

Hannah McGregor

Nicholas van Orden

University of Alberta

30 April 2015

Remediation and the Development of Modernist Forms in *The Western Home Monthly*¹

Introduction

In May of 1901, *The Western Home Monthly* printed a short anecdote, sandwiched between a note to their readers and an advertisement for Lumiere Dry Plates, entitled "That Nothing be Wasted." "There seems to be no limit," it begins, "to the fun to be had with a phonograph. One man writes that he has a perfect record of the barking of his dog, and the dog enjoys hearing the record play as much as his master does" (8). It continues:

This reminds me of the story now going the rounds in the newspapers—perhaps you've heard it—concerning the economics of the pork packing industry in Chicago. Every part and parcel of Mr. Pig is made use of with the greatest of care and ingenuity. ... Even the dying squeals, heretofore wasted on the midnight air, are now preserved on phonograph cylinders. It may be only a story, but it shows the modern tendency of manufacturers to utilize every by-product. (8)

¹ Special thanks first to our collaborators, Harvey Quamen and Matt Bouchard, to the excellent people at University of Alberta Libraries Digital Initiatives, and to EMiC UA. This paper was made possible through the generous support of the Social Sciences and Humanities Research Council of Canada. All images from *The Western Home Monthly* appear courtesy of Peel's Prairie Provinces (peel.library.ualberta.ca), a digital initiative of the University of Alberta Libraries.

With its self-conscious references to media innovation and the circulation of information, and its simultaneous suspicion and embracing of the modern, this anecdote powerfully evokes the magazine's ongoing relation to modernity, one that registers at the levels of cultural critique, self-conscious technological sophistication, and a surprisingly timely remediation of the new media landscape of the moment.

This uneasy combination of the modern and the anti-modern, the conservative and the novel, is characteristic of the category into which *The Western Home Monthly* most obviously fits—the middlebrow. As John Guillory defines it, “[m]iddlebrow culture is the ambivalent mediation of high culture within the field of the mass cultural” (87). Mass cultural is precisely what this magazine wanted to be—a note to the readers in May 1901 claimed “It is our ambition to make *The Western Home Monthly* the representative publication of the great middle classes” (8)—and arguably what it became by the 1930s. Published out of Winnipeg, it never achieved the circulation of New York or London-based magazines, but it certainly made its mark. By Sep. 26, 1932, *Time* magazine could describe it as the most widely-circulating household magazine in Canada, the top of the “Big Five” and the only one *not* published in Toronto (“Press”). With a comparatively modest circulation of 180,000, decidedly non-cosmopolitan origins, and an ongoing interest in the agricultural, it might seem a far stretch to discuss *The Western Home Monthly* as a modern magazine, let alone a modernist magazine. And yet, modernism developed on the pages of this magazine as much as on the pages of any other.

In *Modernism in the Magazines*, Robert Scholes and Clifford Wulfman note that “[m]odernism happened in the magazines all right, but it didn't happen only in the little ones” (41). Scholes and Wulfman were echoed two years later by Andrew Thacker, who, in the introduction to Volume II of *The Oxford Critical and Cultural History of Modernist Magazines*

(focusing on North American periodical production between 1894 and 1960), argues that “the story of how modernism appeared in magazines in North America is a multifaceted one” characterized by “‘cross-fertilization,’ shown in the multiple networks of connection between and across distinct categories of publication” (20). However, the only chapter on Canadian magazines in that volume focuses on English Canadian little magazines, an inclusion that risks reifying these avant garde publications as the sole site of modernist periodical production. We will argue here that if modernism did happen on the pages of a variety of periodicals, traversing national borders as well as the borders of brow and genre, then it is necessary to explore how modernism was happening on the pages of a middlebrow magazine like *The Western Home Monthly*. As scholars of the middlebrow have asserted in recent years, “the middlebrow is a ‘hybrid form’” that often bridged the gaps between “the sentimental and the political, realism and experimentation, and modernity and the Victorian” (Sullivan and Blanch 4). Patrick Collier’s work on the aesthetic continuities between “the image-collages of the *Illustrated London News* and the shaped and gathered fragments of *The Waste Land*” similarly refuses the narrative of radical rupture between the Victorian middlebrow and the modernist avant garde (“Imperial/Modernist” 511). And yet continuities can be remarkably hard to track, especially across the vast archives of long-term serial publication. Fortunately, emergent methods of digitization and distant reading open up periodical archives to new kinds of analysis.

This chapter will proceed in four parts. First, we will articulate our argument for reading *The Western Home Monthly* through the lens of modernism by exploring the links that have been drawn recently between modernism, the middlebrow, and new media studies. Second, we will outline the method through which *The Western Home Monthly* was digitized and the tools we used in our analysis. The third section will demonstrate how our distant reading methods helped

us to better understand the formal dimensions of the magazine, particularly in terms of the influence of advertising and increasing formal fragmentation. In our fourth section we will analyse a single issue of *The Western Home Monthly*, showing how a combination of distant and close readings helps us to understand the place of an agrarian middlebrow magazine within the transnational and intermedial phenomenon of modernist culture.

I. “MODERNISM is a much-abused word”

Even a glancing survey of recent scholarship in modernist studies reveals that the term is still much-debated, and any attempt to summarize these debates exceeds the purview of this chapter. We instead turn to Douglas Mao and Rebecca L. Walkowitz’s account of “The New Modernist Studies,” a field that they define in terms of the expansion of the traditional temporal, geographical, and methodological boundaries of modernism.² Attention to spaces beyond the familiar cosmopolitan nodes of modernist production (New York, London, Paris; 737), to the fluid boundaries “between high art and popular forms of culture” (738), and to the central role that mass media played in the circulation of modernist thought (742) has served to push back against a narrow critical lens that would associate modernism with *Blast* but not *Cosmopolitan*, with *Preview* but not *The Western Home Monthly*.

As the borders of what constitutes modernist cultural production have shifted, so too has our understanding of modernism. As Collier explains in *Modernism on Fleet Street*, modernism has traditionally been defined “as a rejection of mass culture in all its forms” (2), with mass print media like newspapers serving as a perfect foil: “newspapers use a simplified language,

² Our own research emerges from the collaborative Editing Modernism in Canada project (<http://editingmodernism.ca/>), a SSHRC Strategic Knowledge Cluster project that focused on the interwoven concerns of New Modernist Studies, particularly the transnational reach of modernism beyond cosmopolitan centres.

modernism various forms of radical, new, and complex language; newspapers are instantly consumable products of commodity culture, modernism opaque and difficult to consume, and prone to entertain anti-capitalist ideologies from socialism to fascism” (4). This version of modernism, however, is deliberately reductive and relies on the suppression of contradictory literary histories (5) and the denial of other possible relationships between modernism and mass media. The newspaper could just as easily be an analogy for “modernist style: the imagist poetics of the fragment grew up alongside the visually and textually fragmented modern newspaper page” (4). *The Western Home Monthly*, with its roots in the aesthetics of Victorian newspaper printing (its founders had in fact been newspaper printers before turning to household magazines), is a valuable site for exploring the shifting meaning of the fragment.

Other scholars have gone further in refusing to see mass media as modernism’s other, arguing to the contrary that modernism was deeply engaged with mass media. Sean Latham explains that “[t]he writers and artists we now call modernists” were responding to “the ‘technologizing of information’”—he is borrowing Friedrich Kittler’s phrase here—via mass media “in aesthetically innovative ways ranging from the invention of the unconscious to experiments with stream-of-consciousness writing” (1). Jessica Pressman similarly draws on new media scholars (including Kittler, Lev Manovich, and Marshal McLuhan) to define “the late decades of the nineteenth century and the early decades of the twentieth century” as “the classical period of our contemporary technological age,” the period in which “[o]ur modern mediatized consciousness emerged” (4). Modernism, she argues, is deeply engaged with this technological and media shift, “employ[ing] the media of its time to reform and refashion older literary practices” such that “making it new ... is also about using new media” as well as “renovating the past *through* media” (3-4). The formal dimensions of modernism are thus

analogous to the shifting discourse networks of the period, networks that are no longer “based in oral and analog modes of communication” but in the synchronicity and autonomy of technologized information (5). Her conclusion is to the point: “modernism is centrally about media” (4).

While Kittler mentions film, typewriters, and phonographs as examples of the “technologizing of information,” Latham adds magazines to this list, insisting alongside Ann Ardis that we can productively read magazines “as new media technologies” (1). If we can think of magazines as new media, Latham continues, then it is valuable to read them using the vocabulary of new media, particularly in terms of their *affordance* and *emergence*, by which he means the “action possibilities” (1) of “complex systems capable of producing meaning through the unplanned and even unexpected interaction of their components” (3). Latham’s argument, to which we will return in more detail below, is essentially that magazines differ from other print media in their non-linear and intermedial layout; they contain an almost hypertextual quality “that invites the reader to construct connections” within individual issues and across different iterations of the same magazine (2). Rather than treating the magazine as a series of discrete textual objects that can be mined for content—an approach that has dominated in the past, though it has been largely repudiated by media studies—Latham treats it as a system composed of diverse and overlapping units, a “chaotic mixture of advertisements, images, texts, headers, captions, maps, and photographs” (3) that readers can navigate in diverse and unpredictable ways. While Latham’s focus here is on the modernist magazine, these features are arguably even *more* evident in the *middlebrow* magazine. As Sullivan and Blanch argue, the periodical can be seen as “the exemplary middlebrow artifact—the ‘composite text’—demanding and constructing a range of reading practices in response to its discursive collage and complexity” (6).

James Mussell's work on the periodical form enriches this analysis, focusing not only on the complex diversity within individual issues and the qualities of emergence that result, but also on the dynamic of seriality that simultaneously connects and differentiates different iterations of the same periodical:

No single issue exists in isolation, but instead is haunted by the larger serial of which it is a part. This larger serial structure is invoked through the repetition of certain formal features, issue after issue. It insists on a formal continuity, repeated from the past and projected onwards into the future, providing a mediating framework whose purpose is to reconcile difference and present it in a form already known to readers. (6)

This mediating framework, as Mussell describes it, can be structural (recurring features or columns), formal (typeface and page layout), and material (paper quality and print technologies). The identity of a particular periodical is thus abstracted in that it exists outside of any individual iteration, and decidedly material in that it is instantiated within the magazine issues as physical objects (8). If the new medium of the magazine is one of the ways in which modernism was making itself new, that newness could not only be a radical break with the past, but must also rely on the familiarity and repetition that might be associated with more conservative forms. It is in this interplay between sameness and difference, tradition and innovation, that *The Western Home Monthly's* ambivalent relation to modernism becomes clear.

The Western Home Monthly was published by the Stovel Printing Company of Winnipeg, Manitoba, in Western Canada, between 1899 and 1932. Stovel Printing was founded in 1889 by the Stovel brothers, who had moved from Mount Forest, Ontario to Winnipeg four years prior. Stovel quickly became the largest English-language printer and publisher in the city, and was at

the forefront of printing in Western Canada. A 1931 promotional publication produced by Stovel Printing lists their range of print productions, which includes photographs, maps, and catalogues, as well as a variety of magazines, including eight trade publications, *The Nor'West Farmer and Farm & Home*, and *The Western Home Monthly*, which they describe as “a National magazine for the urban home” (Stovel 37). In addition to standing at the forefront of print culture in Western Canada, the magazine also published work by an international collection of authors whose writing introduced many of the aesthetic innovations of realism and modernism to a mainstream readership.

It is possible to explore the shifting, complex relationship of the old and the new, the conservative and the innovative, at the level of form and content across the thirty three year run of the magazine. The early issues, with their ornate mastheads, anonymous editorial voice, and straightforward four-column format, look distinctly Victorian, a sensibility still in evidence as late as 1919. This seeming conservatism is belied by the magazine’s deliberate material and rhetorical incorporation of new media. The remediation of the early twentieth century’s new media landscape within *The Western Home Monthly* is central to our argument. Reading the magazine not as a collection of different texts but as a coherent media object allows us to trace a genealogy of emergent modernist forms across a digitized archive of 24,170 pages and 33,099,536 words—an archive that provides a singular opportunity and a host of methodological challenges.

While at times the magazine’s interest in modernism is linguistically explicit, such as the June 1925 editorial feature on “Modernism” by the Rev. John Mackay, Principal of Manitoba

College in Winnipeg (13)³—from which the title of this section is taken—or the March 1932 article on Canadian painters working in Montparnasse (Dickie 18-19), we are not interested in documenting individual occurrences of words and phrases already generally associated with modernism. Instead we begin from the argument that modernism happened on the pages of magazines as the new medium negotiated the affordances of the periodical and its relation to the new media landscape of the modern era. We are interested in the intersections between new media, advertising, and the fragment across the hypermediated and remediated pages of *The Western Home Monthly*'s digitized archive. To emphasize the differences between the cultural institution that is the print magazine titled *The Western Home Monthly* and the digitized version of this archive, we refer to the magazine as *The Western Home Monthly* and to the digitized version as the *WHM*.

II. Working with the Digital *WHM*

The Western Home Monthly's digitization was made possible through a partnership between Editing Modernism in Canada at the University of Alberta (EMiC UA), the research group with which we are affiliated; the Manitoba Legislative Library, where the magazine's archive is housed; the Digital Initiatives office at the University of Alberta Libraries; and Peel's Prairie Provinces, a U of A Libraries digital collection of pre-1953 materials from Canada's West. The resulting resource is a full-colour digital edition of all extant issues of the magazine; it will be the first household magazine hosted on the Peel's Prairie Provinces website. The digital *WHM*

³ The "modernism" to which Mackay refers is actually one side of a theological debate raging in the Presbyterian church in the 1920s and 30s that distinguished between modernism and fundamentalism as two possible interpretations of scripture; many of the textual occurrences of "modernism" and "modernist" within the pages of *The Western Home Monthly* hold this meaning, indicating why a straightforward search for particular keywords is not the most productive way to engage with the magazine.

contains 348 issues, including: two issues from each of 1901 and 1903, ten issues from 1904, and every issue between 1905 and 1932 except for January 1916, September 1919, and March 1922. A special illustrated issue titled *The 1914 War* supplements the twelve issues from 1915.

Peel's Prairie Provinces outsourced the digitization of *The Western Home Monthly* to Backstage Library Works of Provo, Utah. Backstage provides "preservation-quality digital images" as well as optical character recognition (OCR) and archival-standard metadata markup ("Digitization"). The digital *WHM* that Backstage produced includes 348 PDF and METS files (one for each issue) and 24,170 jp2 and ALTO files (one of each for every scanned page). In total, the digital *WHM* consists of 66GB of data across more than 50,000 files. METS/ALTO, the markup schema used to encode the magazine, is an archival standard originally developed for the digitization of newspapers. The Library of Congress's Metadata Encoding and Transmission Standard (METS) is an XML-based schema used to describe the hierarchical structure of digital objects, to preserve the names and locations of the various files that comprise an object, and to maintain digitization metadata pertinent to that object. The METS schema is roughly analogous to the bindings of a print publication—METS files ensure the structural integrity of a digital object. Whereas METS data determines the structure of the digital object, the Analyzed Layout and Text Object XML schema (ALTO) contains metadata relating to the stylistic elements of the original print material, such as font descriptions, location data (indicating the position of each word or image on the page), and the original textual content. ALTO files are roughly analogous to detailed maps of each digitized page within a digital object. The ALTO schema is also maintained by The Library of Congress. Used together, METS/ALTO metadata is a powerful tool for the arrangement, maintenance, and preservation of digitized objects.

The digital *WHM* METS and ALTO files contain enormous quantities of detailed information (*Figure 1*). In addition to headers that note crucial namespace, measurement, and OCR information, the ALTO files for each scanned page identify individual text blocks (such as columns, article titles, and image captions), individual text lines within each text block, and individual strings of content (words) within each text line. Each block, line, and string is distinguished by a unique identification number. The horizontal position (“HPOS”), vertical position (“VPOS”), width, and height of every element are also noted. Information about the OCR reliability of every character within each word is included within each string tag. The spaces between words are represented by unique “SP” identifiers, with the VPOS, HPOS, and width of each space faithfully recorded. In addition to text blocks, the ALTO files label each illustration and advertisement using “ComposedBlock” identifiers. The size and position of each illustration is indicated. Every ComposedBlock “TYPE=‘Advertisement,’” like the standard text blocks, contains information about all of the lines, strings, and pieces of textual content within each ad. Images within advertisements are not identified by this ALTO XML schema. In total, information about more than 56,000 advertisements, 33,000 illustrations, 24,000 pages, and 33,000,000 words is contained in more than 314,000,000 lines of XML code.

Our first challenge was to extract all of the textual Content from the ALTO files so that we could run topic modeling and text-mining programs on the content of the entire collection. A simple PHP script and some command line wizardry, courtesy EMiC UA collaborator Matt Bouchard, produced 24,170 plain text files comprised of the ALTO files’ “Content” attribute. Further work on the command line, and with Perl (Matt Bouchard again; see Appendix A), concatenated these plain text files into appropriately-named issues and digests (digests being all of the issues from a calendar year). We next combined all of the ALTO files into issues and

digests to facilitate broad analyses of the meta-data (for example, to calculate the number of advertisements or illustrations in a given year). The possibilities for working with these files are seemingly endless. In addition to the work discussed below, location metadata (VPOS and HPOS, which are measured to the tenth of a millimeter) could be used to track the shifting position of advertisements across the issues and digests; changing ratios of text, image, and white-space might be used to construct arguments about design aesthetics; and information about typefaces and font sizes could lead to visualizations of the mediating frameworks that Mussell argues are central to the serial identity of periodicals. As the remainder of this paper will demonstrate, even a comparatively simple use of the affordances of METS/ALTO can reveal valuable new information.

Our goals in this first examination of the digital *WHM* include an analysis of the relationships between emerging media technologies and the development of modernist formal characteristics such as textual fragmentation, collage, and narrative disruption. We considered questions such as, “How did the relationship between modernism and media develop on the pages of *The Western Home Monthly*?” “How can digital reading methods help us discover, understand, and represent this relationship?” and, “What do parallels between the remediation of the magazine (into a digital archive) and remediation *within* the magazine suggest about the interplay between media technologies, representations of visual and textual data, and digital humanities-inflected analyses of literary and cultural artefacts?” To focus this project and to leverage the most basic affordances of the METS/ALTO files, we used text mining data and topic models to juxtapose the appearance of new media in the magazine as a whole to the appearance of those media in advertisements. We also extracted simple statistical data from the ALTO files in order to visualize trends and shifts in the magazine’s publication, layout, and

general form. *Figure 2* compares the total number of pages, advertisements, and illustrations across all 348 issues of the digital *WHM*. Many discussions about the development of new printing technologies, local events in Winnipeg, and world-historical trends emerged from these simple graphical representations. For example, is the precipitous decline in advertising between 1929 and 1930 a reflection of the stock market crash in October 1929? Was the sudden drop in illustrations around 1916 caused by the fire that ravaged the Stovel Printing Company's main plant? Is the steady rise in page numbers during the first years of the twentieth century the result of increased circulation, changes in printing technologies, or easier access to materials (both raw and intellectual)? Many similar questions will hopefully be explored by other scholars with other interests and expertise.

Our decisions to use MALLET for topic modeling and RapidMiner for text mining were influenced by a range of factors, including our experience using the tools, the size of the data, and the research questions that motivated our analysis. Based on a set of user-defined parameters, MALLET (Machine Learning for Language Toolkit) analyzes massive quantities of data and produces "topics"—lists of words that are statistically likely to occur in proximity to each other within a document or across a set of documents. For example, we would expect the words "western," "home," and "monthly" to feature prominently in MALLET's reading of the digital *WHM*—which they do, as *Figure 3* demonstrates. Topics produced by MALLET reveal within minutes patterns and themes across the content of the *WHM* that would have taken months to discern through traditional reading practices (if we had been able to discover them at all). As Robert K. Nelson explains in his excellent "Mining the *Dispatch*" project, "Topic modeling... allows us to step back from individual documents and look at larger patterns among all the documents, to practice not close but *distant reading*, to borrow Franco Moretti's

memorable phrase.” *Figure 3*, for example, gives us a glimpse into the patterns of repetition and difference that characterize the serial identity of *The Western Home Monthly*. The four most statistically relevant topics, which maintain their dominance throughout the 348 issues, focus on the magazine’s identity as well as overarching concerns with time (year/s, time, day), gender (man, woman, girl), labour (make/made), nation, and home. Along the bottom of *Figure 3*, the rising and falling topics of new media emerge and, in some cases, disappear—phonographs, radio, and then film—all of which cluster within a vocabulary of advertising that includes the brand names Kotex, Palmolive, Edison, Heintzman, Crisco, and Quaker. This, in visual form, represents the interplay between the traditional and the modern, between the familiar structures of the magazine and the introduction of new media through the discourse of advertising. The juxtaposition of general trends and specific words produced by MALLETT is echoed in the text-mining data produced by RapidMiner.

RapidMiner is a powerful analytics, text-, and data-mining platform used extensively across a wide-range of industries. Our use of RapidMiner barely scratched the surface of its text-mining capabilities; version 5.3 (freely available through Sourceforge) was more than adequate for our needs. Using the “Process Documents From Files” operator, we broke each of the *WHM* digest files into individual “tokens” (words), transformed the tokens into lower-case letters, removed common stopwords and tokens shorter than three letters or longer than five hundred, and applied the built-in Porter stemming algorithm (which works iteratively through the material to reduce words to their root forms—i.e. “working,” and “worker” to the root “work”). We copied the RapidMiner output into an enormous spreadsheet for ease of access and manipulation. This same process was repeated on the digest files containing only text from the *WHM*’s advertisements. Powerful though spreadsheets are, scanning 2.4 million cells of data to discover

patterns and trends is unmanageable. Another EMiC UA collaborator, Dr. Harvey Quamen, wrote a wonderful PHP script to extract data from our spreadsheets and topic models in a fashion suitable for input into the powerful R statistical software environment (Appendix B). Many of the images presented here were produced using ggplot2, a data visualization package for R. The simpler line and bar graphs were made using Windows Excel's build-in graphing functionality. Combining several reading strategies allowed us to analyse the enormous quantity of material contained in the *WHM*. As Nelson notes, “[t]opic modeling and other distant reading methods are most valuable not when they allow us to see patterns that we can easily explain but when they reveal patterns that we can't, patterns that surprise us and that prompt interesting and useful research questions.” Distant reading tools can also be used to supplement traditional close or attentive reading practices. Topic models might suggest keywords deserving of particular attention; for example, knowing the importance of gendered terms such as man, woman, and girl across the *WHM* corpus alters the significance of these terms when they are read within the context of particular editorials, serialized literary texts, and advertisements. Similarly, RapidMiner's statistical data reveals the emergence of new technologies such as radio or cinema and suggests specific digests or issues to examine using more traditional methods of scholarly enquiry. Collections of material comprised of several million words can only be parsed effectively using a combination of reading practices. Current objections to distant reading as a legitimate mode of critical engagement often fail to consider not only the historical contingency of all reading, but also the benefits of hybrid analyses that embrace distant readings, close readings, and everything in between.

III. Remediation

We have so far argued that the relationship between magazines and modernism is based in materiality and mediation; as Lisa Gitelman argues, the “material properties” of media “do (literally and figuratively) *matter*” (10). Of course, it is not only matter that matters to Gitelman. She is equally interested in what she calls “protocols,” or the “vast clutter of normative rules and default conditions, which gather and adhere like a nebulous array around a technological nucleus” (7). Extrapolating from her examples, a study of magazines as a form of new media might include a consideration of seriality, subscription rates and policies, advertising, circulation networks, reading habits, and print technologies. These features combined constitute the newness and the modernness of the magazine, making it of interest to scholars invested in complicating and expanding our understanding of modernism. But how might we arrive at such a portrait of a historical magazine? And how is this task inflected by digitization and digital or distant reading methods?

A common concern about the study of print culture in digital form is that the process of digital remediation fundamentally shifts the nature of the object of study. The digitized magazine page is not, and cannot be treated as, a proxy or replacement for the material page; what we can say about the digitized page, using digital tools, is not equivalent to what we could say about the material page. We by no means deny this argument. Instead, like Mussell, we’re interested in how digitizing periodicals leads to an opportunity to rethink print differently (3). “What appears to be a deficit,” Mussell concludes, “a misrepresentation, in digital resources, is actually difference, introduced through transformation” (17-18). For researchers, the process of working through this transformation can itself be beneficial. In “Remediation as Reading: Digitising *The Western Home Monthly*,” McGregor argues that digitization affords a unique perspective on periodicals, addressing one of the key challenges in periodical studies, that of dealing with vast

and heterogeneous archives. A digitized periodical can be more easily analyzed for the patterns of repetition and difference that characterize the serial text.

Repetition is intrinsic to how periodicals, including magazines and newspapers, establish their identities across years of shifting content. The repeated features of *The Western Home Monthly*—monthly columns, familiar writers, typefaces and column widths—establish that “formal continuity” that Mussel describes as “a mediating framework whose purpose is to reconcile difference” (6). The digital remediation of the magazine is surprisingly revealing of the magazine’s mediating frameworks; in the terms of Jay David Bolter and Richard Grusin’s theory of remediation, digital remediation emphasizes the relationship between *immediacy* and *hypermediacy*. Immediacy, or “the transparent presentation of the real” (Bolter and Grusin 21) might take the form of those monthly features that become invisible through their familiarity, or the protocols that Gitelman argues successful media must render invisible “in favor of attention to the phenomena, ‘the content’” (Gitelman 6). Conversely, hypermediacy, or “the enjoyment of the opacity of media themselves” (Bolter and Grusin 21), is constituted in moments when the mediated nature of the magazine is foregrounded—often in the process of trying to train readers to use this new media object and, as a result, to see it as immediate once again.

The July 1919 edition of the editorial titled “A Chat With Our Readers” provides a strong example of the interplay between immediacy and hypermediacy. This “Chat” section varies across the many issues in which it appears between 1909 and 1921, sometimes including letters from readers, other times directly addressing readers on topics the editors situate as particularly important. In the July 1919 issue they focus on “The Value of Advertising”:

[M]any of the things we count today as necessities or simple luxuries could not be made and sold at their reasonable prices except as advertising has created a broad market for them, making millions of sales at little prices and little profits.

And so you owe very much to advertising. You owe much to the people of yesterday, who have read and been influenced by past advertising, and so have made possible the economies and varieties, and wide distribution of merchandise that you enjoy. (“A Chat” 1)

This page in July 1919 (*Figure 4*) looks much like the corresponding page in the preceding and following issues. Its formal features constitute those invisible protocols or mediating frameworks that disappear in the process of reading—from the title and volume information in the header, to the page layout, to the blank subscription form at the bottom. These protocols—issue numbering, subscription rates, and layout—can be read as examples of Bolter and Grusin’s immediacy, the naturalized and seemingly transparent presentation of formal media characteristics. In this issue, however, the editors’ explanation about how these dimensions of the magazine as a medium are meant to be read shift the magazine’s advertisements from immediacy to hypermediacy. In the process of trying to teach readers how to use the magazine properly, the editors apply pressure to the medium, rendering it temporarily opaque rather than transparent. As a result it becomes clear to us, for a moment, how strongly the magazine was invested in and moved by advertising. It also gives us an opportunity to investigate what kind of impacts an ever-increasing emphasis on advertising as a medium may have had on the magazine.

The Western Home Monthly underwent a series of important shifts between 1919 and 1920. Full-colour covers were introduced in January 1920 and the use of colour gradually extending into glossy advertisements throughout the magazine. In this period there was also a

rapid increase in features on, and advertisements for, radios. As *Figure 5* demonstrates, the shift away from the phonograph and toward the radio as the dominant new technology occurred rapidly, and was largely driven by advertising, the space in which readers' relationships to new technologies was most actively negotiated. These gestures toward modernity (colour printing, new forms of media) continue to share space with signs of the old; the sheer range of printing techniques at work, and of technologies alluded to or directly mediated within the magazine's pages, make it a fascinating object. The contrasts are often striking, with photo-collages of "Motherhood of Many Lands" (May 1929) followed by woodcut-illustrated advertisements for "Canada's Pulp and Paper Industry" (June 1929). The catalyst for much of this change was advertising, and the mosaic of different printing techniques served in themselves as advertisements for Stovel Printing's ability to keep pace with Winnipeg as the city moved forward into modernity. The advertisements in the magazines are also at the forefront of its modernization—they are the first colour features, the first to incorporate glossy paper, and the first to introduce new media into their hypermediated aesthetics.

Another impact that advertising had on the magazine becomes apparent in a major shift in its layout. At the beginning of 1919 the magazine followed the same layout protocols it had since 1899—editorial features were presented sequentially and in their entirety, extra spaces filled up with miscellaneous items including random photographs and anecdotes likely pulled from other publications. Beginning in October 1919, the magazine's layout underwent a dramatic shift, offering the first few pages of each feature and then continuing the piece on later pages, with the distinctly modernist outcome of fragmented narratives and collage-like spreads. *Figure 6* demonstrates the marked rise of this technique, charted through the magazine's use of the phrase "continued on," which signaled the discontinuous content scattered across an issue. While this

dynamic remained prevalent through the rest of the magazine's run, *Figure 7* offers a clear example of the collage-like pages that resulted from this new non-linear layout. These facing pages show how the accumulation of multiple fragments onto a single page frees up more full-page space for advertisements. The resulting collage brings together a travel narrative about hunting on the Pacific Coast, a whimsical story of fairies bringing seeds to life, an adventure tale set on the Dakota Plains, and an article advising women on comfortable modern dress.

The impact of this shift in page layout was an increased emphasis on what Latham calls "emergence." Emergence, according to N. Katherine Hayles, "is any behavior or property that cannot be found in either a system's individual components or their additive properties, but that arises... from the interaction of a system's components" (qtd. in Latham 3). Rather than reading magazines as "stable object[s] consisting of the same words and images laid out the same way across multiple copies," Latham argues that the magazine's "wide affordances"—how readers can interact with it in unpredictable ways—"generate... multiple different paths that individual readers can then take through a text," paths that "produce the phenomenon of emergence: the creation of meanings and behaviors generated by the multiple ways in which textons [strings of signs] can interact with one another" (4). He concludes that "[t]he modern magazine, in short, is a distinct media form with its own history and trajectory that helped shape not only the aesthetics of literary modernism but the intellectual foundations of hypertext and other distinctive media practice we perhaps mistakenly call 'new'" (4). The wide affordances of non-linear layouts, the resulting emergence, and the aesthetic modernism of *The Western Home Monthly* appear to have been motivated not by philosophical or political attempts to undermine the hegemony of traditional linear narratives, but by the more practical and material concerns of advertising and print technologies.

In Latham's terms, distant reading can help produce different possible "paths" through the digital *WHM*, paths which highlight the meanings and behaviours generated by the multiple ways in which "textons" such as pages, illustrations, and advertisements interact with one another. *Figure 8* shows paths representing the number of advertisements, illustrations, pages, and words (divided by one thousand) annually across the *WHM*. The importance of advertising is immediately evident in this image—ads dominate nearly every year. In general, the number of advertisements, illustrations, pages, and words published change in direct relation to each other, rising and falling roughly in unison. As noted above, world historical events (such the stock market crash) and changes in printing technologies (such as the incorporation of colour images) can be mapped closely onto these paths, dramatic events often precipitating sudden shifts. Combining close and distant reading practices is crucial to interpreting this information and discerning these connections. For example, careful analysis of several issues of the *WHM* around 1925 reveals that the increase in advertising between 1924 and 1925 was likely related to the incorporation of colour advertisements throughout the magazine. Starting with two-tone colours early in the year, advertisements in 1925 quickly developed into full-colour glossy inserts featuring everything from flooring and furniture to bacon and biscuits (*Figure 9*). During that same period the number of pages and words increased in proportion to the increase in advertisements but the number of illustrations stagnated, suggesting that space previously allotted to illustrations was being absorbed by the increase in colour advertising. These closely linked metrics in the mid-1920s are particularly interesting because this period represents arguably the most formally settled period in the magazine's history. Although ads, illustrations, pages, and words were generally directly related in earlier years, their relation was often not proportional—the massive increase in advertising between 1907 and 1908 was accompanied by

only small increases in page numbers and illustrations. From 1922 to roughly 1930, however, the form of the magazine remained relatively stable—*Figure 8* shows all four paths following similar trajectories. This period corresponds almost exactly with the dramatic rise in the fragmentation of the texts within *The Western Home Monthly*—use of the “continued on...” convention increased from seventy four in 1921 to a high of 383 in 1929. This combination of relative formal stability and increased fragmentation suggests that this period in the magazine echoed the development of typically modernist literary forms, particularly narrative fragmentation and the mixing of genres and voices. Many prominent modernist texts featuring these same characteristics were published during the same period, including: Eliot’s *The Waste Land* (1920), Joyce’s *Ulysses* (1922), Woolf’s *To The Lighthouse* (1927), and Faulkner’s *As I Lay Dying* (1930). While it is impossible to claim any direct link between *The Western Home Monthly* and these various writers, strong formal parallels suggest lines of continuity between the middlebrow magazine and literary modernism.

We argued above that the digital remediation of *The Western Home Monthly* puts us in a better position to understand the interplay of immediacy and hypermediacy on the magazine’s pages, and we have shown how this dimension of the magazine was constantly pushed by advertising and the development of new print technologies. As Bolter and Grusin explain, the interplay between immediacy and hypermediacy, those “twin preoccupations of contemporary media” (21), are intertwined with the process of remediation, itself “a defining characteristic” of new media (45). They define remediation as “the representation of one medium in another” (45), and the dialectic between old and new media that results (50). Modern magazines like *The Western Home Monthly* are dense with remediation, incorporating old media like newspapers, woodcuts, and paintings, as well as new media like advertising, photography, radio, and film.

Despite the methodological challenges that an enormous archive like this one poses to us, it also offers an unprecedented opportunity to track the dynamics of remediation across pages, issues, and years. In the process, it allows us to formulate more clearly an argument about how the dynamic processes of remediation led to the adoption of distinctively modernist forms on the pages of a middlebrow magazine. In the final section of this paper, we will turn to a single magazine issue as a case study for reading modernist forms against the dynamic of remediation in *The Western Home Monthly*.

IV. Remediating New and Old Media

In the third instalment of Martha Ostenso's novel *Wild Geese*, printed in the October 1925 issue of *The Western Home Monthly*, the four young protagonists gather for an illicit party:

They sat about in the sitting room for an hour listening to Mark's phonograph. Judith had heard one in the home of an Icelander, but it had a horn and had not produced the alluring music that she listened to now. Her eyes grew dark and absent as she let her emotions drift with the spirit of the dance. A waltz played, and she feared that she would cry before it came to an end.

Lind and Mark danced a little, and Jude watched them enraptured. It was all so new to her, and yet it seemed the thing to which she belonged. (9)

For the sheltered farm girl Judith, the phonograph stands in for all things modern: youth, sexual freedom, art, and the newness of technology and experience. The phonograph, and the world that it signifies beyond the farm and its deprivations, is one of a range of images that signal Judith's temptation away from tradition and familial duty and into an illicit premarital relationship with the city-dwelling Sven.

As old-fashioned as this scene might seem, *Wild Geese* was far from an old-fashioned novel in 1925. Faye Hammill has convincingly demonstrated that Canadian cultural commentators of the 1920s classed the novel as “objectionably realistic” in a period when “realism” tended to be equated “with explicitness about sex” (86). While the dominant popular literary mode of the time was sentimental romance, *Wild Geese* is rife with themes that refuse the sentimental, including “illegitimate sexual relationships, jealousy, cruelty, and the violent overthrowing of patriarchal authority” (87). Stigmatized in its time, the novel has since been consecrated as an early work of prairie realism that helped pave the way for the intertwined movements of realism and modernism in Canadian literature.

Understanding *Wild Geese* as a novel wedded to neither nostalgia nor tradition, the question remains why the phonograph plays such a central role. As *Figure 5* above demonstrates, in 1925 the phonograph was far from a dominant technology, having fallen off drastically in appearances in the magazine as of 1920. The dominant new medium at the time was radio. The scatterplot in *Figure 10* draws on topic modelling results to show with what other terms radio frequently co-occurred. *Figure 10* tells us at least two things: first, that the rise of interest in radio happened quite rapidly, and second, that this rise coincided so neatly with the serialization of three particular novels that the word “radio” ended up clustering with character names. Topics 18, 3, and 0 correspond, respectively, to the novels *Wild Geese* by Martha Ostenso (August 1925-January 1926), *The Flame of Courage* by George Gibbs (February-July 1926), and *Mantrap* by Sinclair Lewis (August-December 1926). In the October 1925 issue of *The Western Home Monthly* the word phonograph appears seven times while the word radio appears 187 times. Beyond its single instance in the novel, “phonograph” appears in two other items: an article entitled “Advances in Radio,” in which it stands in as the medium of the past whose

progression might serve as a predictor for the future of the radio; and an advertisement for J.J.H. Maclean & Co. Limited, a piano and organ manufacturer offering phonographs on clearance. This smattering of references to the phonograph is densely surrounded on all sides by radios—from double-paged colour advertisements to passing notes in the Classifieds, to an abundance of editorial content debating the future of the technology, its appropriate uses and its social impacts. An article comparing Quebec to the rest of Canada lists radio, along with “[j]azz, joy rides, gasoline filling stations fitted out with the pomp of pleasure gardens, [and] cross-word puzzles” as “part of the immense equipment of an overmaterialized civilization” that is corroding our “personal and national potency” (Osborne 13).

In this moment radio is remarkably hypermediated. In their introduction to an article entitled “What a Radio Manufacturer Thinks about Radio Editorials,” the anonymous editorial “we” insists that the era of radios being an experimental, home-made apparatus are long passed; radios have instead become another standard household good, reliably mass produced by corporations (Armour 41a). This overt editorial framing of what radios are and what they mean suggests a moment in which the significance of this new medium is actively contested, a suggestion reinforced by the multiple overlapping and conflicting versions of the radio remediated in this issue alone. A single representative advertisement (*Figure 11*) connects the hypermediacy of the radio to the scene in *Wild Geese* that features Mark and his phonograph.

The two-page full-colour advertisement was new to *The Western Home Monthly* in 1925; it represented an unprecedented level of technological sophistication for the magazine and was among the most expensive advertising spaces they offered. In its combination of text and image, modernity and tradition, this ad for Fada Radios exemplifies almost everything that Richard Ohmann argues is characteristic of modern advertising from the early twentieth century. To

begin, it combines all three of the forms of visual abstraction that Ohmann outlines: the icon, the symbol, and the index (180-84). The icon, the object itself abstracted from the surrounding world, appears in the form of the two radios in the lower left-hand side of the ad. The symbol, or trademark, appears in the form of the grand piano icon that accompanies the slogan “The Grand Piano of the Radio World.” The index, “an image of people, places, or occasions to be somehow associated with the product and its use” (182), effectively ties icon and symbol together, showing how a radio can be like a grand piano. In this stylized illustration, well-dressed men and women sit and stand around an elegantly appointed room observing a concert. The grand piano in this room echoes the symbol below, while the sweeping line that frames the index draws the viewer’s eye back to the brand name, insistently linking the two. The text more clearly explains this link: “To engage artists for the entertainment of one’s friends was once the privilege of only the wealthy. The advent of radio, however, enabled every home to share this pleasure.” The value of modernity is expressed here through an aspirational gesture toward traditional forms of wealth and elegance, conflating democracy with commodity acquisition. At the same time, this image suggests the ambivalence toward modernity that Ohmann argues is typical of advertising of the period (206). While the “now” of the radio is more advanced, more accessible, more available to the reader, the “then” of the elegant gathering is implicitly more desirable. The modern technology that puts that elegance in the reach of the viewer is a sign of the same modernity that makes that elegance a thing of the past.

The similarity between this Fada advertisement and the passage from *Wild Geese* is evident: young people gathered around a scene of music-making in an image that simultaneously harkens to an absent past (the elegant home concert, the newness of the phonograph) and points to the modernity of the present (the new technology of the radio, the rejection of traditional

mores and agrarian lifestyles). Through this consonance of imagery, the outdated medium of the phonograph is drawn into conversation with the contested new medium of the radio, much as the phonograph is used as a parallel for the radio in the magazine's discussions of new technology in the twentieth century. The result is that a seemingly nostalgic scene is pulled into a collage of competing meanings through which multiple possible courses can be charted. The collage-like texture that we discussed above as a property of individual magazine pages post-1919 can thus be read across an entire magazine issue, not only the page. And just as the collage-like page represented a shift to modernist formal aesthetics prompted by print technologies and the dominance of advertisements, here we can see how the aesthetic experimentation at the level of ads and page layout can shift the meaning of literary texts as well. If, as Pressman argues, modernism is invested in "renovating the past *through* media" (4), then the relationship between past and present played out in the remediation of a new medium like the radio constitutes a modernist moment in *The Western Home Monthly*.

There are many such moments to be discovered, and the methods outlined in this paper are key to such discovery. We set out to investigate two questions: what does modernism look like on the pages of *The Western Home Monthly*, and how have digital reading methods helped us to understand it better? Methods of distant reading allow us to navigate the enormity of a resource like the digital *WHM*, finding meaning in its lines of code that in turn lead us back into the magazine itself. As J. Stephen Murphy explains, "visualization is not simply a fast-forward button, doing the work we would otherwise do much more slowly. Visualization also helps researchers direct their research by helping them see relationships among data that would be otherwise obscured" (vii). In our case, modelling the magazine as data has given us a way into an archive that could otherwise be approached only through sampling, allowing us to identify

patterns that in turn have led to answers about how exactly modernism takes shape in *The Western Home Monthly*.

Works Cited

“A Chat With Our Readers: The Value of Advertising.” *The Western Home Monthly* July 1919:

1. Print.

Armour, S.D. “What a Radio Manufacturer Thinks about Radio Editorials.” *The Western Home Monthly* October 1925: 41a. Print.

Bolter, Jay David, and Richard Grusin. *Remediation: Understanding New Media*. Cambridge: MIT Press, 2000. Print.

Northern Electric. “Canada’s Pulp and Paper Industry.” Advertisement. *The Western Home Monthly* June 1929: 43. Print.

Collier, Patrick. “Imperial/Modernist Forms in the *Illustrated London News*.”

Modernism/modernity 19.3 (2012): 487-514. Web. *Project Muse*. 16 Apr. 2015.

---. *Modernism on Fleet Street*. Aldershot: Ashgate Publishing, 2006. Print.

Dickie, Francis. “Montparnasse—Heart of the Artistic World.” *The Western Home Monthly* March 1932: 18-19, 53, 62, 64. Print.

“Digitization Services.” *Backstage Library Works*, 2015. Web. 27 Apr. 2015.

Endres, Kathleen L., and Therese L. Lueck. *Women’s Periodicals in the United States:*

Consumer Magazines. Westport: Greenwood Press, 1995. Web. *Google Books*. 27 Apr. 2015.

Fada Radio. Advertisement. *The Western Home Monthly* October 1925: 44-45. Print.

Gitelman, Lisa. *Always Already New: Media, History, and the Data of Culture*. Cambridge: MIT Press, 2008. Print.

- Guillory, John. "The Ordeal of Middlebrow Culture." Rev. of *The Western Canon: The Books and School of the Ages*, by Harold Bloom. *Transition* 67 (1995): 82-92. Web. *JSTOR*. 25 June 2014.
- Hammill, Faye. "The Sensations of the 1920s: Martha Ostenso's *Wild Geese* and Mazo de la Roche's *Jalna*." *Studies in Canadian Literature* 28.2 (2003): 74-97. Web. 12 July 2012.
- Latham, Sean. "Affordance and Emergence: Magazine as New Media." *What is a Journal? Towards a Theory of Periodical Studies*. MLA Convention 2013, Special Session 384. Web. < www.ru.nl/publish/pages/674401/latham_mla_2013.pdf>. 21 Dec. 2012.
- Mackay, John. "Modernism." *The Western Home Monthly* June 1925: 13, 75-76.
- Mao, Douglas, and Rebecca L. Walkowitz. "The New Modernist Studies." *PMLA* 123.3 (2008): 737-748. Web. *JSTOR*. 2 Mar. 2015.
- McGregor, Hannah. "Remediation as reading: digitising *The Western Home Monthly*." *Archives and Manuscripts* 42.3 (2014): 248-257. Web. *Taylor and Francis Online*. 18 Nov. 2014.
- Murphy, J. Stephen. "Introduction: 'Visualizing Periodical Networks.'" *The Journal of Modern Periodical Studies* 5.1 (2014): iii-xv. Web. *Project Muse*. 9 Mar. 2015.
- Mussell, James. "'In Our Last': the Presence of the Previous in Periodical Form." Ms. Courtesy of the author. Forthcoming in *Victorian Periodicals Review*. 2015. Print.
- Nelson, Robert K. "Introduction." *Mining the Dispatch*. Digital Scholarship Lab, University of Richmond, n.d. Web. 27 Apr. 2015.
- Ohmann, Richard. *Selling Culture: Magazines, Markets, and Class at the Turn of the Century*. London & New York: Verso, 1996. Print.

Osborne, W.F. "Quebec in Relation to Canada." *The Western Home Monthly* October 1925: 13, 69. Print.

Ostenso, Martha. "Wild Geese. Third Installment." *The Western Home Monthly* October 1925: 7-10, 72-79. Print.

Pearson, M. Winston, and William Thompson. "Motherhood of Many Lands." *The Western Home Monthly* May 1929: 12-13, 68-70. Print.

"Press: Maple Leaf Magazines." *Time* 26 Sept. 1932. *Time.com*. Web. 27 Jan. 2014.

Pressman, Jessica. *Digital Modernism: Making it New in New Media*. Oxford Scholarship Online, 2014. Web. 21 Nov. 2014.

Scholes, Robert, and Clifford Wulfman. *Modernism in the Magazines: An Introduction*. New Haven & London: Yale UP, 2010. Print.

Stovel Company. *Historical Outline of the House of Stovel: A Short Historical Sketch Together with Illustrations and Descriptions of the Growth of the Company*. Winnipeg: Stovel, 1931. Web. *Peel's Prairie Provinces*. 27 Apr. 2015.

Sullivan, Melissa, and Sophie Blanch. "Introduction: The Middlebrow—Within or Without Modernism." *Modernist Cultures* 6.1 (2011): 1-17. Web. 2 Mar. 2015.

Thacker, Andrew. "General Introduction: 'Magazines, Magazines, Magazines!'" *The Oxford Critical and Cultural History of Modernist Magazines: Volume II, North American 1894-1960*. Ed. Peter Brooker and Andrew Thacker. Oxford: Oxford UP, 2012. Print.

"That Nothing be Wasted." *The Western Home Monthly* May 1901: 8. Print.

Appendix A

The following command line prompt was used to combine various files throughout the project.

```
cat *.[extension of files to concatenate] >> [name of concatenated file].[extension of concatenated file]
```

The following PHP script was used to extract the text content from the ALTO files.

```
<?php
foreach ($argv as $value) {

$xmlDoc = new DOMDocument();
$xmlDoc->load($value);

$searchNode = $xmlDoc->getElementsByTagName("String");

foreach( $searchNode as $searchNode )
{
    $value_string = $searchNode->getAttribute('CONTENT');
    $file = fopen($value, "a");
    fwrite($file, $value_string.' ');
    fclose($file);
}
}

?>
```

The following command line prompt was used to run the above PHP script across all 24,170 ALTO files. Many thanks to Matt Bouchard for his help developing this prompt.

```
find . -xdev -name "*.xml" -exec php [name of above php script] {} \;
```

The following Perl script was used to combine the individual page text documents into their issues (appropriately named based on their respective directories). Many thanks again to Matt Bouchard for this.

```
#!/usr/bin/perl

use strict;
use warnings;

use File::Find;

find (\&dirs_to_combine, "[path to files]");

sub dirs_to_combine {
if ((-d $_) && ($_ =~ [file name pattern])) {
my $conquer = "cat $File::Find::dir/" . $_ . "/* .txt >> [path to files]" . $_ . ".txt";
`$conquer`;
}
}
```

Appendix B

Dr. Harvey Quamen wrote the following PHP script to extract material from our RapidMiner spreadsheets and convert it into a format suitable for processing in R.

```
<?php

/*****
 *
 *   Harvey Quamen
 *   hquamen@ualberta.ca
 *
 *   This script parses a CSV word frequency file and extracts
 *   frequencies for selected words. It produces a CSV file suitable
 *   for import in R for data visualization.
 *
 *****/

/*****
 *
 *   This is the word frequency file to read; if you use a
 *   different file later, change the name here:
 *****/
$file = '[name of spreadsheet containing RapidMiner data]';

/*****
 *
 *   A couple of settings -- labels for R axes and such.
 *****/
$legend_label = 'Word';
$x_axis = 'Year';
$y_axis = 'Frequency';

//   If the user requested no words, then output a usage message and exit.

if (empty($argv[1])) {
    echo "\nThis script parses a word frequency file and extracts\n";
    echo "each word's frequency to produce a CSV file for import into R.\n\n";
}
```

```

    echo "Usage:\n\n";
    echo "\tphp " . $argv[0] . " word_list [> result_file]\n\n";
    echo "...where word_list is a space-separated list of words to fetch\n";
    echo "frequencies for.\n\n";
    echo "The results will get printed to the screen, but you can choose to\n";
    echo "route the results into a result file with the usual Unix `>` character\n";
    echo "followed by a filename.\n\n";
    exit;
}

//      if we're still here, then we have at least one word
$words = array();
array_shift($argv); //      element 0 is this script's name; axe it
$words = $argv;
//      clean up stray spaces
$words = array_map('trim', $words);

//      open the word frequency CSV file
$handle = fopen($file, 'r');

//      if we couldn't open it, exit with an error
if (!$handle) {
    echo "Could not open '{$file}' for reading.\nExiting.\n";
    exit;
}

//      open the output stream as a file to write to;
//      we'll use PHP's inherent ability to understand
//      the CSV output format.
$csv_output = fopen('php://output', 'w');

//      read the header row; we'll parse dates out of this. Looks like:
//      Word,Total Occurences,Document Occurrences,1901,1903,1904,1905, ...

$header = fgetcsv($handle, 1000);
//      remove the first three columns; we really want only the dates
$header = array_slice($header, 3);

```

```

//      output a new header row for us; variables are defined up top.
//      If any of them contain spaces, e.g., PHP will quote them for us.
fputcsv($csv_output, array($legend_label, $x_axis, $y_axis));

//      Read the data file as a comma-separated file; iterate through
//      the data, extracting yearly frequency counts for the years we
//      want. Each column in the original file generates one point in R,
//      so we need to output one new row for each data point.
//      We'll output this as CSV to the terminal window so that PHP
//      can handle any quoted strings by itself.

while ($row = fgetcsv($handle, 2000)) {
    $word = trim(array_shift($row));
    $total_count = array_shift($row); //      ignore this
    $doc_count = array_shift($row); //      ignore this too
    if (in_array($word, $words)) {
        for ($index = 0; $index < count($row); $index++) {
            $date = (int) $header[$index];
            //      use PHP's inherent ability to write to CSV format
            fputcsv($csv_output, array($word, $date, $row[$index]));
        }
    }
}

fclose($handle);

//      END OF SCRIPT

?>

```

The following PHP script, also from Harvey Quamen, extracts data from Mallet's topic modeling output and converts it into a format suitable for processing in R.

```
<?php
```

```

/*****
*
*   Harvey Quamen
*   hquamen@ualberta.ca
*
*   This script parses a Mallet composition file and extracts
*   composition percentages for selected topics. It produces
*   a CSV file suitable for import in R for data visualization.
*
*****/

/*****
*   This is the Mallet composition file to read; if you use a
*   different file later, change the name here:
*****/
$file = '[name of file]';

/*****
*   A couple of settings -- labels for R axes and such.
*****/
$legend_label = 'Topic';
$x_axis = 'Date';
$y_axis = 'Weight';

/*****
*   More settings -- you might never need to change these.
*   The idea here is to output topic numbers and dates as
*   strings rather than as numbers so R displays them more
*   correctly.
*****/
define('TOPICS_AS_STRINGS', true);

```

```

define('DATES_AS_STRINGS', true);

//      If the user requested no topics, then output a usage message and exit.

if (empty($argv[1])) {
    echo "\nThis script parses a Mallet topic composition text file and extracts\n";
    echo "one topic's composition percentages and sorts them chronologically.\n\n";
    echo "Usage:\n\n";
    echo "\tphp " . $argv[0] . " topic_number [> result_file]\n\n";
    echo "...where topic_number is the numeral corresponding to a topic that has been\n";
    echo "generated by Mallet.\n\n";
    echo "The results will get printed to the screen, but you can choose to\n";
    echo "route the results into a result file with the usual Unix `>` character\n";
    echo "followed by a filename.\n\n";
    exit;
}

//      if we're still here, then we have at least one topic number
$topic_numbers = array();
array_shift($argv); //      element 0 is this script's name; axe it
$topic_numbers = $argv;

//      open the Mallet composition file
$handle = fopen($file, 'r');

//      if we couldn't open it, exit with an error
if (!$handle) {
    echo "Could not open '{$file}' for reading.\nExiting.\n";
    exit;
}

//      open the output stream as a file to write to;
//      we'll use PHP's inherent ability to understand
//      the CSV output format.
$csv_output = fopen('php://output', 'w');

//      read (and abandon) the useless header row

```

```

$header = fgetcsv($handle, 1000, "\t");

//      output a new header row for us; variables are defined up top.
//      If any of them contain spaces, e.g., PHP will quote them for us.
fputcsv($csv_output, array($legend_label, $x_axis, $y_axis));

//      Read the data file as a tab-delimited file; iterate through
//      the data, extracting the parts we want. Remember that Mallet
//      produces data in pairs: topic number followed by tab followed
//      by composition percentage. So when we encounter a topic number
//      we're interested in, also grab the following element which
//      contains the composition percentage.

while ($row = fgetcsv($handle, 2000, "\t")) {
    $row_num = array_shift($row);
    $file = array_shift($row);
    //      ignore results for .DS_Store; Mallet should have ignored it.
    if (strpos($file, '.DS_Store') !== false) continue;
    for ($index = 0; $index < count($row); $index += 2) {
        if (in_array($row[$index], $topic_numbers)) {
            $topic = TOPICS_AS_STRINGS ? "" . $row[$index] . "" : $row[$index];
            $date = DATES_AS_STRINGS ? "" . parseDate($file) . "" :
parseDate($file);
            echo $topic . ' ' . $date . ' ' . (float) $row[$index + 1] . "\n";
        }
    }
}

fclose($handle);

//      END OF SCRIPT

?>

```