Comparative Complementizers in Canadian English: Insights from Early Fiction

Marisa Brook

This paper is posted at Scholarly Commons. http://repository.upenn.edu/pwpl/vol20/iss2/2
For more information, please contact libraryrepository@pobox.upenn.edu.
Comparative Complementizers in Canadian English: Insights from Early Fiction

Abstract
There are five verbs in present-day English that indicate the apparentness of a subsequent finite subordinate clause: seem, appear, look, sound, and feel. These verbs can be linked to the lower clause by one of five comparative complementizers: as if, as though, like, that, and null. Although like is the newest of these variants (López-Couso and Méndez-Naya 2012:177), it is overwhelmingly the predominant one in vernacular Canadian English and as if and as though have become negligible (López-Couso and Méndez-Naya 2012:185). I investigate this rapid lexical replacement with the use of two corpora: the Toronto English Archive (Tagliamonti 2003-2006) and an earlier collection of Canadian writing, drawn from Project Gutenberg Canada, primarily representing the decades between 1860 and 1930. The Toronto data shows a change in apparent time whereby like is overtaking the two remaining variants, that and null. In the earlier written materials, there are only 18 tokens of like, but all of these are in fictional dialogue written to come across as nonstandard; this suggests that like was known as a highly colloquial comparative complementizer until it started catching on across registers. The literality of the subordinate clause proves to be a key aspect of the change: in the earlier material, as if and as though are more likely with fully metaphorical subordinate clauses, that and null more often introduce the concrete ones, and like does not exhibit a clear preference. This semantic conditioning is stable over real time between 1860 and 1930, implying that one advantage that like has had over its moribund competitors is its versatility as far as literality is concerned.

This working paper is available in University of Pennsylvania Working Papers in Linguistics: http://repository.upenn.edu/pwpl/vol20/iss2/2
Comparative Complementizers in Canadian English: Insights from Early Fiction

Marisa Brook*

1 Introduction

Present-day English has a handful of verbs that denote apparentness. Five of these, referred to here as OSTENSIBILITY VERBS,¹ are capable of taking a finite complement clause: seem, appear, look, sound, and feel. The subordinators that link these verbs to the complement clauses are known as COMPARATIVE COMPLEMENTIZERS (Rooryck 2000:48, López-Couso and Méndez-Naya 2012), and in North American dialects, the variants are as if, as though, like, that, and a null form:

(1) It seems (as if/as though/like/that/Ø) there is a problem with the software.

Like is so prevalent in this context in vernacular Canadian English that as if and as though are negligible (López-Couso and Méndez-Naya 2012:185). This is intriguing given that like is the newest variant (2012:177) and that as late as the mid-1980s, the use of like in this context was “often regarded as nonstandard” (Quirk et al. 1985:1175).

The present study investigates the abrupt rise of the comparative complementizer like in Canadian English. It employs both the Toronto English Archive (TEA) (Tagliamonte 2003–2006, Tagliamonte 2006a) and a corpus of texts written by Canadians in the 19th and early 20th centuries, Early Canadian Writings (ECW) (Brook 2011), to examine the roles that have been played by that and Ø in this change.

2 Background

The (ostensibility verb) + (comparative complementizer) paradigm is dichotomous in terms of the historical origins of its verbs. Seem and appear in this context are related to their basic copular forms (Quirk et al. 1985:1174). Their earliest attestations with finite complement clauses in the Oxford English Dictionary (OED) are from the 13th and 14th centuries. Meanwhile, look, sound, and feel are sensory verbs (Quirk et al. 1985:203) used metaphorically, and in the OED they are not found in this function until the 17th and 18th centuries.

Table 1 shows that there are still idiosyncrasies in present-day English in terms of which comparative complementizers combine with which verbs. From a syntactic point of view, seem, appear, look, and sound in this context are all COPY-RAISING VERBS (Rogers 1974); they are capable of taking either an expletive it as matrix subject, or a full NP that is ‘copied’ from a position within the subordinate clause, leaving behind a coindexed pronoun. However, only as if, as though, and like allow copy-raising to occur (Potsdam and Runner 2001:465–466). That and Ø block it, prohibiting a lower NP from being copied upwards into the matrix subject position. For instance:

(2) a. It seems (as if/as though/like/that/Ø) the boys are sick today.
   b. The boys, seem (as if/as though/like) they, are sick today.
   c. *The boys, seem (that/Ø) they, are sick today.

*For their guidance, assistance, and suggestions, I am indebted to Sali Tagliamonte, Naomi Nagy, Bridget Jankowski, Aaron Dinkin, Alexandra D’Arcy, J. K. Chambers, Bronwyn Bjorkman, Derek Denis, Matt Hunt Gardner, my audiences at NAWV 42 (2013) and Change and Variation in Canada 5 (2011), and the University of Toronto Language Variation and Change Research Group. The spoken data in this study are drawn from the Toronto English Archive, funded by the Social Sciences and Humanities Research Council of Canada (Tagliamonte 2003–2006, Tagliamonte 2006a), grant #710-2003-0005.

¹Bender and Flickinger (1999:5) call these “verbs of perception”, while Quirk et al. (1985:1175) exclude seem and appear from that label and informally refer to a relatively large set as “verbs of ‘seeming’”. For the sake of clarity, I avoid both of these terms.
<table>
<thead>
<tr>
<th></th>
<th>like</th>
<th>as if</th>
<th>as though</th>
<th>that</th>
<th>Ø</th>
</tr>
</thead>
<tbody>
<tr>
<td>seem</td>
<td></td>
<td></td>
<td>grammatical</td>
<td></td>
<td>expletive subject only</td>
</tr>
<tr>
<td>appear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ungrammatical</td>
</tr>
<tr>
<td>look</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>animate NP subject only</td>
</tr>
<tr>
<td>sound</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>feel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Potential combinations of ostensibility verbs and comparative complementizers.

*Seem* and *appear* therefore combine with *that* and Ø only when the matrix subject is an expletive; these two complementizers do not permit full NPs in the matrix subject position. This is one of three restrictions on the distribution of *that*/*Ø* as comparative complementizers. Another one is the fact that *look* and *sound* strongly resist *that* and Ø, appearing with it only extremely infrequently even when there is no copy-raising taking place (Huddleston and Pullum 2002:962, López-Couso and Méndez-Naya 2012:187):

(3) *It looks (that/Ø) something weird happened to the water.*

The third restriction is that *feel* can take *that* and Ø only when the matrix subject is animate. *Feel* does not always introduce a subordinate clause with a copy-raised element; its ability to refer to the thoughts of a person in matrix subject position (or, in other words, to assign a theta-role of experiencer to its subject) means that the content-clause need not contain a coreferential nominal:

(4) *She feels (as if/as though/like/that/Ø) things are getting better.*

However, *feel* can acting as a copy-raising verb when the matrix subject is not an experiencer. In this case, as with *seem* and *appear*, *that* and Ø are disallowed when there is copy-raising:

(5) a. *It feels (as if/as though/like) the situation is becoming more complicated.*
   b. *The situation, feels (as if/as though/like) it, is becoming more complicated.*
   c. *The situation, feels (that/Ø) it, is becoming more complicated.*

In spite of these differences when it comes to history and syntactic properties, the comparative complementizers *that* and Ø can substitute for *as if, as though, or like* where possible “without any perceptible change of meaning” (Huddleston and Pullum 2002:9), fulfilling the classic criterion for being part of the same variable context in variationist sociolinguistic study.

3 Methodology

The spoken data in the present study are drawn from the Toronto English Archive (TEA) (Tagliamonte 2003–2006, Tagliamonte 2006a), a corpus of sociolinguistic interviews with 160 speakers native to the Toronto area between the ages of nine and 92. The TEA encompasses 350 hours of recordings and amounts to approximately 1.5 million transcribed words.

The earlier written data (Early Canadian Writings or ECW) (Brook 2011) is compiled from public-domain texts produced between 1830 and 1960, made available in digital form by Project Gutenberg Canada (http://www.gutenberg.ca). To be included, authors needed to have been native speakers of Canadian English, writing alone, who either were born in Canada or moved to Canada before the age of seven and stayed, since according to Chambers (2009:106), “someone coming to a dialect region under the age of 7 will master the dialect like a native”.

In order to function as a real-time corpus, the written material needed to be drawn solely from writers in middle age (Bridget Jankowski, p.c., June 19, 2011); only works written when these authors were between the ages of thirty and sixty were included. Poetry and children’s books were excluded due to the high chance that the language used within them was different in register.

With these constraints on the total output of Project Gutenberg Canada (as of June 2011), the number of discrete works in the ECW corpus comes to 185. The texts are a mix of fiction and non-fiction, with a total word-count of about 13 million.
I extracted every instance of an ostensibility verb immediately followed by both a comparative complementizer and at least most of a finite subordinate clause. Sentences with negative polarity and interrogative contexts were excluded since it was not clear that these would be conducive to the same distribution of variants. Intervening material (e.g., ‘it seems to me that’) was left out on the same basis. Parentheticals (e.g., ‘She knows what she’s doing, seems like’) are excluded since they allow only Ø and like. Other syntactically exotic structures (passivized ‘feel’, ostensibility verbs inside relative clauses, quotatives, etc.) are set aside for similar reasons.

I hypothesize that testing for the correlation of these forms with speaker age (TEA) and with date of publication (ECW) will uncover signs of the change towards like. There ought to be a sex-effect in the TEA whereby women favor the incoming form, as per Labov (2001:501). The ECW materials can be expected to predate most of the shift since they are much older (primarily representing the decades between 1860 and 1930) and since written materials tend to lag behind speech in terms of changes-in-progress in the first place (e.g., Pintzuk 2003:525). That said, I hypothesize that the very first signs of the emergence of like may well be visible in the ECW.

In both corpora, the choice of ostensibility verb is predicted to have an effect on the distribution of the comparative complementizers since the verbs act divergently, as discussed above.

Another potentially relevant constraint in both corpora is the level of metaphoricality of the subordinate clause. The clause can be fully metaphorical (6a), or indisputably concrete and literal (6c); it can also be ambiguous (6b):

(6) a. I feel as though I could eat a boiled alligator this moment.
   (May Agnes Fleming, Magdalen’s Vow, fiction, 1871)
   b. [He] looked as though he might have fallen over a precipice.
   (James De Mille, Lost in the Fog, fiction, 1871)
   c. It looked as if a real Manitoba blizzard was setting in.
   (Nellie L. McClung, The Black Creek Stopping-House, fiction, 1912)

The distinction can be made more apparent by substituting a phrase that explicitly refers to opinion for the ostensibility verb and complementizer. Examples in which the subordinate clause is metaphorical emerge fairly nonsensical in context:

(7) a. #[I suspect that] I could eat a boiled alligator this moment.
   b. ?[My opinion was that] he might have fallen over a precipice.
   c. [The impression we got was that] a real Manitoba blizzard was setting in.

Two hypotheses emerge from this constraint: one, that as if’ and as though would initially exhibit a preference for metaphoricality in the earliest ECW data, while that and Ø would be favored for concrete clauses; and two, that over time all four variants would be subject to the process of semantic bleaching that tends to accompany the later stages of grammaticalization (Hopper and Traugott 2003:94–98), resulting in the loss of this semantic constraint and paving the way for like to introduce subordinate clauses of all sorts.

4 Results and Discussion

4.1 Toronto English Archive: Distributional

Figure 1 shows that like is indeed the dominant variant in Toronto English. As noted by López-Cousío and Méndez-Naya (2012:185), as though is unattested in this context in the TEA, and as if is found only a handful of times. That and Ø remain in use, but at low frequencies. Like represents just under 80% of the Toronto tokens, suggesting that the system is nearing saturation.

Testing the apparent-time trajectories, as seen in Figure 2, reveals that like has been gaining ground at the expense of both that and Ø.
The TEA age-cohort that has the greatest proportion of speakers who use this construction at least once is the 20- to 29-year-old set, as Figure 3 shows. Notably, the effect is driven by the women in this age-group: there are 24 female speakers in the corpus between the ages of 20 and 29, and all but two of them (91.7%) use this pattern at least once. Similarly, Figure 4 shows that there is an inverse relationship between speaker age and the number of times this construction is relied on: every speaker who uses it nine times or more is 20 years old or younger.

To summarize: more younger than older speakers use the (ostensibility verb) + (comparative complementizer) structure; when they do, they tend to use it more often; and for the younger speakers, like is by far the variant of choice.

These results are analogous to the findings of Tagliamonte and Denis (2010) for general extenders (GEs) in Toronto in two ways: it is the young adults who are using GEs the most frequently (2010:349), and within the confines of the structure there is a case of lexical replacement occurring (2010:362), namely stuff usurping things.

It is possible that the increased usage of the (ostensibility verb) + (comparative complementizer) structures among younger speakers reflects a newly increased range of discourse-pragmatic functions. Along these lines, López-Couso and Méndez-Naya (2010) argue that seems like, looks like, and sounds like are becoming grammaticalized as invariant epistemic markers. Speakers under the age of 30 in the TEA do occasionally use these structures simply to put distance between
themselves and their own statements, even if factual and/or pertaining entirely to their own lives:

(8) a. It seems like I was doing a lot of Internet writing.
   (TEA, Narissa Wold, female, 27 years old)
   b. It seems like after I finish my Ph.D., I'll probably stick to music.
   (TEA, Richard Gruensten, male, 27 years old)

Figure 3: Proportion of speakers within each age-cohort in the Toronto English Archive who use the (ostensibility verb) + (comparative complementizer) construction at least once.

Figure 4: Each speaker who uses the construction at least once, according to age (X) and number of tokens used (Y).

An alternative explanation would be that hesitancy in and of itself is subject to age-grading in the TEA. Ostensibility verbs are just one of several strategies that younger speakers use for the sake of hedging and disclaimers:

(9) I haven’t been to Thailand, I haven’t traveled anywhere but like, I’m just saying, it seems that everybody always kind-of makes an effort to relate stuff back to English versus like you might not be able to go somewhere always and be able to have someone translate it to German or, or like Chinese for you all the time. Right?
   (TEA, Brent Kim, male, 21 years old)
4.2 Early Canadian Writings: Distributional

As Figure 5 shows, all four of the comparative complementizers aside from *like* are well-established in the ECW. *Like* is rare but not unattested (n = 18).

![Figure 5: Distribution of comparative complementizers in the ECW (n = 3215).](image)

The apparently erratic nature of the complementizers in the pre-1890 section of the corpus can be attributed to differences in the material available for each decade. The texts from the 1870s are only twelve in number and all fiction (which could be expected to contain more metaphors and thus more *as if* and *as though*); the ones from the 1880s, eleven in total and mostly nonfiction (bound to have more *that* as a result of the author laying out concrete inferences and conclusions). The stability of the complementizer distributions starting in 1890 reflects the greater amount of data and better balance in terms of genre thereafter.

As rare as the comparative complementizer *like* is in this corpus, there are two ways in which

---

2There are also three tokens of the apparent hybrid *like as if*. While these will not be considered further here, they are intriguing in that they suggest a possible intermediate form in the shift towards the adoption of *like* in Canadian English.
its behavior is telling. One is its real-time distribution: *like* is unattested in this corpus before 1896, but after the 1890s, its proportion begins creeping upwards, as Table 2 shows. It is possible that this is the very beginning of the S-curve — or at least of its echo in written media.

<table>
<thead>
<tr>
<th>Decade</th>
<th>up to 1869</th>
<th>1870s</th>
<th>1880s</th>
<th>1890s</th>
<th>1900s</th>
<th>1910s</th>
<th>1920s</th>
<th>1930 onwards</th>
</tr>
</thead>
<tbody>
<tr>
<td>% <em>like</em></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.7%</td>
<td>0.5%</td>
<td>0.5%</td>
<td>1.0%</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

Table 2: *Like* as a proportion of comparative complementizers in ECW, split by decade.

The specific tokens of *like* share several properties. Not only are they all from fiction, but all 18 come from the speech of characters. Many of the examples of the dialogue that *like* appears in are deliberately written to come across as highly colloquial; some of these contain a veritable cornucopia of nonstandard variants across an assortment of variables:

(10) a. Seems like I jest couldn’t live without ye.  
(Charles G. D. Roberts, *Earth’s Enigmas*, fiction, 1896)  

b. Nellie, who’s yer yaller friend over there by the windy, which looks like he’d like to make sassage-meat o’ my head?  
(Charles G. D. Roberts, *Earth’s Enigmas*, fiction, 1896)  

c. “Bon, dat’s fuss rate,” said Baptiste. “Seems lak dat’s make me eat more better for sure.”  
(Ralph Connor, *Black Rock*, fiction, 1898)  

d. I ain’t superstitious, but it looks like luck meant you to have another chance, don’t it?  

This speaks to *like* having been known to speakers of Canadian English from the 1890s onwards. However, it appears to have been considered nonstandard enough that for the entire period covered by ECW (approximately 1830 to 1960), it is found in prose only when authors are very intentionally attempting to employ offbeat dialectical speech in dialogue.

### 4.3 Metaphoricality of the Subordinate Clause

I turn now to the question of the semantic mechanisms that I hypothesized would underlie this shift. As Figure 7 shows, in accordance with the first hypothesis, that *as if* and *as though* become increasingly likely with a higher level of metaphoricality in the subordinate clause, while *that* and Ø are the opposite:

![Figure 7: Comparative complementizers in ECW according to the level of metaphoricality (i.e., non-literality) of the subordinate clauses (n = 3215).](image-url)
However, in contrast to the second prediction, the semantic preferences of the four older complementizers do not neutralize towards the end of the interval covered by the ECW. Their semantic preferences are all stable in real time; as Figure 8 shows, the surviving variants *that* and Ø even show similar patterning in the TEA data from the early 21st century, disregarding the fact that the system has been flooded with *like*:

![Figure 8: Comparative complementizers in TEA according to the level of metaphoricality (i.e., non-literality) of the subordinate clauses (n = 480).](image)

This result thus demonstrates that *like* could not have been ushered in by leveling of this constraint. If anything, *like* could have been able to catch on amid this competition because — uniquely — it is not sensitive to the literality of the subordinate clause. Evidence from this comes from the fact that the tokens of *like* divide into their semantic functions in very similar ways between the two corpora. In other words, semantically speaking, *like* was used largely the same way by fictional characters between 1896 and 1931 as it later was by nonfictional Toronto residents in the early 21st century, in spite of the differences in time, in medium, and in token-count. Figure 9 shows this parallel:

![Figure 9: Tokens of *like* in the ECW (n = 18, spanning 1896 to 1931) and in the TEA (n = 385, 2002 to 2007) divided into three levels of literality of subordinate clause.](image)

4.4 Toronto English Archive: Multivariate

In order to test for the simultaneous effect of these factors on the new variant *like*, a multivariate analysis was performed with Goldvarb X (Sankoff et al. 2005) for applications of *like* in Toronto
incorporating lexical verb, metaphoricality, age, and sex:

<table>
<thead>
<tr>
<th>Corrected mean</th>
<th>0.79</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood</td>
<td>-193.404</td>
</tr>
<tr>
<td>Total N</td>
<td>486</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Factor-weight</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 to 29</td>
<td>0.67</td>
<td>86.6%</td>
<td>226/261</td>
</tr>
<tr>
<td>Under 17</td>
<td>0.33</td>
<td>82.4%</td>
<td>28/34</td>
</tr>
<tr>
<td>30 to 59</td>
<td>0.27</td>
<td>61.5%</td>
<td>56/91</td>
</tr>
<tr>
<td>60 and older</td>
<td>0.09</td>
<td>35.7%</td>
<td>10/28</td>
</tr>
<tr>
<td>Range</td>
<td>58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Metaphoricality of subordinate clause | | |
|--------------------------------------|--|
| Metaphorical                         | 0.86 | 95.2% | 60/63 |
| Ambiguous                            | 0.54 | 81.1% | 120/148 |
| Concrete                             | 0.38 | 74.5% | 205/275 |
| Range                                | 48   |       |       |

<table>
<thead>
<tr>
<th>Lexical verb</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>seem</td>
<td>0.61</td>
</tr>
<tr>
<td>feel</td>
<td>0.46</td>
</tr>
<tr>
<td>Range</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>[0.50]</td>
</tr>
<tr>
<td>Male</td>
<td>[0.50]</td>
</tr>
</tbody>
</table>

Table 3: Multivariate analysis of the factors selected as significant to the selection of like (TEA) as a comparative complementizer linking an ostensibility verb to a finite subordinate clause.

The age effect indicates the expected shift towards like in this context, with the 17-to-29-year-olds at the leading edge of the change in progress. *Seem* appears to be somewhat ahead of *feel* in terms of the adoption of *like*, which is unsurprising given that *feel* supports *that* and Ø in an extra context: when the matrix subject is an experiencer that need not have been copy-raised from the subordinate clause.

Surprisingly, sex is non-significant; the female lead expected given a change-in-progress is not observed. Although the multivariate analysis is consistent with the notion that this social factor has leveled with the change nearing completion, a cross-tabulation of age and sex does not show a straightforward gender effect at any point in apparent time. This is a change that has not been clearly led by the female speakers in Toronto.

The significant effect of metaphoricality exists because the only other variants that remain in the system at non-negligible levels — *that* and Ø — both have a preference for concrete clauses. That is, as Figure 8 shows, *like* has absorbed the task of introducing metaphorical clauses, but both of the variants that combine more readily with concrete clauses are holding out to some extent.

5 Conclusion

In spite of having had only a short career as a comparative complementizer in Canadian English, *like* displays a number of interesting patterns. While the early written texts show that it was a very infrequent and nonstandard variant, I suggest that its lack of sensitivity to the level of literality in the subordinate clause has made it more versatile than all four of the older variants from the beginning, and that this has helped *like* overtake all of its competition. Neither *as if* nor *as though* retains any meaningful level of existence in vernacular Canadian English as represented by Toronto. *That* and Ø survive, but are declining in apparent time; *like* is well on its way to saturating the

---

5 *Look* and *sound* were followed by *that* or Ø only three times out of 119 tokens, with the other 116 all *like*. This makes these verbs too close to categorical to be included in the multivariate analysis (see Guy 1988:131, Tagliamonte 2006b:87). *Appear* is used only once in this context in the Toronto English Archive.
The loss of as if and as though means that semantically, like is the only remaining variant that easily introduces metaphorical subordinate clauses. When it comes to the effect of lexical verb, like has also taken over as the variant of choice following look and sound, since those verbs are almost entirely incompatible with that and Ø.

Both of these divisions found among the older four comparative complementizers — the opposing semantic preferences and the fact that they do not all appear with the same verbs in the same contexts — can likely be attributed to the historical division between the earlier seem and appear and the later look, sound, and feel. That is, what was once a paradigm deeply split by centuries-old differences is now united thanks to a new complementizer variant capable of reaching across constraints both syntactic and semantic.

References


Brook, Marisa. 2011. Looks as if there’s something interesting going on here: Comparative complementizers following perception verbs in Canadian English. Ms., University of Toronto.


