Paul Barrett

Paraphrasing the Paraphrase

OR What I Learned from Reading Every Issue of *Canadian Literature / Littérature canadienne* and *Studies in Canadian Literature / Études en littérature canadienne*

Forty years after its publication, the insights and theoretical implications of Frank Davey's "Surviving the Paraphrase" have become truisms of Canadian literary study. Davey critiques thematic criticism for its acts of paraphrase wherein literary texts are reduced to a single theme or statement. According to Davey, the thematic "critic extracts . . . the paraphrasable content and throws away the form. He attends to the explicit meaning of the work and neglects whatever content is implicit in its structure, language or imagery" (3). Davey attacks thematicism's basis in "Arnoldian humanism," which he identifies as an anti-literary "tradition in which the artist speaks, unconsciously or consciously, for the group." (2). Indeed, in the wake of Davey's publication, it became a customary rhetorical move in Canadian literary scholarship to declare the seriousness of one's study by aligning oneself, even implicitly, with Davey and dismissing thematic criticism in its historical and contemporary forms. Whereas thematic criticism once represented an effort to "institutionalize Canadian literature and to establish a close rapport between it and what was assumed to be the character of the Canadian nation-state" (Kamboureli 19), critics such as Smaro Kamboureli and Russell Brown have since shown that "it has not been so much thematicism that has governed Canadian criticism but the critics' obsession with the idea of it" (Kamboureli 20). Brown explains that "while this critical approach largely disappeared, *attacks* on Canadian thematic criticism continued and became more public" (658; emphasis original). He describes the "emergence of an orthodoxy" both within and beyond the academy where "accusing a work of thematicism was a useful

way to dismiss a critical study" (664). Moreover, as Imre Szeman points out, "there are far fewer texts of 'thematic criticism' than one might imagine from all the worries about it" (30). As such, a return to Davey, paraphrase, and anti-thematicism might seem a retrospective reflection upon another time in Canadian critical writing. However, despite the waning import of thematicism to contemporary debate, Davey's dismissal[1] of the merits of thematic criticism remains a negative structuring presence within the field and an implicit starting point for many contemporary critics.

What if the critical consensus is wrong? What if Davey's attack on paraphrase is misdirected and the intellectual tradition that grounds anti-thematicism is not assured? More accurately, what if there is value in the paraphrase of literature, genre, canon, and criticism that has gone heretofore unremarked upon by virtue of its identification as the kind of thematicism that Davey warns us against? What critical possibilities emerge from the polemical assertion that paraphrase provides unique insights that are otherwise obscured by Davey's combination of close reading, new criticism, and attention to the dialectical emergence of form and content? What new observations can we make about Canadian literature if we adopt, even as a negative hypothesis, a criticism of paraphrase in order to survive "Surviving the Paraphrase" and come out from under Davey's shadow?

My provocation, and the reason I open with this discussion of paraphrase, is that emerging computationally assisted forms of textual analysis blur the lines between received notions of close reading and paraphrase in a manner that rebukes much of Davey's critique. These two techniques make it possible to place paraphrase and close textual analysis into a troubled, yet productive, dialogue. In what follows, I engage in digital forms of close and distant reading of two major Canadian literary journals to demonstrate how the emergence of the methods of digital humanities (DH)[2] provides new opportunities for literary and cultural paraphrase by disrupting the distinction between paraphrase and close reading. This disruption is most apparent in the use of topic modelling algorithms to *read* a large corpus of texts.

Topic modelling is a computational method developed first by computer scientists which has become aligned with Franco Moretti's concept of "distant reading," whereby the critic uses an algorithm to identify subtle themes in a large corpus of documents through analysis of word repetition, collocation, and documents' shared terms. Topic modelling takes as its input a large collection of texts and provides, as output, a series of topics (defined as a collection of keywords relevant to that topic) that describe, in varying

degrees of proportionality, that collection of documents. Topic modelling is a kind of *automated paraphrase* as it begins with the mathematical assumption that a collection of documents can be interpreted as a statistical distribution of topics or themes. Indeed, Russell Brown's description of thematic argument as a "reading act that follows from observing the existence of patterns that seem to have significance or to delineate a range of significance" (672) is also a fair description of topic modelling. The advantage of topic modelling, however, is that the themes are derived from reading a broad corpus and not projected outward from a selected group of texts. Advocates such as Mark Steyvers and Tom Griffiths argue that it provides a necessary wide view of a discourse or corpus which can guide deeper interpretations and which would be otherwise unavailable to an individual reader.

Where some forms of DH scholarship align more closely with traditional literary analysis (such as text encoding or tracking an author's use of a particular term) and can be conceived of as merely using new tools to automate traditional methods of analysis, topic modelling is more controversial because it is an allegedly non-directed, hypothesis-free form of analysis. One does not use topic modelling software to search for a given theme or topic in a collection of works but rather the critic runs the algorithm over the corpus and sees what terms group together.[3] Skeptics argue that this method transforms the critic from exegetical ponderer or directed inquisitor into something akin to a beachcomber seeking fortune with a metal detector. Furthermore, the algorithmic complexity of topic modelling, with its basis in Bayesian statistical modelling,[4] is unfamiliar to most literary scholars. Topic modelling thereby becomes something of a theoretical and methodological black box, and critics are left to ponder the linguistic and symbolic artifacts that the model generates.

From a technical perspective, topic modelling attempts to "infer the underlying topic structure" (Blei "Modeling" 10) of a collection of documents.[5] The model typically uses a form of Bayesian probabilistic inference known as Latent Dirichlet allocation in order to assign words to topics and thereby topics to documents. In the most general of terms, where familiar forms of probability model a phenomenon (such as the rolling of a die), Bayesian probability attempts to model states of knowledge or belief; it is also unique in its capacity to cope with uncertainty in probability. Within a Bayesian framework, a topic is formally defined as a "distribution over a fixed vocabulary" (Blei "Probabilistic" 78), and the topic modelling process attempts to identify how topics are distributed in a corpus. David Blei explains:

the goal of topic modeling is to automatically discover the topics from a collection of documents. The documents themselves are observed [in the sense of being known data], while the topic structure—the topics, per-document topic distributions, and the per-document per-word topic assignments—is *hidden structure*. The central computational problem for topic modeling is to use the observed documents to infer the hidden topic structure. This can be thought of as "reversing" the generative process—what is the hidden structure that likely generated the observed collection? ("Probabilistic" 78; emphasis original)

The language of "discover" and "hidden structure" can be misleading, suggesting that these topics are *there* in the documents and not observations or interpretations of the texts. This controversial point brings into sharp relief the simultaneous technical, hermeneutical, and exegetical implications of topic modelling. Yet such figurative language is simply Blei's means of explaining the relationship between the documents, the known elements, and the unknown topical structure. The task of the topic modelling algorithm is to analyze the known elements in order to "discover," or more accurately, assert, the hidden topical structure.

From a hermeneutical perspective, topic modelling practices precisely the forms of paraphrase that Davey critiques. In its attendance to content over form, "The movement . . . is towards paraphrase—paraphrase of the culture and paraphrase of the literature" (Davey 3). Topic modelling's hypothesis, that documents exhibit some identifiable topic, risks advancing the same pernicious assumptions involved in paraphrase that Cleanth Brooks warns against; namely, that texts "constitute a 'statement' of some sort, the statement being true or false" (196). Indeed, the model assigns a percentage score for a given topic in a particular document as a measure of the degree to which the document exhibits that statement. Topic modelling therefore leaves itself open to Davey's critique of paraphrase as it "extract[s] . . . the paraphrasable content and throws away the form," and in this sense the "critical process produced by these assumptions is reductive" (3). In fact, topic modelling goes further than reducing a text to declared themes by actually reducing the text to its mere linguistic elements.

To better demonstrate the utility and hermeneutic value of this form of paraphrase for Canadian literary criticism, I have engaged in a topic modelling analysis of two major Canadian literary journals. I have collected digitized editions of every issue of the journals *Canadian Literature/ Littérature canadienne* and *Studies in Canadian Literature/Études en littérature canadienne (SCL/ÉLC)* and analyzed these collections using the topic modelling software MALLET.[6] Topic modelling these two exemplary

journals provides a meaningful topography of the field of Canadian literary scholarship that can be employed to identify subtle themes in both scholarly discourse and literary production. These journals are particularly important starting points for distant reading of the field of Canadian literature, as an analysis of their content provides a critical view of the changing shape of Canadian literary scholarship over the past sixty years. *Canadian Literature* began as a journal in 1959 and *Studies in Canadian Literature* was launched in 1976, so it is expected that the journals' varied histories would be reflected in their topics. For instance, a temporally organized topic model of these journals would provide a means of comparing trends within a corpus across time periods. While it is beyond the scope of this article, topic modelling enables readers to trace the frequency and usage of a critical term or discourse to identify its shifting registers across journals and time periods. Furthermore, topic modelling also enables readers to identify individual texts that are central to a literary field but whose influence has gone unnoticed in subsequent scholarship. Finally, my analysis of *Canadian Literature* and *SCL/ÉLC* will identify unlikely connections between authors/ texts as well as previously unobserved thematic and formal connections between texts that suggest new groupings of genre organized along thematic and aesthetic lines.

One substantial difference between this topic modelling experiment and the type of paraphrase that Davey critiques is that I am analyzing the field of scholarly work while Davey excoriates the paraphrasing of literary works themselves, particularly as those acts of paraphrase purport to identify "'our imaginative life,' . . . 'national being,' and . . . 'cultural history'" (2). I am not engaged in a topic modelling analysis of the entirety of Canadian literature (yet), but rather of a particular corpus of the field of scholarly analysis commenting upon that literature. My project might well be considered as *paraphrasing the paraphrase* and therefore might also be considered annoyingly "extra-literary" rather than "anti-literary" (2). Yet this act of critical paraphrase does engage Davey's thesis, as the division between literary culture and its criticism is always porous—particularly in Canada, where literary culture is so dependent on academic and government institutional support. Canadian literature is shaped in its dialogue with, interpretation of, and resistance to criticism of that literature; indeed, critics shape what they purport to interpret. My goal is therefore not to necessarily prove Davey wrong so much as to say that while Davey found thematic searching problematic in critical writing, I find it helpful in meta-critical

writing, particularly as digital forms of analysis blur the distinctions between paraphrase and close reading.

Finally, my decision to paraphrase the paraphrase is also motivated by the feasibility of topic modelling literary journals. Articles in *Canadian Literature* and *SCL/ÉLC* are available in digital format and can be converted into a format suited to topic modelling. Digitizing the entirety of the field of Canadian literary production is not only practically impossible but also would require arbitrary, yet binding, decisions about which texts to include within the borders of the canon. Such a paraphrase of the literature itself would be motivated more by the impossible desire for completeness, to be able to view the entire field at once, and thereby settle the question of interpretation by making some final decision about Canadian literature's actual meaning. In place of that project of Potemkin completeness, my paraphrasing the paraphrase is wilfully arbitrary and selective, yet this arbitrariness is both an unavoidable limitation of the project as well as a necessary condition for asserting the exegetical dimension of digital forms of reading and interpretation.

Two major objections to DH work are that it transforms exegetical discovery and critical analysis into a degraded scientism and that it naturalizes an analytic framework that is deeply indebted to forms of neoliberal and corporate rationalism. Stanley Fish rebukes DH methods precisely because they are "dictated by the capability of the tool. . . . Because the patterns are undetectable, you don't know in advance what they are and you cannot begin your computer-aided search . . . in a motivated—that is, interpretively directed—way" (n. pag.). Fish argues that where traditional forms of analysis are motivated by an interpretive hypothesis, digital forms rely merely on what patterns the tool detects; as such, interpretation devolves into pattern detection. However, in Fish's traditional form of exegesis, his analysis does not begin with the "interpretively directed" series of questions but rather with the act of reading. The digital pattern detection that he critiques is analogous to the act of reading in that both frame their exegetical objects in a manner that enables the critic to pose the interpretive questions that constitute the critical act. The first step in both traditional and digital forms of analysis is to read the text, and therefore the difference between the two modes of interpretation is one of degree, not kind.

Tom Eyers critiques the method of DH work, arguing that it practices a kind of "abstract objectivity," and is part of a general "proliferation of positivist methods . . . another face of the neo-liberalization and corporatization of the

university" (n. pag.). Eyers worries that DH replaces theoretical and methodological frameworks with "quantitative fireworks" that offer a mirage of "alluring transparency afforded by the sharp technological lenses" (n. pag.). Contrary to Fish's and Eyers' assertions, I argue that DH work need not be an instrumentalist, pseudo-objective process of "running the numbers"; rather, digital tools are decidedly "motivated" and "interpretively directed." In the development and selection of the tool, the critic makes a number of interpretive decisions that affect the results. Furthermore, these tools are embedded in networks of power that structure the way in which knowledge is constructed. Indeed, selecting a particular tool requires justifying one framework of interpretation out of a number of possible choices.[7] As such, where traditional forms of literary interpretation often occlude the "frame of analysis," the novelty of digital humanities modes of reading renders the analytic framework all the more explicit.

Both Fish and Eyers fall prey to what Alan Liu identifies as the "fallacy that there are immaculately separate human and machinic orders, each with an ontological, epistemological, and pragmatic purity that allows it to be brought into a knowable methodological relation with the other" (416). For Liu, hermeneutics does not occur at the border between machine and interpretation—a dyad that keeps the purity of the machine intact—but rather pervades all DH critique such that the ideological content and textuality of the machine, the algorithm, and the tool are foregrounded and rendered sites of interpretation. This runs contrary to Fish's and Eyers' suggestions that DH work transforms exegesis into a degraded kind of scientism that affirms some banal thesis. In fact, the opposite hypothesis appears true: DH work brings into view the forms of power, symbolic capital, and representation of the machine, corroding its alleged instrumentality and neutrality in order to render visible its ideological frames and mechanisms of operation.[8]

In place of this machinic fallacy that conceives of the text and the machine as "immaculately separate," I follow Tanya Clement in her use of the notion of "differential reading" as a kind of dual-focused lens of interpretation suitable to topic modelling and DH more generally. Clement argues that differential reading "positions close and distant reading practices as both subjective and objective methodologies" (n. pag.), where the DH critic oscillates between the computer-assisted forms of reading that treat texts as objects and the close reading practices that reframe objective data as subjective interpretation. Within this methodology the critic moves

dialectically between close and distant reading, reading the singular passage against the grain of the corpus and vice versa. Susan Brown describes this as "[w]orking at the gap between humanities research questions and digital humanities development," which "allows digital tools and research results to emerge from a dialectical relationship, allowing the research process to change in concert with the production of new modes of engaging in research" (218). Brown's notion of "[w]orking at the gap" employs humanistic inquiry to defamiliarize the digital tool by rendering it textual. Differential reading enables close and distant, human and machine, forms of reading to reframe one another dialogically while contending with the seeming incommensurability of both as a condition of interpretation. As such, differential reading identifies both the limitations of the tool as well as the exegetical weaknesses of close reading and paraphrase.

In what follows, I model a differential reading of the two journals under consideration in order to demonstrate how topic modelling can uniquely identify meaningful links between texts that would go unobserved with a traditional close reading practice. First, I compare the topic models of *Canadian Literature* against those of *SCL/ÉLC* to identify differences between the two journals. The most dominant topics in the respective journals may indicate the unstated editorial and ideological frameworks and guidelines of these journals.[9] Secondly, I combine the two journals into one corpus and compare the various models of the combined journals, asking what differences emerge when comparing a smaller topic model to a larger model. Finally, I use the topic models themselves as starting points for close reading of particular articles to assess the capacity of topic modelling to offer grounds for unique observations of a literary field.

Comparing the ten most dominant topics in forty-topic models of *Canadian Literature* and *SCL/ÉLC* reveals a number of almost universal literary topics with little identification of Canada as organizing space or term (indeed, the term "Canada" is not present in either model) (see Table 1). A number of terms are shared between the most dominant topics of both journals: "world," "time," "human," "place," "sense," "mind," "man"; these might be thought of as a kind of basic vocabulary of literariness itself.[10] Where *SCL/ÉLC*'s second-most dominant topic links notions of "literary," "culture," and "writing" with "nation," "discourse," "theory," and "politics," no such explicitly national and cultural concerns emerge in *Canadian Literature*, where the closest approximation is in the language of "history," "writing," "words," and "power."

Table 1

The ten most dominant topics in *Canadian Literature* and *SCL/ÉLC* using a forty-topic model.

Topic	Canadian Literature	SCL/ÉLC
1	life man make kind men find things book day good literature war human thing give love world people mind	world time place nature sense human reality experience vision art words things real present mind man meaning word power
2	time world sense art image past life nature fact reality experience point meaning human tradition individual order made dream	literature literary cultural culture writing social critical history english political contemporary role national nation criticism texts discourse theory politics
3	history language family toronto culture writing mother political english place works critical words past subject present community white power	character fiction story characters time life fact toronto past part events future scene lives instance figures david short contrast
4	book writers literary writing years published time literature university public people long part national made writer small great history	narrative identity story readers past read suggests subject desire experience relationship process reading york act difference argues position notes
5	woman part end notes body collection figure women complex letter margaret opening clear simply michael final water forces bear	toronto years century made published american time british women author early letters country great young bentley public states moodie
6	story stories fiction children young characters american world read people time back home david play life hero black place	day white life water home death night head man earth dark dead house city long light back eyes river
7	death man love review vision land book nature life place world sun theme literary present imagination earth images romantic	life father love man woman mother death narrator men young family child wife desire children women order society husband
8	french social society century review english british man historical god author indian study london west novels art modern political	people writing book kind write things writer read writers back lot time thing story english stories place good sense
9	story narrative stories book life women readers social identity cultural collection reading fiction characters narrator home lives author father	language body subject words space female writing symbolic voice dance photograph bodies speaking order women theory flow trans feminist
10	poem poetry words page lines poetic line verse language god water sound word voice atwood images back read landscape	poem poetry speaker poetic smith lines line love stanza scott katie verse crawford page alfred poetics persona max pratt

Ordering the topics according to their presence in the corpus (moving from more to less prevalent topics) roughly aligns with a gradual shift from universalist notions of the literary to more precise and interesting groupings of terms specific to the field and the journal. *Canadian Literature* is predominantly concerned with what might be described as the more formal dimensions of textuality, with topics exhibiting words such as "fiction," "literary," "present," "images," "romantic," "identity," "modern," and "political." *SCL/ÉLC* moves far more rapidly from notions of "literary" and "cultural" to specific images of "day," "dead," "light," "body," and "dance." The last topics for both journals in Table 1 are concerned with verse, yet the topics begin to reveal the specific way in which poetry has been analyzed in the respective journals: Atwood is dominant in *Canadian Literature*, whereas essays in *SCL/ÉLC* are more attentive to Crawford's *Malcolm's Katie* and the work of F. R. Scott. To move from these cursory observations to a more meaningful differential reading of the field, I trace a key term in the model back to the articles themselves.

Combining the two journals provides an opportunity to compare the individual models with a synthesized model that reads both journals as one corpus, arguably providing a more comprehensive view of the literary terrain. With a selection of eighty-five topics, Table 2 shows the ten most dominant topics and their corresponding probability of the combined journals:

Table 2
The ten most dominant topics of both journals in an eighty-five-topic model.

Probability	Topic Contents
0.88179	time words sense point language experience meaning present part order reality nature kind structure process writes works act literary
0.60057	place life world past time home land sense landscape living city back space memory great change toronto myth modern
0.58315	good make matter free found made full century day personal half nature makes long find reason true original case
0.55607	narrative ed subject writing reading suggests discourse read argues voice narrator texts representation process order desire position narratives means
0.54824	world art man human vision artist image eye mind fact give hand images snow eyes imagination early appears reality

Table 2 continues on next page

Table 2, continued

The ten most dominant topics of both journals in an eighty-five-topic model.

Probability	Topic Contents
0.54799	toronto london public york letter thomas figure marriage friend family success ottawa popular private james middle mary heart career
0.5449	story life characters fiction father stories mother child narrator young children character family narrative lives novels readers death person
0.52523	love death woman house man book night life earth poem light voice sun face fire men image head eyes
0.45362	history political social culture cultural historical people world power community society american past colonial politics important role century post
0.44536	life man english great john country french men american society people social time early war british george history fact

The second topic's attention to "place," "home," "land," "landscape," "space," "toronto," and "myth" suggests that these articles engage Northrop Frye's infamous question of the relationship between Canadian identity and space. Based on the dominance of this topic (as represented by its probability), we could posit that either Frye's intuition—that Canadian writers are concerned with the experience of alienation from the surrounding space—is correct or that he continues to frame debates in the field. A temporally organized topic model, one which compares the dominant topics by year or decade, might offer a historical periodization by revealing whether Frye's influence has waned in the decades subsequent to his writing. Regardless, the topic model has therefore provided new questions with which to return to the corpus; the next step is to move from the topics to the articles themselves to understand the particular composition of each topic.

The presence of the term "modern" in the second topic suggests intriguing possibilities concerning the function of modernism and modernity in shaping this broader concern with space. "modern" did not appear in the forty-topic model of *SCL/ÉLC* and is only present in the eighth-most dominant topic in the *Canadian Literature* model. The elevation of this term to the second topic of the combined model renders it a point of discursive confluence between the journals. While not all instances of the use of the term "modern" refer to aesthetic modernism, there may be evidence of a possible reframing of Canadian modernism as decidedly spatial. To what extent, for instance, does this

combining of attention to domestic and rural spaces alongside questions of the modern accord with Glenn Willmott's thesis that Canadian modernist texts tend to deconstruct the movement from country to city (152), or Dean Irvine's argument that Canadian modernism is characterized by A. J. M. Smith's notion of "eclectic detachment" (9)? Tracing this thread to the articles themselves[11] reveals a recurrence of the words "land," "landscape," "space," and "memory," all engaged with questions of the modern. A few key quotations from these articles[12] reveal a repeated concern with modernity as it is staged in the sites of rural Newfoundland, the historic and contemporary prairies, the creeks of Saskatchewan, the rivers of Japan, and the Trans-Canada highway. Indeed, Kristen Warder begins her article with reference to Robert Wardaugh's recent question "When is the prairie?"—a question that is both an acknowledgement of and rejoinder to Frye's question "Where is here?" Warder's claim that Shane Rhodes' writing is "dismantling the facile binary of rural pastoralism and urban modernity" (8) supports Willmott's thesis and reveals a spatial modernist dimension in Rhodes' writing. Kathy Mezei's argument that "[o]ut of this bricolage of elements and carefully contrived dissonance [Anne] Wilkinson constructs a dwelling place for her poetic imagination" (164) together with her observation that Wilkinson "speaks continually of her psychic and physical dislocation" (173) may link A. J. M. Smith's notion of "eclectic detachment" to Frye's concern with space. In addition to providing a counter-corpus demonstrating unexpected links between a diverse range of texts, this move from distance to close reading (and even closer when we return to the novels and poems themselves armed with new interpretive frameworks) and reading these spatial articulations of the modern alongside each critic's engagements with the aesthetic and political dimensions of modernism provides new insight into Canadian articulations of what Chana Kronfeld terms "marginal modernisms."[13] A differential reading of this topic reveals new stagings of the articulation of modernism in Canada in the rural sites of Newfoundland, Saskatchewan, and the Prairies. Furthermore, there are also unexpected transnational connections that recast these spaces not as degraded facsimiles of Europe but as linked via a kind of affective chain of equivalencies with the forests of Japan. By returning to the texts themselves armed with these new questions, the interested critic could identify alternative routes by which these marginal modernisms have circulated across Canada and transnationally.

In addition to comparing the topics to the articles themselves, changing the number of topics produces new models and thereby further grounds for differential readings. In a one-hundred-topic model of the two journals, topic

twenty is about Davey himself: his name is a keyword in this topic alongside others such as "Bowering," "Frye," "Prairie," "Vancouver," "Science," "Tish," "Ostenso," "Macphail," and "Leseur." It is encouraging that the topic modelling algorithm managed to identify a "Tish" topic and the mention of Ostenso, Macphail, and Leseur provoke an analysis of Davey, his contemporaries, and these earlier poets. Tempering that enthusiasm, however, the topic also includes the keywords "Roche," "Jalna," and "Whiteoak," suggesting a link between the largely forgotten Mazo de la Roche's *Jalna* novels and the aforementioned Tish poets; this is more likely mere coincidence than evidence of a meaningful pattern. Of course, this is precisely the function of differential reading: to move from the distant reading of the topic to investigate whether the topic presents a new and insightful frame for understanding the texts themselves or whether its grouping of texts and terms is little more than coincidental. In this case, the latter is likely the case, particularly given that a slight reduction in the number of topics to eighty replaces the references to *Jalna* with "Mouré," "Garde," "Olson," and "Body," all of which suggest far more provocative possibilities for a differential reading of the Tish poets.

To return to my original series of questions, while topic modelling may not get us out from under Davey's shadow entirely, it does demonstrate the manner in which DH work destabilizes received notions of reading as well as the relation between close reading and paraphrase. Davey's critique of paraphrase and its attendant heresies is less convincing when viewed through the framework of differential reading which conceives of close and distant reading as dialectically informative. Topic modelling forms of paraphrase generate alternate corpora which will inform directed close readings across a discourse that would have been impossible with traditional forms of reading and research. If we follow Russell Brown's argument that "[c]ultural generalizations are only heuristic tools" (668) and that "[a] statement of theme can be thought of as the creation of a metonym that enables discussions of texts and permits useful comparisons between . . . texts" (673), then the value of topic modelling as a kind of reflexive paraphrase becomes apparent. Topic modelling is *willfully reductive* in the ways that Davey warns against, yet it reduces only in order to foreground themes that would go otherwise unnoticed by an individual reader. This is not the classic form of thematicism and paraphrase that aimed to excavate a national identity or in which "the artist speaks, unconsciously or consciously, for the group" (Davey 2). As such, topic modelling is neither indebted to nor a symptom of the "Arnoldian humanism" (2) that Davey sees in these acts of thematic paraphrase. Topic modelling paraphrases cultural

production not to assert a stable cultural or national identity but to open and re-examine received notions of identity, culture, and textuality. Paraphrase thereby becomes a tool to raise new critical questions while avoiding the banal conclusions of thematicism. A differential reading practice of these themes requires that critics read the theme against the individual texts in order to assess the efficacy of the model. Furthermore, differential reading enables a reflection on methodological practices that subjects the algorithm and the digital tool to a hermeneutics of suspicion.

In this respect, the digital humanities *is*, as its critics suggest, a Trojan horse. However, it need not be a Trojan horse in the sense that Fish and Eyers imagine, nor for what Len Findlay calls "the neoliberal arts" (n. pag.), wherein the humanities become a mere training ground in the knowledge economy. Instead, digital humanities forms of reading can be a Trojan horse where a new form of humanistic critique is smuggled into the processes of computation and positivism, where the ontological certitude of data is challenged, and the roles of human inquisitor and computational tool are re-examined. I agree with Diana Brydon's assessment that the "humanities need a new humanism" (47), one that responds to the limitations of humanist inquiry by not abandoning the categories of humanism or the humanities but by subjecting them to the very processes of critique they engender; the digital humanities may provide an inroad into just such a reflexive form of humanism. Furthermore, in many respects, Canadian literature is an effective field for "working at the gap" between the digital and the literary, particularly as it nurtures a recurring crisis of its own existence and is regularly worrying the terms of its own enunciation. Susan Brown argues that "[m]arginality, liminality, and hybridity— all of which concern gaps and unstable affiliations with identity and community—provide valuable vantage points for engagement with shifting technologies" (219); in this sense, Canadian literary study's marginality and continuous state of crisis makes it a productive place to assess the crisis of the humanities, digital and otherwise.

NOTES

1 Russell Brown's useful analysis of the critique of thematicism in Canadian literature in the 1970s and 1980s shows that while Davey's text is exemplary, his work is one of many contemporaneous critiques. See, for instance, Barry Cameron and Michael Dixon's "Mandatory Subversive Manifesto: Canadian Criticism vs. Literary Criticism" in a special issue of *Studies in Canadian Literature.*
2 Of course the digital humanities is a diverse and contested field that includes text mining, markup, corpora analyses, and computationally-assisted forms of reading (to name only a few of its methods).

3 This is a somewhat necessary oversimplification. The critic does not merely run the algorithm, but must make a number of critical decisions that will determine the structure and content of their model. The first crucial decision is the particular algorithm to use in the creation of their model: Latent Dirichlet allocation is the most common, but other options include Probabilistic Latent Sampling Indexing, Non-Negative Matrix Factorization, and Gibbs Sampling. Secondly, the critic must choose the number of topics that they want the model to produce.

4 For a comprehensible and clear introduction to Bayesian probability and "data science" in general see Grus (2015).

5 For a reasonably clear introduction to topic modelling, see Blei's "Probabilistic Topic Models" (2012). The video and slides of Blei's lecture at the Machine Learning Summer School (2012) and his lecture at Google Tech Talks provide lucid explanations of topic modelling for a general technical audience. For an explanation of topic modelling from a humanities perspective see Ted Underwood's excellent "Topic modeling made just simple enough."

6 I am deeply indebted to the editors and staff of both journals for making these archives available online and thereby enabling my research. It would be impossible to engage in this form of research were these journals not already digitized and easy to access. I am hopeful that this project and subsequent distant reading projects will encourage other journals to digitize their archives and make them available.

7 Perhaps the most significant technical and methodological decision for topic modelling is the user's selection of the number of topics that the modeller should generate. The user does not merely request that the software investigate a corpora and return the implicit topics of that corpora. Rather, the user inputs the number of topics the model should generate; this choice has a profound effect on the content of the topics. Choosing too few topics results in a small number of overly general topics, while choosing too many results in a large number of overly specific topics. While there are mathematical methods for assessing the most appropriate number of topics for a given corpus, there is also a degree to which the user selects the number of topics that will generate the output that best suits the user's hypothesis. The user massages the results in order to reduce the amount of topic leak (where similar terms appear in multiple topics) as well as chimera topics (where multiple topics are incorrectly grouped together into one). These seemingly banal choices that maximize topic coherence deny the possibility that these anomalies may reflect a meaningful pattern within the corpora. Indeed, the fetishizing of coherence and consistency (see Newman, Bonilla, and Buntine) within topic modelling work may foreclose the interpretation of trends and topics that appear incoherent but are in fact meaningful textual characteristics. Identifying such algorithmic constraints of topic modelling and their theoretical and methodological implications is a necessary condition of differential reading as it provides a framework for understanding how technical conditions shape digital reading practices. In addition to these technical decisions, the selection of corpora also has methodological implications. *Canadian Literature* and *Studies in Canadian Literature* are my chosen journals to study because they are both the most representative of Canadian literature as a field as well as the easiest to digitize. Other journals such as *English Studies in Canada*, *Canadian Poetry, The Canadian Review of Comparative Literature*, *ARIEL, Queen's Quarterly*, and *University of Toronto Quarterly* were not singularly focused on the field of Canadian literature or were simply too difficult to digitize. *Essays on Canadian Writing* is an obvious exclusion from my project; however, digital editions of this journal prior to 1993 are not readily available. Each of these limitations indicates both the experimental and incomplete nature of this project while gesturing towards future work that may provide a deeper understanding of the field.

8 For examples of this kind of work see Bethany Nowviskie's "Digital Humanities in the Anthropocene" (2015), Roopika Risam's "Beyond the Margins: Intersectionality and the Digital Humanities" (2015), Tara McPherson's "Why Are the Digital Humanities So White? or Thinking the Histories of Race and Computation" (2012), Johanna Drucker's "Humanities Approaches to Graphical Display" (2011), Miriam Posner's "What's Next: The Radical, Unrealized Potential of the Digital Humanities" (2015), and the work featured on the *Postcolonial Digital Humanities* website.

9 Of course these models do not account for the manner in which editorial agency steers the content and theme of the journals. Nor do they account for the practical limits of production, changes in the peer review system, where authors choose to submit their work, and other important dimensions of how a journal comes to exhibit particular thematic content. A compelling extension of my research would organize topic models along editorial lines in an effort to identify changes in topical content resulting from changes in editors and to read those models against stated editorial guidelines.

10 Topic models rely on a list of stopwords: words that are ubiquitous functional words with a structural, rather than a lexical, function (e.g., "the," "and," "it,") should be excluded from the model. One might argue in favour of adding this generic literary language to the stopword list in order to foreground the more unique and substantive terms of the field, particularly given the prevalence of terms like "story," "writing," "literary," and so forth. A comparative reading of a model that includes these terms alongside one which excludes them might produce compelling results; however, I have included these terms in my topic models in order to provide as accurate a depiction of the two journals as possible.

11 Some of the articles for which this topic is the most represented include Kristen Warder's "(Un)Settling the Prairies: Queering Regionalist Literature and the Prairie Social Landscape in Shane Rhodes' *The Wireless Room*," J. N. Nodelman's "Gabrielle Roy's *La route d'Altamont* and Canadian Highway Narrative," Wanda Campbell's "Every Sea-Surrounded Hour: The Margin in Maritime Poetry," Kathy Mezei's "Home, the Unhomely, and the Everyday in Anne Wilkinson," Gregory Maillet's "In a Boat on the River Nowhere Writing Home: The Spiritual Poetic of Tim Lilburn," Lyle P. Weis' "Bipolar Paths of Desire: D.C. Scott's Poetic and Narrative Structures," and Issue 71 of *Canadian Literature*.

12 Warder: "Rhodes challenges the nostalgic identity of the prairies as a place still dominated by a traditional way of life by dismantling the facile binary of rural pastoralism and urban modernity; this rural area is not as pristine or traditional as it first appears. Even in the tranquil countryside seemingly 'laid out in pre-history,' modern life intrudes, engendering a rebirth in the landscape" (8); Nodelman: "This sense that a coherent enunciation of some kind lies off to the side of the road, just a little out of reach from inside the car, is of paramount importance regarding how narrative reworkings help produce the articulated spaces of modern highway travel" (221); Mezei: "Like other modernist women poets, however, her poetry at times appears to waver between the language and subject of domesticity expressed through the everyday and the decorative and the mythic and symbolic as well as the abstract, historic, and monumental. Her ambivalence towards the domestic and the everyday reflects that of modernism" (160); Maillet: "This breathless awareness is not modern, though, and section two concludes with 'Could be Feb. 14, 1244, Could be North Japan,' another statement of the affinity between the creeks and rivers of Saskatchewan and medieval Japan, as Lilburn becomes a poet of Zen paradox, 'under the ground, moving / my arms to the stars'" (9-10).

13 Curiously, if one shifts the focus of reading by increasing the number of topics to one hundred, the most dominant topic becomes an amalgam of the two most dominant topics in the eighty-five-topic model and the term "modern" does not appear until the eighty-

third topic. This significant movement of a single term, from the second to eighty-third most-represented topic, indicates how selecting the number of topics has a substantial effect on the results of the model and the subsequent hermeneutic possibilities that the model presents. This confirms the intuition that topic modelling does not provide evidence or data for an argument but rather a particular perspective by which to playfully mis-read a corpus. A differential reading that wished to trace this thread would compare the articles which best represent this topic with those that best represent the "modern" topic of the eighty-five-topic model to understand this shift in terms.

WORKS CITED

Blei, David. "Modeling *Science*." *Columbia University*. Web. 15 Nov. 2015.

—. "Probabilistic Topic Models." *Communications of the ACM* 55.4 (2012): 77-84. Web. 20 Nov. 2015.

Brooks, Cleanth. *The Well Wrought Urn: Studies in the Structure of Poetry*. New York: Harcourt Brace, 1947. Print.

Brown, Russell Morton. "The Practice and Theory of Canadian Thematic Criticism: A Reconsideration." *University of Toronto Quarterly* 70.2 (2001): 653-89. Print.

Brown, Susan. "Don't Mind the Gap: Evolving Digital Modes of Scholarly Production Across the Digital-Humanities Divide." *Retooling the Humanities: The Culture of Research in Canadian Universities*. Ed. Smaro Kamboureli and Daniel Coleman. Edmonton: U of Alberta P, 2011. 203-32. Print.

Brydon, Diana. "Do the Humanities Need a New Humanism?" *Retooling the Humanities: The Culture of Research in Canadian Universities*. Ed. Smaro Kamboureli and Daniel Coleman. Edmonton: U of Alberta P, 2011. 233-62. Print.

Cameron, Barry, and Michael Dixon. "Mandatory Subversive Manifesto: Canadian Criticism vs. Literary Criticism." *Studies in Canadian Literature/Études en littérature canadienne* 2.2 (1977): 137-45. Print.

Campbell, Wanda. "'Every Sea-Surrounded Hour': The Margin in Maritime Poetry." *Studies in Canadian Literature/Études en littérature canadienne* 33.2 (2008): 151-70. Print.

Canadian Literature / Littérature canadienne. Archives. *Canadian Literature: A Quarterly of Criticism and Review*. https://canlit.ca/full-issue/back-issues/. Web. 20 Dec. 2016.

Clement, Tanya. "Text Analysis, Data Mining, and Visualizations in Literary Scholarship." *Literary Study in the Digital Age: An Evolving Anthology*. Ed. Kenneth M. Price and Ray Siemens. *MLA Commons*, 2013. Web. 15 Oct. 2015.

Davey, Frank. "Surviving the Paraphrase." *Canadian Literature/littérature canadienne* 70 (1976): 5-13. Print.

—. *Surviving the Paraphrase*. Winnipeg: Turnstone, 1983. Print.

Drucker, Johanna. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5.1 (2011): n. pag. Web. 20 Dec. 2016.

Eyers, Tom. "The Perils of the 'Digital Humanities': New Positivisms and the Fate of Literary Theory." *Postmodern Culture*. 23.2 (2013): n. pag. Web. 1 Dec. 2015.

Findlay, Len. "Academics on Empty? Academic Freedom, Collegial Governance, and the Failure of Academic Leadership in Canada." *Athabasca University*. Web. 5 Dec. 2015.

Fish, Stanley. "Digital Humanities and the Transcending of Morality." Opinionator*, New York Times*. 9 Jan. 2012. Web. 10 Sept. 2015.

Godard, Barbara. "Crawford's Fairy Tales." *Studies in Canadian Literature/Études en littérature canadienne* 4.1 (1979): 109-38. Print.

Grus, Joel. *Data Science From Scratch*. Sebastopol, CA: O'Reilly Media, 2015. Print.

Irvine, Dean. Introduction. *The Canadian Modernists Meet*. Ed. Irvine. 1–17. Ottawa: U of Ottawa P, 2005. Print.

Kamboureli, Smaro. "Introduction: Shifting the Ground of a Discipline: Emergence and Canadian Literary Studies in Canada." *Shifting the Ground of Canadian Literary Studies*. Ed. Kamboureli and Robert Zacharias. Waterloo: Wilfrid Laurier UP, 2012. 1-36. Print.

Kronfeld, Chana. *On the Margins of Modernism*. Oakland: U of California P, 1996. Print.

Liu, Alan. "The Meaning of the Digital Humanities" *PMLA* 128.2 (2013): 409-23. Print.

Maillet, Gregory. "In a Boat on the River Nowhere Writing Home: The Spiritual Poetic of Tim Lilburn." *Studies in Canadian Literature/Études en littérature canadienne* 30.1 (2005): 226-44. Print.

McPherson, Tara. "Why Are the Digital Humanities So White? Or Thinking the Histories of Race and Computation." *Debates in the Digital Humanities*. Ed. Matthew K. Gold. Minneapolis: U of Minnesota P, 2012. 139-60. Print.

Mezei, Kathy. "'And We Are Homesick Still': Home, the Unhomely, and the Everyday in Anne Wilkinson." *Studies in Canadian Literature/Études en littérature canadienne* 30.1 (2005): 160-80. Print.

Mimno, David et al. "Optimizing Semantic Coherence in Topic Models." *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (2011): 262-72. Web. 5 Oct. 2015.

Moretti, Franco. "Conjectures on World Literature." *New Left Review* 1 (2000): 54-68. Web. 1 Oct. 2015.

Newman, David, Edwin V. Bonilla, and Wray Buntine. "Improving Topic Coherence with Regularized Topic Models." *Advances in Neural Information Processing Systems*. 2011. Web. 21 Nov. 2015.

Nodelman, J. N. "Gabrielle Roy's *La route d'Altamont* and *Canadian Highway Narrative*." *Studies in Canadian Literature/Études en littérature canadienne* 33.1 (2008): 211-28. Print.

Nowviskie, Bethany. "Digital Humanities in the Anthropocene." *Digital Scholarship in the Humanities*. 30.1 (2015): 4-15. Print.

Posner, Miriam. "What's Next: The Radical, Unrealized Potential of the Digital Humanities." *Miriam Posner Blog*. 27 July 2015. Web. 20 Dec. 2016.

Risam, Roopika. "Beyond the Margins: Intersectionality and the Digital Humanities." *Digital Humanities Quarterly* 9.2 (2015): n. pag. Web. 20 Dec. 2016.

Steyvers, Mark, and Tom Griffiths. "Probabilistic Topic Models." *Handbook of Latent Semantic Analysis*. Ed. Thomas K. Landauer et al. Routledge, 2014. 424-40. Print.

Studies in Canadian Literature / Études en littérature canadienne. Archives. *SCL/ÉLC*. https://journals.lib.unb.ca/index.php/SCL/issue/archive. Web. 20 Dec. 2016.

Szeman, Imre. "The Persistence of the Nation: Interdisciplinarity and Canadian Literary Criticism." *Essays on Canadian Writing* 65 (1998): 16-37. Print.

Underwood, Ted. "Topic modeling made just simple enough." *The Stone and the Shell*. 7 Apr. 2012. Web. Oct. 15 2015.

Warder, Kristen. "(Un)Settling the Prairies: Queering Regionalist Literature and the Prairie Social Landscape in Shane Rhodes's *The Wireless Room*." *Studies in Canadian Literature/Études en littérature canadienne* 32.2 (2007): 115-31. Print.

Weis, Lyle P. "Bipolar Paths of Desire: D.C. Scott's Poetic and Narrative Structures." *Studies in Canadian Literature/Études en littérature canadienne* 12.1 (1987): 35-52. Print.

Willmott, Glenn. *Unreal Country*. Montreal: McGill-Queen's UP, 2002. Print.

Wood, Brent. "No-Man's Land: Mythic Crisis in Gwendolyn MacEwen's *The T.E. Lawrence Poems*." *Studies in Canadian Literature/Études en littérature canadienne* 29.2 (2004): 141-62. Print.