Oliver5Oliver	OLIVER	In Oliver's Army
C1983Oliver6Oliver	OLIVER	Severest Friend
C1983Oliver7Oliver	OLIVER	Oliver to His Army
C1983Oliver8Oliver	OLIVER	Oliver Speaks to His Countrymen
C1983Oliver9Oliver	OLIVER	An Expert Teacher
C1983VoiceofChildren1Child	CHILD	Coal Dust

Table 4 Texts drawn from *Cromwell*

Our file name	Voice	Poem
MUAA11982Moloney10Moloney	MOLONEY	Moloney enters into a Dialogue concerning the Listowel water supply
MUAA11982Moloney10Woman	WOMAN	Moloney enters into a Dialogue concerning the Listowel water supply
MUAA11982Moloney1Moloney	MOLONEY	Moloney Up and At It
MUAA11982Moloney2Moloney	MOLONEY	Moloney at the Wake
MUAA11982Moloney3Moloney	MOLONEY	Moloney Meets Miss Immaculata Mullally
MUAA11982Moloney4Moloney	MOLONEY	Moloney's Revenge
MUAA11982Moloney5Moloney	MOLONEY	Moloney Recalls the Marriage of the Barrell Muldoon
MUAA11982Moloney7Moloney	MOLONEY	Moloney Remembers Timmy Thankgod

MUAAI1982Moloney8Moloney	MOLONEY	Moloney and the Dust
MUAAI1982Moloney9Moloney	MOLONEY	Moloney Remembers the Resurrection of Kate Finucane

Table 5 Texts drawn from *Moloney Up and At It*

Our file name	Voice	Poem
PMA1995Ace1Ace	ACE	Ultimate Peace
PMA1995Ace1Narr	NARRATOR 2	Ultimate Peace
PMA1995Ace2Ace	ACE	The only, the only
PMA1995Ace2Narr	NARRATOR 2	The only, the only
PMA1995Ace3Ace	ACE	Conversation with an eggshell
PMA1995Ace3Narr	NARRATOR 2	Conversation with an eggshell
PMA1995Ace4Ace	ACE	Sniff
PMA1995Ace4Narr	NARRATOR 2	Sniff
PMA1995Ace5Ace	ACE	A drop from Ace's Nutbook
PMA1995Ace6Ace	ACE	Whore
PMA1995Ace6JaneyM	JANEY-MARY	Whore
PMA1995Ace7Ace	ACE	Generation Map
PMA1995Ace7Narr	NARRATOR 2	Generation Map
PMA1995JM10JaneyM	JANEY-MARY	Knock
PMA1995JM10Narr	NARRATOR 2	Knock
PMA1995JM2Narr	NARRATOR 2	Question to a rolled-over poet
PMA1995JM2JaneyM	JANEY-MARY	Question to a rolled-over poet
PMA1995JM1JaneyM	JANEY-MARY	Some nights
PMA1995JM3JaneyM	JANEY-MARY	Great lover
PMA1995JM4Ace	ACE	If me granny could see me now (maybe she does)
PMA1995JM4JaneyM	JANEY-MARY	If me granny could see me now (maybe she does)
PMA1995JM6JaneyM	JANEY-MARY	Druggie
PMA1995JM6Narr	NARRATOR 2	Druggie
PMA1995JM7Ace	ACE	Facing faces
PMA1995JM7JaneyM	JANEY-MARY	Facing faces
PMA1995JM8Ace	ACE	Don't blame
PMA1995JM8JaneyM	JANEY-MARY	Don't blame
PMA1995JM9JaneyM	JANEY-MARY	whore
PMA1995JM5JaneyM	JANEY-MARY	Encouragement
PMA1995JM5Ace	ACE	Encouragement
PMA1995TheMistake1Narr	NARRATOR 2	The Mistake
PMA1995TheMistake1TheMist	THE MISTAKE	The Mistake
PMA1995TheMistake2TheMist	THE MISTAKE	What it must feel like
PMA1995TheMistake3Narr	NARRATOR 2	The heart of the matter
PMA1995TheMistake3TheMist	THE MISTAKE	The heart of the matter
PMA1995VoiceOfChildren5Child	CHILD	Shy Child

Table 6 Texts drawn from *Poetry My Arse*

Appendix B Representative output tables using books as categories

The tables in this section are specific to the discussion of 3.1 in which the four books are taken as categories, and each poem within each book is considered. The question is which book does the file individuating a voice most resemble? That is the question answered on each row. For example, Table 7 shows the results for *Moloney Up and At It*. The first entry in a row is the name of a file corresponding to a voice within the book. The second column indicates the category that this file is most similar to, and the third and fourth columns provide the statistical P -values (from the Mann-Whitney test) associated with the probability of erroneously rejecting the null hypothesis associated with the comparison. The null hypothesis in this case is that the overall rank similarity of this file with respect to the category provided by the poems in the book *Poetry My Arse* is due to random chance. The first row indicates that one can have a great deal of confidence in rejecting the null hypothesis, that one has an estimated chance of less than 0.05% ($p < 0.0005$) of being wrong in rejecting it to conclude that there is something rather significant about the similarity between the file and the other book. At each row one is considering the file with respect to the category in the second column versus the complement of that category (the overall rank similarity of everything else); thus, a given file can have equally significant similarity to more than one category (as does MUAAl1982Moloney6Woman). As indicated in the main text, this is a striking result in that *Moloney Up and At It* does not demonstrate a strong internal identity. Instead its contents cluster with the other book.

Following the Bernoulli schema one can evaluate the homogeneity of the category. This is best understood in terms of testing a coin for whether it is a fair coin. If one flips a coin 20 times, one expects, if it is fair, to have roughly as many instances of heads as tails. However, if one repeats that experiment many times, if one conducts 100 experiments consisting of 20 coin flips each, one shouldn't be surprised if in 3 of the experiments, heads came up 15 out of the 20 tosses in each experiment. However, there is only a 3% chance of this happening. This is related to the significance values and confidence about rejecting the null hypothesis of a random effect as associated with the Mann-Whitney test, discussed above, and with inferential statistics in general. The Bernoulli schema applies the same reasoning to n -sided coins (polyhedra). This can be used to assess homogeneity. The coin in use for this research has four sides (a regular tetrahedron): each of the categories is a possible outcome. The number of tosses is determined by the number of files within the category. Thus, with a four sided coin, and the number of files that *Moloney Up and At It* is divided into, the chance of precisely one category to be most appropriate for all of the files is vanishingly small. This reasoning shows the category of poems provided by *Moloney Up and At It* to be extremely homogeneous, but with respect to the category of poems from *Poetry My Arse*. This is a surprising outcome. Yet, it is related to the heterogeneity of *Poetry My Arse* as discussed below.

Filename	Results		
	Assigned	P Value >	P Value <
MUAAl1982Moloney10Moloney	PMA	0.0	0.0005
MUAAl1982Moloney10Woman	PMA	0.0	0.0005
MUAAl1982Moloney1Moloney	PMA	0.0	0.0005
MUAAl1982Moloney1Woman	PMA	0.0	0.0005
MUAAl1982Moloney2Moloney	PMA	0.025	0.05

MUAAI1982Moloney3Moloney	PMA	0.0	0.0005
MUAAI1982Moloney3Woman	Moloney	0.025	0.05
MUAAI1982Moloney3Woman	PMA	0.0	0.0005
MUAAI1982Moloney4Moloney	PMA	0.0005	0.001
MUAAI1982Moloney5Moloney	PMA	0.0	0.0005
MUAAI1982Moloney6Woman	PMA	0.025	0.05
MUAAI1982Moloney6Woman	Cromwell	0.025	0.05
MUAAI1982Moloney7Moloney	PMA	0.0025	0.005
MUAAI1982Moloney7Woman	PMA	0.0	0.0005
MUAAI1982Moloney8Moloney	PMA	0.0	0.0005
MUAAI1982Moloney9Moloney	PMA	0.0	0.0005

Table 7 Results for *Moloney Up and At It*

The poems of *Cromwell* are split between just two categories. The homogeneity with respect to their source is significant ($p < 0.01$). It is interesting that *Poetry My Arse* is the only other possibility; however, it doesn't reach statistical significance. Looking within the voices, it is clear that BUFFUN is fairly unique to *Cromwell* ($p < 0.01$), but OLIVER is divided in similarity between the two books.

Filename	Results		
	Assigned	P Value >	P Value <
C1981Buffun10Buffun	PMA	0.025	0.05
C1981Buffun1Buffun	PMA	0.0025	0.005
C1981Buffun2Buffun	Cromwell	0.01	0.025
C1981Buffun3Buffun	Cromwell	0.005	0.01
C1981Buffun4Buffun	Cromwell	0.0025	0.005
C1981Buffun5Buffun	Cromwell	0.005	0.01
C1981Buffun6Buffun	Cromwell	0.025	0.05
C1981Buffun7Buffun	Cromwell	0.0025	0.005
C1981Buffun8Buffun	Cromwell	0.005	0.01
C1981Oliver10Oliver	Cromwell	0.01	0.025
C1981Oliver1Oliver	Cromwell	0.001	0.0025
C1981Oliver2Oliver	PMA	0.001	0.0025
C1981Oliver3Oliver	PMA	0.0	0.0005
C1981Oliver4Oliver	Cromwell	0.01	0.025
C1981Oliver5Oliver	Cromwell	0.01	0.025
C1981Oliver6Oliver	PMA	0.025	0.05
C1981Oliver6Oliver	Cromwell	0.01	0.025
C1981Oliver7Oliver	PMA	0.1	0.25
C1981Oliver7Oliver	Cromwell	0.01	0.025
C1981Oliver8Oliver	Cromwell	0.0005	0.001
C1981Oliver9Oliver	PMA	0.0	0.0005
C1981VoiceofChildren1Child	Cromwell	0.01	0.025

Table 8 Results for *Cromwell*

Applying comparable reasoning to the other two books, one notes again (Table 9) that OZZIE is a distinctive voice for *The Book of Judas* ($p < 0.01$) but that the AUTHOR OF THE

LETTERS has a voice quality that makes it more similar to the poems of *Cromwell* overall ($p < 0.05$). As an entire category, the poems within *The Book of Judas* are not homogenous, rather they split along voice lines. Exactly the same holds for *Poetry My Arse* (Table 10 and Table 11). *Poetry My Arse* exhibits the least homogeneity as a book, it has the greatest number of constituent clusters.

Results			
Filename	Assigned	P Value >	P Value <
BOJ1991Ozzie11	Cromwell	0.01	0.025
BOJ1991Ozzie11	BOJ	0.001	0.0025
BOJ1991Ozzie1	BOJ	0.005	0.01
BOJ1991Ozzie2	BOJ	0.0	0.0005
BOJ1991Ozzie3	BOJ	0.0	0.0005
BOJ1991Ozzie4	BOJ	0.001	0.0025
BOJ1991Ozzie5	Cromwell	0.025	0.05
BOJ1991Ozzie5	BOJ	0.005	0.01
BOJ1991Ozzie6	BOJ	0.001	0.0025
BOJ1991Ozzie7	BOJ	0.001	0.0025
BOJ1991Ozzie8	BOJ	0.01	0.025
BOJ1991Ozzie9	BOJ	0.005	0.01
BOJ1191VoiceofChildren2Child	Cromwell	0.001	0.0025
BOJ1991VoiceofChildren3Child	PMA	0.0	0.0005
BOJ1991VoiceofChildren3Narr	Moloney	0.01	0.025
BOJ1991VoiceofChildren4Child	PMA	0.025	0.05
BOJ1991authorofletters10Author	PMA	0.005	0.01
BOJ1991authorofletters11Author	Cromwell	0.025	0.05
BOJ1991authorofletters13Author	PMA	0.025	0.05
BOJ1991authorofletters13Author	Cromwell	0.025	0.05
BOJ1991authorofletters1Author	Cromwell	0.001	0.0025
BOJ1991authorofletters2Author	PMA	0.005	0.01
BOJ1991authorofletters3Author	PMA	0.0	0.0005
BOJ1991authorofletters4Author	PMA	0.005	0.01
BOJ1991authorofletters4Author	Cromwell	0.025	0.05
BOJ1991authorofletters5Author	Cromwell	0.001	0.0025
BOJ1991authorofletters6Author	Cromwell	0.01	0.025
BOJ1991authorofletters7Author	Cromwell	0.0	0.0005
BOJ1991authorofletters8Author	PMA	0.005	0.01

Table 9 Results for *The Book of Judas*

Results			
Filename	Assigned	P Value >	P Value <
PMA1995Ace1Ace	PMA	0.0	0.0005
PMA1995Ace1Narr	PMA	0.0005	0.001
PMA1995Ace2Ace	Cromwell	0.01	0.025
PMA1995Ace2Narr	Cromwell	0.001	0.0025
PMA1995Ace4Ace	PMA	0.0	0.0005
PMA1995Ace4Narr	PMA	0.0	0.0005
PMA1995Ace5Ace	Cromwell	0.005	0.01

PMA1995Ace6Ace	Moloney	0.0	0.0005
PMA1995Ace6Ace	Cromwell	0.0025	0.005
PMA1995Ace6JaneyM	Moloney	0.005	0.01
PMA1995Ace6JaneyM	PMA	0.001	0.0025
PMA1995Ace7Ace	Moloney	0.0	0.0005
PMA1995Ace7Ace	PMA	0.01	0.025
PMA1995Ace7Narr	PMA	0.005	0.01
PMA1995JaneyMary10JaneyM	PMA	0.0	0.0005

Table 10 Results for Poetry My Arse

Results			
Filename	Assigned	P Value >	P Value <
PMA1995JaneyMary10Narr	Moloney	0.0	0.0005
PMA1995JaneyMary10Narr	Cromwell	0.05	0.1
PMA1995JaneyMary1JaneyM	Moloney	0.0	0.0005
PMA1995JaneyMary1JaneyM	PMA	0.05	0.1
PMA1995JaneyMary2JaneyM	Moloney	0.05	0.1
PMA1995JaneyMary2JaneyM	PMA	0.0	0.0005
PMA1995JaneyMary2Narr	PMA	0.005	0.01
PMA1995JaneyMary2Narr	Cromwell	0.1	0.25
PMA1995JaneyMary3JaneyM	Moloney	0.001	0.0025
PMA1995JaneyMary4Ace	PMA	0.0	0.0005
PMA1995JaneyMary4JaneyM	PMA	0.0	0.0005
PMA1995JaneyMary6JaneyM	PMA	0.0	0.0005
PMA1995JaneyMary6Narr	Cromwell	0.0025	0.005
PMA1995JaneyMary7Ace	PMA	0.0	0.0005
PMA1995JaneyMary7JaneyM	Moloney	0.0	0.0005
PMA1995JaneyMary7JaneyM	PMA	0.01	0.025
PMA1995JaneyMary8Ace	Moloney	0.005	0.01
PMA1995JaneyMary8Ace	PMA	0.0	0.0005
PMA1995JaneyMary8JaneyM	Moloney	0.01	0.025
PMA1995JaneyMary8JaneyM	PMA	0.0	0.0005
PMA1995JaneyMary9JaneyM	Moloney	0.0005	0.001
PMA1995JaneyMary9JaneyM	PMA	0.025	0.05
PMA1995JaneymMary5Ace	PMA	0.0	0.0005
PMA1995JaneymMary5Janey	Moloney	0.025	0.05
PMA1995JaneymMary5Janey	PMA	0.0	0.0005
PMA1995TheMistake1Narr	PMA	0.01	0.025
PMA1995TheMistake1Narr	Cromwell	0.05	0.1
PMA1995TheMistake1TheMistake	Cromwell	0.01	0.025
PMA1995TheMistake2TheMistake	Moloney	0.025	0.05
PMA1995TheMistake3Narr	Cromwell	0.001	0.0025
PMA1995VoiceofChildren5Child	Moloney	0.05	0.1
PMA1995VoiceofChildren5Child	Cromwell	0.025	0.05

Table 11 Results for PMA (continued)