SYNTACTIC CATEGORIES INFORMING VARIATIONIST ANALYSIS:
THE CASE OF ENGLISH COPY-RAISING

by

Marisa Alana Brook

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Linguistics
University of Toronto

© Copyright by Marisa Alana Brook 2016
Syntactic categories informing variationist analysis: The case of English copy-raising

Marisa Alana Brook
Doctor of Philosophy
Department of Linguistics
University of Toronto
2016

Abstract

This thesis uncovers and investigates two levels of syntactic change progressing in tandem in Canadian English. One involves “comparative complementizers” (Rooryck 2000:48; López-Couso and Méndez-Naya 2012a), which link perception verbs to finite subordinate clauses: *It seems (like/as if/as though/that/Ø) she’s getting better quickly*. The predominant variant in Canadian English is *like* (López-Couso and Méndez-Naya 2012; Brook 2011a, 2014) – an incoming form that now represents 68.2 percent of comparative complementizers across corpora of sociolinguistic interviews from Ontario (Tagliamonte 2003-06, 2006, 2007-10, 2010-13, 2013; Tagliamonte and Denis 2014). Looking beyond the variable context, as per Aaron (2010), suggests that a second, broader change is occurring. The entire *seem*/*like* structure is overtaking Subject-to-Subject raising (*seem*/*to VP*). Younger speakers are increasingly saying *It/she seems like she’s getting better quickly* rather than *She seems to be getting better quickly*.

I find that the key variationist concept of orderly heterogeneity (Weinreich et al. 1968) applies readily to the covarying syntactic constructions despite their abstract nature. I also address the question of whether one level of change has led to the other. A methodologically comparable corpus from York, England (Tagliamonte 1996-1998, 1998) shows a low proportion
of *like* (13.2 percent) and no broader-level change; I hypothesize that the broader-level shift depends on a certain threshold of *like* in the community grammar.

Since *like*, unlike its remaining covariants, allows copy-raising (Rogers 1974, Horn 1981, Asudeh 2002, Asudeh and Toivonen 2007, Gisborne 2010, inter alia), it gives speakers the choice of a matrix NP or matrix expletive subject. Experimental/acquisitional research indicates that these two options correspond to different evidential/epistemic values (Rett and Hyams 2014); I argue that with a high proportion of *like* they together comprise a clear, streamlined binary system that makes Subject-to-Subject raising much less useful in comparison, accounting for its emerging decline.
Dedication

To my parents, my siblings, and the memory of my grandparents.
Permissions statements

Permission to use the Belleville Oral History Project was granted by Professor Sali A. Tagliamonte (University of Toronto) through arrangements with the Hastings County Historical Society in May 2007.

Permission to use the Toronto English Archive, the Southeastern Ontario English Corpus, the Northern Ontario English Corpus, the York English Archive, and the Roots Corpus was granted by Professor Sali A. Tagliamonte (University of Toronto).
Acknowledgments

Success has something major in common with language: it doesn’t grow well in a vacuum. Accordingly, I owe a lot to a large number of people.

At the centre of it all is Sali A. Tagliamonte, whose research lured me to variationist sociolinguistics from a distance, and who proved to be the ideal supervisor up close. It was Sali who pointed out in October 2010 that I could work on comparative complementizers for my MA Forum paper, and then suggested in October 2013 that I take the analysis farther for my doctoral dissertation. Sali, you have been inspiration, mentor, editor, cheerleader, collaborator, confidante, kickass boss, and a constant source of irrepressible enthusiasm. Thank you so much for everything.

Diane Massam is to thank for one of the central ideas in the present work: the intuition that copy-raising structures might be taking over from raising ones. Diane and I share a keen interest in subordinate clauses of all sorts, but I owe her nearly all of my understanding of the theoretical background covered in this study. I’d also like to thank Diane for the wonderful company and for being a crucial source of knowledge on academic writing and argumentation more generally. Diane, you’ve been absolutely invaluable in so many ways! Thank you!

It has been a huge privilege to get to work with Jack Chambers, the dialectological living-legend of the department. Where do I begin? Jack, your wisdom, patience, humour, and encouragement have done so much to help me along, and you have contributed so much to my analysis and to my knowledge and argumentation skills to begin with. Thanks a million, eh?

I’d like to thank Aaron Dinkin and Naomi Nagy for being my ‘internal-externals’ and for contributing a considerable amount to my analysis/interpretation, always pushing me to take things a little farther than I already had, and fearlessly working around my inordinate number of allergies in order to feed me hamentashen/latkes/pasta/brownies/etc. I have also been deeply fortunate to have had Belén Méndez-Naya as my external examiner, and I am very grateful to her for her extensive input and prodigious knowledge of the ins and outs of comparative complementizers, as well as historical English more generally!

This thesis has also benefitted substantially from the input of the following colleagues, in alphabetical order: Bronwyn Bjorkman, Heather Burnett, Ailís Cournane, Alexandra D’Arcy, Derek Denis, Matt Hunt Gardner, Larry Horn, and Bridget Jankowski. (Does anyone know where the rest of the alphabet went?)

I’m grateful to my audiences at CVC 5 (2011), NWAV 42 (2013), LSA 2015, and CVC 9 (2016) for providing feedback on various stages of this project and/or the MA Forum paper that I’m building on here. I’d also like to thank the ever-lively Language Variation and Change Research Group at the University of Toronto and the members of two LIN1256 seminars – Winter 2014 and Winter 2016 – for the same.
I want to take a minute here to thank the research assistants from the Language Variation and Change Lab in both Ontario and the United Kingdom who have done the interviews and transcribed the data. Thank you all so much for your hard work.

I would also like to express my gratitude to Elizabeth Cowper, Yoonjung Kang, and Keren Rice for their thoughtful general guidance and support over the years as I’ve made my way through graduate school.

I’m honoured to have been a part of a wonderful Ph.D. cohort. Radu Craioveanu, Ross Godfrey, Yu-Leng Lin, Iryna Osadcha, Chris Spahr, and Élodie Thomas: I’m looking at six amazing people from all over the world (six different native countries!), and trying to work out how to express how much I’ve learned from each of you. I think I’ll just have to leave it at that.

On top of all the people listed above, around the department these days I’m thankful for the friendship of Mary Aksim, Emily Blamire (you always make my day better!), Joanna Chociej, Emily Clare, Clarissa Forbes, Shayna Gardiner (my most supremely excellent gym-buddy!), Erin Hall, Chris Harvey, Beth Houze, Alexah Konnelly (suuuuuush!), Ross Krekoski (w00t!), Ruth Maddeaux (you are lovely and full of spirited ideas), Jessica Mathie (our masterful LGCU co-president!), Zoë McKenzie, Emilia Melara, Gloria Mellesmoen, Dan Milway, Safieh Moghaddam, Mercedeh Mohaghegh, Alex Motut (I admire so many things about you!), Katharina Pabst, Virgilio Partida Penalva, Jessamyn Schertz, Maksym Shkvoretz, Jim Smith (“code all the things!”), Brianne Süß, Becky Tollan, Nick Welch, and Tomohiro Yokoyama. (Anyone not listed: apologies. I don’t know you as well as I know these people. But hi!)

Jill Given-King and Mary Hsu: what would our department ever do without you? Both of you have fielded an incredible number questions from me (sometimes absurd and/or panicky ones) and have readily helped me figure out the answers right away in spite of the fact that I’m sure you had much better things to do. Thank you so much. I also want to be sure to thank Claudia and the rest of the janitorial staff for picking up after all of us. We appreciate it!

To my closest friends from outside this department – Eve Honeywell (and Mike Tomlinson), J. A. Macfarlane, Becca McLeod, Chelsea Powell, Bill Price (and Sally Kim), Caroline Sheehan, and Meredith Snyder – I’m so thankful to know each of you. You are all amazing people and I treasure my friendships with each of you so deeply.

Thanks to my other local friends and to the Toronto/Hamilton branches of my family for doing so much to help me feel at-home in Toronto from the beginning. And, over at Student Health Services, thanks to Dr. Susan Harrison and Dr. Kiran Clair for the meticulous, thoughtful care even in the face of the mildly staggering number of chronic-health issues that I have. Also, special thanks to Elissa at Counselling and Psychological Services for helping me get a grip on my obsessive-compulsive disorder. It’s been such a huge help.
From my undergrad days, I’d like to thank Wayne Harbert, Carol Rosen, Abby Cohn, and John Whitman at Cornell University for all their assistance and advice (and wonderful classes). I also owe a debt to the University of Victoria Department of Computer Science (particularly Mary Sanseverino and Valerie King) for their earlier mentorship and warm introduction to academia over the summer of 2003. I hope you can all forgive me for shifting gears slightly.

I probably wouldn’t be here without having been championed by a number of stellar teachers from earlier on – particularly in the two subjects that were most often my favourites, English and mathematics (which in retrospect seems suggestive). I’m especially grateful to Kathy Roth, Miriam Stanford, Linda Rajotte, Susan MacDonald, and Jim Williams. Not to mention Kirsten Davel, for introducing her AP Human Geography class in 2003-04 to language families (and for insisting that we consider taking at least one linguistics class in university).

None of the above would have been possible without the unyielding support of my family from the very beginning and all the way through. My parents have provided me with as many extracurricular learning opportunities as possible, and have gone out of their way to accommodate basically every passionate interest of mine over the years. Sure, my mother taught six-year-old me the word ‘syllable’ as a (correct) answer to a long and convoluted question I had about what “parts of words” were called. And my father is the reason why eleven-year-old me received an exuberant email with the subject-line "Latin and Greek plurals!" partway through a school day after asking about ‘-us’ and ‘-i’ in the morning. But these were just a couple of examples of the many, many things I have always been encouraged to inquire about. The fact that I’ve spent a little over a decade earning all three of my degrees in the same field belies the sheer range of other interests I’ve encountered and pursued, always with the encouragement of my parents. I’m immensely grateful for having always had my curiosity both taken seriously and celebrated. Okay, so a few of the questions I’ve been asking lately have been kind of technical and ended up requiring about two hundred pages just for an attempt at an answer; however, the spirit of things is the same, and the fact that I’m still doing this owes a lot to Mom and Dad.

My brother and sister have long been steadfast supporters, wonderful company, and – in idiosyncratic ways – freakin’ hilarious people. Thanks, siblings, for being yourselves, putting up with an interminable series of spontaneous lectures about my field (along with other types of know-it-all oldest-sibling obnoxiousness), and reminding me not to take myself too seriously. No one else does quite as good a job when it comes to that last thing, and trust me, it’s really important.

I’d also like to thank every company out there that makes really good peach-flavoured iced tea. In general I don’t much approve of corporate sponsorship, but it’s tough not to acknowledge the de facto case in my own life. On the plus side, if I have to be addicted to something, it might as well be this.
# Table of Contents

## 1 Introduction

1.1 Variation and subordination .......................... 1
1.2 Comparative complementizers ......................... 2
   1.2.1 Introduction ...................................... 2
   1.2.2 Variation and change ............................ 3
1.3 The role of the present work .......................... 3

## 2 Background: Syntax and variation

2.1 Syntax: copy-raising ................................. 8
   2.1.1 Introduction ...................................... 8
   2.1.2 Which verbs are part of the set? ................. 11
   2.1.3 Is it raising or something else? ................ 12
   2.1.4 Is the embedded coreferential pronoun necessary? 20
   2.1.5 Theta-role assignment ............................ 22
   2.1.6 Why are *that* and Ø different from the other complementizers? 26
   2.1.7 Disambiguating *feel* + subordinate clause .... 28
2.2 Change and variation ................................ 30
   2.2.1 Introduction ...................................... 30
   2.2.2 Historical trajectory ............................. 32

## 3 Epistemicity, evidentiality, and grammaticalization

3.1 Introduction ........................................ 36
3.2 Epistemic and evidential markers ..................... 37
   3.2.1 Definitions ....................................... 37
   3.2.2 Distinction ....................................... 38
   3.2.3 The *I think* class of epistemic markers ....... 39
   3.2.4 *Seem, appear, look, and sound* as evidential markers 41
   3.2.5 *Seems like, looks like, and sounds like* as epistemic markers 46
3.3 Grammaticalization ................................ 46
   3.3.1 Introduction ...................................... 46
   3.3.2 Key principles of grammaticalization .......... 48
   3.3.3 Grammaticalization as applied to epistemic and evidential markers 51
   3.3.4 Controversy ....................................... 52
   3.3.5 Grammaticalization as applied to comparative complementizer structures 54
   3.3.6 What about *feels like* or even *I feel like*? 57
3.4 Conclusion ........................................... 58

## 4 Methodology ........................................ 59
4.1 Data ................................................ 59
   4.1.1 Canada ............................................ 59
   4.1.2 United Kingdom ................................. 60
4.2 Variable context .................................... 61
   4.2.1 Broad definition ................................. 61
   4.2.2 Justification ..................................... 62
4.3 Extraction ........................................... 65
   4.3.1 AntConc search .................................. 65
4.3.2 Exclusions

4.4 Coding

4.4.1 Dependent variable
4.4.2 Independent linguistic variables
4.4.3 Independent social variables
4.4.4 Statistical analysis

4.5 Conclusion

5 Results and discussion: Overall variation

5.1 Canada (Ontario)

5.1.1 Overall distributional results
5.1.2 Apparent time
5.1.3 Verb
5.1.4 Matrix subject
5.1.5 Metaphoricality
5.1.6 Tense
5.1.7 Polarity
5.1.8 Sex
5.1.9 Clara
5.1.10 Belleville 1975
5.1.11 Random forests and conditional inference trees
5.1.12 Multivariate analysis
5.1.13 Summary: Canada

5.2 United Kingdom (York)

5.2.1 Overall distributional results
5.2.2 Apparent time
5.2.3 Metaphoricality
5.2.4 Verb
5.2.5 Matrix subject
5.2.6 Tense
5.2.7 Sex
5.2.8 Mini-studies of rural places in the UK
5.2.9 Summary: York

6 A second level of change

6.1 Introduction

6.1.1 A puzzle
6.1.2 A new solution

6.2 Methodology

6.2.1 Finite and non-finite structures
6.2.2 Dividing up the finite tokens
6.2.3 Extraction and exclusions
6.2.4 Coding

6.3 Results: Ontario

6.3.1 Finite versus non-finite: Overall
6.3.2 Finite versus Subject-to-Subject raising: Apparent time
6.3.3 Comparative study of two approaches to the same data
6.3.4 Sex
6.3.5 Multivariate analysis
List of Tables

Table 2.1: Corpora in which comparative complementizers have been studied. 31
Table 3.1: Two ways in which perception verbs have been described as pragmatic markers. 41
Table 3.2: Epistemic and evidential complement-taking phrases. 44
Table 4.1: Sources of Canadian data. 60
Table 4.2: Sources of British data. 61
Table 4.3: Combinations considered part of the variable context of the present study. 61
Table 5.1: Overall distribution of comparative complementizers in Ontario. 75
Table 5.2: Canada tokens split by complementizer and verb. 78
Table 5.3: Factors selected as significant for the complementizer like in Ontario. 100
Table 5.4: Overall distribution of comparative complementizers in York. 104
Table 6.1: Ontario speakers with the highest individual token counts. 117
Table 6.2: The three variants under the envelope-of-variation division of the finite tokens. 125
Table 6.3: The three variants under the type-of-complementizer division of the finite tokens. 126
Table 6.4: Overall tokens of non-finite and finite structures in Ontario. 129
Table 6.5: Factors selected as significant for the [permitted copy-raising] structure in Ontario. 138
Table 6.6: Overall tokens of non-finite and finite structures in York. 141
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Overall distribution of comparative complementizers in Ontario.</td>
<td>76</td>
</tr>
<tr>
<td>5.2</td>
<td>Ontario tokens in apparent time.</td>
<td>76</td>
</tr>
<tr>
<td>5.3</td>
<td>Ontario tokens by verb.</td>
<td>78</td>
</tr>
<tr>
<td>5.4</td>
<td>Ontario tokens in apparent time, split by verb.</td>
<td>79</td>
</tr>
<tr>
<td>5.5</td>
<td>Ontario tokens by matrix subject.</td>
<td>84</td>
</tr>
<tr>
<td>5.6</td>
<td>Ontario tokens in apparent time, split by matrix subject.</td>
<td>85</td>
</tr>
<tr>
<td>5.7</td>
<td>Cross-tabulation of matrix subject type and lexical verb in Ontario.</td>
<td>88</td>
</tr>
<tr>
<td>5.8</td>
<td>Ontario tokens by metaphoricality.</td>
<td>89</td>
</tr>
<tr>
<td>5.9</td>
<td>Ontario tokens in apparent time, split by metaphoricality.</td>
<td>90</td>
</tr>
<tr>
<td>5.10</td>
<td>Ontario tokens by tense.</td>
<td>91</td>
</tr>
<tr>
<td>5.11</td>
<td>Ontario tokens in apparent time, split by tense.</td>
<td>92</td>
</tr>
<tr>
<td>5.12</td>
<td>Ontario tokens by polarity.</td>
<td>92</td>
</tr>
<tr>
<td>5.13</td>
<td>Ontario tokens in apparent time, split by polarity.</td>
<td>93</td>
</tr>
<tr>
<td>5.14</td>
<td>Ontario tokens in apparent time, split by speaker sex.</td>
<td>94</td>
</tr>
<tr>
<td>5.15</td>
<td>Clara’s proportions of comparative complementizers in real time.</td>
<td>95</td>
</tr>
<tr>
<td>5.16</td>
<td>Random forest analysis like versus that and Ø in Ontario.</td>
<td>98</td>
</tr>
<tr>
<td>5.17</td>
<td>Conditional inference recursive partitioning tree with year of birth for Ontario.</td>
<td>99</td>
</tr>
<tr>
<td>5.18</td>
<td>Overall distribution of comparative complementizers in York.</td>
<td>104</td>
</tr>
<tr>
<td>5.19</td>
<td>York tokens in apparent time.</td>
<td>106</td>
</tr>
<tr>
<td>5.20</td>
<td>York tokens by metaphoricality.</td>
<td>108</td>
</tr>
<tr>
<td>5.21</td>
<td>York tokens by verb.</td>
<td>109</td>
</tr>
<tr>
<td>5.22</td>
<td>York tokens by matrix subject.</td>
<td>111</td>
</tr>
<tr>
<td>5.23</td>
<td>York tokens by tense.</td>
<td>112</td>
</tr>
<tr>
<td>5.24</td>
<td>York tokens by sex.</td>
<td>112</td>
</tr>
<tr>
<td>5.25</td>
<td>York tokens in apparent time, split by speaker sex.</td>
<td>113</td>
</tr>
<tr>
<td>6.1</td>
<td>Each speaker in the Ontario data by age and number of tokens.</td>
<td>117</td>
</tr>
<tr>
<td>6.2</td>
<td>Preliminary division of the data into two levels of variation.</td>
<td>120</td>
</tr>
<tr>
<td>6.3</td>
<td>Proportion of variable and non-variable speakers across all corpora.</td>
<td>123</td>
</tr>
<tr>
<td>6.4</td>
<td>Envelope-of-variation division of the finite tokens.</td>
<td>125</td>
</tr>
<tr>
<td>6.5</td>
<td>Type-of-complementizer division of the finite tokens.</td>
<td>126</td>
</tr>
<tr>
<td>6.6</td>
<td>Overall finite and nonfinite tokens in Ontario.</td>
<td>130</td>
</tr>
<tr>
<td>6.7</td>
<td>Finite and nonfinite tokens in Ontario with envelope-of-variation division.</td>
<td>132</td>
</tr>
<tr>
<td>6.8</td>
<td>Finite and nonfinite tokens in Ontario with type-of-complementizer division.</td>
<td>134</td>
</tr>
<tr>
<td>6.9</td>
<td>Cross-tabulation of sex and type-of-complementizer variants in Ontario.</td>
<td>136</td>
</tr>
<tr>
<td>6.10</td>
<td>Type-of-complementizer variants in Ontario, split by sex.</td>
<td>137</td>
</tr>
<tr>
<td>6.11</td>
<td>Clara’s type-of-complementizer variants in real time.</td>
<td>139</td>
</tr>
<tr>
<td>6.12</td>
<td>Overall finite and nonfinite tokens in York.</td>
<td>142</td>
</tr>
<tr>
<td>6.13</td>
<td>Finite and nonfinite tokens in York with envelope-of-variation division.</td>
<td>142</td>
</tr>
<tr>
<td>6.14</td>
<td>Finite and nonfinite tokens in York with type-of-complementizer division.</td>
<td>143</td>
</tr>
<tr>
<td>6.15</td>
<td>Cross-tabulation of sex and type-of-complementizer variants in York.</td>
<td>143</td>
</tr>
<tr>
<td>7.1</td>
<td>Superimposition of proportions of like and [permitted copy-raising] in Ontario.</td>
<td>155</td>
</tr>
<tr>
<td>7.2</td>
<td>Evidential/epistemic stance of three constructions.</td>
<td>166</td>
</tr>
<tr>
<td>7.3</td>
<td>Replication of Liberman (2013) with Google Ngram Viewer.</td>
<td>169</td>
</tr>
<tr>
<td>7.4</td>
<td>Replication of Squires (2013) with Google Ngram Viewer.</td>
<td>169</td>
</tr>
<tr>
<td>7.5</td>
<td>Synthesis of the above two graphs with Google Ngram Viewer.</td>
<td>169</td>
</tr>
<tr>
<td>7.6</td>
<td>Several like collocations with Google Ngram Viewer.</td>
<td>170</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Variation and subordination

Subordination is a breeding-ground for linguistic variation and change (see e.g. Algeo 1988:22). Well-studied subordinator variables in English include complementizers (Rissanen 1991, Palander-Collin 1997, Tagliamonte and Smith 2005), restrictive relative-clause markers (Quirk 1957, Guy and Bayley 1995, Romaine 1982, Tagliamonte et al. 2005), quotative verbs (Butters 1982; Tannen 1986; Blyth et al. 1990; Tagliamonte and D’Arcy 2004, 2007; Buchstaller 2011), and verb-complementation strategies (Algeo 1988, Mair 2002). Pockets of variation can be identified even structures that are ordinarily categorical. For instance, infinitival complements of verbs are almost always introduced by to, but there are two alternatives among select verbs. A few have Ø as an alternative to to, e.g. help (Kjellmer 1985; Mair 1995, 2002); others can take and, e.g. be sure (Pullum 1990:222, Huddleston and Pullum 2002:1302) and try (Quirk et al. 1985:978-979, Huddleston and Pullum 2002:1302, Hommerberg and Tottie 2007, Ross 2013, Brook and Tagliamonte to appear).

What this suggests is that – within English and possibly cross-linguistically – categoricity among subordination strategies is the exception. Accordingly, all types of subordination, however minor or low-frequency, might prove to be conducive to variation and even change. Some collocations that originate as subordination strategies are susceptible to abrupt grammatical change that yields pragmatic markers, e.g. methinks (Palander-Collin 1997) and I think (Thompson and Mulac 1991, Kärkkäinen 2003, Tagliamonte and Smith 2005, Denis 2015). The general morphosyntactic field surrounding the clausal boundary is full of variables to
study, any of them potentially a source of considerable insight into the diachronic development of grammatical structures.

1.2 Comparative complementizers

1.2.1 Introduction

Within the present-day English subordinator inventory, one small subsystem worthy of closer attention is the set of complementizers that can link any of the verbs seem, appear, look, sound, and feel to finite clausal complements. Five such complementizers are recognized in the literature: as if, as though, that, Ø, and like. The variants as if, as though, and like are typically called COMPARATIVE COMPLEMENTIZERS\(^1\) (Rooryck 2000:48; López-Couso and Méndez-Naya 2012a:172); the more familiar that and Ø complementizers can also follow these verbs, albeit with additional restrictions to be covered in detail.

The various combinations of verbs and complementizers have emerged along two separate lines, historically speaking (Brook 2011:6-8). The copulas seem and appear plus finite subordinate clauses are attested from the 13\(^{th}\) and 14\(^{th}\) centuries in the Oxford English Dictionary, mostly with that and Ø as their complementizers. Look, sound, and feel – perception verbs able to handle metaphorical propositions – are found with finite subordinate clauses starting only in the 17\(^{th}\) and 18\(^{th}\) centuries and primarily with as if. This likely accounts for some enduring asymmetries, such as the fact that look and sound are generally incompatible with that and Ø (Huddleston and Pullum 2002:962, Gisborne 2010:276, López-Couso and Méndez-Naya 2014:39). Similarly, it is most likely as a result of the verbs’ split historical emergence that that and Ø prefer concrete subordinate clauses, while as if and as though prefer metaphorical ones (Brook 2011a, 2014).

\(^1\) Other terms for these include “comparative subordinators” (López-Couso and Méndez-Naya 2012b:315), “minor declarative complementizers” (López-Couso and Méndez-Naya 2012a:172), “conjunction[s]” (D’Arcy 2007:392), and “subordinating conjunction[s]” (Gisborne 2010:239). Note that what Bácskai-Atkári and Dekány (2014:204-205) refer to as a “comparative complementizer” appears to be something different.
Nonetheless, I refer to *seem, appear, look, sound,* and *feel* together as a set (and label them *OSTENSIBILITY VERBS* as in Brook 2014, since there is no consensus as to what they ought to be called and terms such as ‘perception verbs’ are ambiguous). The intersection of two properties unites these five verbs in present-day English to the exclusion of all others. Semantically, they all lend a denotation of apparentness to the subordinate clause. Syntactically – at least with the complementizers other than *that* and Ø (Huddleston and Pullum 2002:962, Gisborne 2010:275) – the five verbs permit an optional syntactic transformation known as **COPY-RAISING**, whereby a nominal from the lower clause is ‘raised’ to the matrix subject position and the original position is realized as a coreferential pronoun (Rogers 1971, 1972, 1974a, 1974b; Horn 1981; Heycock 1994; Potsdam and Runner 2001; Asudeh 2002, 2012; Asudeh and Toivonen 2007, 2012; Landau 2009, 2011; Gisborne 2010:280; Mack 2010; Kim 2014):

(1a) It LOOKS LIKE Rachel won the game.²
(1b) Rachel LOOKS LIKE she won the game.

### 1.2.2 Variation and change

There has been little work so far on the complementizers following ostensibility verbs from the perspective of variation and change. López-Couso and Méndez-Naya (2010, 2012a, 2012b, 2014) have conducted both synchronic and diachronic studies of *as if, as though, like,* and *that.* Notably, they find effects of both variety and register (2012a). A noteworthy result is that *like* is by far the most popular comparative complementizer in spoken Canadian English; by the 1990s, it also shows signs of catching on in written American English, but it takes its time to appear in spoken British English and has not yet established itself in British written materials (López-Couso and Méndez-Naya 2012a:185). These results suggest that *like* is the incoming form across

---

² I capitalize the verb and comparative complementizer as long as the example or token falls into the variable context (see Chapter 4 for methodological details).
all three varieties and that the United Kingdom is lagging behind Anglo-America when it comes to the adoption of the *like* complementizer.

I have previously established that *like* is indeed the innovative variant in Canadian English (Brook 2011a, 2014): it is increasing straightforwardly in apparent time in the Toronto English Archive (Tagliamonte 2003-2006, 2006). My earlier conclusion was that *like* has essentially unified the set of complementizers after ostensibility verbs: it always allows copy-raising and can handle both concrete subordinate clauses (2a) and metaphorical (2b) ones, which makes it more versatile than any of the competing variants (Brook 2011a, 2014). This work elaborates on those findings considerably.

(2a) This guy LOOKS LIKE he’s pretty worn-out.
(2b) This guy LOOKS LIKE he swallowed a live crocodile tail first.

1.3 The role of the present work

This thesis addresses a need for several types of further investigation, among them: greater geographic breadth in complementizer analysis across comparable corpora of sociolinguistic interviews; more consideration of the role played by semantic/pragmatic factors such as evidentiality; and an examination of the role played by related syntactic structures. Chapter 2 introduces the background from syntax and variation (copy-raising; synchronic and diachronic analysis of these five complementizers), and Chapter 3 covers the background from semantics and pragmatics (epistemicity, evidentiality, and the possibility of grammaticalization producing pragmatic markers from the collocations *seems like*, *looks like*, and *sounds like*). After the methodology (described in Chapter 4), the subsequent two chapters provide separate sets of results and discussion. Chapter 5 extends my previous analyses geographically,3 encompassing multiple varieties of both Canadian and British English through methodologically comparable

---

3 Aside from the inclusion of the extra corpora, the two most substantial overall points of departure are that my earlier analyses did not include tokens with negative polarity of the matrix clause, but did include interviewer tokens.
synchronic corpora of sociolinguistic interviews. Then, Chapter 6 helps resolve a loose end from my prior research (Brook 2011a, 2014). An unexpected discovery was that the number of tokens of the entire subordinate clause structure with these verbs and complementizers shows an age-effect in the Toronto English Archive: younger speakers are using them appreciably more (Brook 2011a, 2014). I have previously proposed two possibilities to account for this: grammaticalization, which tends to go hand-in-hand with an increase in frequency (Hook 1991); and age-graded hedging in general that happens to involve these structures among many others. Chapter 6 introduces a third partial reason for the age-effect, which proves to be crucially implicated in the overall picture of the change involving the complementizer like. It also helps to account for why López-Couso and Méndez-Naya (2012b:329) have the impression that the use of the entire structure “seems to be on the increase in the present day”. I show that at least for the verb seem, the entire finite structure [seem* + complementizer + finite subordinate clause] is catching on in apparent time at the expense of the closest alternative: the infinitival structure [seem* + infinitival subordinate clause], which is conventionally referred to as Subject-to-Subject raising (Rosenbaum 1967; see also Davies and Dubinsky 2004). In other words, (3a) is gaining at the expense of (3b):

(3a)  She/it SEEMS LIKE she’s doing really well.
(3b)  She SEEMS to be doing really well.

This finding reveals that the shift towards the comparative complementizer like is actually part of a two-tiered change. It reaffirms a major point established by Aaron (2010): that looking beyond the conventional variable context can bring essential insights to the analysis of a change-in-progress. On top of that, Chapter 6 considers what it means for finite and non-finite structures to be covarying, and employs the methodology of comparative sociolinguistics (Tagliamonte 2002) to compare and contrast multiple ways of partitioning the same data. The outcomes also yield an elegant account of why in colloquial Canadian English, like appears to
have overtaken *as if* and *as though* much sooner than it took over from *that* and Ø (Brook 2011a, 2014): because *like* competes directly with *as if* and *as though*, whereas *that* and Ø appear to be protected by an intermediate step in terms of the change. The results also show that syntactic categories can be usefully integrated into a variationist analysis, that core variationist ideas about changes-in-progress can apply readily to a change involving entire syntactic structures, and that doing both is indispensable when it comes to the case of comparative complementizer structures in present-day English.

Chapter 7 considers the implications of a change involving two levels of morphosyntactic variation. It examines several major larger issues: whether a causal link between the two changes can be established; whether a change from infinitival to finite structures has any cross-linguistic precedents; how much of a problem it is for grammaticalization theory, which predicts the opposite (Hopper and Traugott 2003:175); and whether there is any influence from another major change involving *like*: namely, the *be like* quotative (Butters 1982; Blyth et al. 1990; Romaine and Lange 1991; Ferrara and Bell 1995; Tagliamonte and Hudson 1999; Tagliamonte and D’Arcy 2004, 2007; etc.) My main conclusions are multifaceted. I defend the notion that the change from infinitival to finite subordination after *seem* is a direct consequence of the change towards *like*, and that *like* must reach a certain proportion of the comparative complementizers in the community grammar before the whole finite structure (3a) can take over from the infinitival one (3b). Once the complementizers have been levelled by *like*, the structure levels the types of subordinate clauses that follow it. I speculate that this is partially attributable to the fact that with *like*, copy-raising is optional but not obligatory. I draw on literature from language acquisition and experimental methods (Rett et al. 2013; Rett and Hyams 2014) to argue that the evidential statuses of the two possibilities for matrix subject – copy-raised NPs and expletives – differ considerably; that a comparative-complementizer system consisting exclusively of *like* makes the encoding of evidentiality completely predictable by the type of matrix subject; and
that infinitival complements of *seem* not only have no role to play in the updated system but are inflexible enough in comparison to be far less useful.

The individual discoveries underlying this conclusion owe themselves to the empirical approaches of variationist sociolinguistics, but they draw on several other subfields of linguistics for explanation, and thus have ramifications for each. The sum of it is that a versatile morphosyntactic variant has entered a system of finite subordination, saturated and levelled it, led to the whole structure to take over from a more distantly related subordinate structure, and streamlined the expression of evidentiality in present-day Canadian English.
Chapter 2

Background: Syntax and variation

2.1 Syntax: copy-raising

2.1.1 Introduction

The variation under study involves the verbs *seem, appear, look, sound, and feel* when followed by a finite subordinate clause as a complement. Five complementizers from present-day English can link one of these verbs to a finite subordinate clause: *as if, as though, that, Ø, and like*. Not all twenty-five combinations are acceptable; some of them, such as *look/sound that/Ø, are downright ungrammatical (see section 3.2.2, and also Huddleston and Pullum 2002:962, Gisborne 2010:276, López-Couso and Méndez-Naya 2014:39). The nature of the subject also has a major effect on grammaticality: all of the verbs are capable of taking either a matrix expletive or a matrix NP, but there are restrictions on these that differ subtly between the verbs – to be explored in this chapter and subsequent ones.

At the centre of the matter is an optional syntactic alternation that all of these verbs can take part in. It has traditionally been known as *copy-raising* based on its conventional analysis as the movement of an embedded subject NP to the matrix subject position, leaving behind a coreferential – and bound (Lappin 1983:240) – pronominal ‘copy’ of itself:

(4a) It SEEMS LIKE [the girl] wants to try out the diving board.
(4b) [The girl] SEEMS LIKE [she] wants to try out the diving board.

---

4 Importantly, these are in fact complements and not adjuncts. The most straightforward piece of evidence is that deleting them renders the sentence ungrammatical, as in *He seems* (McCawley 1998:149; Asudeh 2002; López-Couso and Méndez-Naya 2012a:174-175). Other evidence for the complement status of these subordinate clauses is covered by Quirk et al (1985:1183), Bender and Flickinger (1999), Asudeh (2002:5-6), and López-Couso and Méndez-Naya (2012b:316).
A raising-based analysis posits that in sentence (4b), the NP *the girl*, subject of the subordinate clause, has moved upwards to the matrix subject position, with the coreferential copy *she* left in its original position in the subordinate clause.

Copy-raising is not uncommon cross-linguistically⁵ (Massam 1985, Potsdam and Runner 2001:453, Asudeh 2002, Asudeh and Toivonen 2007, Gisborne 2010:269, Asudeh 2012:363). In English, however, copy-raising is found with only a handful of verbs, and there is agreement in the literature that the alternation in this language has received comparatively little attention (Potsdam and Runner 2001:453, Asudeh and Toivonen 2012:321-322; Kim 2014:168).

In terms of its properties, copy-raising is an interesting hybrid. As with *that/Ø* complementation in general (5a), copy-raising (5b) involves a tensed (finite) clausal complement (see e.g. Asudeh and Toivonen 2007:2-3). This means that some of the characteristics of *that/Ø* complementizers and comparative complementizers overlap (see López-Couso and Méndez-Naya 2012b:311 for a list of pertinent references).

(5a) I heard *(that/Ø)* there’s going to be a parade today.
(5b) It seems *(that/Ø/like/as if/as though)* there’s going to be a parade today.

However, copy-raising involves movement of an NP. It is atypical among movement strategies in English in that it leaves behind a pronominal copy. Syntactic movement out of subordinate clauses in English almost always leaves behind syntactic gaps, not pronouns. Forcing gaps to be realized as coindexed pronouns (usually referred to as RESUMPTIVE PRONOUNS) tends to render English sentences ungrammatical:

(6a) *This is the headline that ____ could make me do a double-take.*
(6b) *This is the headline that *it* could make me do a double-take.*

---

⁵ For lists of other languages in which copy-raising has been identified and studied, see Potsdam and Runner (2001:453), Polinsky and Potsdam (2006:181), and Asudeh (2012:363).
Resumptive pronouns are generally viewed as impermissible – or at least unsystematic\(^6\) – in English (Heycock 1994:291-292, Gisborne 2010:280, Landau 2011:783, Asudeh 2012:41). Although they are readily produced by English speakers in the context of syntactic islands and the like (Kroch 1981, Sells 1984, Prince 1990, Erteschik-Shir 1992, Asudeh 2012:40-44), resumptive pronouns tend to seem quite awkward in English (Harbert 2007:458), and are arguably best viewed as “a processing strategy” (Asudeh 2012:41) employed by the speaker in order to avoid losing track of the referent.

While this means that the coreferential pronouns left behind in the subordinate clause by copy-raising might well be syntactically related to resumptive pronouns (McCloskey and Sells 1988; Boeckx 2003; Polinsky and Potsdam 2006:174; Asudeh 2012), it is still the case that they are an anomaly. It is unexpected that a class of verbs in English that can instigate NP movement to the matrix subject position would also permit overt – and systematic (Heycock 1994:291, Potsdam and Runner 2001:461) – pronouns at the original position.

This atypical property of copy-raising verbs, at least by the standards of English, makes the comparative complementizers as if, as though, and like a challenge to classify syntactically (Gisborne 2010:273). Unsurprisingly, there are a number of controversies associated with the syntactic analysis of copy-raising. I will review these one at a time. The results from the present study are unable to speak to all of them directly, but an overview of the contentious issues associated with the syntactic analysis of copy-raising will help inform the results:

- which verbs are true copy-raising verbs (2.1.2);
- whether copy-raising is actually raising (i.e. movement) or something else (2.1.3);
- whether there absolutely has to be a coindexed pronoun in the embedded clause (2.1.4);
- how theta-role assignment works with copy-raising (2.1.5); and

---

\(^6\) Though see Pullum (2008) and Brook (2009, 2011b) on colloquial relative clauses introduced by a semantically bleached where, e.g. There are organs in France where they are tuned as high as 56 Hz (2011b:6). These do consistently support resumptive pronouns.
why that and Ø act so differently from as if, as though, and like (2.1.6).

I conclude section 2.1 with an explanation of why feel is a special case with respect to the data in this study (2.1.7).

2.1.2 Which verbs are part of the set?

One issue with no current consensus is which particular verbs ought to be counted as engaging in copy-raising – and what the set ought to be labelled. Rogers (1971, 1972, 1974a, 1974b) considers seem, appear, look, sound, feel, taste, and smell, but assigns only the sensory verbs look, sound, feel, taste, and smell – to the category of “flip perception verbs” (i.e. when there is an NP in the matrix subject position, it is the percept rather than the perceiver). Quirk et al. (1985:1174-1175) call seem, appear, look, sound, feel, taste, and smell “verbs of seeming”, but elsewhere (Quirk et al. 1985:1033) refer to seem, appear, look, sound, and feel alone as “perception” verbs. Bender and Flickinger (1999:5), focusing on argument structure rather than copy-raising, use the term “verbs of perception” for look, seem, appear, sound, feel, and strike (as in ‘this strikes me as if...’). Gisborne (2010:241) includes only the sensory verbs (look, sound, and feel) but acknowledges that appear acts similarly. Landau (2011) agrees with the initial impulse of Rogers (1974a) in assigning seem, appear, look, sound, feel, taste, and smell to the category of perception verbs. Asudeh (2012:326) argues that only seem and appear count as “[t]rue copy-raising in English”, and argues that look, sound, and feel are a different class (2012:327). There is almost no agreement.

The option of simply referring to “copy-raising verbs” in the present work is less than appealing. This is because the variable context is defined as any of these five verbs plus a complementizer plus a finite subordinate clause. While most of the sentences that fit this template are either potential or realized examples of copy-raising – that is, they are either (4a) or (4b) – there is an additional class thanks to the verb feel that do not fall into either category because they do not involve copy-raising at all (see section 2.1.7). Therefore, here, as elsewhere
I refer to *seem, appear, look, sound, and feel* together as *OSTENSIBILITY VERBS* in order to sidestep the ambiguity of each of the established names that include this set. Note that I do this merely for the sake of convenience, and not in order to disregard the possibility of subtle syntactic differences within the set. Indeed, the lexical effects are both intrinsically interesting and a crucial source of insights into the patterning of the complementizers, so these will not be glossed over. I refer to these five verbs as a set, but openly acknowledge that they are far from uniform.

2.1.3 Is it raising or something else?

The first authors to draw scholarly attention to what we now know as copy-raising were Postal (1971:163, 1974a:268) and Rogers (1971, 1972, 1974a, 1974b). Rogers (and subsequent authors) have also informally referred to copy-raising as the RICHARD transformation, after some of Rogers’ early example sentences, e.g.:

\[(7a)\] It smells as though Richard smokes.
\[(7b)\] Richard smells as though he smokes.

The derivation of such pairs proposed by Rogers (1974a:72) is as follows: the noun phrase from the embedded clause is copied into the matrix subject position, and the coreferentiality is what triggers the pronominalization of the lower copy.

Rogers considers (1974a:74-75), but then rules out (1974a:80-88), an alternative analysis in which the matrix subject is generated *in situ* – which does not depend on a syntactic transformation that derives a sentence such as (7b) from one such as (7a). Drawing on evidence presented in Rogers (1971), he argues that the Richard rule does a better job of accounting for a)

---

7 While *taste* and *smell* do show the same syntactic transformation (Rogers 1974a, etc.), they are comparatively infrequent in the context of finite subordination. They also tend to be much more literal and are not often used to indicate general apparentness. This is in contrast to *look* and *sound*, which *can* refer to sensory information but are not obligated to do so (Rogers 1974a:91, Heycock 1994, etc.) and to *feel*, which can be used to express “propositional attitudes”, i.e. opinions (Asudeh and Toivonen 2012:325), rather than literal examples of something being physically felt through touch. Thus, *taste* and *smell* will not be considered further here.
copy-raising of the expletive *there* (e.g. *there looks like there’s going to be a riot*) and for b) the fact that idioms retain their idiomatic sense even when a copy-raising verb intervenes (e.g. *the cat looks like it’s out of the bag*) (Rogers 1974a:83). He acknowledges, however, that the case for Richard and against the analysis of *in situ* NP generation is not clear-cut (1974a:91-92), especially given the lack of complete synonymy between the pairs linked by the transformation (see section 3.2.2 for more details).

Rogers points out (1974a:90-91) that the verbs *look* and *sound* might be either literal sensory descriptions or more abstract impressionistic observations with a meaning along the lines of *seem*: “the *seem* sense [of *look* and *sound*] is not restricted to physical perception at all, much less a specific sensory mode” (Rogers 1974a:91). He observes that his argumentation favouring the Richard transformation – *there*-raising and idiom reconstruction – applies much more readily to the *seem* meaning than to the literal senses of *look* and *sound*. That is, under a literal reading of *look*, it is unclear what *there looks like*… would mean, and even if *the cat looks like it’s out of the bag*, it is unlikely to be literally so. Therefore, the argument for Richard (the transformational analysis) works for the abstract *look* and *sound*, but whether it applies to their literal uses is not necessarily established.

Rogers concludes, though (1974a:92), that the raising arguments involving *there* and idioms are his strongest, and that discarding them would be senseless. The issue of synonymy, Rogers concludes, is not particularly troubling. He described how other, more unequivocal cases of raising do not require perfect synonymy, so the slight differences in meaning between the sentences with and without optional copy-raising are not necessarily a problem for the Richard analysis (Rogers 1974a:93; see also discussion by Mack 2010:155-158).

---

8 Several works acknowledge inter-speaker variation associated with the acceptability of *there* as a copy-raised subject (e.g. Horn 1981:355, Potsdam and Runner 2001, Asudeh 2002).
The issue that would go on to cause some major—and as-yet-unresolved—divisions between scholars involves two more marked classes of construction that appear to be related to copy-raising: I will be referring to these together as NON-CANONICAL COPY-RAISING. These are the situations in which an NP matrix subject corresponds either to a pronoun inside the embedded clause other than the subject—as in (8) to (11)—or to nothing within the embedded clause at all, as in (12):

(8) Bill: SOUNDS LIKE Martha hit him over the head with the record.
(9) The roach LOOKS to me LIKE Abbie gave it to Myrna.
(10) Mary APPEARS AS IF her job is going well.
(11) That book SOUNDS LIKE everyone thinks it should be banned.
(Potsdam and Runner 2001:456)
(12) The orchestra SOUNDS to me LIKE Mehta is having a good night.
(Rogers 1974a:100)

Rogers (1974a:97) concludes that the Richard transformation, “if it exists at all, does not apply to [surface] non-subjects of like-complement sentences” as in (8) to (11). He then considers embedded clauses in which there is no coreferential pronoun in the first place, as in (12)—arguing that these also “cannot be the result of a copying rule such as Richard, since there is no NP in the like-complement to have been copied” (1974a:100). The cases such as (12), although they tend to be “difficult to comprehend in the absence of explanatory context...nevertheless appear to be grammatical” (1974a:100). They necessitate something other than the Richard rule in order to account for the existence and location of the matrix subject NP, but Rogers (1974a) is unwilling to overturn the Richard transformation entirely. He sees it as a crucial part of the explanation for copy-raising when the verb has a ‘seem’ meaning (that is, seem itself plus the non-literal look and sound). Ultimately, Rogers settles on a compromise involving “dual generation”, i.e. two pathways of derivation, but he does not spell out the specifics and acknowledges that it is somewhat inelegant (Rogers 1974a:101-102).

Horn (1981:355), like Rogers (1974a), briefly entertains the possibility that all matrix subjects in copy-raising sentences are generated in situ rather than raised, but backs away from
committing to that conclusion in the face of evidence for raising in at least some cases. Building on Rogers (1974a), Horn argues that three types of raised subjects – the idioms, the *there* expletives, and certain uses of the weather *it* – support a raising analysis:

(13) The cat SOUNDS LIKE it’s out of the bag.
(14) There LOOKS LIKE there’s going to be a riot.
(15) It LOOKS LIKE it’s raining harder than it really is.

Given the assumptions that that idioms are generated all together in a single location (see also Postal and Pullum 1988, Kim 2014:170), that the expletive *there* is never generated at its surface position, and that the weather *it* when it permits a *de re* meaning (i.e. one of mistakenness) is not an expletive, Horn (1981:356) concludes, like Rogers (1974a) that raising is a necessary part of any analysis of the copy-raising transformation and related structures.⁹

Heycock (1994) agrees with Rogers (1974a) that non-canonical copy-raising involving embedded objects rather than subjects cannot reasonably be interpreted as raising. For one thing, objects are not plausibly anaphors for matrix subjects, especially in the cases where another NP intervenes, such as *everyone* in (16a) and (16b):

(16a) That book SOUNDS LIKE everyone will want to buy it.
(16b) That book SOUNDS LIKE everyone thinks it should be banned.
(Heycock 1994:290)

That is, proposing a binding relationship between *the book* and *it* in these sentences is problematic for Heycock because the non-co-referential subject *everyone* ought to disrupt the binding connection (see also Cowper 1992:148-149). Heycock concludes, therefore, that *that book* and *it* are not bound and that *it* is not comparable to a trace. This means that there has not been movement upwards of *that book*.

For Heycock, this and other evidence represents a fatal blow to the raising account for *any* cases of apparent copy-raising at all, even the canonical, well-behaved ones with embedded

---

⁹ Kim (2014:187) points out that the idioms only work if the coreferential pronoun is in the subject position of the subordinate clause; they fail otherwise. While this is not inconsistent with Horn’s account, it does suggest that raising, if it is occurring, is limited to the canonical cases.
subjects. Heycock refers to an argument based on Lappin (1984:239-240) that unlike with ordinary Subject-to-Subject raising, a matrix subject with a universal quantifier such as everyone would be forced to adopt a wide scope over a verb such as sound with a complementizer. In other words, it sounds like everyone arrived at the same time is grammatical, but #Everyone sounds like he arrived at the same time is iffy (Heycock 1994:290-291). She appears to have been the first to argue that what we refer to as copy-raising need not depend on movement in any case.

Potsdam and Runner (2001) also argue against a movement-based account. They highlight the Minimalist raising analysis by Ura (1998:458-460) of copy-raising in Igbo – said to be analogous to what occurs in English. The question for Ura is why raising of an embedded nominal is possible given that the nominal receives Case at its lower position, and why it is an option in the first place if it is unnecessary. Ura addresses these problems by proposing the following Minimalist derivation (as laid out by Potsdam and Runner 2001:458-460):

1. The embedded clause is assembled with the DP in the lower position.
2. The embedded DP checks the lower clause’s strong EPP feature (a crash would occur immediately otherwise), but not its Case or its phi-features.
3. The matrix clause is added and the DP raises to the matrix subject position – motivated by the unchecked strong EPP feature of the matrix clause – and leaves behind a trace. The raised DP also checks the Case and phi-features of the matrix clause.
4. An idiosyncratic operation (“Rule S”) steps in to spell out the trace in the embedded clause, realizing it as a pronoun. The pronoun checks the Case and phi-features of the embedded clause that were left unchecked earlier.

Potsdam and Runner (2001) raise several concerns with this account. First, they find Rule S to be a clumsy graft onto Minimalist theory – partially undermining the very purpose of the Minimalist framework. Another issue is that Potsdam and Runner find it too difficult to overlook the proposal’s violation of the Tensed S Condition (Chomsky 1973), which disallows
movement out of a tensed clause and which seems to hold across English more generally (Potsdam and Runner 2001:460).

Potsdam and Runner favour a non-raising analysis, in which the embedded clause is formed with the pronoun already as is and then the matrix NP subject is merged into its position (2001:462-463). This account necessitates a base-generated A-chain linking the pronoun and the NP, which is mildly unconventional but not outside the boundaries of what Minimalism allows (2001:463), especially if this A-chain could be convincingly linked to the Agree relation (2001:465). A matrix NP that has been generated in situ accounts for the nominal being more readily interpretable in the upper position than in the lower one, as evidenced by an asymmetry in scope when quantifiers are in the matrix subject position:

(17) Two people SEEM LIKE they have won the lottery.
    = Two people are simultaneously in the position of appearing to have individually won the lottery (two people interpreted at upper position).
    ≠ It appears that the lottery has been won by a pair of people (two people interpreted at lower position) (adapted from Potsdam and Runner 2001:464).

The notion of the chain cannot, however, hold for the non-canonical cases of copy-raising (Potsdam and Runner 2001:463), so the authors argue that “[t]he non-subject construction necessarily involves a thematic use of the predicate and thus is not copy-raising” (2001:454). This leaves Potsdam and Runner in a position similar to that of Rogers (1974a): they have acknowledged the non-canonical cases, but their analysis does not really address them.

Asudeh (2002) dismisses Ura (1998) for the same reasons as do Potsdam and Runner (2001): for him, Rule S is a type of “construction-specific, ad hoc transformation [that] is clearly undesirable and does not fit into current linguistic theory” (2002:3). However, Asudeh is not receptive to Potsdam and Runner’s A-chain idea either; he finds it no more attractive than Ura’s Rule S.
Matushansky (2002) and Gisborne (2010) both advocate for a non-raising account. Matushansky’s position is peremptory: that a copy-raising structure with a matrix NP “is not a raising construction, given that the matrix subject can be connected with nouns other than the subject of the embedded clause” (2002:221); for the same reason, Gisborne calls copy-raising “entirely unnecessary” (2010:279). Both of them thus agree with the concerns of Heycock (1994), of Potsdam and Runner (2001), and even of Rogers (1974a): they independently find the possibility of non-canonical examples with no embedded pronoun enough to dismiss the notion of a syntactic transformation entirely. Gisborne favours an Xcomp analysis that does not depend on movement (2010:279), and then attributes the presence of the lower-clause pronouns to two factors: like/as if/as though being semantically associated with similarity/comparison (as per Gisborne 1996), and pragmatic factors (Gisborne 2010:281).

For Asudeh and Toivonen (2007, 2012) and Asudeh (2012), the non-canonical cases of copy-raising (the cases of the coreferential pronouns being lower-clause objects or not being there at all) are not necessarily a challenge to the core of a raising-based analysis of copy-raising; rather, they represent points on a spectrum of grammaticality. Individual speakers then have different dividing-lines between the permissible and non-permissible cases (Asudeh 2012:327). A speaker who accepts a copy-raised embedded object is also going to accept a copy-raised subject; a speaker who accepts an embedded clause with no coreferential pronoun at all will accept both of the less-marked types (Asudeh 2012:327-329).

With the non-canonical cases assigned to an implicational hierarchy of acceptability across speakers rather than being treated as evidence against raising, it is unsurprising that Asudeh’s account (2012) of canonical copy-raising is raising-centered. For Asudeh, the clinching argument

---

10 It is true that part of what makes the examples without an embedded pronoun work is via “bridging reference”, i.e. a very clear semantic link between the matrix subject and the proposition in the subordinate clause (Gisborne 1996, 2010:281; see also Heycock 1994:292, Landau 2009, 2011:809, Rett and Hyams 2014:176, Kim 2014:184). That said, it can be shown that there is no subtle/covert syntactic attachment of a coreferential NP, since relativization fails (Mack 2010:154): *the soup that Maude has been at the cooking sherry again.*
is the ability of copy-raising verbs to take expletives (both *it* and *there*), which he sees as “a key piece of evidence that copy-raising can take a non-thematic subject and therefore really is a kind of raising” (2012:328).

Asudeh argues (2012:330) that raising accounts are not challenged by the existence of embedded *objects* that appear to raise. Non-subject copy-raising, he argues, is nonsensical. Since “raising is [necessarily] a relation between a higher grammatical function and an embedded subject”, trying to shoehorn any other type of embedded nominal into an account of the copy-raising operation is something he finds to be misguided (2012:330). Second, he argues that if the matrix NP in non-canonical copy-raising is generated *in situ*, it should receive a theta role at its higher position, which ought to obviate the need for a coreferential pronoun in the lower clause (2012:330). Also, if the matrix NP is generated *in situ*, it is unclear why a class of speakers surveyed by Asudeh and Toivonen (2007, 2012) accept the non-canonical cases with a coreferential pronoun in object positions, but *not* the ones that completely lack coindexed pronouns altogether. These would be indistinguishable if the matrix NP were generated at its surface position.

Asudeh (2002, 2012) is very comfortable with the raising account of canonical copy-raising – at least for *seem* and *appear* – and even claims that even *expletive* subjects of copy-raising verbs have been raised – twice, one step at a time – from lower in the clause (2002:10, 2012:336): first to the subject position of the embedded clause, then from there to matrix subject position.

The novel approach of Landau (2011) is to suggest, instead of movement, a null operator that binds the copy pronoun. This suggestion is a bit of a hybrid in that it tidily takes care of the problem presented by the non-canonical cases that lack embedded pronouns – these are merely propositional and lack an operator altogether (2011:808) – but it is also consistent with evidence against movement (Heycock 1994, Potsdam and Runner 2001). It suggests that CPs are not
automatically predicative but can become as such with the addition of the operator (Landau 2011:811).

While Landau’s analysis accounts for both canonical and non-canonical copy-raising (i.e. embedded clauses with and without pronouns), what is unclear is how to reconcile the evidence for movement (the matrix subject position being non-thematic with seem and sometimes look/sound; idioms that survive being split by seem* like; copy-raising of there; the weather it) (Rogers 1974a, Horn 1981, Asudeh 2002, 2012) with the arguments against (scope asymmetry indicating greater interpretability at the upper position; the violation of the Tensed S Condition; traces being unpronounced in English otherwise) (Heycock 1994, Potsdam and Runner 2001).

The issue of whether or not copy-raising is indeed raising – even if limited to the canonical cases – is currently open, and is also not the only lively debate that has grown out of research on copy-raising.

2.1.4 Is the embedded coreferential pronoun necessary?

On one side of this issue are Rogers (1974:99), who acknowledges only briefly that a copy-raising sentence without a pronoun could be grammatical given enough context, and Matushansky (2002), who admits that the “pronoun seems near-obligatory” (2002:211). On the other is Heycock (1994:293), who sees no reason why the pronouns should ever be considered necessary. Likewise, Landau (2011) argues that seem and appear do not necessarily require the copy pronouns at all, citing several online examples.11 Gisborne (2010: 280) agrees with Landau (2011) in that he has identified examples of seem with no embedded pronouns.

---

11 The present study agrees with Gisborne (2010) and Landau (2011) in this respect: one of the tokens with no coreferential pronoun is with appear: “I seem to have had a natural kind-of voice that just APPEARED AS IF I had-had train- a lot of training.” (Kathleen Murphy, F 87, Cumnock). That said, the number of non-canonical tokens of copy-raising is very small opposite canonical examples. While low frequency should not be mistaken for ungrammaticality, this does suggest that copy-raising sentences are much more likely to be well-behaved than to be atypical. It supports the suggestion of Landau (2011:785) that perhaps “copies are not necessary in all contexts, [but] there are contexts where they are strongly preferred” and the conclusion of Kim (2014:184) that there is a hierarchy of acceptability and the “highe[st] frequency cases are those where the matrix subject is coindexed with the subject or its specifier of the highest embedded clause”.

20
In between these positions are Asudeh (2002, 2012) and Asudeh and Toivonen (2007), who consistently maintain that “[seem and appear] require a pronominal copy in their complements, while [look, sound, and feel] do not” (Asudeh 2002:2) – that the “perceptual resemblance verbs do not require copy pronouns at all” (Asudeh 2012:331).

Landau makes a proposal tied to epistemic/evidential status of the verb: “[copy-raising] examples whose subject can be construed as a [source of perception] allow a copy-less complement…[but]examples whose subject cannot be construed as a [source of perception] require a [coindexed pronominal] copy in their complement” (2011: 787). Landau points out the similarity between copy-raising and Proleptic Object constructions (Landau 2011:808-809; see also Higgins 1981, Massam 1985, Davies 2005) in which a nominal object towards the beginning of a clause appears to be non-thematic but also requires a coindexed pronoun inside the lower clause. These are found in e.g. Madurese (Higgins 1981, Davies 2005) and in Hebrew, as in this example from Landau (2011:808):

(18a) xašavnu al Gil še-mašehu nora kara lo
we.thought about Gil that-something terrible happened to.him

(18b) *xašavnu al Gil še-mašehu nora kara
we.thought about Gil that-something terrible happened

(adapted from Landau 2011:808)

Landau concludes (2011:809-811) that two types of structure templates exist: propositional CPs, which are ordinary CP complements (e.g. copy-raising with a perceptual source in the matrix subject); and predicative CPs, which are ungrammatical without a copy and which turn into predicates by means of a copy combining with a null operator that binds it (e.g. Proleptic Object constructions, or copy-raising with something other than a perceptual source in the matrix subject).

Based on corpus findings from the Corpus of Contemporary American English (COCA; Davies 2008–) and the Corpus of Historical American English (COHA; Davies 2010–), Kim
(2014:184) suggests a hierarchy of accessibility based on Ariel (1990), as follows, with each step more acceptable and more frequent:

\[
\text{subject} > \text{specifier of the subject} > \text{(prepositional) object} > \text{context-provided inferred individual}
\]

He also proposes a “Perceptual Characterization Condition” (Kim 2014:183) whereby the matrix NP in a copy-raising structure is the topic and is “perceptually characterized” by the subordinate clause. This allows for non-canonical cases of copy-raising, as long as they refer to something that is a property of the matrix NP. It rules out \*\text{Bill looks as if Mary is intelligent} since Mary being intelligent does not speak to anything about the matrix NP \text{Bill} (Kim 2014:183).\footnote{One imagines that a very contorted scenario could be set up to make this sentence acceptable, but in that case the situation would necessarily make something about Mary being intelligent point to a \textit{characteristic} of Bill. This, in my view, constitutes support for Kim’s Perceptual Characterization Condition.}

Under this interpretation, the frequent presence of a coreferential pronoun – especially a subject – in the subordinate clause naturally falls out of the requirement that the subordinate clause refer to a characteristic of the matrix NP being perceived.

**2.1.5 Theta-role assignment**

The traditional view has been that copy-raising verbs such as \textit{seem} do not assign a theta-role to their matrix subjects (Potsdam and Runner 2001:454-456; Asudeh and Toivonen 2007:3; Gisborne 2010:279), hence why the non-thematic expletive \textit{it} can serve in that position (Lappin 1984:238). With the expletive, the only theta-role assigned by \textit{seem} is of percept (or proposition), assigned to the entire subordinate clause (see e.g. Cowper 1992:72-73, Gisborne 2010:269). A nominal will receive its own theta-role at the lower site (Potsdam and Runner 2001:453), from the verb-phrase inside the embedded clause.

\begin{equation}
(19) \text{It SEEMS THAT Jessie has eaten the cake.}
\end{equation}

(Cowper 1992:72).

In other words, what is doing the ‘seeming’ in (19) is not \textit{Jessie} herself but the notion of \textit{Jessie} having eaten the cake: the theta-role for \textit{seem*} is assigned to the entire lower clause at once, or to
the combination of the nominal copy and the proposition (Cowper 1992:72-73, Gisborne 2010:269). This lack of theta-role assignment to the matrix subject position is what allows matrix expletive subjects – necessarily lacking a theta-role – to be felicitous with copy-raising verbs (Horn 1981, Cowper 1992:77-79). It also accounts for the general – if not invariable – oddness of attempts to use copy-raising verbs *without* a coindexed nominal in the lower clause:

(20a) John eats like there is no tomorrow.
(20b) * John SEEMS LIKE there is no tomorrow.

(Kaplan-Myrth 2000, as cited by Potsdam and Runner 2001:455)

The question that this raises is where the theta-role for a copy-raised matrix NP comes from (Lappin 1983:240). If *seems* is unable to assign it this theta-role itself, then the theta-role must emerge from inside the embedded clause. Under this view, a copy-raised subject has received its theta-role at the lower site and then been copied upwards, along with a copy of the theta-role, into a non-theta position; this means that an argument chain has a single shared theta-role (see e.g. Lappin 1983:240).

Several accounts of theta-roles and copy-raising are more nuanced in that they focus on the subtle polysemy of the perception verbs, especially *look* and *sound*, that was originally described by Rogers (1974a:90-91). Heycock (1994) frames her theta-role discussion around an example:

(21a) Your car SOUNDS LIKE it needs tuning very badly.
(21b) (X is sitting in Y’s car and is listening to the engine and wincing.)
(21c) (In a long distance telephone call, Y has just described to X the bizarre noises that Y”s car is making.)

(Heycock 1994:289)

The crucial difference between interpretations (21b) and (21c) is whether “the matrix subject is...interpreted as the source of the sensory impressions” (Heycock 1994:289). In (21b), the car...

---

13 As with copular and other more basic forms of the verbs (e.g. *She seems eccentric*), a copy-raised matrix subject is a *percept*, i.e. the target of perception (see Quirk et al. 1985:203-204; Gisborne 2010:4-8). That is, the raised NP is not experiencing the *seeming*, but is part of the source of the observation in the entire clause. This will prove to be implicated in the analysis in Chapters 6 and 7.

14 Lappin points out (1984:240-241) that this presents a problem for the Theta Criterion (Chomsky 1981:335), which insists on a single theta-role for each single argument. He goes on to suggest slight amendments to the Theta Criterion (1984:245-246). Potsdam and Runner (2001:462) mention that in Minimalist analyses the problem is moot since Minimalism does not have a Theta Criterion or the equivalent.
itself is the source of the observation; the car is literally *sounding like* the description in the embedded clause. In (21c), however, speaker X’s statement is secondhand, based on information relayed by speaker Y, and *the car* is not literally *sounding like* anything.

The difference between these interpretations, semantically speaking, is in “the type of evidence motivating belief in the truth of the proposition denoted by the subordinate clause” (Heycock 1994:289). That is, the distinction is one of epistemicity/evidentiality (see also Chapter 3, as well as Landau 2011). Heycock analyzes this distinction in terms of differing theta-role assignment (1994:288-289). In (21b), according to her analysis, the verb *sound* assigns a theta-role (presumably of percept) to the matrix subject. In (21b), it does not.15 This analysis is intuitively appealing: is the car doing the *sounding* (as in 21b) or is it not (as in 21c)?

Potsdam and Runner (2001) dismiss the non-canonical cases of copy-raising as not copy-raising at all, in part because they feel that in these cases, the matrix subject is thematic, i.e. the matrix verb necessarily assigns a theta-role to the matrix subject (Potsdam and Runner 2001:454). Their A-chain analysis of canonical copy-raising involves a single theta-role, as above. Asudeh (2012:329) points out that this makes Potsdam and Runner’s analysis more analogous to raising (one theta-role) than to control (two).16

Gisborne (2010:247) follows Heycock (1994) and argues for two classes of *sound* and *look*, “distinguished by whether or not they assign a semantic (or thematic) role to [the matrix NP],” but depending primarily on semantics – specifically, on the evidential status of the verb:

---

15 Heycock acknowledges that the distinction is often subtle, and easier to discern with *sound* than with *look* (1994:289).

16 Cross-linguistically, however, *copy-control* may be something entirely different from even non-canonical cases of copy-raising (Polinsky and Potsdam 2006).

17 Notice that both classes would fall under the aegis of Kim’s (2014:183) Perceptual Characterization Condition.
An example that Gisborne provides is:

(22) Peter SOUNDED to me LIKE he had the spirit of Paganini inside him. (Gisborne 2010:273).

Here, if Peter is literally producing sound that brings Paganini to mind for the speaker, the sense is of the first evidential use for Gisborne. If the speaker is describing a secondhand account of a masterful and exuberant violin performance on the part of Peter, the sense is of the second evidential use.

Asudeh (2012), building on work by Asudeh and Toivonen (2007, 2012) argues that look, sound, and feel “are semantically contentful enough to take a thematic subject argument” (Asudeh 2012:331) – that is, that they can be generated in situ and do not require coindexed pronouns at all. This is closely intertwined with his argument that these three perception verbs are syntactically distinct from seem and appear.

Kim (2014:185) argues that the matrix subject of a copy-raising verb is inherently “ambiguous with respect to the theta-role bearing”. He points to the arguments of Horn (1981), as well as the fact that the expletive can be deleted, for evidence that there is no assignment of a theta-role to the matrix subject. However, he defends the notion that in the case of non-canonical copy-raising – where the matrix NP subject is not coindexed with the subject of the subordinate clause – the verb must be assigning it a theta-role (2015:186). Although his account is non-raising based, there is a way in which his analysis agrees with that of Asudeh (2012): he considers the three perception verbs look, sound, and feel to be distinct from seem in terms of argument structure, though not as thoroughly different from appear (2014:196). This will be reviewed in Chapter 7, since the analysis in the two chapters before that has implications for whether this is changing.
Otherwise, since there are very few examples of non-canonical copy-raising in the data under study for the present work, I will assume that matrix NP subjects are in non-thematic positions (except for with feel: see section 2.1.7).

2.1.6 Why are \textit{that} and \textit{Ø} so different from the other complementizers?

Whatever the syntactic/transformational nature of what we call copy-raising, it is indisputable that this process occurs only with the comparative complementizers – \textit{as if}, \textit{as though}, and \textit{like} (Potsdam and Runner 2001:465). Another way of putting this is that the complementizers \textit{that} and \textit{Ø} do not permit a referential matrix subject (Maling 1983, Lappin 1984:245-246, Heycock 1994:295; Gisborne 2010:275). However, even with expletive subjects, \textit{that} and \textit{Ø} are ungrammatical after \textit{look} and \textit{sound} (Huddleston and Pullum 2002:962, Gisborne 2010:276, López-Couso and Méndez-Naya 2014:39), which is further evidence that \textit{that} and \textit{Ø} are set apart from the comparative complementizers in terms of their properties.

In an analysis that involves raising, the movement of the NP upwards out of the embedded clause must be blocked by \textit{that} and \textit{Ø}.\footnote{This is also mentioned by Huddleston and Pullum (2002:962) – in different terms but with an equivalent example.}

(23a) It \textit{SEEMS} (LIKE/AS IF/AS THOUGH/\textit{THAT}/\textit{Ø}) John is ill.
(23b) John \textit{SEEMS} (LIKE/AS IF/AS THOUGH) he is ill.
(23c) * John \textit{SEEMS} (\textit{THAT}/\textit{Ø}) he is ill.
(adapted from Potsdam and Runner 2001:465)

While this principle might be expected to permit the occasional exception along the lines of (23c) in spontaneous speech, there are no examples of anomalous copy-raising over a comparative complementizer \textit{that} in the present data; with \textit{that} and \textit{Ø} after the copy-raising verbs, the matrix subjects are all expletives (\textit{it} or null). The one example of \textit{appear} that has a referential subject is a token of \textit{as if} (see footnote 11 on page 20). More strikingly, a search of the Corpus of Contemporary American English (Davies 2008–), which has more than 500 million words, on 2 November 2014 found no unequivocal cases of “\textit{seems that}” or “\textit{seemed that}”
introducing a subordinate clause that had matrix NP subjects rather than expletives – a powerful testament to copy-raising being ruled out with the complementizer that.

In order to account for that and Ø blocking copy-raising but the other comparative complementizers allowing it, Lappin’s early analysis (1984) suggested that a that-clause is a complete proposition incapable of accepting external arguments, but that as if ones are capable of defining their complement clauses as either complete (in which case the matrix subject is an expletive) or incomplete (in which case the lower subject NP is copied upwards) (Lappin 1983:247-248).

Heycock (1994:295) makes the observation that this idea of Lappin’s is not a solution so much as a “restate[ment] of the original problem” – the original problem being why both matrix subjects and copy-raised NPs are possible.

Two approaches to this problem appear to exist. One is syntactic: Heycock’s own analysis, building on Maling 1983, concludes that while that and Ø are true complementizers, as if, as though, and like are prepositions (Heycock 1994:295; see also Potsdam and Runner 2001:465-466, Asudeh 2002:6-8). This would make copy-raising – whatever it is – permissible with PP complements but not CP ones. This idea appears to have garnered widespread acceptance regardless of the related raising/non-raising controversy (see Potsdam and Runner 2001:466, Asudeh and Toivonen 2007, 2012, Asudeh 2012:331).

My search was limited in that it examined only that rather than both that and Ø. It also overlooked the possibility of ambiguous if subjects that might have been somewhat referential. Still, the result is compelling.

Maling (1983) originally argued that like is a sort of “transitive adjective”; however, Asudeh (2012:333-334) points out that this is implausible given how non-adjectival the behaviour of like (and as) is in English more generally.

Whether as if and as though have internal structure (e.g. as being a preposition and if being a complementizer) is also a question worth asking (Huddleston and Pullum 2002:971, Asudeh 2012:333-334). However, I follow Bender and Flickinger (1999:15) and Kim (2014: 178) who both argue that it is a single unit. Likewise, in Bácskai-Atkári and Dekány’s (2014:201, 208) analysis of a nearly identical structure in Hungarian, the equivalent of as if is “base-generated as a single – morphologically complex – head” of the embedded clause. This form stabilized in Hungarian and endures to this day (Bácskai-Atkári and Dekány 2014:208). In fact, the process appears to be iterative in Hungarian, which now also permits the equivalent of ‘as that if’ on top of an earlier grammaticalized ‘that if’ (Bácskai-Atkári and Dekány 2014:216-217).
The other approach is semantic/pragmatic: for instance, Matushansky (2002:221) argues that
*like* with a matrix subject “require[s] direct visual perception”, but that *that* is necessarily a value
judgment rather than a perception of something (and is thus presumably incompatible with a
matrix NP). There are links between this idea and the differing epistemic/evidential status of
matrix NPs versus expletives in the context of copy-raising verbs, which will be further
explored in the background in Chapter 3 and prove to be vital to the analysis in Chapter 7. For
now, it can be reasonably expected that *that* and Ø are something quite different from *as if, as
though, and like*. A partial account of this from the standpoint of the changes under study in the
present work will reveal itself in Chapter 6.

### 2.1.7 Disambiguating feel + subordinate clause

*Feel* is capable of taking an expletive as a matrix subject, which is straightforward:

(24)    It FEELS LIKE the chair’s about to catch on fire.

However, when *feel* has an NP matrix subject, there is the potential for ambiguity due to
subtle polysemy of the verb. *Feel* is capable of supporting copy-raising just as *look* and *sound* are,
but it is also very often used to introduce “propositional attitudes” (Asudeh and Toivonen
2012:325) – that is, views and beliefs belonging to an animate entity in the matrix subject
position. Syntactically, these two uses can look very similar on the surface (see also López-
Couso and Méndez-Naya (2012a:188-189):

(24a)    The chair FEELS LIKE it’s about to catch on fire.
(24b)    I FEEL LIKE the chair’s about to catch on fire.

The crucial difference between these is that in (24a) – copy-raising – *the chair* is not receiving
its theta-role from the matrix subject. That is, *the chair* is not doing any of the feeling – as its lack
of animacy may independently suggest. Rather, (24a) is a case in which the speaker is giving a
report on a physical perception involving *the chair*. In (24b), two major differences are apparent:

---

22 I am very grateful to Bronwyn Bjorkman (p.c., 10 April 2014) for a discussion that advanced my understanding of
this matter considerably.
a) there is no place for a nominal coindexed with matrix subject I to appear in the subordinate clause; and b) the lower clause is reporting directly on the internal thoughts of the matrix subject I. That is, there is no copy-raising occurring in (24b); the verb *feel* is assigning a theta-role of EXPERIENCER to the matrix subject. The matrix subject I is indeed doing the feeling in this case, and did not get into the matrix subject position via copy-raising (for related discussion, see Rogers 1974a:48, Gisborne 2010:6).

The presence of a coreferential nominal in the subordinate clause does not automatically mean that a case of *[feel + subordinate clause]* is copy-raising. If experiencers (animate matrix subjects doing the feeling) happen to be referring back to themselves, there is coindexation, but it is coincidental and not indicative of copy-raising. It is simply a reflection of an experiencer report that happens to involve the matrix subject’s own self, as in (25b):

(25a)  I FEEL LIKE it’s going to rain tonight.
(25b)  I FEEL LIKE I might get caught in the rain tonight.

Even aside from that, caution is also needed due to the potential for ambiguity between copy-raising and self-referential experiencer reports:

(26)  I FEEL LIKE I’m about to catch on fire.

Under the copy-raising reading (meaning 1), the speaker is reporting on incoming sensory information. They have discovered through direct physical examination that their skin is extremely hot. This is comparable to example (24a) of the chair feeling extremely hot; the difference is only that the object of investigation is, in this case, the speaker’s own body.

The propositional reading – meaning 2 – is more metaphorical, and is a report on the speaker’s sensations and thoughts. It could, for instance, be paraphrased with ‘I think’ or ‘I suspect’. Note that while it may still reflect feeling physically overheated, if so it comes from the speaker’s internal experience rather than an external investigation.

---

23 Horn (1981:355) makes a similar point with reference to *it* subjects with *look* and *sound*. 
Another way to clarify this ambiguity is to highlight a case in which the ambiguity disappears, such as with a subject whose thoughts are not particularly accessible:

(27) The baby feels like she’s about to catch on fire.

Only meaning 1 – copy-raising – is felicitous here. That is, the speaker is likely in physical contact with the baby and reporting that she is very warm. The speaker cannot use ‘feel’ in the sense of meaning 2 (the experiencer interpretation) because they are unable to access the referent (the baby)’s thoughts. The speaker could be e.g. placing a hand on the referent’s forehead, but could not be reporting directly on what she is thinking.24

As will be laid out in Chapter 4, the present study defines its envelope of variation as any of the verbs *seem, appear, look, sound,* and *feel* when followed by a finite subordinate clause. For the first four verbs, this means that matrix NPs are either expletives or copy-raised subjects; there is no conflation with other uses of these verbs. *Feel* is the exception. Nearly all of the tokens of it with a finite subordinate clause have an experiencer subject rather than a copy-raised one. For this reason, the data will often be split by verb in order to filter out differences stemming from the divergent properties of (most of the) examples of *feel.* While excluding *feel* outright would not be unjustifiable, it turns out to be following the other verbs closely in terms of the change towards *seem,* which suggests that it is part of the same change as what is affecting *seem* regardless of its divergent theta-role properties. See Chapter 5 and Chapter 7 for related results and discussion.

### 2.2 Change and variation

#### 2.2.1 Introduction

Table 2.1 shows corpora in which the complementizers after ostensibility verbs have been studied; these encompass American, British, and Canadian varieties of English.

---

24 The speaker could be *inferring* or just guessing what the baby is thinking, but that is a bit of a stretch under most circumstances.
López-Couso and Méndez-Naya (2012a) analyse as if, as though, like, and that in multiple corpora: for British English, they rely on LOB/FLOB (written, 1960s and 1990s) (Hofland et al. 1999) as well as the Diachronic Corpus of Present-day Spoken English (DCPSE) (spoken, 1960s and 1990s) (Aarts and Wallis 2006). For American English, they use the Brown and Frown corpora (written, 1960s and 1990s) (Hofland et al. 1999). For their study of these complementizers in Canadian English, they use the Toronto English Archive (TEA) (spoken, 2000s) (Tagliamonte 2003-2006, 2006). In previous work I have also studied the variation in the TEA, but I have excluded the it-is-as-if construction (Declerck 1992) and taken the Ø variant into account, meaning that my results are not directly comparable to those of López-Couso and Méndez-Naya (2012a).

I have also done a study of comparative complementizer variation in historical written Canadian English, based on a corpus compiled from Project Gutenberg in 2011 that I have labelled Early Canadian Writings (ECW) (Brook 2011a, 2014). The other diachronic study of comparative complementizers that has been done is that of López-Couso and Méndez-Naya (2012b), who use the Helsinki Corpus of English Texts and the ARCHER corpus to trace the development of as if and as though (and like).

<table>
<thead>
<tr>
<th>Corpora</th>
<th>Variety</th>
<th>Medium</th>
<th>Dates</th>
<th>Studied by</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOB/FLOB</td>
<td>British</td>
<td>written</td>
<td>1960s; 1990s</td>
<td>López-Couso and Méndez-Naya 2012a</td>
</tr>
<tr>
<td>DCPSE</td>
<td>British</td>
<td>spoken</td>
<td>1960s; 1990s</td>
<td>López-Couso and Méndez-Naya 2012a</td>
</tr>
<tr>
<td>Brown/Frown</td>
<td>American</td>
<td>written</td>
<td>1960s; 1990s</td>
<td>López-Couso and Méndez-Naya 2012a</td>
</tr>
<tr>
<td>TEA</td>
<td>Canadian</td>
<td>spoken</td>
<td>2000s</td>
<td>López-Couso and Méndez-Naya 2012a; Brook 2011a, 2014</td>
</tr>
<tr>
<td>Helsinki</td>
<td>historical</td>
<td>written</td>
<td>850 to 1710</td>
<td>López-Couso and Méndez-Naya 2012b</td>
</tr>
<tr>
<td>ARCHER</td>
<td>British/American</td>
<td>written</td>
<td>1650 to present</td>
<td>López-Couso and Méndez-Naya 2012b</td>
</tr>
<tr>
<td>ECW</td>
<td>Canadian</td>
<td>written</td>
<td>1860-1930</td>
<td>Brook 2011a, 2014</td>
</tr>
</tbody>
</table>

Table 2.1: Corpora in which comparative complementizers have been studied.

The present work is, to my knowledge, the first to examine apparent time analyses of comparative complementizers across multiple varieties. It is also the first to explore the broader
syntactic neighbourhood (complementation strategies with *seem*) and to study some of the implications of a system saturated by one variant (*like*). More details about the role played by the present work can be found in section 1.3.

### 2.2.2 Historical trajectory

The complementizers come from two distinct historical sources. The older ones, *that* and Ø, are first attested in the Oxford English Dictionary with *seem* and *appear* in the 13th and 14th centuries, though *as though* is also attested in this context at this time (OED, “*seem*, v.2”). The other verbs are first attested in subordinate strategies later. *Look* is occasionally found with a plain *as* plus a subordinate clause in the 15th and 16th centuries; *feel as if* is found in the 16th century; *look as if* is first attested in 1590 (OED, “*look*, v.”), followed by *sound as if* in the 17th century:

(28) It SEMET wel DAT ðe spies ben, And in-to ðis lond cumen to sen. (c. 1250)
(29) Wende forþ in þi way, It SEMES ASTOW were wode, To wede. (c. 1320)
(30) þat it may APERE PAT þe presciene is signe of þis necessite. (c. 1374)
(31) And al be it so þat it SEME THAT thou art in siker place, yet shaltow alwey do thy diligence in kepynge of thy persone. (c. 1386)
(32) When he was borne,.she FELTE AS IF halfe her herte had bene borne and departed from her body. (1534)
(33) Which SOUNDETH AS IF they had said, he should come down. (1651)
(34) It LOOKS AS IF Physicians learnt their Gibberish for no other purpose, than to embroil what they do not understand. (1700)

(OED: “*seem*, v.2”; “*appear*, v.”; “*look*, v.”; “*sound*, v.1”; “*feel*, v.”)

Complementizers that can follow ostensibility verbs predate the extant variants: Middle English had a particle *swilce* (or *swelce*) that could follow verbs analogous to *seem* and introduce a finite subordinate clause (López-Couso and Méndez-Naya 2012b:312,315). The use of *as if* and *as though* as comparative complementizers seems to have begun in the 13th century (López-Couso and Méndez-Naya 2012b:314), this function “very likely derived from the original adverbial use of the connectives” (López-Couso and Méndez-Naya 2012b:322), along the lines of *He eats as if he’s never seen utensils in his life*. *As though* is attested from earlier than *as if*, with four tokens in the Middle English portion of the Helsinki Corpus (López-Couso and Méndez-
Naya 2012b:320), but shortly after *as if* begins appearing in Early Modern English, it becomes the more popular of the two (López-Couso and Méndez-Naya 2012b:320).

*Like* is the newest variant. López-Couso and Méndez-Naya (2012b:316) find an example dating back to 1864 in the Corpus of Historical American English (Davies 2010–). The oldest instances that I have found are in the Oxford English Dictionary, both with *look*; they come from 1866 and 1890, respectively, and from American newspapers:

(35) It LOOKS LIKE they are really humbugging the poor darkey. (1866)
(36) It LOOKS LIKE they are going to run right square into a warmer country before they can be checked. (1890)
(OED, “*look*, v.”)

*Like* was well-enough established by the 1970s for Rogers (1974a:54) to find it intuitively “more freely-occurring” than either *as if* or *as though*. It was becoming popular in spoken American English of the 1980s (Quirk et al. 1985:1033) and 1990s, even in written corpora (López-Couso and Méndez-Naya 2012a:185). By the first decade of the 21st century, it had become very frequent in spoken Canadian English, to the point at which it had overtaken *as if* and *as though* (López-Couso and Méndez-Naya 2012a:185; Brook 2011a, 2014) and was also winning out over *that* and Ø in apparent time (Brook 2011a, 2014). In contrast, the change in the United Kingdom has been delayed; *like* is not found in this context at all in the spoken British data from the 1960s, though by time of the spoken data from the 1990s it represents 35 percent of the data (López-Couso and Méndez-Naya 2012a:185).

Written media have also lagged behind speech: as late as the mid-1990s, Fee and McAlpine (1997:304) advise that after verbs such as *look*, “*like* is used as a conjunction in speech [but] in writing it should be replaced with *as if*”. In all of the written corpora studied by López-Couso and Méndez-Naya (2012a:185), *like* is more common in “imaginative prose...particularly fictional dialogue”. Indeed, in historical written Canadian data, fictional dialogue is the only
context in which the complementizer *like* appears at all (Brook 2011:45-46). It still shows signs of being the most casual variant, while *as if* acts the most formally and is the most likely variant in informational writing (López-Couso and Méndez-Naya 2012a:185-186).

Even recent linguistics literature that touches on comparative complementizers shows a general cross-Atlantic split in terms of the reception to *like*. This reflects the fact that the change is more advanced in North America than it is in the United Kingdom (López-Couso and Méndez-Naya 2012a:185). British sources that are either older to begin with or written by elderly writers are noticeably more resistant. Quirk et al. (1985) acknowledge the comparative complementizer *like* in two places, but are wary of it:

“[*Look, sound, and feel*] are also used informally, especially in American English, with *like* in place of *as if*” (Quirk et al. 1985:1033).

“The verbs of ‘seeming’ ...are complemented by an adverbial clause beginning *as if* (or less frequently *as though*) in sentences such as the following: Jill looked as if she had seen a ghost; *It seems as if* the weather is improving. [...] After the same verbs, one also frequently hears clauses introduced by *like*, but these are often regarded as nonstandard: *It seems like the weather is improving*” (Quirk et al. 1985:1175).

Twenty years later, at least some British authors still consider *like* to be nonstandard. Although Gisborne (2010:267-281) is untroubled by it, in the opinion of Mair and Leech (2006:318), “a rather small number of alleged syntactic innovations are strongly stigmatized...includ[ing] the use of *like* as a conjunction (as in *And it looks like we could even lose John*).” Likewise, Huddleston and Pullum (2002:1158) acknowledge a “quite strong tradition of prescriptive opposition” to this *like*.

Contemporaneous perspectives from the other side of the Atlantic Ocean provide a sharp contrast and suggest, in accordance with Quirk et al. (1985:1033), Huddleston and Pullum (2002:1158), and López-Couso and Méndez-Naya (2012a:177) that *like* is much more accepted in North America. Asudeh and Toivonen (2007:2) say only that *like* seems to be slightly less formal...
than *as if* and *as though*. This presents a contrast with the evaluation of *like* as “nonstandard” (Quirk et al. 1985:1175) and “strongly stigmatized” (Mair and Leech 2006:318) from the United Kingdom. Even more striking is that D’Arcy (2007:392) considers a token of *feel like* plus a finite subordinate clause to exemplify one of several “largely unremarkable uses” of the word *like*, distinct from “vernacular uses/functions” such as the quotative *be like*, the approximative adverb, and the discourse particle. She acknowledges “sectors of the population who continue to advocate the use of ‘as (though)’ in speech” (2007:413), but the complementizer *like* is well-established enough not to be included among the innovative uses of the word.

This disparity makes a cross-Atlantic study of vernacular speech especially necessary, in order to address the question of just how far behind British English is when it comes to the change towards *like*, and to investigate whether the underlying grammatical conditioning shows any meaningful differences. The fact that a lag can be inferred even from the linguistics literature suggests that it is considerable. Chapter 5 will test this hypothesis; Chapters 6 and 7 will go on to find crucial insights in the ability to compare the results from Canadian English to those from British English.
Chapter 3
Epistemicity, evidentiality, and grammaticalization

3.1 Introduction

As Chapter 2 indicates, the semantic/pragmatic characteristics of ostensibility verbs are often inextricable from their syntactic/lexical functions. The interpretation of the results in the present work will build on previous work associated with both lines of research. In particular, the previous finding that NP subjects with ostensibility verbs are necessarily marked for direct evidentiality (direct perception of that NP) (Rett et al. 2013, Rett and Hyams 2014) will play a large role in Chapters 6 and 7.

Recent work has also raised the possibility that the collocations seems like, looks like, and sounds like (with a matrix expletive, either overt or null) are undergoing grammaticalization and taking on new uses as epistemic markers (Kärkkäinen 2003:45; López-Couso and Méndez-Naya 2010, 2014), and this also has implications for my findings. The present work – focused on two levels of morphosyntactic change – is not well-positioned to evaluate this hypothesis directly, but it will turn out to inform the results.

This chapter begins with definitions of epistemic and evidential markers, including examples. Then, I turn to how these concepts have been applied to ostensibility verbs – in particular, the differing epistemic/evidential status of matrix expletives and matrix NPs. This background will go on to play a crucial role in the interpretation found in Chapter 7.

---

27 Some caution is warranted here. Even if grammaticalization as a framework is accepted (see section 3.34), whether its conventional sense actually applies to pragmatic change is unclear; for instance, recent research on the diachronic trajectories of two classes of pragmatic markers in Canadian English reveals minimal support for gradual grammatical development (Denis 2015).
3.2 Epistemic and evidential markers

3.2.1 Definitions

Epistemicity and evidentiality are closely-related concepts referring to types of pragmatic information about the utterances being produced. **EPISTEMIC MARKERS**, or **EPISTEMIC MODALS**, reduce the level of speaker commitment to the following proposition (Palmer 1986; Thompson and Mulac 1991a:243, 1991b:314; Bybee et al. 1994; Kärkkäinen 2003). The prototypical example of an epistemic marker in present-day English is *I think* (Thompson and Mulac 1991, Kärkkäinen 2003, Denis 2015); rather than literally meaning ‘I am thinking’ or ‘my opinion is’, it is frequently used to “express a mild degree of speaker commitment” (Thompson and Mulac 1991:249), i.e. to put some distance between the speaker and their assertions. Other examples of epistemic markers from present-day English (adapted from Thompson 2002:138) include *I believe, I bet, I guess, I hope, I imagine, I realize, I know, I suppose, I’m sure, I think, I wonder, I’m afraid that, I’m convinced that, and it’s possible that.*

**EVIDENTIALS** or **EVIDENTIAL MARKERS** “indicate how and to what extent speakers stand for the truth of the statements they make…indicate both source and reliability of the information” (Rooryck 2001:125; see also Aikhenvald 2004, Matthewson 2010). Evidentials express whether the subsequent assertion is something that the speaker knows, and if so how they have learned it. Examples (again adapted from the list by Thompson 2002:138) are *it has been said, I can tell, I experience, I find, I find out, I heard, (3rd person) knows, I read, this shows, and this suggests.*

3.2.2 Distinction

---

28 Alongside pointing out the many other names for these, Denis (2015:152) is careful to draw the distinction between ‘epistemic’ (pertaining to the speaker’s knowledge) and ‘doxastic’ (pertaining to the speaker’s beliefs). As other authors do (see e.g. Thompson and Mulac 1991b:317, Cappelli 2007:9, Kärkkäinen 2003:17), I will consider both of these to be covered by ‘epistemic’, but see Denis (2015:152, note 1) for relevant references.

29 A more formal semantic definition of epistemicity relies on e.g. “a quantification over epistemically accessible possible worlds” (Matthewson 2010:334), but the looser one is adequate for the purposes of the present work.
Epistemicity is generally referred to as a type of modality (see e.g. Bybee et al. 1994:Chapter 6), but whether evidentiality qualifies as such a thing – and/or whether it is just a specialized type of epistemicity – is contentious30 (see e.g. discussion and references in Rooryck 2001:125, Kärkkäinen 2003:Chapter 2, Matthewson 2010, Gisborne 2010:Chapter 7, and Rett and Hyams 2014:176). Likewise, the distinction between epistemic and evidential markers is not clear-cut (Kärkkäinen 2003:19). Even if the two classes are treated as separate, Matthewson (2010:335) points out that some analyses derive evidentials from epistemic modals, and some derive epistemic modals from evidentials. Accordingly, Matthewson herself has considered the possibility that the set of epistemic modals and the set of evidentials are ultimately indistinguishable (2010:335).

Either way, it is the case that epistemicity and evidentiality are both centred on subjectivity (Lyons 1977, Palmer 1986, Kärkkäinen 2003:19). Gisborne (2010:240) points out that the extent to which epistemic markers and evidentials differ depends on “whether speakers attenuate their commitment to a proposition when they use an evidential construction”. In other words, does anything used to express evidentiality necessarily convey epistemicity as well? Gisborne himself (2010:242) concludes that it does, pointing to the analysis of Rooryck (2001:125), in which “evidentiality marks both the source and the reliability of a proposition” and the analogous view of Faller (2002:79), in which “one’s judgment of the truth of a proposition is at least in part influenced by one’s source of information”. He concludes that there is “overlap between the

30 Complicating the matter is the fact that evidentiality in particular is realised very differently across languages and dialects (Aikhenvald 2004, Rett and Hyams 2014).
categories ‘evidential’ and epistemic’...this overlap gives rise to a measure of ambiguity”. The ambiguity, in turn, can lead to semantic change\textsuperscript{31} (Gisborne 2010:242).\textsuperscript{32}

Several authors (e.g. Brinton 2008, Rodríguez Louro and Harris 2013) group epistemic and evidential markers together in discussing grammaticalization. While this is not necessarily misguided, I will need to rely on a more precise distinction as a result of there being two separate ways in which the literature has discussed epistemic/evidential factors with respect to ostensibility verbs.\textsuperscript{33} Therefore, I follow the (mostly discrete) definitions for these concepts laid out in section 3.2.1. The results from this study suggest that at least in the case of ostensibility verbs, epistemicity and evidentiality are either different in the first place or sufficiently far apart on a continuum that treating them as divergent is unproblematic.

3.2.3 The I think class of epistemic markers

English has several epistemic adverbials, such as maybe and perhaps (Thompson and Mulac 1991b:313, Kearns 2007, Rodríguez Louro and Harris 2013:419), but another familiar set of epistemic markers in present-day English are the ones that I refer to as the I THINK CLASS. These are structures that have a first-person subject and then a verb that refers, in its most literal form, to mental activity. When used to indicate speaker commitment, I think, I believe, I guess, I reckon, and I suppose all fall into this category (see e.g. Thompson and Mulac 1991a, 1991b; Biber et al. 1999:667-669; Thompson 2002; Kärkkäinen 2003:38-39; Rodríguez Louro and Harris 2013; Denis 2015). Although there may be slight differences in meaning within the set – for instance, the

\textsuperscript{31} Note the connection here to the idea of ambiguity as a prerequisite for reanalysis in general (see e.g. Harris and Campbell 1995:51).

\textsuperscript{32} Epistemic markers may also be able to turn into evidentials over time (De Haan 2001; Gisborne 2010:241) and vice versa (Gisborne and Holmes 2007). Needless to say, if at least one of these things is the case, the task of disentangling epistemic markers and evidentials becomes even more complicated and nebulous.

\textsuperscript{33} López-Couso and Méndez-Naya (2014:38) make much the same point in saying that the verbs seem, appear, look, and sound “combine the notions of epistemicity and evidentiality”.
conviction of Thompson and Mulac (1991b:325) that “I think is a stronger assertion of belief than I guess” – they all function similarly.

Collocations such as I think qualify as epistemic markers not only when they take complements, but also when they are found elsewhere in the sentence. Clause-finally, for instance, markers of the I think class are typically referred to as PARENTHEticalS or more precisely EPISTEMIC PARENTHEticalS (EPs)\(^34\) (Thompson and Mulac 1991a:239, 1991b:317; Thompson 2002:143; Kärkkäinen 2003:39; López-Couso and Méndez-Naya 2010, 2014; Rodríguez Louro and Harris 2013:416; Denis 2015).

Epistemicity is centered on speaker commitment and stance and subjectivity (Kärkkäinen 2003:22). Almost by definition, then, the subjects of the I think class in practice are almost always first-person – occasionally second-person in the case of questions (Thompson and Mulac 1991b:322). Referring to the commitment-level of a third-person subject to an utterance is likely to be nonsensical (Palmer 1986:51; Thompson and Mulac 1991a:243, 1991b:322-323; see also Rodríguez Louro and Harris 2013:419). Kärkkäinen (2003:43) makes a point of describing “hearsay evidentials” e.g. she said or they told me, but these are separate from the I think class.

With the exception of feel (see section 2.1.7), ostensibility verbs differ from the I think class in two major ways: one, if they have first-person nouns (or any other full NPs) in subject position, these will have undergone copy-raising from the lower clause and are not experiencers the way the subject I is in the I think class. Two, these verbs do not directly refer to mental processes; even when they have first-person subjects, seem, appear, look, and sound are not about thoughts, or any other kind of internal reactions. Rather, they are about general apparentness.

\(^{34}\) Kärkkäinen (2003:39) points out that the ‘parenthetical’ category encompasses several non-epistemic pragmatic markers, e.g. you know and I mean.
That said, there are two ways in which these verbs more can be – and are – described as pragmatic markers, depending on subject and/or complement. These are shown in Table 3.1 below.

With full NP subjects and any type of complement (i.e. not just finite subordinate clauses), the perception verbs seem, appear, look, and sound have been described as evidentials (see e.g. Gisborne 2010:Chapter 7); I explore this idea in more detail in section 3.2.4. Quite apart from this, the same verbs with finite subordinate clauses and expletive subjects – in which case they are not evidentials and do not commit to an interpretation involving direct perception, as section 3.2.4 will make clear – have been interpreted as being emerging markers of epistemicity (see e.g. Kärkkäinen 2003:45, López-Couso and Méndez-Naya 2010, 2014); I discuss this in section 3.2.5.

<table>
<thead>
<tr>
<th>Perception verb +</th>
<th>Finite subordinate clause</th>
<th>NP/AdjP/etc. complement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full NP subject</td>
<td></td>
<td>Evidential markers</td>
</tr>
<tr>
<td>Expletive subject</td>
<td></td>
<td>New epistemic markers?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ungrammatical)</td>
</tr>
</tbody>
</table>

Table 3.1: Two ways in which the perception verbs seem, appear, look, and sound have been described as pragmatic markers depending on their subjects and complements.

### 3.2.4 *Seem, appear, look, and sound* as evidential markers

Thompson (2002:138) compiles a list of evidential markers in English that includes both seem and looks like. She is not the only scholar to have categorized these verbs this way. With NP subjects, seem, appear, look, and sound are sometimes described as evidentials regardless of the type of complement (Gisborne 2010:Chapter 7). That is, even if the complement is (e.g.) an adjective phrase (37) or a noun rather than a finite subordinate clause, a nominal subject of a

---

35 *Appear* is being neglected in this chapter. It generally receives far less attention in the literature than *seem* does, and the present work is no exception in this regard. There are so few tokens of *appear* in the data under study (6 out of 1285 finite tokens) that it is tempting to conclude that *appear* is quite stylistically marked for formality and therefore unlikely to be found in sociolinguistic interviews (see also Brook 2011a:47).
perception verb is interpreted as the percept, i.e. what is being perceived (see e.g. Rogers 1974:38-41, Quirk et al. 1985:203-204) – in this case, the entire clause expresses evidentiality. This function is both unavailable and irrelevant when the matrix subject is an expletive (as in 40), as an expletive cannot be the target of perception.

(37) They SOUND LIKE they don’t know what they’re talking about.
(38) They sound weird.
(39) It SOUNDS LIKE they don’t know what they’re talking about.
(40) * It sounds weird (where it is an expletive).

With nominals as percepts (as in they in 37 and 38), the evidential status of the verbs does not depend on what follows the verb (Gisborne 2010:Chapter 7).36 Regardless of the type or length of the complement, sound or seem (or look or appear) with an NP subject fundamentally indicates how the speaker has acquired the information37 – usually through direct perception of the NP subject, or possibly by receiving a report about the NP (also directly), rather than by inference (Gisborne 2010:Chapter 7; Rett et al. 2013).

With NP subjects, look and sound are more precise in their evidentiality than seem (and appear) since they denote the exact sensory apparatus underlying the perception (Rogers 1974:91; Heycock 1994:289; see also Rett et al. 2013, Rett and Hyams 2014:182-183). Seem and appear, on the other hand, are “the most bleached38 perceptual source verbs...[they] impose no substantive restriction on the sensory or mental mode through which the stimulus is perceived” (Landau 2011:788). However, this does not invalidate seem’s claim to being an evidential marker

---

36 There may be a single exception. Rett and Hyams (2014:177) point out that Grimm (2010) argues for Subject-to-Subject raising being ambiguous in terms of direct/indirect evidentiality. This will prove to be central to the argumentation in Chapter 7.

37 Thanks to Aaron Dinkin (p.c., 23 June 2016) for pointing out that this is true only when no alternative perceiver is specified. That is, I sound weird implicates the speaker as perceiver, but I sound weird to Molly specifies someone else in that role.

38 Note that this does not necessarily indicate a belief that seem and appear were once less bleached – simply that they are semantically weak now.
with an NP subject; the only major semantic difference is that *seem* “is neutral with respect to evidence source”,\(^{39}\) i.e. particular sensory mode (Rett et al. 2013; see also Heycock 1994:289).\(^{40}\)

Copy-raised subjects with *seem, appear, look, or sound* are NP subjects, which makes them sources of perception; copy-raising therefore involves evidential uses of the verbs according to Gisborne (2010:Chapter 7). Copy-raising is optional, but can only take place if the source of the observation within the subordinate clause is direct perception of the NP to be copy-raised (Rett et al. 2013, Rett and Hyams 2014). For this reason, Rett et al. (2013) and Rett and Hyams (2014:177) see copy-raised NPs as a small, self-contained system of grammatically encoded evidentiality in present-day English.

This last idea is not uncontroversial. Rett et al. (2013) acknowledge that other scholars believe that “evidentiality is neither [embedded in the grammar] nor obligatory” in modern-day English. However, it seems reasonable to say that that copy-raising either *is* a small evidential system in present-day English or represents the closest match that present-day English has to such a thing. The results of this work speak directly to the reorganization of this system, and the notion of ‘variation’ between matrix NPs and matrix expletives for copy-raising structures being governed by direct perception will prove to be highly important. Both will be addressed in Chapter 7.

An expletive matrix subject of a copy-raising verb, on the other hand, does not lead to an evidential construction and does not commit to representing direct perception; research in both theoretical linguistics (Rogers 1974:77-78; Asudeh and Toivonen 2012) and L1 acquisition (e.g.

---

\(^{39}\) Gisborne (2010:247) is somewhat less confident; he proposes that *look and sound* are much more “straightforward” in their behaviour when it comes to indirect evidence.

\(^{40}\) Similar is a reviewer comment mentioned by López-Couso and Méndez-Naya (2014:58, note 8): that perhaps *look* and *sound* have enough semantic content to be more in need of the complementizer *like* than *seem* and *appear* are. My sense is that this may well account for why *look* and *sound* started taking comparative complementizers along the lines of *as if* in the first place, but that it does not have much of an effect on the modern synchronic level.

(41a) John looks like he’s sick.
(41b) It looks like John is sick.

As the argument goes (Rogers 1974:76-78, Asudeh and Toivonen 2012:331, Rett et al. 2013, Rett and Hyams 2014:177), a speaker saying that John looks like he’s sick is reporting having seen John and spotted evidence of sickness affecting him. However, a speaker saying that it looks like John’s sick has not necessarily seen John at all; this sentence is necessarily ambiguous when it comes to the speaker’s source of evidence. As Rett et al. (2013) note, any number of things other than direct perception could suggest the notion of John being ill, and some of these are indirect: a doctor’s note asking John to be excused from work/school, a lot of tissues and cough-drop packages in John’s wastebasket, etc.

Rett et al. (2013) and Rett and Hyams (2014:177) lay out the differing evidential properties of (41a) and (41b) as follows in Table 3.2. The key is not so much that expletive subjects are versatile – or “unmarked for [directness of] evidentiality” (Rett et al. 2013) – but that copy-raised subjects are not. They unequivocally indicate direct evidence.

<table>
<thead>
<tr>
<th></th>
<th>Copy-raised subject</th>
<th>Expletive subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct evidence</td>
<td>acceptable</td>
<td>acceptable</td>
</tr>
<tr>
<td>Indirect evidence</td>
<td>unacceptable</td>
<td>acceptable</td>
</tr>
</tbody>
</table>

Table 3.2: Epistemic and evidential complement-taking phrases – adapted from Rett et al. 2013, Table 1 and Rett and Hyams 2014:177, Table 1.

The perception experiment performed by Rett et al. (2013) and Rett and Hyams (2014) with adult L1 speakers of English supports this interpretation. The results show that although copy-raised subjects “are acceptable (to some degree) in direct scenarios, they are absolutely

---

41 Rogers (1974) labels this a presupposition; Asudeh and Toivonen (2012:334-335) argue that it is better considered an entailment; and Rett and Hyams (2014:178) see it as fundamentally non-truth-conditional.
unacceptable in indirect scenarios” (Rett et al. 2013). That is, John looks like he’s sick cannot be used to comment on e.g. John’s wastebasket in the absence of evidence taken from the perception of John himself.

In their follow-up corpus study focusing on child language (Rett et al. 2013, Rett and Hyams 2014), there were likewise no cases in which children were observed to use a copy-raised subject when they had only indirect evidence (Rett et al. 2013).  

This is consistent with the resolution of what Asudeh and Toivonen (2007:7-8, 2012:331-335) refer to as “the puzzle of the absent cook”. They intuit that Tom seems like he’s cooking cannot be uttered felicitously if the only evidence for the statement is someone standing in Tom’s kitchen and seeing a meal in the process of being made (2010:335):

In sum, the solution to the puzzle of the absent cook is that a copy-raising subject is interpreted as the [percept]...and ascribing the role of [percept] to the subject is infelicitous if the individual in question is not perceivable as the source of the report.

In other words, Tom is a copy-raised NP in the sentence Tom seems like he’s cooking, but that necessarily means the evidence must have come from the direct perception of Tom; the reason why the sentence does not work in context is that the perception of Tom in Asudeh and Toivonen’s scenario is based only on indirect evidence and inference.

The other part of why this represents a “puzzle” for Asudeh and Toivonen (2007, 2012) is the question of how Tom could ever be infelicitous in this context if the subject NP does not receive a thematic role from the verb seem. They resolve it by proposing that the copy-raised NP subject does “entail a source of perception, but that this source is not connected to an argument or thematic role” (2012:335) – that percepts, like overt perceivers e.g. me in it seems to me, “are not theta roles in the sense of the Theta Criterion” (2012:369).

42 Curiously, Asudeh and Toivonen (2012) report that 6.35% of their survey respondents find copy-raised subjects grammatically unacceptable even in the most unmarked and canonical examples. It is tempting to wonder whether these speakers turn to other structures when they want to convey direct evidentiality.
Those last nuances aside, these insights from psycholinguistic and theoretical studies point towards a copy-raised NP is necessarily a source of direct perception, whereas a matrix expletive with a copy-raising verb can be used for direct evidentiality in this way but need not be; it is equally applicable in the case of indirect evidence, hearsay, guessing, etc.

3.2.5 *Seems like, looks like, and sounds like as epistemic markers*

A different matter is whether *seems like, looks like, and sounds like* – with expletive subjects rather than copy-raised NPs – are evolving into epistemic markers via grammaticalization (Kärkkäinen 2003:39; López-Couso and Méndez-Naya 2010). While the collocations *(it) seems like, it looks like, and it sounds like* did not originate as descriptions of the speaker’s mental activity the way *I think* and *I guess* did, they share the syntactic property of being able to take a finite subordinate clause. They can also likewise be used as hedges (Liberman 2013), and there is considerable overlap between hedging and epistemicity (see e.g. Kärkkäinen 2003:22; López-Couso and Méndez-Naya 2014:55; Denis 2015:206-207).

The next section defines grammaticalization, explores why it is not uncontrovertial to apply it to the matter of epistemic markers, and then reviews what has been said about the grammaticalization of both *I think* class epistemic markers and *seems like* ones. The present work does not evaluate these claims – some of the sources of evidence are outside the variable context – but the background will play a role in the analysis, especially in Chapter 7.

3.3 **Grammaticalization**

3.3.1 **Introduction**

GRAMMATICALIZATION (alternatively GRAMMATICALIZATION or occasionally GRAMMATIZATION) is used here to refer to the diachronic process of “grammatical meaning develop[ing] out of lexical meaning” (Bybee and Pagliuca 1985:59).43 This might involve either a single lexical item

---

43 A survey of overlapping recent definitions can be found in Norde (2009:6).
taking on new grammatical functions or “a frequently-used sequence of words or morphemes becom[ing] automated as a single processing unit”\textsuperscript{44} (Bybee 2003; see also Kuryłowicz 1965). In both cases, there is “reanalysis of one kind of pattern...as another kind of pattern” (Thompson and Mulac 1991b:324).

Grammaticalization is not comprised exclusively of modern studies (see discussion by Lehmann (1995[1982]:Chapter 1), but the study of it in recent times is generally said to have been pioneered by Meillet (1948[1912]). Along with analogy, Meillet coined the term ‘grammaticalization’ to describe one of two ways in which new grammatical units develop. More recently, seminal texts on grammaticalization have included those of Kuryłowicz (1965), Lehmann (1995[1982]), Heine et al. (1991), Hopper and Traugott (1993/2003), and Bybee et al. (1994).

The pathways of change along which grammaticalization develops are conventionally labelled ‘clines’. Perhaps the most rudimentary of these is the one proposed by Givón (1979:208-209), who proposes the following tendency – or “meta-cline”, as Norde (2009:55) puts it:

\textbf{discourse > syntax > morphology > morphophonemics > zero}

Similar are clines that are described by Hopper and Traugott (1993), e.g.:

\textbf{content item > grammatical word > clitic > inflectional affix}

How pragmatics fits into a cline of grammaticalization is a bit nebulous. Some analyses have included later-stage grammatical and pragmatic change under the label of grammaticalization (see Thompson and Mulac 1991b:325), but other authors are less comfortable with this (see discussion by Norde 2009:20-21), and a few even use a separate label – ‘pragmaticalization’ or ‘pragmaticization’ – for changes that reach into the realm of pragmatics (see e.g. Traugott 1982, Brinton 1996, Wischer 2000, Rodríguez Louro and Harris 2013). Since grammaticalization

\textsuperscript{44} This is comparable to what López-Couso (1996) calls a “fossilized expression”.

47
involving comparative complementizers is not especially dramatic and since the related literature (see section 3.3.5) refers to it as ‘grammaticalization’, I will do the same.

3.3.2 Key principles of grammaticalization

Several authors have established central principles associated with grammaticalization, notably Hopper (1991) (see also discussion by Thompson and Mulac 1991b, Heine et al. 1991:20, Bybee et al. 1994, Denis 2015). While these lists tend to display considerable overlap, different authors make different choices.45 To start, I paraphrase the principles proposed by Hopper (1991):

_Divergence:_ When a lexical item becomes a more grammatical entity, the language may retain the old lexical use as well as the new grammatical one.

_Layering:_ Older functions may continue to exist alongside their more developed counterparts: at any given time, newer and older ‘layers’ of functions may coexist.

_Specialization:_ The greater the degree of grammaticalization, the smaller the number of related forms that are affected.

_Persistence:_ A form undergoing grammaticalization is likely to retain traces of the earlier meaning of the lexical form, at least up to a point.

_Decategorialization:_ Grammaticalizing forms tend to become harder to classify in terms of word class, and when the classification can be pinned down, it will usually be something more functional or offbeat than everyday nouns and verbs.

Other aspects of grammaticalization that are frequently referred to include the following:

_Semantic change_ (see e.g. Bybee et al. 1994:6, Thompson and Mulac 1991a: 239): This underlies Hopper’s principle of persistence. Often referred to as “bleaching” (Givón 1975) or “semantic reduction” (Bybee et al. 1994:6), the semantic changes associated with

---

45 Heine and Reh (1984:16) even assemble different sets of principles for different types of grammatical change.
grammaticalization typically\textsuperscript{46} involve the (partial or complete) loss of the lexical meaning. This semantic reduction “correlates with a generalization of the contexts in which the [unit] can be used” (Bybee et al. 1994:6).

**Phonological reduction** (see e.g. Bybee et al. 1994:6): As a form undergoes grammaticalization, it tends to become truncated, or become unstressed and lose phonetic detail. For instance, vowels can become centralized and consonants may be deleted. The result is that the unit is both shorter and less-complex in terms of the articulatory effort. Multi-word pragmatic markers in the process of grammaticalizing “often become fused (e.g., kind of, sort of > kinda, sorta)” (Denis 2015:9).\textsuperscript{47}

**Unidirectionality:** A classic aspect of traditional approaches to grammaticalization (e.g. Traugott 1982, Heine et al. 1991, Bybee et al. 1994:12-14) is that grammaticalization proceeds in only one direction as a result of semantic/pragmatic properties of language change. This notion quickly became contentious. In particular, unidirectionality does not seem readily applicable to pragmatic change (Ocampo 2006; see also discussion by Norde 2009:23).\textsuperscript{48} Note, for instance, that a change from syntax to pragmatics defies the cline proposed by Givón (1979:208-209), which begins with a step from “discourse” to “syntax”.

Receptiveness to counterexamples has accrued only slowly. Norde (2009:1) highlights the fact that the term ‘degrammaticalization’ for a hypothetical opposing process was coined only as part of an assertion that it simply did not exist (Lehmann 1995[1982]). Naturally, this drove some researchers to seek ways of disproving the notion (Norde 2009:1). By the 1990s, Heine et al. (1991:212) dismissively describe these as “exceptions to the unidirectionality principle”. Hopper and Traugott (1993:126) admit that unidirectionality “cannot be regarded as an absolute

\textsuperscript{46} But not always (Hopper and Traugott 1993:87-93; see also Denis 2015:5-6).

\textsuperscript{47} Related, on a morphosyntactic level, is clipping (see e.g. Tagliamonte 2012:79-80).

\textsuperscript{48} The development of discourse markers may be fundamentally different from the development of grammatical units to begin with (Brinton and Traugott 2005).
principle”, and ten years later, the same authors acknowledge the existence of counterexamples but maintain that these “are sporadic, whereas the evidence for unidirectionality is systematic and cross-linguistically replicated”. Even Norde (2009), who undertakes an intensive and comprehensive study of degrammaticalized phenomena, acknowledges that there is an asymmetry in terms of directionality (Norde 2009:53):

I think it would be fair to dismiss unidirectionality as an absolute principle. But as much as I agree with [several scholars who advocate taking all types of change into account], it is also evident that grammaticalization is far more frequent, and far more cross-linguistically regular, than is degrammaticalization...it cannot be denied that directional tendencies exist.

More recent work by Diertani (2011) on specifically morphological grammaticalization/degrammaticalization suggests that the processes are one and the same, and that the higher prevalence of grammaticalization over its opposite is to some degree an illusion and a side-effect of the availability of contexts in which each can take hold.

**Frequency increase:** With reference to work by Hook (1991), Thompson and Mulac (1991b:319) claim that “an increase in a[n] item’s text frequency is an important concomitant of its grammaticization”. In other words, as an item takes on new functions, its frequency as a proportion of the words in a data-set can be expected to climb.

**Gradualness** (see e.g. Lichtenberk 1991, Hopper and Traugott 1993:3, Bybee et al. 1994:4, Brinton and Traugott 2005, Traugott and Trousdale 2010, Denis 2015:13): Another traditional principle of grammaticalization currently under dispute is the notion that grammaticalization occurs slowly, involving many situations “with older and newer forms coexisting side by side” (Norde 2009:16), as in the notion of layering. Some scholars use ‘gradualness’ to refer to changes that start out only slowly or with uncertainty on the part of the speech community (see Norde 2009:16-17 for discussion), but recently, a number of studies of discourse-pragmatic change have uncovered very little support for gradual grammatical development, at least in the case of the specific variables they examined (Tagliamonte and Denis 2010; Pichler and Levey 2011;
Denis 2015). More broadly, one criticism of grammaticalization theory in general is based on the idea that gradual, continuous change is an illusion based on an overly language-centered view; it is unrealistic from the standpoint of how transmission works between successive generations of language learners (Janda 2001, Joseph 2004).

3.3.3 Grammaticalization as applied to epistemic and evidential markers

The process of grammaticalization is often invoked to account for the emergence of epistemic/evidential markers out of lexical items (see e.g. Thompson and Mulac 1991b:314). When it comes to the *I think* class of epistemic marks, several subject-verb combinations that originally referred to literal mental activities on the part of the speaker have since taken on speaker-centred epistemic/evidential meanings (Thompson and Mulac 1991a, 1991b, Traugott 1995, Rodríguez Louro and Harris 2013:420).

Syntactically, the reinterpretation of *I think* and *I guess* as “unitary epistemic morphemes” (Thompson and Mulac 1991b:315) makes them more adverbial than they were as literal collocations (Thompson and Mulac 1991b:318). This leads to a partial breakdown in the barrier between the main clause and the subordinate clause (Thompson and Mulac 1991a:249; see also Aijmer 1997, Thompson 2002). In turn, this accounts for a low proportion of the complementizer *that* alongside grammaticalizing epistemic markers such as *I think* and *I guess* (Thompson and Mulac 1991a:247-248, Tagliamonte and Smith 2005), and for the related ability of *I think* and *I guess* and the like to be used parenthetically elsewhere in the clause (Thompson and Mulac 1991b:317, Aijmer 1997; see also Rodríguez Louro and Harris 2013:420). *I think* is sometimes even said to have grammaticalized so much that it is pragmatically comparable to markers without epistemic functions e.g. *you know* and *I mean* (Kärkkäinen 2003:184).

---

49 Most of these verbs refer to thoughts, and *reckon* originally meant ‘estimate’ or ‘count’ (Rodriguez Louro and Harris 2013:416-417).

50 This is not unprecedented: there is the analogous case of *methinks* in the 15th and 16th centuries (López-Couso 1996, Palander-Collin 1997, Wischer 2000).
From a semantic/pragmatic point of view, Traugott (1995:32) refers to this type of emergence of speaker-centred epistemic markers as *subjectification*, and defines it as the development of a grammatically identifiable expression of speaker belief or speaker attitude to what is said. It is a gradient phenomenon, whereby forms and constructions that at first express primarily concrete, lexical and objective meanings come through repeated use in local syntactic contexts to serve increasingly abstract, pragmatic, interpersonal, and speaker-based functions.

Rodríguez Louro and Harris (2013:420) present a similar argument, also with reference to the *I think* class of epistemic markers:

The evolution of meaning from a source gram expressing cognition (e.g. *think*) or calculation (e.g. *reckon*) to the encoding of epistemicity and evidentiality illustrates the shift from the concrete, lexical and objective to the abstract, pragmatic, interpersonal and subjective.

### 3.3.4 Controversy

Grammaticalization is not an inevitable part of morphosyntactic change (see e.g. Tagliamonte and Denis 2010, Pichler and Levey 2011, Denis 2015). Even when a case of it seems well-established, the grammaticalization framework in and of itself is contentious. It is possible that conventional grammaticalization, even if it exists and is sound, simply does not capture grammatical shifts in epistemic markers very well (Heine 2014, Denis 2015).

More broadly, grammaticalization has drawn a considerable amount of criticism, both from authors receptive to the framework (e.g. Diewald 2010) and from those who reject the very premise (e.g. Campbell and Janda 2001; Miller 2002; Joseph 2001, 2003, 2004). In different places, grammaticalization – in theory and/or in practice – has been criticized as, for example:

*Narrow and limited:* Rissanen (1997) declares that research projects involving grammaticalization “are often based on rather meagre empirical evidence and they tend to concentrate on individual linguistic elements” (Rissanen 1997). A related issue is that some studies rely on synchronic information alone; Thompson and Mulac (1991b:324) acknowledge that their analysis of grammaticalization has this weakness, and that their data “does not allow us to say anything definitive about whether the synchronic [patterning]...has a diachronic
parallel”. Rissanen (1997:1) also pushes for “a more comprehensive diachronic approach to the study of grammaticalization”. Ideally, an evaluation of grammaticalization ought to have a lot of time-depth and a lot of data (see e.g. Denis 2015).

**Redundant, uneconomical, and/or epiphenomenal:** Norde (2009:29) summarises the critical position of Newmeyer (1998:235) as follows:

Since semantic change, phonetic reduction, and reanalysis, Newmeyer argues, may occur independently from one another, there is no reason to assume separate dynamic force (i.e. grammaticalization) when these changes happen to co-occur. For the same reason, he argues, there is no need for a ‘grammaticalization theory’.

Several scholars working from the perspective of historical linguistics have similar major problems with grammaticalization: that grammaticalization can be accounted for in its entirety by mechanisms such as reanalysis, borrowing, extension, analogy, sound change, and semantic/lexical change (Newmeyer 1998, 2001; Campbell 2001; Campbell and Janda 2001; Janda et al. 2001; Joseph 2004). Diewald (2010) acknowledges this as one of three major “problem areas” in grammaticalization studies. Campbell (2001:158) summarizes his argument by asserting that grammaticalization has no independent status of its own. Cases of grammaticalization are explained adequately by the other mechanisms of linguistic change, and grammaticalization explains nothing by itself but must rely on these other mechanisms and kinds of linguistic change.

For instance, bleaching may be characteristic of semantic change in general rather than grammaticalization (Campbell 2004:296), and phonological change may be characterized by a tendency towards reduction regardless of whether there is grammatical change occurring (Campbell 2001:121, 2004:297). How to define grammaticalization given that it appears to apply to most cases of language change – and given that ‘grammar’ is difficult to define in the first place – is another issue (Campbell 2001:155).

While these objections from both within and outside the area of grammaticalization studies do not make grammaticalization inherently useless as a framework, they do necessitate caution (see e.g. Norde 2009:52). While referring to grammaticalization may be useful as a shorthand, or
as a way of locating changes with key commonalities (Campbell 2001:158), it should not be assumed to be something distinctive.

### 3.3.5 Grammaticalization as applied to comparative complementizer structures

Several scholars whose work touches on the *seems like* constructions assume that grammaticalization has occurred somewhere along the way. For instance, Gisborne (2010:269) describes the comparative complementizers (*as if, as though, and like*) as “an example of the kind of minor construction that arises in the process of grammaticalization.” Even more specifically, Kärkkäinen (2003:45), in a survey of epistemic markers in American English, points out that when *seem, look, and sound* are used epistemically in her data, “the construction is usually *not* followed by the complementizer *that* (... it is more likely to be followed by *like*)... which implies that such phrases may also be grammaticalizing into unitary epistemic phrases”.

*Seems like, sounds like,* and *looks like* are slightly different from the *I think* class of epistemic markers. The collocations that are potentially grammaticalizing into what López-Couso and Méndez-Naya call “*like*-parentheticals (2014:41-42) are the ones with expletive subjects – overt or null (López-Couso and Méndez-Naya 2010; 2014:41-42) – rather than first- or second-person pronominal subjects as in the case of *I think* (Rodríguez Louro and Harris 2013:419; see also Thompson and Mulac 1991b). In this case, the fact that this is the speaker’s viewpoint is implicit (*it seems like*) rather than overt (*I think*). The shift from their literal denotations of ostensibility to speaker-centered impressions and evaluations qualifies them as a case of SUBJECTIFICATION (see e.g. Langacker 1985, Traugott 1989:34, López-Couso 2010), a process whereby structures describing the external world come to have internal meanings associated with subjective evaluation.

Another dissimilarity between the grammaticalizing *seems like* structures and the *I think* class is the presence of the complementizer in the former. The loss of the complementizer *that* after *I think* is very regularly attributed to grammaticalization of the *I think* construction
(Thompson and Mulac 1991a, 1991b, etc.); however, when it comes to the potential grammaticalization of seems like, looks like, sounds like, and so on, the complementizer is part of the grammaticalizing form (López-Cuso and Méndez-Naya 2010, 2014:47,52). Commonalities are that the units are short and fossilized, and that the grammaticalizing collocation has the verb “in the present indicative active” tense (Rodríguez Louro and Harris 2013:419).

When it comes to syntactic flexibility, grammaticalization might lead in one of two ways (Traugott 2010). It can result in a narrowing range of positions in the clause (e.g. Palander-Collin 1998:421) or quite the opposite – grammaticalizing forms “gain[ing] syntactic freedom” (Rodríguez Louro and Harris 2013:419; see also Thompson and Mulac 1991b:317) such that they can move around “as other epistemic particles in English do, such as maybe” (Thompson and Mulac 1991b:326). Traugott (2010) places pragmatic markers within the second category. Accordingly, López-Cuso and Méndez-Naya (2014) investigate like-parentheticals that can appear in places other than clause-initial position, and look for three criteria in order to ascertain whether grammaticalization has been taking place:

**Decategorialization**: Like-parentheticals have become detachable from the sorts of clauses that they started out introducing, and do not even always require the overt expletive; this makes them increasingly adverbial, along the lines of the shift from it may be that to maybe (2014:47). The prosody of like-parentheticals also indicates that they are independent (2014:48).

**Layering**: Along with like-parentheticals, there are still finite subordinate clauses introduced by seems like and related forms (2014:48).

**Fossilization**: The like-parentheticals show less variability in terms of tense, aspect, polarity, etc. (2014:49-51).

**Fusion**: The complementizer like is “bound to the verb” in the parentheticals; nothing intervenes (2014:52).
Increasing speaker-centeredness: These like-parentheticals encode “attitudes of the speaker, source of information, text-organization and speaker-hearer interaction” (2014:54).

The variable context of the present study thus exists in a range towards the end of a cline of grammaticalization. It can be assumed that the seems like, looks like, and sounds like collocations when used as subordinators are an older layer beneath the like-parentheticals that exist at the forefront of grammaticalization (López-Couso and Méndez-Naya 2014). There may also be multiple layers of grammaticalization found within the data to be examined here, since the finite subordinate structures “with it-deletion (e.g. Looks like I finished my paper) pattern with parentheticals as regards morphosyntactic fixation, which could indicate that they have already taken the first step down the path of grammaticalisation” (López-Couso and Méndez-Naya 2014:57) – presumably towards increasingly epistemic, rather than literal, uses of the structures.

Again, the present study is not well-positioned to evaluate grammaticalization much further than this; as Chapter 4 specifies, parentheticals and other offbeat grammatical uses of the collocations are excluded. Moreover, the data are almost exclusively synchronic, whereas a proper study of grammaticalization necessitates diachronic data in order to evaluate each criterion over real time (Denis 2015:167).

Worth noting, though, is the fact that like taking over from as if and as though (López-Couso and Méndez-Naya 2012a, Brook 2011a, 2014) suggests specialization of a comparative complementizer that might well pave the way for the parenthetical uses of seems like, sounds like, and looks like, especially as these units can appear clause-finally in a way that as if and as though are more resistant to (Everyone these days has a smartphone, seems as if.)

---

51 There is a fine line between grammaticalization and simple lexical replacement that makes phrases appear to be more invariant. General extenders (GEs) are another case of pragmatic markers that have been interpreted through the lens of grammaticalization (e.g. Cheshire 2007) but have also been argued to be a case of lexical replacement rather than true grammatical change (Tagliamonte and Denis 2010, Pichler and Levey 2011, Denis 2015).
The fact that *like* is the newest complementizer (López-Couso and Méndez-Naya 2012a:140, 2012b:316; Brook 2011a:10) means that there is a risk of conflating new *functions* with new *forms.* *Like* has certainly brought in parentheticals aside from *it seems* and *it appears* (López-Couso and Méndez-Naya 2014), but epistemic uses of the ostensibility verbs predate its ubiquity in North American English. Kärkkäinen (2003:38) files *it seems/appears that* under epistemic phrases, and even Palmer (1986) notes that *it appears that* can be used to lessen the commitment of a speaker to their statement. This does not necessarily mean that *seem, look, sound,* and *feel* have *always* functioned just as they do now, disregarding the complementizer and the parenthetical usages. However, it does leaves open the possibility that these verbs have *long* been able to convey epistemicity and that the arrival of *like* is, to some degree, coincidental. Grammaticalization of ostensibility verbs may have occurred in tandem with the emergence of *like* as a comparative complementizer in a relatively short time (López-Couso and Méndez-Naya 2012a:18; Brook 2011a, 2014), but phenomena that are attributable entirely to the latter are not automatic evidence for the former.

Still, the fact that the present variable context is not necessarily an endpoint along a cline of grammaticalization will prove to be important, because it means that it is a potential intermediate stage from the standpoint of grammaticalization. I revisit this notion in Chapter 7.

### 3.3.6 What about *feels like* or even *I feel like?*

Much less attention has been paid to whether *I feel like* or *it feels like* might be developing into an epistemic marker as well. Earlier, I have mentioned this as a possibility (Brook 2011a:65). Thompson (2002:138) briefly refers to it in that she includes the verb *feel* in her list of epistemic markers in present-day English.\(^{52}\) Denis (2015:169) includes it as one of the members of the *I think* class of epistemic markers.\(^{53}\) *Feel* is a unique hybrid: a verb that assigns a theta-role of

---

52 It is worth pointing out that she does not elaborate on the structures she has in mind, and therefore is not necessarily referring to *feel* plus a finite subordinate clause.

53 See also Quirk et al. 1985:202-203.
experiencer, like the literal *think* (see section 2.1.7) but can also take comparative complementizers (*as if, as though, and like*). This may be surprising given that with a first-person pronoun and a finite subordinate clause, *feel* is not so far removed from the epistemic functions of *I think, I guess, I reckon*, etc. This raises the question of whether *I feel (like)* has been changing more like *seem* or more like *think*; I will revisit the matter in subsequent chapters.

### 3.4 Conclusion

Epistemic markers reduce the speaker’s commitment to an assertion; evidential markers specify the source of information. Epistemic markers in English include the *I think* class, and also *seems like, looks like, and sounds like* with expletive subjects, after undergoing a slight grammatical change from literal ostensibility verbs to markers of increased speaker distance/hesitancy (López-Couso and Méndez-Naya 2010, 2014). Evidentiality in English is not deeply embedded in the grammar but can be indicated by matrix NP subjects and sensory verbs (*look, sound, feel*) regardless of complement type (adjective, noun phrase, subordinate clause), at least aside from Subject-to-Subject raising (Grimm 2010, Rett and Hyams 2014).

The framework of grammaticalization is controversial for multiple reasons, but *seems like, looks like, and sounds like* as epistemic parenthetics fit into it (López-Couso and Méndez-Naya 2010, 2014). What this means for the present study is that a) ostensibility verbs with finite subordinate clauses are an intermediary step on a cline of grammaticalization; and b) the whole finite structure examined here could represent a range of subtly different points along such a cline. The present study is not set up to address these possibilities directly, but they will be kept in mind as two levels of change are examined (in Chapter 5 and 6), and return in Chapter 7 amidst a discussion of the implications of the findings.
Chapter 4
Methodology

4.1 Data

The data analyzed in this thesis are all drawn from large corpora of sociolinguistic interviews (see e.g. Labov 1984) representing vernacular speech, collected between 1995 and 2014 in Canada (Ontario) and the United Kingdom (various).

4.1.1 Canada

The Canadian English data are drawn primarily from the Toronto English Archive (Tagliamonte 2003-2006) and two corpora of outlying dialects: Southeastern Ontario (Tagliamonte 2007-2010), encompassing three smaller communities to the southeast, and Northern Ontario (Tagliamonte 2010-2013), which includes data from six outlying towns, primarily along the province’s Highway 11. All of these corpora are comprised of sociolinguistic interviews (Labov 1984), each about one hour in length; the sample-designs for each corpus are roughly balanced for age and sex. Another source of data from Ontario is the Belleville Oral History Archive,54 collected in 1975 by members of the Youth Opportunity Program. While this is an oral-history project rather than a collection of modern sociolinguistic interviews, the types of questions are similar to those asked in a sociolinguistic interview and previous work has found the Belleville data to be consistent with modern apparent time results from the same community (Denis 2015; Brook in prep.). The Belleville 1975 data thus serves as a real-time comparison to the later synchronic data.

54 Permission to use the Belleville Oral History Project was obtained by Sali A. Tagliamonte through arrangements with the Hastings County Historical Society in May 2007.
<table>
<thead>
<tr>
<th>Location</th>
<th>Year</th>
</tr>
</thead>
</table>
| Toronto             | 2002-07
| Toronto English Archive (TEA) | 2002-2007\(^{55}\) |
| Southeastern Ontario|        |
| Belleville          | 2007-08|
| Lakefield           | 2007-08|
| Burnt River         | 2007-08|
| Northern Ontario    |        |
| North Bay           | 2009-10|
| Timmins             | 2010-11|
| Kirkland Lake       | 2011   |
| Temiskaming Shores  | 2011   |
| Haliburton          | 2012   |
| Almonte             | 2012-13|
| Beaverton           | 2013   |
| Archival data from Belleville |    |
| Belleville Oral History Archive | 1975 |

Table 4.1: Sources of Canadian data: locations and dates of collection.

4.1.2 United Kingdom

The British data include data from the York English Corpus (Tagliamonte 1996-1998), collected in and around the city of York; and the Roots Corpus (Tagliamonte 2001-2003), including speakers mostly over the age of 60 in rural areas around the UK. As with the Canadian data, these are standard sociolinguistic interviews lasting about an hour each. Aside from the age restriction within in the Roots Corpus – and the fact that there are more male than female speakers from Northern Ireland (see e.g. Tagliamonte and Smith 2005:298) – the samples of speakers in the British data are more or-less balanced according to age and sex.

\(^{55}\) Though note that annual interviews with one speaker are ongoing as part of a panel study. See section 4.1.12.
## Variable context

### Broad definition

The dependent variable in this study is the comparative complementizer, with variants *as if*, *as though*, *like*, *that*, and Ø. The variable context under study is defined broadly in terms of syntactic slots; anything that matches the options in Table 4 is included, minus the exceptions listed in 4.3.2.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Verb</th>
<th>Complementizer</th>
<th>Complement</th>
</tr>
</thead>
<tbody>
<tr>
<td>It</td>
<td>seem*</td>
<td>that</td>
<td>finite subordinate clause</td>
</tr>
<tr>
<td>Ø</td>
<td>appear*</td>
<td>Ø</td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>look*</td>
<td>as if</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sound*</td>
<td>as though</td>
<td></td>
</tr>
<tr>
<td></td>
<td>feel*56</td>
<td>like</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Combinations considered part of the variable context of the present study.

Apparent copy-raising where the coindexed nominal in the subordinate clause is not a subject (Potsdam and Runner 2001; Landau 2009), or where there is no coindexed nominal at all (Heycock 1994:294, Potsdam and Runner 2001:457-458, López-Couso and Méndez-Naya

---

56 Note that this includes the past tense form *felt*.
are included. There is no need to exclude these, especially as they turn out to be almost unattested in practice. Of the unambiguous tokens of copy-raising, there is one that corresponds to a non-subject nominal in the relative clause, and one that does not contain a coreferential nominal.57

Similarly, for now I am not concerned with the semantic character of the verbs. It is recognized that look and sound in particular are often ambiguous between the description of a literal sound/appearance and a more abstract epistemic report. Chapter 7 will revisit the notion of epistemicity and evidentiality in order to address this. I have otherwise disregarded this distinction in the present study; that is, instances of look and sound that are literal are not distinguished from the ones that are semantically “quite bleached” (Asudeh and Toivonen 2012:235), especially as a large proportion of tokens of each are ambiguous in this regard. For this reason, all tokens of look and sound with referential subjects are counted as copy-raising and considered to have the same structure. The only potential problem is feel, since tokens of it that fit into Table 4.3 might be true copy-raising or might be simply expressing “propositional attitudes”, i.e. opinions (Asudeh and Toivonen 2012:325).58 All tokens of both uses of feel are included, but the difference between feel and the other verbs will be revisited as part of the discussion of the results.

4.2.2 Justification

Not all of the combinations of verb and complementizer are grammatical: look and sound do not accept that/Ø (Potsdam and Runner 2001:465-466, Huddleston and Pullum 2002:962, Gisborne 2010:276; see also Brook 2011a, 2014:2). There are also numerous subtler restrictions: because of the blocked copy-raising, seem and appear both appear with that/Ø only when the

57 These are both nebulous cases that have played an important role in theoretical analyses of copy-raising (see Chapter 2), but they are very rare in the present data.
58 For more about untangling the two syntactic constructions involving feel, see 2.1.7.
subject is an expletive; and since *feel* that/Ø requires an experiencer as matrix subject, an expletive in this context will be ungrammatical for most speakers.

Nonetheless, for the primary set of results I will combine all of the structures under one variable context, for two reasons. One is that there is no immediate reason to separate them, either by verb or by complementizer. When it comes to the verbs, Quirk et al. (1985:1175) group *seem, appear, look, sound, and feel* as “verbs of ‘seeming’” that take a subordinate clause introduced by *as if or as though* (or *like*). This suggests that the verbs are close enough in terms of both semantics and syntax. Likewise, López-Couso and Méndez-Naya (2012a:178) include this entire set of verbs and even include some examples of *it + BE + like*.

Some authors argue for splitting the complementizers. Gisborne (2010:273) finds *as if, as though*, and *like* interchangeable but describes them as a “unique word class” that “does not behave like *that*” (Gisborne 2010:274). Similarly, López-Couso and Méndez-Naya (2012a) turn to *that* only after an initial analysis and disregard the Ø tokens.

However, Huddleston and Pullum (2002:962) argue that “with *seem and appear* the content clause [introduced by *that*]...can be replaced without any perceptible change of meaning by a phrase introduced by *as if (or as though or like)*”. Likewise, Quirk et al. (1985:1175) state that *appear and seem* when followed by a *that*-clause have a “similar meaning” to the five verbs with *as if/as though/like* clauses.

The classic description of a variable is “two or more ways of saying the same thing” (see e.g. Labov 1972). One of the problems that this presents is how to apply it to morphosyntactic variation, in which synonymy is much harder to establish (Tagliamonte 2006b:71-72). A more developed definition of the linguistic variable (Tagliamonte 2006b:76) involves it exhibiting the following:

a. synonymy or near-synonymy (weak complementarity)
b. structurally embedded, i.e. implicated in structural relations with other elements of the linguistic system, e.g. the phonemic inventory, phonological space, functional heads, grammatical subsystems, etc.
c. correlation with social and/or linguistic phenomena
The first of these criteria is met by the statements of Huddleston and Pullum (2002:962) and Quirk et al. (1985:1175) that the that-clause is similar or identical in meaning to the ones introduced by as if, as though, or like. The second requirement is satisfied because of the nature of comparative complementizers: they are a small, closed class of items following a handful of verbs to help express the apparentness of the proposition in the subordinate clause. Correlations with social/linguistic factors will be demonstrated in Section 3. The split historical origins of the complementizers that follow seem, appear, look, sound, and feel are not necessarily a problem for defining the envelope of variation. Tagliamonte (2006b:72) notes that “[w]hen the linguistic variable lies beyond phonology, the variants may not be similar at all. They may have entirely different lexical sources as well as different histories in the language”.

The difference between a copy-raised subject and the nominal left in situ – It seemed like she understood versus She seemed like she understood – also merits attention, especially as López-Couso and Méndez-Naya (2012a:180) treat these separately. Potsdam and Runner (2001:454) state that the two versions are “synonymous”. Rogers (1974:76) disagrees outright, but then clarifies that the differences are likely to be epistemic/evidential in nature (see also Rogers 1974:91). Horn (1981:354) also disagrees about synonymy, but for reasons very similar to those of Rogers (1974): Horn has the sense that transformation shifts the epistemic nature of the sentence slightly. This is not enough of a disparity to present a concern to the present analysis.

López-Couso and Méndez-Naya (2012a:188; 2012b:326) suspect that the choice of complementizer, especially after seem, depends on the degree of commitment that the speaker has towards the proposition in the subordinate clause, with verbs such as seem entailing “a medium degree of commitment” and as if and as though doing more to lessen this than that and Ø (López-Couso and Méndez-Naya 2012b:326). Given that like has saturated the system in Toronto (López-Couso and Méndez-Naya 2012a; Brook 2011a, 2014), this raises the question of
whether the epistemic value of *seems that* is being lost entirely or whether *like* has simply adopted it and taken on a greater range of potential epistemic values. I will return to this question in Chapter 7.

Tagliamonte (2012:4) points out that “[a]lthough some variants may differ subtly in meaning and distribution, if they are part of a linguistic variable they will be members of a structured set in the grammar.” In sum, in the analyses that follow, the five verbs and five complementizers will be treated as a single variable context in order to establish a baseline. While this is something of an oversimplification, it is necessary as a first step. I leave the matter of how many different overlapping variables are truly represented among the ostensibility verbs plus complementizers to future research.

4.3 Extraction

4.3.1 AntConc search

I downloaded all of the transcripts available as of March 2014 in the corpora described above (section 4.1) housed at the Language Variation and Change Lab at the University of Toronto. I used AntConc (Anthony 2012) in order to extract tokens.

I searched for the following strings. The wildcard symbol enabled me to find instances of the verb across tenses at once; the only exception was the irregular form *felt*, which I covered in an additional search:

- seem* \(\rightarrow\) seem, seems, seeming, seemed
- appear* \(\rightarrow\) appear, appears, appearing, appeared
- look* \(\rightarrow\) look, looks, looking, looked
- sound* \(\rightarrow\) sound, sounds, sounding, sounded
- feel* \(\rightarrow\) feel, feels, feeling
- felt

I sorted the results by alphabetical order of the word to the right, saved the output to text files, and retained all of the instances of the following collocations:

- seem* like
- appear* like
- look* like
- sound* like
- feel* like
I examined the output examples one by one in order to filter out exclusions (see section 4.3.2). In the case of *seem*, *appear*, and *feel*, tokens with no overt complementizer are expected. *Look* and *sound* may also appear very occasionally with a zero complementizer (Huddleston and Pullum 2002:962; López-Couso and Méndez-Naya 2012a:186). Therefore, it was necessary to read through the rest of the instances of the verbs in order to find the tokens with a Ø comparative complementizer.

### 4.3.2 Exclusions

Searching for the strings above yielded a number of examples that fell outside the variable context and thus needed to be excluded:

**NP/adjectival/gerund complements:** These are not finite subordinate clauses.

(42) Yeah you feel- it feels like um, cold water on your back ’cause it slowly trickles down, yeah.
    (Alicia Karan, F 24, Toronto)

(43) I felt like rolling down the window and screaming at the guy next to me, ”What the hell’s going on?!”
    (Richard Yonson, M 49, Toronto)

**Intervening indirect objects:** While these are very much a subset of ostensibility verbs (see e.g. Rogers 1974:54, Gisborne 2010:273), examples of *seem* with an indirect object intervening between the ostensibility verb and the complementizer favour *that* and Ø rather than *like* (Kärkkäinen 2003:38-39).

(44) It SEEMS to me THAT there are stranger and stranger people that you can come across on the subway.
    (Diana Tsang, F 24, Toronto)

---

59 Speakers from the data used for the present work are referred to by pseudonym, sex, age, and hometown.
A search of the data found in the present study finds 55 examples of *seem* to me + [finite subordinate clause], all with *that*/Ø. The only instances of *like* in this context are after three of five tokens of *look*/*sound* to me + [finite subordinate clause]. Because these may represent an offshoot structure, because the token numbers are small, and because they are almost all represented by *that*/Ø rather than *like*, there is very little more to be said about the structures with the extra dative argument and I exclude them from the present analysis. Future research could examine these separately in order to determine how similar the change is to the one described in Chapter 5.

**Noun forms of feel, feeling, etc.** As these are not verbal, they do not form part of the variable context. Although *that*, Ø, and possibly *like* can follow the noun forms of *feel*, *feeling*, etc., in this case all of the complementizers introduce content clauses after a noun, and these are adjuncts (and hence can be deleted without compromising grammaticality) rather than complements.\(^{60}\)

(46) It didn’t have that intimate feel that the Forum used to have, so.
    (Caleb Slade, M 64, Belleville 2007-08)

(47) It took me a long time to sort of get over the feeling that somehow it was my fault.
    (Christy Biggs, F 33, York)

**Quotatives:** While these are an interesting use of perception verbs (see e.g. Blondeau and Nagy 2008, Davydova and Buchstaller 2015:452) and will be revisited later in conjunction with the development of the quotative *be like* (see Chapter 7), they do not involve copy-raising and are therefore excluded here:

(48) I really don’t like to spend it, so when I go shopping I FEEL LIKE, “What am I doing? I’m not going to buy anything, so there’s no point in being here.”
    (Dana Naskauskaas, F 17, Toronto)

(49) And I liked the singing, but – and then, um I FELT AS IF, “Oh well.”
    (Violet Marshman, F 71, Maryport)

---

\(^{60}\) Similar is the one nominal example found by López-Couso and Méndez-Naya (2012a:183) of the noun *look* and the complementizer *as though*: *But there was a look about her mouth as though she were tasting lemons.*
**Ostensibility verbs inside relative clauses:** Syntactic gaps are ungrammatical with several complementizer variants and are attested only with Ø.

(50) Yes, you can often spot the ones that you FEEL will always be a reserved type of person.  
(Elaine Chapman, F 55, York)

**Ambiguously placed Ø complementizers:** When it is impossible to tell whether a discourse marker is intervening material in the higher clause or a left-periphery item in the lower one, the sentence was excluded. Although this means that the zero tokens will be underrepresented in the analysis, there were only a handful of examples (N = 4).

(51) But she began to sort of FEEL you know she could do more with her own life.  
(Derek Burns, M 60, York)

**Parentheticals:** These are relevant to the question of whether some of the [verb + complementizer] combinations are undergoing grammaticalization (see Chapter 6, cf. López-Couso and Méndez-Naya 2010). However, they are not structures that take subordinate clauses, and they do not exhibit variation beyond like and Ø. They are thus excluded from the present study.

(52) But the damage is so much more worse, IT SEEMS.  
(Richard Seppi, M 36, Toronto)

(53) When was it? Years and years ago, it SEEMS LIKE. At school.  
(Nick Hudson, M 17, York)

**Interruptions, restarts, etc.:** Cases in which the subordinate clause was unclear were excluded due to the difficulty in categorizing them.

(54) Was de-- denied them. Which I FEEL um, THAT er, THAT um took away from them.  
(Rose Murdoch, F 69, Maryport)

### 4.4 Coding

### 4.4.1 Dependent variable

The dependent variable in the present study is the set of comparative complementizers that link ostensibility verbs to finite subordinate clauses: *as if, as though, like, that*, and Ø. Examples
can be found in (55) to (59). The British data holds an extra variant: *as* on its own (N=2) as in (60).

(55) That SOUNDS **AS IF** it was very unkind to me, but I can understand it now.
    (Sue Evans, F 69, York)

(56) It **LOOKS AS THOUGH** you’re looking in the mineshaft.
    (Connie O’Brien, F 68, Kirkland Lake)

(57) I like to sleep there a lot because I **FELT LIKE** that was my home.
    (Alice Faddei, F 22, Timmins)

(58) but it **APPEARS Ø** it was a contract they had
    (Brian Whiting, M 82, Maryport)

(59) It **SEEMS THAT** everybody always kind of makes an effort to relate stuff back to English.
    (Brent Kim, M 21, Toronto)

(60) I never **FELT AS** I was doing anything you know, I got the sugar.
    (Helen Philips, F 79, Maryport)

4.4.2 Independent linguistic variables

**Lexical verb:** Since not all of the verbs combine grammatically with all of the complementizers (see section 4.2.2), an effect of verb is expected, as was found by López-Couso and Méndez-Naya (2012a:187) and Brook (2011a, 2014).

**Subject of ostensibility verb:** This study combines examples of matrix expletives, copy-raised subjects, and – in the case of *feel* – base-generated experiencer subjects. While it is not clear that we can expect an overall effect of expletive versus NP subject – in my previous study of a similar data-set, subject type was not significant (Brook 2011a:61) – subject will probably interact with verb since *feel* selects its arguments so differently from the other verbs (see section 2.1.3).

(61) I **FELT Ø** it was rather a shame that they didn’t take over that place
    (Daisy Smith, F 69, York)

(62) And ah it **LOOKS LIKE** ah- there might be a little argument going on or something.

---

61 The *as* on its own is attested as a historical form. There are two very early tokens of *look as* in the OED from the 15th and 16th centuries ("look, v."). *Looks as* was also used in a play by British author Samuel Taylor Coleridge (López-Couso and Méndez-Naya 2012b:315-316). Both could be linked to the complementizer *as* and/or relativizer *as* (see e.g. Kolbe 2011).
(Bobby Hunter, M 72, Kirkland Lake)

When an unpronounced subject could be determined to be an expletive (i.e. if there were no other possible sources of third-person verb-agreement), it was coded as such. If an unpronounced subject could equally represent an expletive or a third-person null pronoun (see e.g. Harvie 1998), the token was coded as ambiguous. There were a small number of tokens of null subjects that represent ellipsis or PRO, as in (65); these were coded as if they were their antecedents.

Verb tense: These five verbs appear with four types of inflection: present, past, gerund, and uninflected (i.e. after modals or other verbs). This was not significant in the earlier study of comparative complementizers in Toronto (Brook 2011a:61), but it will help ascertain whether grammaticalization of *seems like*, *looks like*, and *sounds like* into epistemic markers is indeed taking place (López-Couso and Méndez-Naya 2010, 2014).

(63) Fortunately, it LOOKS LIKE they’re cleaning it up a bit.  
(Jack Mosun, M 83, Toronto)

(64) I SEEMED AS THOUGH I was always working. But anyhow I got through life all right. Yeah.  
(Roy Burke, M 85, Lakefield)

(65) Sometimes we describe ah ah the local twang as SOUNDING LIKE you’ve got marbles in your mouth.  
(Scott Bramhall, M 35, Belleville)

(66) You needed to have that experience to FEEL LIKE you were worthy of that kind of love,  
(Kathy Thompson, F 31, Toronto)

Polarity: My earlier study of comparative complementizers in the TEA disregarded polarity by excluding cases of matrix negation – for instance, (67):

(67) It doesn’t SEEM LIKE people worried as much as they worry now.  
(Melissa Fenchurch, F 30, Kirkland Lake)

The present study includes such examples. Negation accompanying a copy-raising verb tends to originate from within the subordinate clause; as with the matrix subjects, negation is raised from inside it (Lakoff 1968:43; Rogers 1974:55; Gisborne 2010:278). Rogers (1974) labels this “Negative Transportation”, while Quirk et al. (1985:1033-1034) opt for “transferred
negation”. (68a) shows a sentence after this transformation, and (68b) shows the sentence before raising.\(^{62}\) A pair of sentences related by transferred negation “are at least close paraphrases” of each other (Rogers 1974:56).

(68a) It doesn’t look like it’s going to rain.
(68b) It looks like it isn’t going to rain.
(examples from Quirk et al. 1985:1033-1034)

In this study, there is no reason to set aside *It doesn’t look like it’s going to rain* but include *It looks like it isn’t going to rain* if they are transformationally equivalent.\(^{63}\) While Quirk et al. do acknowledge the way in which “intuitions may differ as to whether and to what degree two sentences with differently placed negatives are synonymous” (1985:1034), they are concerned about this mainly with respect to other types of complementation, e.g. complements following adjectives such as likely and probable.

**Level of metaphoricality of the subordinate clause:** López-Couso and Méndez-Naya (2012b:326) point out that *as if* and *as though* are closely associated with hypothetical proposition. Gisborne (2010:273) argues for a stronger position: that after verbs such as sound or seem, the complementizers like, as if, or as though are “restricted to [their] hypothetical meaning and it seems as though the resemblance being described is always rather abstract”. Gisborne presents the following sentence to exemplify this.

(69) Peter sounded to me like he had the spirit of Paganini in him.
(Gisborne 2010:273)

Gisborne’s example is an excellent case of a highly metaphorical subordinate clause after like. In practice, however, this is only one extreme of the range of metaphoricality found in this context. In previous work (Brook 2011a, 2014), I discovered that although like, as if, and as though do certainly favour the metaphorical, they are not restricted to it. *That* and Ø show the opposite

---

\(^{62}\) For evidence, see Lakoff (1969) and discussion by Rogers (1974).

\(^{63}\) Transferred negation is not limited to ostensibility verbs: it is also found with e.g. say, think, imagine, and more (Rogers 1974:56; Quirk et al. 1985:1033-1034).
pattern: they are typically introducing a very literal proposition, but are sometimes found with metaphorical clauses as well. Metaphoricality thus turns out to be a significant conditioning factor operating on the comparative complementizers in the Toronto English Archive (Brook 2011a, 2014). I therefore expect this to be true of the data from elsewhere in Ontario and also from the United Kingdom.

With concrete clauses, the [verb + comparative complementizer] structure can be replaced with another phrase referring explicitly to opinion.

(70) I FEEL LIKE I’ll be more social next year.
(Carrie Tippman, F 18, Temiskaming Shores)

(71) I have the opinion that I’ll be more social next year.

(72) LOOKS LIKE Bob Rae is going to be the Premier.
(Benedetto Battistoni, M 74, Timmins)

(73) I believe that Bob Rae is going to be the Premier.

At the other end of the spectrum are the examples more comparable to Gisborne’s: the metaphorical subordinate clauses. These constructions lose their meanings and might even become nonsensical when a phrase referring to thought or opinion is substituted:

(74a) And they put this coffin on my shoulder and it just FELT LIKE York Minster had hit me.
(Neil Thomas, M 62, York)

(74b) # And they put this coffin on my shoulder and I just became convinced that York Minster had hit me.

(75a) Yeah, there’s- there’s a grade six. That makes us- that makes me FEEL LIKE I’m in grade two.
(Christopher O’Neil, M 19, Toronto)

(75b) # Yeah, there’s- there’s a grade six. That makes us- that gives me the impression that I’m in grade two.

The third category in terms of metaphoricality, in the middle of the spectrum, is the ambiguous cases. These encompass both the examples that might be concrete or might not, and

---

64 Whether this falls out of epistemicity is a question worth asking. It will be partially addressed in Chapter 7.

65 Or even to fact, in some cases.
the ones that are relatively concrete but use idiomatic or otherwise non-literal language inside their proposition:

(76) You LOOK LIKE you’re going to cry. 
(Arthur Sorlee, M 49, Temiskaming Shores)

(77) Instead of saying nothing. ‘Cause I FEEL LIKE it's like dead air or something. Yeah, in a way. 
(Elizabeth McKinley, F 19, Toronto)

4.4.3 Independent social variables

Age. With respect to this particular variable, speaker age has previously been shown to be a significant factor in Toronto, pointing towards a change-in-progress wherein the like complementizer is the innovative form (Brook 2011a, 2014). The results from the corpora studied by López-Couso and Méndez-Naya (2012a:185) suggest that the same shift towards the adoption of the comparative complementizer like is occurring in the United Kingdom, albeit with some delay since like was not found at all in the spoken British English of the 1960s.

There is no reason to expect that the shift towards this like is not also progressing in the other, more remote communities in Ontario; like has been attested in written Canadian English at least as far back as 1896 (Brook 2011a) and has been prevalent enough in Toronto English for long enough that there has been ample opportunity for it to diffuse outward, geographically speaking. The Belleville Oral History Archive materials from 1975 will provide a real-time comparison on Canadian English and attest to how far along the change had proceeded by the time of the interviews.

---

66 The data in the main study were collected over about 15 years, meaning that year of birth would be slightly more accurate – and in order to err on the side of caution, that is what I use in the statistical analysis later in this chapter. The distributional results combine Canadian data collected between 2002 and 2013, one community at a time; however, the change is not rapid enough for this to be a major concern to me. I did not find a statistically significant effect of community within the Canadian data with respect to the introduction of like and thus combined the data (see Chapter 5). Using date of birth would also open up a can-of-worms in terms of what the best dividing-lines between binned age categories would be; this is certainly a valid line of inquiry but for the most part (with the exception of Figures 5.16 and 5.17) it lies outside the scope of the present work.
Sex: While ordinarily it is reasonable to anticipate that a linguistic change will be led by women (Labov 2001:501), no significant sex effect was found in the Toronto data (Brook 2011a, 2014), so there may not be any sex effect in the Ontario data more broadly. It is possible that the influence of sex simply levels as the change approaches the top of the S-curve (cf. Gardner et al. 2013 for be like in Toronto, which appears to have the same pattern). However, an alternative option is that the change was never female-led to begin with. The latter interpretation is supported by a previously-reported finding (Brook 2014:9) that a cross-tabulation of age and sex does not uncover a history of a female lead in apparent time. That said, comparisons with the behaviour of like according to gender in outlying Ontario communities, as well as the data from Belleville in 1975, may help illuminate the lack of a sex effect in Toronto.

4.4.4 Statistical analysis

In Chapters 5 and 6, I employ Goldvarb X (Sankoff et al. 2005) for distributional analysis and Rbrul (Johnson 2009) for mixed-effects models taking into account the potential random effect of individual speaker. Some specialized types of graphs are to be produced using R (R Core Team 2016).

4.5 Conclusion

This concludes the sources of data, the data extraction, conclusions, and coding for dependent and independent variables. Chapter 5 presents the results in accordance with the methodology laid out in this chapter. Chapter 6 examines a different level of change; it has its own separate methodology section. Then, Chapter 7 considers the two sets of results together and provides further interpretation.
Chapter 5
Results and discussion: Overall variation

5.1 Canada (Ontario)\(^{67}\)

5.1.1 Overall distributional results

It has previously been reported that the comparative complementizer system of Toronto is dominated by like (López-Couso and Méndez-Naya 2012; Brook 2011a, 2014). Table 5.1 and Figure 5.1 show that this is true across the province of Ontario more generally at the beginning of the 21\(^{st}\) century. Like represents 68.2 percent of the data; the null complementizer is in second place with 15.9 percent, followed closely by that with 14.3 percent. As if and as though are attested but very infrequent, each represented by a mere 8 tokens out of a total of 1010.\(^{68}\)

<table>
<thead>
<tr>
<th></th>
<th>like</th>
<th>Ø</th>
<th>that</th>
<th>as if</th>
<th>as though</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>68.2%</td>
<td>15.9%</td>
<td>14.3%</td>
<td>0.8%</td>
<td>0.8%</td>
<td>100%</td>
</tr>
<tr>
<td>N</td>
<td>689</td>
<td>161</td>
<td>144</td>
<td>8</td>
<td>8</td>
<td>1010</td>
</tr>
</tbody>
</table>

Table 5.1: Overall distribution of comparative complementizers across seem, look, sound, and feel\(^{69}\) in Canadian English.

---

\(^{67}\) Although the Canadian data come from a dozen different locations in the province of Ontario (see Chapter 4), I combine them here (but leave out Belleville 1975; I will examine it separately). The different varieties show considerable parallelism. A mixed-effects logistic regression conducted in Rbrul (Johnson 2009) shows that when speaker is taken into account, the only statistically significant difference is Belleville 2007-08, which is an outlier. Given that there are only 44 tokens from Belleville and that 14 of these are from a single speaker with a penchant for feel* that/Ø, Joyce Dowell (F 57), I acknowledge but largely disregard the possibility that Belleville itself is lagging behind the change. I also ran models looking for interactions between location and each of the other fixed effects; none came out significant.

\(^{68}\) Note that a direct comparison between these results and those of López-Couso and Méndez-Naya (2012a:185) is not possible since their data also takes into account the “it-is-as-if” construction (Declerck 1992:223), which leads to differing numbers of tokens with as if, as though, and like.

\(^{69}\) I exclude appear from the Canada analysis entirely because there are only two tokens. The verb seems to be limited to more formal registers (see also Brook 2011:47). This, in my view, accounts for why Matushansky (2002:228) briefly expresses the opinion that appear “does not allow the complementizer like in most dialects” and why López-Couso and Méndez-Naya (2014:39) conclude that it “is only found with [as if, as though, or like] in exceptional cases”.

75
Figure 5.1: Overall distribution of comparative complementizers across all ostensibility verbs other than appear in Canadian English (N = 1010).

5.1.2 Apparent time

Looking at like, that, and Ø in apparent time shows that speakers of different ages in Ontario have very different complementizer systems. Figure 5.2 reveals that like is the incoming form, showing a monotonic increase in apparent time.\(^\text{70}\)

\(^{70}\) It is arguably odd that there is no peak visible in apparent time, as there is expected to be even for a morphosyntactic change (see Tagliamonte and D’Arcy 2009). Perhaps the saturation point has led to a ceiling effect. Also worth pointing out is the fact that there are fewer tokens for the under-17 group (N = 62) than for any of the other groups (N > 100 in each case).
Like is gaining at the expense of both that and Ø. The zero form – more popular than that among elderly speakers – has dropped to negligible levels among the youngest cohort. That has been declining somewhat more gradually, but still falls below 10 percent of the tokens used by the youngest group.\footnote{It is conceivable that a written token of the comparative complementizer like could actually be a null complementizer followed immediately by a discourse like, though across a greater variety of complementation strategies, Blondeau and Nagy (2008:302) find that such cases are “actually quite rare”. Reassurance that zero tokens are not being wrongly categorized as like complementizers comes from the fact that in Figure 5.2, among speakers under 45 – who probably have the largest number of syntactic slots available for discourse like (D’Arcy 2005) – that and Ø are found patterning very close together.} As if and as though are represented in equal number but are very rare, and thus are not included here.

In sum, Figure 5.2 reveals that speakers under the age of 30 seldom use anything other than like as a comparative complementizer after these four verbs, suggesting that the change has reached the point of saturation. As if and as though appear to have long since disappeared from this context in vernacular Ontario English; and among younger speakers, that and Ø have declined to the point that the youngest speakers hardly use them in this context.

5.1.3 Verb

Given that the four verbs differ in terms of the complementizers they accept in particular circumstances – or in the first place (see Chapter 2) – it is imperative to examine the role being played by the factor of lexical verb in the change-in-progress towards like. Look and sound are claimed to be incompatible with that and Ø (Huddleston and Pullum 2002:962, Gisborne 2010:275), and although occasional anomalous tokens do exist (López-Couso and Méndez-Naya 2012a:186), the prediction is that only as if, as though, and like can be expected to appear after these two verbs. Given that as if and as though are already known to be rare in this data-set, like ought to be virtually the only variant that continues to appear after look and sound. Seem and feel, on the other hand, ought to retain variation with that and Ø.
Table 5.2 shows a breakdown of the tokens of comparative complementizers per verb; Figure 5.3 is a set of stacked bar-graphs showing proportions of the comparative complementizers per verb, with horizontal width of the bar corresponding to the relative frequency of the verb in the data-set.

<table>
<thead>
<tr>
<th></th>
<th>like</th>
<th>that</th>
<th>Ø</th>
<th>as if</th>
<th>as though</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>seem</td>
<td>138</td>
<td>22</td>
<td>27</td>
<td>6</td>
<td>3</td>
<td>196</td>
</tr>
<tr>
<td>look</td>
<td>148</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>152</td>
</tr>
<tr>
<td>sound</td>
<td>42</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>feel</td>
<td>361</td>
<td>122</td>
<td>131</td>
<td>1</td>
<td>4</td>
<td>619</td>
</tr>
<tr>
<td>total</td>
<td>689</td>
<td>144</td>
<td>161</td>
<td>8</td>
<td>8</td>
<td>1010</td>
</tr>
</tbody>
</table>

Table 5.2: Canada tokens split by complementizer and verb.

Figure 5.3: Ontario tokens split by complementizer and verb of the subordinate clause (N = 1010), with horizontal bar span corresponding to relative verb frequency.

*Look and sound* are almost categorically followed by *like*, which is expected given that *as if* and *as though* are negligible in number and both of these verbs resist *that* and Ø (Huddleston and Pullum 2002:962, Gisborne 2010:276, López-Couso and Méndez-Naya 2014:39). There are
three unexpected tokens of *look*\(^72\) along the lines of the one reported by López-Couso and Méndez-Naya (2012a:186), e.g.

(78)  I uh went out in the backyard. He was laying there. He LOOKED Ø he was half asleep.
(Shayna Jones, F 54, Toronto)

*Seem* supports variation between all five variants, but it has a greater proportion of *like* than *feel* does; of the verbs, *feel* retains the greatest proportion of the competing variants *that* and Ø.

*Feel* appears to be behind *seem* with respect to the change towards *like*; the next step is to examine the rise of *like* with these verbs in apparent time. Figure 5.4 shows the proportion of *like* split by verb in apparent time. Notably, *look* and *sound* both reach categorical status; the holdout tokens of *as if* and *as though* with these verbs, along with the three anomalous tokens of *look*\(^*\) Ø, are found only in speakers in the older cohorts.

![Figure 5.4: Proportions of *like* (opposite all other variants combined) in apparent time in Canada after each verb (N = 1010).](image)

Figure 5.4 also shows that, indeed, *seem* has been ahead of *feel* when it comes to the change toward *like*; *feel* has caught up only once *seem* has hit saturation, among the speakers under the age of 30. *Like* clearly reached *look* and *sound* first, but some caution is necessary here. There is a

---

\(^72\) They are all in the past tense, which suggests a syntactic niche for this anomalous combination in Canadian English. Whether this has any historical precedent is a question for future research.
potential conflation between a progression of *like* through the verbs and the order in which *like* has taken over from its covariants. Results for *look* and *sound* such as those in Figure 5.4 could simply reflect *like* having overtaken *as if* and *as though* before *that* and Ø, with the verb effect merely an epiphenomenon resulting from the fact that *that* and Ø are (mostly) unattested with *look* and *sound*. Chapter 6 will present evidence that this is the case, rather than the distinction between *look/sound* and the other verbs being inherently a lexical effect.

Assuming that *like* is the newest variant (see e.g. Brook 2011:10), this suggests that the change from *as if/as though* to *like* happened relatively early, while the change from *that/Ø* to *like* has been more recent but is nearing completion. Subsequent parts of this chapter, as well as Chapter 6, will revisit and uphold this idea.

Unaffected by this potential conflation are both *seem* and *feel*, since they each accept all five of the variants. In Figure 5.4 it is possible to see both some differentiation in apparent time, but it is also apparent that the two verbs are undergoing the same change towards *like*. Evidence that *feel* is closely following *seem* when it comes to the adoption of *like* is potentially surprising given how different *feel* is from the other verbs in terms of its argument structure and theta-roles (see section 2.1.7). For *seem*, *like* is a more flexible alternative to *that* and Ø: *like* allows copy-raising of an NP through it, while *that* and Ø do not (see section 2.1.6). However, for *feel*, *like* does not have this advantage. This is because *feel* in this data-set is hardly ever functioning as a copy-raising verb (see section 2.1.7). When *feel* has an NP matrix subject, this is almost exclusively an experiencer rather than a copy-raised argument of the lower clause; in these cases, *that* and Ø serve as well as *like* does. Yet Figure 5.4 demonstrates that *feel* has been undergoing the same change in terms of the comparative complementizers as *seem* has. Although *feel* still retains each of *that* and Ø at a rate of about 20 percent (as Figure 5.3 shows), the shift towards *like* with *feel* looks comparable to the shift towards *like* with *seem*. This finding reinforces the notion that including them within the same variable context is justifiable (see
3.2.2). That said, given how different the subtle syntactic properties of these two verbs are even when they are both taking finite subordinate clauses, their sheer similarity in terms of the change is an intriguing puzzle. It could be cast as an issue of form/function asymmetry. In terms of form – surface syntactic construction – these two verbs are essentially identical. They both have a matrix subject, a comparative complementizer, and a finite subordinate clause. When it comes to function, though, they are divergent. Parallelism in terms of the change towards *like* is not guaranteed. A legitimate question is what has ensured that *feel* would follow *seem*.

In my view, there are two elements involved. The first is that *as if* and *as though* have been replaced by *like* across the board as complementizers;\(^74\) the fact that *feel* is somewhat different from *seem* in terms of its arguments and their theta-roles is not enough to counteract this change in a single holdout context. This does not address why *feel* has kept up with *seem* when *as if* and *as though* have disappeared from the vernacular and *like* is competing only with *that* and Ø, but it is a start. Beyond that point, once *like* has overtaken *as if* and *as though*, the verbs *look* and *sound* are essentially categorical in being followed by *like*. This is the point at which the possibility of levelling arises. Since frequent collocations can help “propel...the new option...through the grammar” (Tagliamonte and Smith 2005:307), as in the case of the emergence of the Ø complementizer opposite *that* in late Middle English and Early Modern English (Tagliamonte and Smith 2005), it is reasonable to suppose that a frequent *look*\(^*\) *like* and *sound*\(^*\) *like* have encouraged the use of *like* after the very similar verb *seem* in the 20\(^{th}\) century;

\(^73\) One major question that arises from this is as follows: how different do syntactic structures have to be before they cannot influence each other in terms of variation and change? Although I do not have a ready answer to this question, Chapter 6 will show that the divergence can be more dramatic than this one. See also Chapter 7, as well as Dinkin (forthcoming) and Maddeaux and Dinkin (2015) on different functions of *like* as nearly simultaneous incoming forms.

\(^74\) This could be true even beyond ostensibility verbs. The *it-is-as-if* pattern (Declerck 1992; López-Couso and Méndez-Naya 2012a:180-181) is becoming *it-is-like* in Toronto English (Brook 2011a:69-71). Other places where *as if* and *as though* might be in the process of being replaced by *like* are: a) after the verbs *taste* and *smell*; and b) after a range of verbs, as adjunctive clauses of manner (Bender and Flickinger 1999; López-Couso and Méndez-Naya 2012a:173). See Chapter 7 for some additional thoughts on this.
and that all three of these influenced feel* to pair increasingly often with like rather than the less versatile that and Ø slightly later. The syntactically/semantically neighbouring verbs have some points of divergence – but not enough to be immune to patterns of change affecting the others.

The other reason that I suspect for like catching on with feel as easily as with seem is that like does in fact have advantages over that/Ø for feel. When it comes to feel, that and Ø are incompatible with all subjects that are not experiencers (see section 2.1.7). This includes true copy-raised subjects, as in (79b) (since that and Ø block copy-raising); it also includes expletives, along the lines of (79d), since an expletive cannot possibly be an experiencer:

(79a) She feels that things are looking up.
(79b) * Her forehead feels that it’s burning up.
(79c) It feels like things are looking up.
(79d) * It feels that things are looking up.

Like is the only one of the remaining colloquial variants that can appear in contexts such as (79b) and (79d); and it handles (79a) and (79c) just as easily. As if and as though can do so as well, but they are no longer active variants on the colloquial level, and that and Ø are comparatively restricted. For both feel and seem – in completely opposite ways – like is the only non-negligible variant whose grammaticality does not depend on the identity of the matrix subject.75

In spite of the fact that seem is farther ahead than feel in terms of the change towards like, seem is the verb that hosts the greatest number of complementizer variants in the Ontario data; this is consistent with what López-Couso and Méndez-Naya (2012b:327) find in diachronic corpora of English. To put this another way, in the Canadian data, feel seldom appears with either as if and as though anymore in spite of their grammatical acceptability. On the other hand,

75 This means that for the Canadian data, there are a number of contexts that are in effect categorical now that as if and as though are no longer a non-negligible part of the vernacular complementizer inventory after these verbs. These could be defined by the intersection of verb and subject, for instance: look, sound, NP + seem, and expletive + feel. Future studies of complementizers after ostensibility verbs that find similarly low proportions of as if and as though may want to keep this in mind, as the situation in early-21st-century Canadian English is close to being a combination of variable and categorical contexts all thrown in together. While this could be argued to be a fleeting effect on the way to saturation of all of the contexts by like, it is worth keeping in mind. (Thanks to Aaron Dinkin for a relevant discussion of this point.)
seem appears with all five comparative complementizers – including the single example of as though in the Toronto subcorpus, in the speech of an elderly man:

(80) He walked down Ashdale Avenue, walk up the walk, right away it SEEMED AS THOUGH he- he probably walked in- in and threw his lunchbox down and walked out the backdoor, got in the Model-T and then off they went to Sunnyside every Friday night! (Garry Haslem, M 85, Toronto)

However, the retention of as if with seem may be something of an illusion thanks to a cluster of unexpected tokens in young people. While three of the tokens of as if are, as expected, uttered by speakers over the age of 70, the other five are surprising in that they are produced by female speakers between the ages of 16 and 24. The purpose that seem* as if serves for these much younger speakers is likely to be different than the role that it is playing for elderly speakers; it may reflect prescriptive norms (see e.g. Fee and McAlpine 1997) or the influence of written registers.\footnote{A search of the Toronto Internet Corpus (Tagliamonte 2016) for complements of seem yields two tokens of seem* that, one token of seem as though, and no others. This speaks to a register difference: seem* like is ubiquitous in speech but more difficult to find in written conversation.} A necessary acknowledgment is that it is possible that this is merely an individual effect amplified by a large amount of data; four of the five tokens are from the same speaker, Clara Felipe, who has been interviewed almost every year between the ages of 16 and 29 (Tagliamonte 2012:274-276; Wagner and Tagliamonte to appear). However, it is telling that all of Clara’s tokens of as if (81 to 84) were produced between the ages of 16 and 19 as opposed to more recently:

(81) I think he still like, tries to like, make it SEEM AS IF he's cool, like to me.  
Clara Felipe, F 16, Toronto

(82) Or it SEEMS AS IF he never left and like, I never talked to him.  
Clara Felipe, F 16, Toronto

(83) So it kind-of makes it SEEM AS IF she doesn't acknowledge that other people do hard things too.  
Clara Felipe, F 17, Toronto

(84) I don't wanna make it SEEM AS IF he is bothering me.  
Clara Felipe, F 19, Toronto
I propose that *seem* as if is a hyper-standard fossilized form when it is used by younger speakers – which makes the fact that it is younger women using it unsurprising given the sensitivity that women are thought to have to linguistic prestige (Labov 1994, 2001:266, 291; Eckert 2000:192). The contexts here are not particularly high-register, but it is notable that three of Clara’s tokens, in spite of being from separate interviews, follow the verb *make*. This suggests that at least for her, *make* might be serving as a syntactic niche in which a fossilized *seem as if* (with an expletive subject) is holding on; Clara does not use *make* like this with any of the other verbs or any of the other comparative complementizers. (For more details on the Clara case-study, see section 5.1.9.)

5.1.4 Matrix subject

Figure 5.5 shows that *like* is the primary variant across all subject types, but that matrix expletives are the most conducive to *like*.

![Graph showing the use of like, that, and Ø across different subject types](image.png)

*Figure 5.5: Canada tokens with like, that, and Ø, split into matrix subject types. Ambiguous subjects are excluded (N = 930).*

Similarly, an apparent time view of the proportions of *like* in each context (Figure 5.6) shows that *like* consistently has a higher rate of use with expletives than with pronouns across apparent time.
Figure 5.6: Proportions of "like" with three types of matrix subjects in apparent time in Canada (N = 863). Ambiguous subjects are excluded – as are other NP subjects since there are few of these.

It is unsurprising that expletives are in the lead; if *seems like*, *looks like*, and *sounds like* are undergoing grammaticalization (see e.g. López-Couso and Méndez-Naya 2010, 2014), expletives – overt or otherwise – can be expected to be leading the change. The surprise is that first-person pronouns are not set apart from other pronouns given that *I feel like* is an experiencer phrase that has been observed to be becoming more frequent (see e.g. Squires 2013) and might be undergoing grammaticalization the way *I think* has.\(^77\)

It is not surprising that tokens with full NP matrix subjects (i.e. nouns rather than pronouns) show such a high proportion of *like*. For most of the verbs, noun phrases occur in the matrix subject position only when there has been copy-raising (see Chapter 2). *That* and Ø block copy-raising (see Huddleston and Pullum 2002:962, Gisborne 2010:275); and while *as if* and *as though* are just as capable of supporting copy-raising as *like* is, they are rare in this data-set. This leaves *like* as the only non-negligible comparative complementizer in Toronto English that permits copy-raising:

\(^{77}\) This is a possibility that will be revisited in Chapter 7.
(85a)  *He seems that/Ø he knows what he’s doing.
(85b)  He seems like he knows what he’s doing.

The question arises of why the nominal (and pronominal) matrix subjects show any variation between like and the other options. That and Ø are still found in non-negligible proportions across the board when it comes to the breakdown by matrix subject type.

The resolution to this issue is that feel is both the most prevalent ostensibility verb and the one with atypical syntactic/semantic conditions on the matrix subject relative to the other verbs (see section 2.1.7). Feel, when referring directly to the thoughts of the matrix subject, does not involve copy-raising at all. In this way, a pronoun or an NP serves as an experiencer; feel can easily take that or Ø when it has an NP subject that is not copy-raised.

The other characteristic of the verb feel that helps account for the patterns found in Figure 5.5 is the fact that *it feels that/Ø is ungrammatical, meaning that any tokens of feel with an expletive subject are categorically represented by like (aside from anomalies and holdout tokens of as if or as though). Since *look that/Ø and *sound that/Ø are claimed to be ungrammatical regardless of subject type (Huddleston and Pullum 2002:962, Gisborne 2010:275), and usually are in this data set as well (see Table 5.2 and Figure 5.3), virtually all of the tokens of expletive-it + (verb) + that/Ø have to be due to seem alone.

A speaker cannot utter seem* that/Ø if they have already chosen a referential matrix subject (a pronoun or other NP). The opposite is true of feels that/Ø; because expletives and copy-raised subjects are both ruled out with that/Ø, the matrix subject needs to have been an animate noun (i.e. an entity capable of feeling). With like, neither of these restrictions applies; it seems like, someone seems like, it feels like, and someone feels like are all equally acceptable. It is therefore little wonder that there are such high proportions of like in the expletive contexts.

The ways in which feel is quite different from seem and the other copy-raising verbs has major implications for the analysis of subject type. Feel has already been shown to be conducive
to the same range of comparative complementizers and even to be very comparable to *seem* in terms of the apparent time shift towards *like* in Ontario (see section 5.1.3). However, it is something of an imposter among the ostensibility verbs for the aforementioned reason that it largely does not trigger copy-raising. It *can* do so, but there are only a few incontrovertible examples of this. (In the Toronto data, for instance, out of 300 tokens of *feel* only two of them are clearly copy-raising.) For instance, in (85) it is indisputable that *feel* is acting as a copy-raising verb since swings are inanimate and therefore not capable of doing the ‘feeling’ (i.e. not capable of accepting the theta-role of experiencer):

(86) And- and the new swings they just put in, they- those swings FEEL LIKE they’ve been lowered two feet.
(Christopher O’Neil, M 20, Toronto)

Most of the NPs that accompany *feel* in the Toronto data are either pronouns or circuitous nominals referring back to the speaker (“part of me”; “the skeptic in me”). Neither of these categories can be indisputably shown to be copy-raising rather than a mere statement about the feelings of an animate matrix subject. The conclusion is that there is very little incontrovertible copy-raising with *feel* occurring in this data-set; most of the tokens of this verb are with experiencer subjects, particularly those referring back to the speaker. A distribution of the different subject types across the verbs (with no split by complementizer), as in Figure 5.7, immediately reveals how different *feel* is from the other verbs in this regard.
Figure 5.7 shows that while *seem*, *look*, and *sound* are most often found with expletives,78 *feel* appears predominantly with first-person pronouns. This is to be expected. *Feel* is the single verb of this set that bestows upon its subject the theta-role of experiencer, and the only one that can readily accept a subordinate clause without a pronominal copy of the matrix subject, as in (87c):

(87a) Rachel feels like she is always on top of things.
(87b) Rachel seems like she is always on top of things.
(87c) Rachel feels like it’s raining outside right now.
(87d) * Rachel seems like it’s raining outside right now.79

*Feel* is, unsurprisingly, well-suited to referring directly to a person’s emotions. While the expletive constructions *it seems/looks/sounds* do reflect the speaker’s opinion, *I feel* refers to them directly, whereas the 1st-person *I seem/look/sound* reflect apparentness and 2nd-person *you*

---

78 At first glance, it seems odd that *look* and *sound* are not more similar to each other in spite of their comparable semantic/syntactic profiles. In particular, *look* has a higher proportion of expletive subjects than *sound* does. If expletive subjects are indicative of grammaticalization (López-Cuso and Méndez-Naya 2010, 2014), then perhaps (it) *looks like is farther ahead in this regard in Canadian English than (it) sounds like.*

79 There is disagreement in the literature as to whether sentences with a copy-raising structure but without a coindexed pronoun in the subordinate clause are acceptable – and, if so, whether this threatens movement-based accounts of what we call ‘copy-raising’ (see Chapter 2 for more details references). The present study does not address this controversy: there are very few cases of copy-raising without a coindexed pronoun in the present dataset, so in practice, *feel* is the only verb actively supporting such subordinate clauses.
*seem/look/sound* are mere inferences (see Chapter 3). This syntactic and semantic difference results in *feel* favouring different types of matrix subjects – the ones whose feelings/thoughts need to be talked about. *I feel* like, in particular, is frequent, which suggests that it might be becoming more of an epistemic parenthetical along the lines of *I think*. See Chapter 7 for more details.

### 5.1.5 Metaphoricality

I have previously established (Brook 2011a, 2014) that metaphoricality plays a major role in the comparative complementizer system of the Toronto English Archive. Figure 5.8 shows that this is the case across Ontario. Across three levels of concreteness, the proportions of tokens with *that* are inversely correlated with metaphoricality, as are the proportions of tokens with Ø. Conversely, the opposite is true of the *like* tokens: they represent nearly all of the metaphorical tokens, then smaller proportions with the ambiguous and the concrete tokens.

![Figure 5.8: Canada tokens with like, that, and Ø, split by complementizer and metaphoricality of the subordinate clause (N = 994).](image)

In historical written Canadian English (Brook 2011), *that* and Ø show a preference for concrete clauses while *as if* and *as though* prefer the metaphorical ones; all four of these complementizers show stable semantic preferences in real time between 1860 and 1930 (Brook...
2011:52-60, 2014:9). Although *like* is not nearly as sensitive to the metaphorical-concrete distinction as the other four variants are, with the decline of *as if* and *as though*, *like* is left as the only non-negligible comparative complementizer that combines readily with metaphorical subordinate clauses.\(^{80}\)

A split of the tokens according to metaphoricality in apparent time, in Figure 5.9, reveals that *like* is consistently found at a higher rate with metaphorical clauses than with ambiguous ones, and at higher rate with ambiguous clauses than with concrete ones. While it is not assured that the ambiguous clauses comprise a uniform set, it is interesting to note that they act as an intermediary group in the change.

![Figure 5.9: Proportions of like (opposite that and Ø together) in apparent time in Ontario after all verbs and before three different semantic categories of subordinate clause (N = 994).](image)

For speakers over the age of 30 in particular, the level of metaphoricality of the subordinate clause has a major effect on the proportion of *like*. Since *as if* and *as though* once had a preference for metaphorical clauses and *that* and Ø the opposite (Brook 2011a, 2014), Figure 5.9 is essentially another way of exploring how *like* first overtook *as if* and *as though*, then began

---

\(^{80}\) Thanks to Aaron Dinkin for some very interesting discussion that helped this point emerge.
replacing *that* and Ø. An explanation for why this occurred in this order – and why it was the metaphorical clauses that shifted towards *like* first – can be found in Chapter 6.

### 5.1.6 Tense

As with matrix subject, there are parallels to findings for complementizers more generally as reported by Tagliamonte and Smith (2005:301): present-tense contexts are more advanced than past when it comes to the change towards the innovative complementizer, as Figure 5.10 shows.

![Figure 5.10: Ontario tokens with like, that, and Ø, split according to tense of the ostensibility verb (N = 917). Other, rarer contexts are excluded here due to the fact that dividing them into four age-categories results in some underpopulated cells.](image)

The apparent time breakdown in Figure 5.11 shows that the effect has been consistent over time except for the youngest age cohort.
5.1.7 Polarity

Polarity turns out to look very much like tense, both synchronically and diachronically. Positive contexts show a higher proportion of like than negative ones do; in apparent time the effect is small but consistent over time except in the youngest age group.
The fact that there are any tokens of that and Ø with negative polarity on the matrix level is arguably unexpected. If that and Ø block copy-raising of an embedded nominal to the matrix subject position, it stands to reason that with these verbs they would also block “transferred negation” (Quirk et al. 1985:1033-1034; see also Gisborne 2010:278) from the lower to the upper clause, and thus should not occur in the context of negative polarity at all.

This prediction turns out to be true in the Canada data with only two exceptions: one case of seem that with negation transferred above it, and one case of seem Ø with the same. All of the other cases of negative polarity with that or Ø (N = 36) are with the verb feel, which again shows itself to be syntactically distinct from the other verbs. There are only two exceptions to the notion of that and Ø blocking transferred negation as well as copy-raising.\(^1\)

While cases of negation plus feel still do qualify as transferred negation from the embedded clause, this is not a problem for the analysis; the other major class of matrix verbs permitting transferred negation (i.e. aside from the ostensibility verbs) are verbs indicating opinion, e.g.

\(^{1}\) I do not pursue further the possible ability of that and Ø to block negation in the present work, but it is a potential avenue for future research in corpus research, syntactic theory, and copy-raising.
believe, expect, figure, reckon, think, and suppose (Quirk et al. 1985:1033). Feel fits into this set in terms of its semantics and theta-role assignment; and in the case of the other verbs, clearly that and Ø do not act as a barrier to negation-raising.

With only two exceptions, these data suggest that seem that and seem Ø block both copy-raising and transferred negation. Since negation can raise over any of these constructions’ components in isolation – seem with as if, as though, or like; that and Ø when the verb is feel or a similar verb of opinion (Quirk et al. 1985:1033) – the implication is that only the syntactic combination of seem and that/Ø erects a barrier to raising in general.

5.1.8 Sex

The sex effect is inconsistent in apparent time. In apparent time, as Figure 5.14 shows, there is no particular sex-related pattern associated with the rise of the comparative complementizer like. It is curious to see that there is no clear female lead from the beginning as per Labov’s Principles 3 and 4 (Labov 2001:274-275,292-293). Unlike most, this is not a clearly female-led change, at least in Ontario.

Figure 5.14: Proportion of tokens with like in apparent time in Ontario, split by sex.
5.1.9 Clara

As mentioned in Section 5.1.6, four out of eight tokens of *as if* in the Canadian data come from speaker Clara Felipe. Clara was born in 1986 and has been interviewed by her older sister nearly every year since 2002, resulting in a real-time longitudinal study of a single speaker of Toronto English through adolescence (Tagliamonte 2012:274-276; Wagner and Tagliamonte to appear).

Clara’s proportions of comparative complementizers in real time are shown in Figure 5.15. Notably, her use of *like* starts out quite low at the ages of 16 and 17, but then jumps up to match her age-cohort and stays more or less stable thereafter.

![Figure 5.15: Clara’s proportions of comparative complementizers in real time according to time of interview (N = 52).](image)

When it comes to the *be like* quotative, Clara shows what would be predicted by Labov (2001): incrementation in real time followed by stabilization (Tagliamonte 2012:274-276). The results for her complementizers are less straightforward, though the shift from 2002-03 to the more recent time periods is not necessarily incompatible with incrementation. An important

---

82 I am grateful to discussions arising from Sali A. Tagliamonte’s LIN1256 seminar on lifespan change in the Winter 2016 semester that contributed to this section and to the other parts of the present work that involve the Clara data.
caveat is that Clara uses only 6 tokens of ostensibility verbs followed by finite subordinate clauses in 2002-03, but it is interesting that *like* starts out so infrequent. If the interview situation was unfamiliar enough at first to cause Clara to retreat into more standard forms, that could help account for the patterns.

One surprise is that Clara does not use the Ø form at any age.\(^83\) Her third variant, as noted in section 5.1.3, is actually the rare *as if*. Given that her four tokens are all with *seem*, all with an expletive *it*, and mostly immediately after the verb *make*, I have already noted (see section 5.1.3) that this particular syntactic context might be a fossilized hyperstandard niche for *as if*. The fact that Clara does not use *make* with any of the other ostensibility verbs or comparative complementizers is telling. So is the fact that all four of her tokens of *as if* occur between the ages of 16 (2002) and 19 (2005); Clara leaves *as if* behind after that. From 2006 onwards, *like* accounts for 91 percent of Clara’s comparative complementizers (41 out of 45 tokens between the ages of 19 and 27 inclusive). Her most frequent construction in later years (ages 21 to 27 inclusive) is *I feel like* (*N* = 30 out of 36); for more about that, see Chapter 7.

### 5.1.10 Belleville 1975

There are 14 tokens of ostensibility verbs with finite subordinate clauses in the Belleville 1975 data. Most have the verbs *seem* and *feel* with the complementizers *that* and Ø. The one exception is the single token of *look* – which is also the single token of *like*:

\[(88)\] He brought it in and one of the- one of the chief or somebody said, "It LOOKS LIKE the horse's been drugged."

(Bill Hurst, M 67, Belleville 1975)

Combining the 2002-2013 Ontario data yielded only one community whose low proportion of *like* made it a statistical outlier: Belleville (see section 5.1.1). However, it was not the case that

---

\(^83\) This leads back to the question of whether the zero form is becoming ungrammatical for some younger speakers in Canadian English (first raised in Brook 2011:33), i.e. whether ‘It seems the complementizer is missing’ is self-referential.
Belleville in 2007-08 had more *as if* or *as though* than the rest of the province; the difference was that it had more *that* and Ø. Likewise, there are no tokens of *as if* and *as though* from 1975. Even in data from this many decades ago, the token of *look* from Belleville has a *like* complementizer rather than an *as if* or an *as though*. Given the dearth of these in 2007-08, and assuming that *like* is the newest variant (Brook 2011a, 2014a), this suggests that *like* took over from *as if* and *as though* in vernacular Canadian English fairly early, and quickly.

5.1.11 Random forests and conditional inference trees

I return now to the results from the province of Ontario more generally. Given the number of factors, especially linguistic factors, that have been seen to condition the variation, the next step is to ascertain their relative contributions. As a prelude to a multivariate analysis, I conduct an exploration of the variation using R (R Core Team 2016), starting with a random-forests analysis that uses the cforest package (Hothorn et al. 2006; Strobl et al. 2007, 2008). A single conditional inference tree reveals the most major significant splits between subsets of the data; random forests run through many conditional inference trees to arrive at an estimate of the relative importance of each of a set of independent variables (see e.g. Baayen 2008, Tagliamonte 2012:152-155, Tagliamonte and Baayen 2012). Using an individual conditional inference tree is of interest primarily when it comes to apparent time; it will reveal the points that divide significantly different age-cohorts. Since conditional inference trees are capable of handling continuous factors, absolute year of birth (YOB) is used for both trees and forests instead of a 4- or 5-way binned division of speaker age.

The results from the random forests analysis are shown in Figure 5.16. Here I have excluded *as if* and *as though*, all tokens with *look* and *sound* since those are nearly categorical (Guy 1988:131, Tagliamonte 2006b:87), plus the three communities with relatively few tokens (Almonte, Beaverton, and Haliburton), and tokens of ambiguous linguistic factors other than metaphoricality (N = 715).
The random forests analysis reveals that year of birth is by far the most important factor in the Canadian data when it comes to comparative complementizer variation. Subject and verb (limited to *seem* and *feel*) are the most important of the linguistic factors but only slightly more so than tense and metaphoricality; sex is also less important but not by much. Smallest of all, as suggested by the apparent time analysis in Figure 5.12, is the effect of polarity.

![Variable importance graph](image)

*Figure 5.16: Random forest analysis of most important factors involved with variation of like versus that and Ø together as comparative complementizers.*

The statistical prominence of year of birth, coupled with the monotonic results in apparent time as seen in section 5.1.2, clearly indicates that the shift towards *like* is a dramatic change-in-progress; in comparison, the other factors have a modest influence on the variation.

A single conditional inference tree is capable of dividing a data-set into subsets of a continuous variable according to where the largest significant differences are along that
dimension. In this case, it can find the ‘break points’ according to year of birth\(^{84}\) when it comes to the rise of *like* (in this case, with *seem* and *feel* alone); these will provide an indication of how to divide up the speakers according to binned year-of-birth groups for the multivariate analysis to follow. The results are shown in Figure 5.17:

Figure 5.17: Conditional inference recursive partitioning tree with year of birth (continuous). Data-set same as for Figure 5.16 (N = 715).

Figure 5.17 shows that the major break-point when it comes to the proportions of *seem like* and *feel like* is the year 1960; the sets of speakers born before (or in) and after this year are significantly different with an effect size beyond all other divisions based on year of birth. If we assume an S-curve typical of linguistic change (Osgood and Sebeok 1954, Weinreich et al. 1968, Bailey 1973, Kroch 1989, etc.), the primary split should correspond to the midpoint of the change.

The secondary break-points in Figure 5.17 are 1941 and 1973, meaning that the largest secondary significant splits in the use of *seem like* and *feel like* are found between groups on either side of these birth years. Even the oldest speakers in the corpus tend to have *look like* and

\(^{84}\) See also Gardner et al. 2013, who use a similar strategy to investigate the *be like* quotative.
sound like at high rates, and those verbs are excluded here; the findings captured in Figures 5.16 and 5.17 reflect only the comparative complementizer like spreading to seem and feel.

5.1.12 Multivariate analysis

Due to the need to take multiple competing constraints into account, and to control for the effect of the individual speaker, I conduct a mixed-effects logistic regression in Rbrul (Johnson 2009) with speaker included as a random effect. Fixed effects included in the model were verb, matrix subject, metaphoricality level of the subordinate clause, tense, polarity, speaker sex, and speaker age (divided into four cohorts as in most of the factor-by-factor analysis). Again, I exclude the verbs look and sound since they are too close to categorical for a multivariate analysis (Guy 1988:131, Tagliamonte 2006b:87). The results are shown in Table 5.3.

<table>
<thead>
<tr>
<th>Factor-weight</th>
<th>%</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year-of-birth cohort (p = 2.64e-23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1974 onwards</td>
<td>0.90</td>
<td>81.8%</td>
</tr>
<tr>
<td>1961 to 1973</td>
<td>0.56</td>
<td>56.1%</td>
</tr>
<tr>
<td>1942 to 1960</td>
<td>0.46</td>
<td>38.4%</td>
</tr>
<tr>
<td>Before 1942</td>
<td>0.09</td>
<td>12.5%</td>
</tr>
<tr>
<td>Range</td>
<td>81</td>
<td></td>
</tr>
</tbody>
</table>

| Matrix subject type (p = 5.91e-09) |     |         |
| Expletives                              | 0.78| 71.1%   | 166    |
| 2nd-person pronouns                    | 0.53| 66.0%   | 47     |
| Noun phrases                            | 0.53| 53.3%   | 30     |
| 1st-person pronouns                    | 0.52| 61.8%   | 400    |
| 3rd-person pronouns                    | 0.18| 33.3%   | 72     |
| Range                                   | 78  |         |

85 That said, even without look and sound, the data here are not an ideal match for a multivariate analysis. Because [NP + seem] is almost categorically represented by like, and [expletive + feel] is almost categorically the opposite, the split in the data goes well beyond a mere subject-verb interaction. A better approach would be to run separate analyses of [expletive + seem] and [NP + feel] and then compare them to each other. The practical barrier to that in the present work is that there are not enough tokens of the former for a multivariate analysis (N = 142), and while there are enough tokens of the latter (N = 567), it would not mean very much in isolation. The matter is let for future work with larger data-sets, and this acknowledgment is left as a warning sign that this might be an issue. (Thanks to Aaron Dinkin for helping me spot and understand this problem. Any errors here are my own.)
Metaphoricality of subordinate clause ($p = 2.39 \times 10^{-06}$)

<table>
<thead>
<tr>
<th>Type</th>
<th>Value</th>
<th>Proportion</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metaphorical</td>
<td>0.90</td>
<td>92.5%</td>
<td>53</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>0.31</td>
<td>63.4%</td>
<td>287</td>
</tr>
<tr>
<td>Concrete</td>
<td>0.20</td>
<td>54.7%</td>
<td>375</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td></td>
<td>59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex ($p &lt; 0.05$)</th>
<th>Value</th>
<th>Proportion</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.59</td>
<td>64.3%</td>
<td>484</td>
</tr>
<tr>
<td>Male</td>
<td>0.41</td>
<td>54.1%</td>
<td>231</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td></td>
<td>18</td>
</tr>
</tbody>
</table>

Table 5.3: Factors selected as significant for the choice of the complementizer like (versus that and Ø together) for the verbs seem and feel in Canadian English as produced by Rbrul (Johnson 2009), with individual run as a random effect. Non-significant factors – verb, tense, and polarity – are not included here.

Age is divided 4-way according to birth year of speaker, in accordance with the conditional-inference tree in Figure 5.17. The effect comes out as highly significant and very large, with the constraint-ranking showing a linear age-effect: as with the random forest analysis, this is evidence of a change-in-progress with seem and feel.

The largest of the effects among the linguistic factors in the multivariate analysis is that of matrix subject. As Figure 5.7 suggests, expletives favour like; they are the context that is leading the change. First- and second-person pronouns are barely distinguished from each other and from full NPs; it is the third-person pronouns that are disfavouring like and retaining the greatest proportion of that/Ø.

Metaphoricality is the next-largest of the significant linguistic factors. As a result of the preference that that and Ø have for concrete clauses, like is the one variant left in the system that does not resist metaphorical ones, and accordingly the proportion of like is higher with increasing metaphoricality of the subordinate clause. Verb, tense, and polarity were non-significant, showing that feel is statistically indistinguishable from seem in terms of the change towards like.

The one surprising result given the factor-by-factor apparent time analysis is the effect of sex: overall, a significant difference emerges whereby women are leading as per Labov (2001:501). Although the apparent time trend in Figure 5.14 is not consistent or obvious, taking
the linguistic effects into consideration yields a significant effect that suggests the like is used significantly more by women than by men, and suggests a subtle underlying female lead for the change.

5.1.13 Summary: Canada

The dominant comparative complementizer in Canadian English (as represented by multiple sites in the province of Ontario) is like; speakers under the age of 30 hardly use the four covariants at all. The large and significant effect of age in apparent time indicates that like represents a change-in-progress: it is coming in at the expense of that and Ø. Among speakers under the age of 30, like is nearing the saturation point within the set of complementizers. There are a few tokens of as if and as though, but these either are produced by elderly speakers or represent an apparently fossilized (make it) seem\(^*\) as if collocation produced by young women as a very occasional alternative to seem\(^*\) like in speech. That and Ø remain in the system to a much greater degree than as if and as though, but they are already low-frequency and are on the decline in apparent time. The only verbs with which they appear at non-negligible levels are seem and feel, and even then they are restricted by subject: to expletives with seem and to referential subjects (i.e. everything except expletives) with feel. For look and sound – and for [NP + seem] and [expletive + feel] – like is the only complementizer available on the vernacular level. Overall, like seems to have displaced as if and as though fairly early, and with that, has become the dominant complementizer everywhere as if and as though used to be found in variation with each other and with like. The sum is that for all of the verbs except seem, expletives lead to like almost categorically; and for all of the verbs except feel, NPs yield like as well. That said, in spite of the differences, seem and feel are going through the change towards like in very comparable ways; the difference between these verbs is small and non-significant. What this suggests is that like is displacing that and Ø equally after it seems and after NP feels. In both cases, that and Ø
both exhibit a preference for concrete subordinate clauses, so the only variant left in the system that pairs easily with the highly metaphorical ones is *like*.

The verb *appear*, while it has the same syntactic properties as *seem*, is subject to register: it is reasonably common in historical written Canadian English up to 1930 (N = 185; Brook 2011:31), but is not often found in colloquial speech of present-day Ontarians of any age (N = 2).

Sex comes out as a significant predictor of *like* usage in Ontario, with female speakers using more *like*. Although the apparent time results are not consistent, in the context of a change-in-progress towards a nonstandard feature, a significant effect of sex with female speakers leading suggests a classic sex-effect of women leading a change from below as per Labov's (2001:292-293) Principle 4.

### 5.2 United Kingdom (York)

#### 5.2.1 Overall distributional results

There are two reasons to expect that there will be much less *like* in the United Kingdom than in Canada and that if there is a shift towards the comparative complementizer *like* in apparent time in the York English Archive, it will be far behind what is occurring in Canada.

One reason is that this was earlier reported for the LOB, FLOB, DCPSE–LLC, and DCPSE-ICE-GB corpora by López-Couso and Méndez-Naya (2012). The authors find that there are no examples of the complementizer *like* in spoken British data collected in the 1960s, and that by the 1990s *like* still represents a mere 35 percent of the tokens (López-Couso and Méndez-Naya 2012a:185). This indicates that the change in the UK is well behind what has been occurring in Canada.

The other reason for the expectation of a British lag is that, as discussed in section 2.2.2, older British sources and/or elderly British linguists commenting on the comparative complementizer *like* tend to single it out as a target of prescriptive disapproval: this *like* is
“nonstandard” (Quirk et al. 1985:1175), or even “strongly stigmatized” (Mair and Leech 2006:318; see also Huddleston and Pullum 2002:1158) in the United Kingdom. American and Canadian sources, as noted, do not label the variant this way; neither does the British but younger Gisborne (2010). Such opinions suggest that the change towards *like* is newer in the UK than it is in Canada.

Indeed, it is immediately clear from Table 5.4 and Figure 5.18 that although there are relatively few tokens of comparative complementizers from York in the late 1990s (N = 144), together they represent a system that is very different from that of the province of Ontario.\(^8^6\)

<table>
<thead>
<tr>
<th></th>
<th>like</th>
<th>that</th>
<th>Ø</th>
<th>as if</th>
<th>as though</th>
<th>as</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>13.2%</td>
<td>28.5%</td>
<td>29.2%</td>
<td>11.1%</td>
<td>17.4%</td>
<td>0.7%</td>
<td>100%</td>
</tr>
<tr>
<td>N</td>
<td>19</td>
<td>41</td>
<td>42</td>
<td>16</td>
<td>25</td>
<td>1</td>
<td>144</td>
</tr>
</tbody>
</table>

Table 5.4: Overall distribution of comparative complementizers across all verbs in the York English Archive (N = 144).

As anticipated, the change towards *like* is considerably less advanced in the UK. *Like* in York ranks a mere fourth when it comes to the frequency of the comparative complementizers. *As if* and *as though*, now so rare as to be nearly unattested in spoken Canadian English, are found in

---

\(^8^6\) For the sake of precision, it is worth mentioning that there are two factors involved: one is the geographic (and associated socio-political and/or cultural) distance. The other is that the York English Archive was collected in the mid-1990s, which puts it at seven to fifteen years behind the corpora from Ontario.
larger numbers in York – though they are both less frequent than *that* and Ø, each of which comprises almost 30 percent of the tokens. Although York is not necessarily analogous to an earlier stage of Canadian English, it does provide a look at how the complementizers following ostensibility verbs might have behaved before *like* overtook its competing variants.

Another interesting difference between Canada and the UK when it comes to comparative complementizers is that there is an additional variant among the York tokens as compared with the Canadian data: *as* on its own. This variant has been reported historically, both from the 15th and 16th centuries (OED, “look, v.”) and from the very beginning of the nineteenth century (López-Couso and Méndez-Naya 2012b:316). To my knowledge, this use of *as* has not been reported before in modern corpora of vernacular speech; that said, there is only a single token of it in York (with a second one in the rural UK data; see section 5.2.9).

5.2.2 Apparent time

Figure 5.19 shows the variants in apparent time in York. For the speakers 30 years old and above, the system is more or-less stable, particularly for *that* and Ø. Among the speakers 29 years and younger, *like* suddenly accounts for 12 out of 16 tokens (75 percent). The total N for this age group is low, but nonetheless this is a much larger proportion of *like* than in any of the older cohorts in York. The system appears to have been radically altered in very little time. As in North America, *like* is coming in, but here the change is far more dramatic and has occurred in the presence of more active covariants for *like* to compete with.
While *like* is attested as a comparative complementizer in the three older age cohorts, it is very low-frequency in each. There are probably several reasons for a pattern of this nature. There may be some degree of age-grading: prescriptive opposition and a history of major stigma associated with the comparative complementizer *like* in the UK (Mair and Leech 2006:318; see also Huddleston and Pullum 2002:1158) might make it saliently nonstandard and likely to be abandoned in the context of the workplace as young people move into careers. However, this is likely not the only factor involved. Given that there were signs of *like* entering British English between the 1960s and 1990s (López-Couso and Méndez-Naya 2012a:185), there is ample room for change-in-progress as well as age grading. A sudden dramatic leap in the proportion of a variant from nearly negligible levels points towards the diffusion of that variant from elsewhere (Tagliamonte 2012:57). It is possible that young people in York have picked up this *like* through diffusion – perhaps from the nearest large cities as per the Cascade Model of dialectology (see e.g. Trudgill 1974, Labov 2003).

---

87 For similar results in the York English Archive, see Denis (2011) on general extenders and Gardner et al. 2013 on the *be like* quotative.
5.2.3 Metaphoricality

The rates of all five variants in the UK, split by the level of metaphoricality of the subordinate clause, are shown in Figure 5.20. A caveat is that only 143 tokens split between five active variants and three semantic classes means that the Ns of individual cells are limited. However, it can be seen that for the three variants that are also found in non-negligible numbers in Ontario, the patterns are the same: the proportion of *like* is directly correlated with metaphoricality, and *that* and Ø each show the opposite pattern. In spite of its very low frequency among everyone above the age of 30 in York, *like* is the most popular variant before metaphorical subordinate clauses – though there are only 13 tokens in this context. The ambiguous context shows no particular preference for any complementizer; each of the five has a rate under 25 percent. Concrete contexts, exactly as in Canada, strongly favour *that* and Ø. The others variants are rare here.

Beyond those, the effect of metaphoricality on *as if* and *as though* can also be examined given that these variants are still represented in vernacular speech in York. *As if* behaves the same as *like*: increasing metaphoricality makes it increasingly likely. The one surprise is *as though*. While my previous work on Canadian writing in the 19th and early 20th centuries (Brook 2011a, 2014a) found *as though* patterning the same as *as if* (albeit at lower rates), *as though* does not clearly correlate with metaphoricality in York speech of the mid-1990s. Again, there are relatively few tokens of metaphorical subordinate clauses, but *as though* appears to be underrepresented in this context.
Figure 5.20: All tokens in York except the one instance of as, split by complementizer and metaphoricality of the subordinate clause (N = 143).

Regardless of the status of *as though* when it comes to metaphoricality, it is clear that four of the variants have preferences in terms of the levels of metaphoricality of the subordinate clause, and that the three variants also found at non-negligible levels in Ontario are all patterning the same way when it comes to this linguistic effect. Chapter 6 provides an account of why *like* can reasonably be expected to take over from *as if* and *as though* earlier than it from *that* and Ø – across all dialects.

### 5.2.4 Verb

Figure 5.21 shows the pattern by lexical verb in the York data. Because *look* and *sound* do not accept the complementizers *that* and Ø (Huddleston and Pullum 2002:962, Gisborne 2010:275), these verbs are almost categorically followed by the complementizer *like* in Canada, where *as if* and *as though* are only seldom found in vernacular speech (Table 5.2, Figure 5.1). In York, however, *as if* and *as though* remain in the system at non-negligible levels. This means that *look* and *sound* still show variation in terms of the complementizers with which they appear.
The Canadian results (see section 5.1.3) indicated that *like* had caught on with the verbs *look* and *sound* earlier than with *seem* and *feel*. This appears to be the case as well in York; these are the verbs that have the highest proportions of *like*. *Look* prefers *as though* over *as if* and *sound* is the opposite, but the proportion of *like* is similar for the two verbs.

These results also suggest a resolution to the unexpected finding that in York *as though* does not show a straightforward correlation (in either direction) with metaphoricality. It is the most prevalent complementizer with *seem*, which is more semantically bleached than *look*, *sound*, or *feel* (Landau 2011:788). Perhaps the local prevalence of *seem as though* has made this complementizer variant insensitive to the metaphorical status of the subordinate clause. That said, this is not the only potential account of the results for *as though* in York. An alternative (or overlapping) possibility is that *like* has been replacing *as though* before *as if*.88 Possible evidence for the latter comes from how *as if* is lower frequency than *as though* overall, but the opposite is true in the context of the metaphorical clauses. *Like* is entering the system primarily through these contexts, according to Figure 5.20. Perhaps *like* has been replacing *as though* in this context

---

88 This would be consistent with the conclusion of López-Couso and Méndez-Naya (2012a:185-186) that on a scale of formality, *like* is the most casual, *as though* is intermediary, and finally *as if* is the most formal.
before *as if*. However, this is difficult to reconcile with the results by verb; as seen in Figure 5.21, *seem* *as though* and *look* *as though* continue to be attested.

Curiously, *that* and Ø are hardly found with *seem* in York. There are two tokens of *seem* Ø, but no instances of *seem that*. In fact, the only verb that shows variation between *that* and Ø in York is *feel*. This points to an interaction between verb and metaphoricality in York. For concrete clauses, locals rely mostly on *feel that*/Ø; for metaphorical ones, they prefer *seem*, *look*, and *sound* with one of *as if*, *as though*, and *like*. Since this almost restricts the concrete clauses to NP subjects, it is possible that concrete subordinate clauses go hand-in-hand with experiencer subjects, at least locally. Either way, this set of verbs – and their complementizers – is more deeply divided than it is in North America. My earlier conclusion (Brook 2014:10) that *like* has unified the system in Toronto sets up the hypothesis that the distinction between *feel* and the other verbs will fade in York as *like* continues to take over from the other variants. A follow-up study of complementizers with newer data would establish whether *like* has done this to any extent within the last twenty years.

### 5.2.5 Matrix subject

In Figure 5.22, the York tokens are divided by category of the matrix subject. The major observation when it comes to *like* is that the change is being led by expletive contexts, as it is in Canada (see Figure 5.7). *Like* is already the most popular complementizer after expletive matrix subjects, but in all of the other subject contexts, other complementizers are much more prevalent.
First-person-pronoun contexts show remarkably little *like* – only a single token, from a 29-year-old speaker. However, there is an interaction between matrix subject and verb for all of the pronouns, unsurprisingly. The fact that *that* and Ø are well-represented here suggests that most of these tokens are with the verb *feel* (unsurprisingly, given that *feel* is likely to take first-person pronouns as subjects), which is lagging well behind the other verbs in terms of the change towards *like*, as seen in section 5.2.4. Indeed, there are 51 tokens of 1st-person pronoun contexts in York, and 50 of them are the verb *feel*. Most of the second- and third-person pronouns (23 out of 30 in combination) are also *feel*.

In other words, *I feel like* (N = 1) has not yet caught on in York as of 1997-98. Other pronouns appear to be ahead of first-person pronouns when it comes to the change before *like*, but really this is an effect of lexical verb in disguise.

### 5.2.6 Tense

The effect of tense, shown in Figure 5.23, is minimal. While there are more tokens of *like* in the present, making the direction of the effect the same as in Canada, this might simply be due to the lack of *feel like* in York.
Figure 5.23: All tokens in York except the one instance of as, split by complementizer and verb tense (N = 143).

5.2.7 Sex

Figure 5.24 shows the tokens split by sex of speaker. Like is used at higher rates by male speakers overall; conversely, female speakers use more as if and as though than male speakers do. However, the apparent time results in Figure 5.25 suggest that there is little difference within any given age-group. The only cohort that uses a non-negligible proportion of like is the youngest one, and there is no female lead for them. There is no evidence that the change towards like in York is female-led.
Figure 5.24: All tokens in York except the one instance of as, split by complementizer and by sex of speaker (N = 143).

Figure 5.25: Proportions of like, versus all other variants (except as), in apparent time in York, split by speaker sex (N = 143).

5.2.8 Mini-studies of rural places in the UK

Cumnock: There are 36 tokens of comparative complementizers, only two of which are like. 31 of the tokens, including both of the likes, are with the verb feel; the other verbs are rare or unattested here.
Maryport: This location has a token of as on its own, along with 30 tokens of other variants across a mix of the verbs. 11 tokens have as if and 9 as though. There are no instances of like.

Wheatley Hill: Although there are 29 tokens of comparative complementizers, 16 of them come from the same speaker, Pat Dickson (F 41), and twelve of those are feel* as though in various tenses and linguistic contexts. However, there are five tokens of like, found across several different speakers and next to three of the verbs (look, sound, and feel). Like has more of a foothold in Wheatley Hill than it does in either of the other rural locales in the UK surveyed. This may reflect the town’s status as a suburb for city commuters.

5.2.9 Summary: York

The complementizers of York in the mid-1990s look very different in terms of their proportions from those of the Canadian data examined earlier in this chapter. Since as if and as though are still active, York has five variants rather than three. Like exists locally, but for everyone above the age of 30 it is a marginal variant. In apparent time, for the speakers under 30, like is predominant variant by far. This suggests that it has reached York through diffusion, rather than via parent-to-child transmission within the community. Effects of metaphoricality and of matrix subject do seem to be present in spite of the possibility of diffusion, but both of these can be seen to be an underlying lexical effect rather than the independent influence of these two factors, so whether they are really in operation in the system is questionable. Feel, which is responsible for almost all of the tokens of that and Ø in York, is more set apart from the other verbs here than it is in Canada; the influx of like in Ontario has brought feel in line with seem and the other verbs in apparent time, but there is not yet enough like in the system in York for this to have occurred. Since like seems to be replacing as if and as though more readily than that and Ø, like is showing up first in metaphorical contexts — feel is mostly concrete — and with expletives, since pronouns are also found mostly with feel. When it comes to speaker sex, there is no female lead in York. In spite of that, the change towards like is underway locally and the
sudden age effect looks very much like diffusion from a larger city in the UK. While York is behind Ontario in terms of the shift towards *like*, its complementizer situation provides insight into how an earlier stages of change might have looked as *like* entered the picture.
Chapter 6
A second level of change

6.1 Introduction

6.1.1 A puzzle

An unexpected extra finding came out of my earlier work on complementizer variation after ostensibility verbs in the Toronto English Archive (Brook 2011a, 2014): the numbers of tokens of the entire [verb + complementizer + finite subordinate clause] structure were not evenly distributed in apparent time. Rather, there was an inverse correlation between age and average number of tokens, suggesting that younger people in Toronto are using these constructions more than their older counterparts are. Figure 6.1 shows that this is true across Ontario. A small caveat is that these numbers are raw, i.e. not normalized.\(^{89}\) The chart assumes that the interviews are all precisely the same length and contain the same amount of speech recorded per speaker, which might be slightly inaccurate in some cases. Nonetheless, a pattern is evident in Figure 6.1 whereby the average number of tokens per age-group generally increases with decreasing age of speakers in the corpus up to the 17- to-19-year-olds.

This is not necessarily a unique finding: López-Couso and Méndez-Naya (2012b:329) briefly report that the use of comparative complementizers “seems to be on the increase in the present day”. The question this raises is whether there is a change-in-progress occurring even aside from the matter of the innovative *like* complementizer.

\(^{89}\) Thanks to Derek Denis for pointing this out.
Also suggestive are the speakers with the highest token counts; Table 6.1 displays these individuals. All but one of them are between the ages of 17 and 24; all but one are female. These findings alone cannot be used to diagnose a change-in-progress; however, they hint at the possibility of a change led by young women (Labov 2001:275, 292-293).

<table>
<thead>
<tr>
<th>N</th>
<th>Name</th>
<th>Location</th>
<th>Sex</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>Alice Faddei</td>
<td>Timmins</td>
<td>F</td>
<td>22</td>
</tr>
<tr>
<td>20</td>
<td>Tammy Chan</td>
<td>Toronto</td>
<td>F</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>Erin Tam</td>
<td>Toronto</td>
<td>F</td>
<td>19</td>
</tr>
<tr>
<td>17</td>
<td>Elizabeth McKinley</td>
<td>Toronto</td>
<td>F</td>
<td>19</td>
</tr>
<tr>
<td>17</td>
<td>Carrie Tippman</td>
<td>Temiskaming Shores</td>
<td>F</td>
<td>18</td>
</tr>
<tr>
<td>14</td>
<td>Joyce Dowell</td>
<td>Belleville</td>
<td>F</td>
<td>57</td>
</tr>
<tr>
<td>12</td>
<td>Jennifer Parkins</td>
<td>Toronto</td>
<td>F</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td>Stacy Yeh</td>
<td>Toronto</td>
<td>F</td>
<td>18</td>
</tr>
<tr>
<td>11</td>
<td>Phoebe Castillo</td>
<td>Toronto</td>
<td>F</td>
<td>24</td>
</tr>
<tr>
<td>11</td>
<td>Mitchell Pounding</td>
<td>North Bay</td>
<td>M</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 6.1: Ontario speakers with the highest individual token counts.

I have previously considered two explanations for the pattern in Figure 6.1. One is what has been suggested by López-Couso and Méndez-Naya (2010, 2014): that *seems like*, *looks like*, and *sounds like* are becoming grammaticalized as invariant markers of epistemicity (see Chapter 4
for more background). As grammaticalization tends to entail an increase in raw frequency (Hook 1991), an increasing range of grammatical functions could account for the increased numbers of tokens among younger speakers in the TEA.

The other possible explanation that I have entertained (Brook 2011a, 2014) is that ostensibility verb structures are part of a larger phenomenon. If hedging strategies in general are age-graded such that younger speakers tend to express more hesitancy about their own opinions than older adults do, then perhaps *seems like, looks like, and sounds like* in their function as epistemic markers (see Chapter 3) are just one of many strategies that younger speakers use to put more distance between themselves and their statements. 90 This could be either in the context of a sociolinguistic interview, or in general since younger adults do not have the same life experience as older ones and may feel less confident about their own opinions.

Neither of these possibilities is inherently misguided or necessarily wrong. However, the present chapter will introduce evidence that points towards a third, separate reason for the overall uneven distribution of tokens by age: an apparent time trade-off between two types of clausal complements after the verb *seem*.

### 6.1.2 A new solution

Chapter 5 covered variation between five complementizers only in the context of finite subordination: either where there is a matrix expletive, as in (89a), or when an NP has been copy-raised, as in (89b) (or where *feel* has an experiencer subject). These, together, represent the options for finite subordination (i.e. tensed subordinate clauses) after the verbs *seem, look, sound,* and *feel*. With the verbs *seem* and *appear*, there is one additional option in terms of a subordinate clause: an *infinitival* (non-tensed) complement (89c) – SUBJECT-TO-SUBJECT RAISING.

(89a) It seems (like/as if/as though/that/O) Rachel won the game.
(89b) Rachel seems (like/as if/as though) she won the game.
(89c) Rachel seems PRO to have won the game.

90 See Brook 2011a:41-42 for examples. This possibility lends itself well to future research.
Copy-raising, as in (89b), is known to be related to Subject-to-Subject raising, as in (89c) (Rosenbaum 1967; see also Davies and Dubinsky 2004, López-Couso and Méndez-Naya 2012b:325). It is an alternative way of taking the proposition *seeming to have won the game* and applying it to *Rachel*. Both structures involve a nominal originally associated with an embedded clause that moves upwards to occupy the matrix subject position – the difference being that in the case of copy-raising, even if it is not really movement (see section 2.1.3 for an overview of this controversy), the operation is optional.

This chapter shows that in Ontario English, finite and non-finite subordination are closely enough related to be able to covary as whole structures. After the verb *seem*, there is an inverse relationship in apparent time between the finite subordination structures – (89a) and (89b) together – and Subject-to-Subject raising – (89c). More specifically, the finite subordination structures are increasing in frequency at the expense of Subject-to-Subject raising. This account is not necessarily incompatible with either of the earlier possibilities I proposed in an attempt to account for the uneven numbers of finite tokens in apparent time – grammaticalization and age-graded hedging – but it has a larger amount of explanatory power, at least in this data-set.

The possibility of variation and/or change occurring between finite subordinate structures and Subject-to-Subject raising does not invalidate the analysis conducted in Chapter 5 (or in prior analyses of complementizers after the verbs *seem, appear, look, sound*, and *feel*). It just proposes that the non-finite contexts, at least after *seem*, comprise a variant of a secondary, higher level of variation.91 Chapter 5, along with earlier work (Brook 2011a, 2014), was a thorough investigation of the variation within the ‘finite’ category. The present chapter considers change on the level above, starting by specifying two variants. Subsequent analysis

---

91 Hopper and Traugott (2003:175) predict the existence of variation like this after *seem*, though they anticipate the opposite direction of change, i.e. “(It) *seems that he is right* > *He seems to be right*.”
and interpretation will refine this somewhat (see section 6.2.3), but Figure 6.2 shows where it will be necessary to start in order to establish a sense of the variation.

![Diagram showing a preliminary division of data into two levels of variation](image)

Figure 6.2: Schematic of a preliminary division of the data into two levels of variation (i.e. two variables), one nested inside the other. On the higher level are two variants: Subject-to-Subject raising and finite subordination. Finite subordination here corresponds to a set of five variants as analyzed in Chapter 5 and in my previous work (Brook 2011a, 2014).

## 6.2 Methodology

### 6.2.1 Finite and non-finite structures

A concern when it comes to setting up the variable context is the fact that not all of the verbs participate in this variation between finite and non-finite (Matushansky 2002:228; Mack 2010:Chapter 4; Landau 2011:784). *Seem* and *appear* do so readily, but *feel* does not and the other verbs are more nebulous; the acceptability of *look to V* and *sound to V* appears to be “lexically and dialectally conditioned” (Mack 2010:179). Algeo (1988:23) finds *look to be* acceptable in the United Kingdom, where it is “especially characteristic of sports language”, but not in the United States. Matushansky (2002:228) considers both of them infelicitous. Mack (2010:179) accepts that *look* and *sound* may be permissible in some regions:

(90a) The cake seems to be incredible.
(90b) The cake appears to be incredible.
(90c) % The cake looks to be incredible.
(90d) % The cake sounds to be incredible.
(90e) * The cake feels to be incredible.
(90f) (adapted from Mack 2010:179)
My own intuitions likewise suggest a range of results when it comes to appending a non-finite subordinate clause to the verbs:\footnote{Gisborne (2010:243-244) also addresses this briefly, though he notes only a single example from the British press with \textit{look to be} and does not say whether he would accept an analogous example with \textit{sound}.}

\begin{itemize}
\item[(91a)] This seems to be a problem.
\item[(91b)] This appears to be a problem.
\item[(91c)] ? This looks to be a problem.
\item[(91d)] * This sounds to be a problem.
\item[(91e)] * This feels to be a problem.
\end{itemize}

A quick search of the Toronto English Archive data finds one token of \textit{feel* to} and three tokens of \textit{appear* to}; all of the rest of the non-finite tokens (N = 248) are with \textit{seem}. Because \textit{appear} is so infrequent in this data with finite subordination, the only verb productive enough in both cases – finite and non-finite subordinate clauses – is \textit{seem}.

Restricting the study in this chapter to \textit{seem} risks compromising the extent to which this broader look at the syntax can explain the apparent time patterning of ostensibility verbs more generally. However, it is a necessity, on two counts. One is that there are not enough tokens of the alternation for any of the other verbs. The other is that Hopper and Traugott (2003:175) anticipate change involving finite subordination versus Subject-to-Subject raising after \textit{seem}, but they do not indicate whether they believe it also affects \textit{appear}.

The next challenge in defining the envelope of variation is that the covarying constructions need to be defined without regard for whether they are \textit{internally} conducive to variation. The covariant of Subject-to-Subject raising is not any particular complementizer, but the entire finite subordinate template, regardless of how it is realized. One variant is the finite structure, which is realized in various ways (five complementizers, two types of matrix subjects); and the other is Subject-to-Subject raising, which shows no internal variation at all since the infinitival marker is always \textit{to} (see Figure 6.2). On the level of the entire construction – at least for the preliminary analysis – this does not matter. While the finite subordinate clause is a CP (or sometimes
perhaps a clause-containing PP – see section 2.1.6) and the non-finite one is an IP (see e.g. Matushansky 2002), the criterion of near-synonymy (see Tagliamonte 2006:76) is fulfilled since both structures are strategies for taking a particular proposition and applying it to the subject. Kim (2014:168), for instance, quickly concludes that when it comes to “truth-conditional meaning, both the typical [Subject-to-Subject] raising and [copy-raising structures]...are synonymous to their counterparts with the expletive subject.” In other words, both (92a) and (92b) are synonyms of (92c).

(92a) He seems PRO to be ill.
(92b) He seems like he’s ill.
(92c) It seems like he’s ill.

Copy-raising (92b) and Subject-to-Subject raising (92a) being structurally related to each other follows from the common syntactic properties (raising, complementation). The minor differences – the fact that copy-raising is optional, the divergence in terms of whether the subordinate clause is tensed, and the differing realizations of the lower coindexed nominal (PRO versus overt pronoun) – are not enough to obviate synonymy. Hopper and Traugott’s conviction (2003:175) that one of these structures is giving way to the other is, therefore, not inherently far-fetched. The only conflict is that they predict the opposite direction of change – (92c) to (92a).

Another necessity is to establish that the finite and non-finite seem patterns are actually variable within speakers. Supertokens (see Tagliamonte 2006b:96) exist, as in (93a) though establishing interchangeability (93b) requires a bit of tweaking in accordance with the slightly different syntactic properties as listed above:

(93a) Now it SEEMS LIKE that’s really changed, like a lot of people SEEM TO be having casual sex. (Erin Tam, F 19, Toronto)
(93b) Now that SEEMS TO have really changed, like a lot of people SEEM LIKE they’re having casual sex.
A non-negligible number of speakers show variation between the finite structures and Subject-to-Subject raising. There are 290 speakers across all of the corpora in the study who use at least two tokens of *seem* + any subordinate clause, whether finite or Subject-to-Subject raising. Figure 6.3 shows that nearly 40 percent of these speakers (113 out of 290) show variation between *seem* to V and *seem* + a finite subordinate clause.

![Figure 6.3](image)

*Figure 6.3: Of the 290 speakers who use at least two tokens of *seem* + any subordinate clause, the proportion who show variation between the finite and non-finite subordination strategies and the proportion who are categorical one way or the other.*

This establishes that a sizeable proportion of the speakers in the corpora alternate between the two strategies and the overall patterning is not due to every speaker having a preferred variant that they individually use all of the time, without there being any intraspeaker variation (see Tagliamonte 2006:96).

### 6.2.2 Dividing up the finite tokens

The final question is whether the finite tokens really do represent a uniform ‘variant’ structure opposite the non-finite ones as in Figure 6.2, or whether dividing them up would make more sense. As it stands, all of the finite tokens are combined into one variant. While this is necessary as a first step to establish the patterning, leaving the finite tokens as they are – completely undifferentiated in spite of possibly containing two or more distinct sets depending on specific realizations (e.g. matrix verb, complementizer) – runs the risk of overlooking
meaningful patterns. Yet given that the realization of both the matrix subject and the complementizer vary within the set of finite subordinate clause tokens, there is no obvious best way of going about splitting the finite tokens.

Given this, the best strategy is to compare different approaches. The methodology of comparative sociolinguistics can be used not only to compare similar analyses of different data (Tagliamonte 2002, 2012:Chapter 6), but also to evaluate separate analyses of the same data (see e.g. Tagliamonte and Poplack 1993). I thus treat the matter of how to best categorize the finite tokens opposite the non-finite ones as an empirical question.

One way of dividing up the finite tokens is by the type of subject: expletives (it or null) versus NP subjects. I refer to this division as the ENVELOPE-OF-VARIATION DIVISION, since it dictates which range of variation is available where. With an expletive, that and Ø are permitted; with a copy-raised subject, that and Ø are not found (see Chapter 2, and also Huddleston and Pullum 2002:962, Gisborne 2010:275). The envelope-of-variation division is essentially the same as the approach taken by López-Cuso and Méndez-Naya (2012a, 2012b) for their complementizer study: they examine variation when the subject is an expletive, and when it is an NP. For the purposes of the present study, alongside Subject-to-Subject raising as the third variant construction, the envelope-of-variation division is schematized in Figure 6.4, with the finite subordinate clauses divided into two ‘variants’ opposite Subject-to-Subject raising.
Figure 6.4: Three variant constructions under the envelope-of-variation division of the finite tokens: the split depends on the nature of the matrix subject.

The envelope-of-variation division thus leads to three variants, defined as follows:

<table>
<thead>
<tr>
<th>Expletive finite</th>
<th>NP finite</th>
<th>Subject-to-Subject raising</th>
</tr>
</thead>
<tbody>
<tr>
<td>it seem* as if</td>
<td>NP seem* as if</td>
<td>NP seem* to VP</td>
</tr>
<tr>
<td>it seem* as though</td>
<td></td>
<td></td>
</tr>
<tr>
<td>it seem* like</td>
<td>NP seem* as though</td>
<td></td>
</tr>
<tr>
<td>it seem* that</td>
<td>NP seem* like</td>
<td></td>
</tr>
<tr>
<td>it seem* Ø</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2: The three variants under the envelope-of-variation division of the finite tokens.

A different way of dividing up the tokens is according to whether copy-raising is possible. Because this depends on which complementizer is present, this approach is tantamount to which complementizer appears in the token. For this reason, I refer to the second strategy to dividing up the finite tokens as the TYPE-OF-COMPLEMENTIZER division. However, the crucial difference is of underlying syntactic properties: whether copy-raising is forbidden (*that/Ø) (Huddleston and Pullum 2002:962, Gisborne 2010:275) and where it is permitted (regardless of whether it is actually realized). Figure 6.5 lays this out:
With the type-of-complementizer division of the finite tokens, the three variants are as follows:

<table>
<thead>
<tr>
<th>Copy-raising blocked</th>
<th>Copy-raising permitted</th>
<th>Subject-to-Subject raising</th>
</tr>
</thead>
<tbody>
<tr>
<td>it seems that</td>
<td>it seems as if</td>
<td>NP seem* to VP</td>
</tr>
<tr>
<td>it seems Ø</td>
<td>it seems as though</td>
<td></td>
</tr>
<tr>
<td></td>
<td>it seems like</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP seems as if</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP seems as though</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP seems like</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3: The three variants under the type-of-complementizer division of the finite tokens.

Hopper and Traugott (2003:175) anticipate that the raising tokens such as *He seems to be right* are taking over from *(It) seems that he is right*, but it is not clear whether they expect a loss of *all* of the finite tokens with any complementizer, or just the subset with *that* and Ø. This approach to dividing up the finite tokens tests the latter directly.

The difference between these two ways of spitting up the finite tokens in the data is quite subtle. However, it can be illuminated by drawing attention to what is being held constant. In
the envelope-of-variation approach, matrix subject is controlled for, which means that all of the expletives are in one variant but a few of the *complementizers* are found in both. Conversely, in the type-of-complementizer approach, a particular complementizer will be found in only one variant, but there are expletive *subjects* in both. In the type-of-complementizer approach, *it seems like* is put with *NP seems like* because in both cases, copy-raising is permitted. The fact that it is unrealized in one case is irrelevant.

With the variable context restricted to finite and non-finite tokens of *seem*, and these three strategies to defining the variants laid out (undivided finite tokens; envelope-of-variation division; type-of-complementizer division), the subsequent step is to outline the extraction and coding of the non-finite tokens, which are newly introduced and thus not already covered in Chapter 4.

### 6.2.3 Extraction and exclusions

The non-finite subordinate clauses are all introduced by the infinitival marker *to*. Given that the analysis is restricted to the verb *seem*, only a single new search in AntConc (Anthony 2012) was necessary:

*seem* to  
*seem to, seems to, seeming to, seemed to*

However, not all of the instances of ‘*seem* to’ found in the search fall within the variable context. The following two types of sentences were excluded:

**Cases of ‘to’ that are not part of an infinitive:** These include dative arguments of *seem*, which are not part of the grammatical context under consideration.

(94) It seems to me the younger- young people do that.  
(Dirk Brooks, M 73, Toronto)

**Elisions, or sentences cut off before the embedded verb:** These are not in full variation with the finite subordinate clauses: consider the lack of a parallel between (96b) and (97b).

(95) Well they seemed to, I like-  
(Murray Bruce, M 53, Toronto)
6.2.4 Coding

All of the tokens of finite subordination and Subject-to-Subject raising after *seem* were coded for two social factors and three linguistic ones:

**Age:** Whether there is change-in-progress is the primary matter of interest. If there is a trade-off in apparent time between these and the raising structures, coding for age will reveal it.

**Sex:** The variation between the complementizers in Ontario in Chapter 5 turned out to be conditioned by a significant sex effect consistent with Labov’s Principles 3 and 4 (Labov 2001:274-275, 292-293). If there is a change-in-progress affecting larger syntactic structures, then female speakers might be leading it in Ontario as well.

**Finite or non-finite:** This is the basic split between the finite subordinate clauses after *seem* (the ones with the comparative complementizers, regardless of type of subject or complementizer, all together) and Subject-to-Subject raising.

93 While “...but it seems like *it*” would be a grammatical resolution of this sentence, this would not qualify as a finite subordinate clause.
Type-of-complementizer division: As per Figure 6.5 and Table 6.1, this is the other way of splitting the finite tokens. The variant constructions are [blocked copy-raising], [permitted copy-raising], and [Subject-to-Subject raising].

6.3 Results: Ontario

6.3.1 Finite versus non-finite: Overall\textsuperscript{94}

Table 6.4 shows the overall distribution of finite subordinate clauses and Subject-to-Subject raising with seem in the Ontario data. Tokens of Subject-to-Subject raising are more than four times as prevalent as the finite ones.

<table>
<thead>
<tr>
<th></th>
<th>Finite</th>
<th>Subject-to-Subject raising</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>158</td>
<td>577</td>
<td>735</td>
</tr>
<tr>
<td>%</td>
<td>21.5%</td>
<td>78.5%</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.4: Overall tokens of non-finite (seem* to) and finite (seem* + comparative complementizer) in Ontario.

6.3.2 Finite versus Subject-to-Subject raising: Apparent time

A breakdown in apparent time shows that the finite structures after seem are catching on at the expense of the Subject-to-Subject raising. While Ontario speakers over the age of 60 use the finite subordinate structures only occasionally, younger speaker are increasingly likely to say It/He seems like he’s ill rather than He seems to be ill.

\textsuperscript{94} As in Chapter 5, a mixed-effects logistic regression performed in Rbrul (Johnson 2009) that takes speaker into account as a random effect yields no significant effect of location. Therefore, I combine the Ontario data.
This result shows a trade-off in apparent time that strongly suggests the structures are covarying. While there is no adolescent peak, as in Chapter 5, again there are very few tokens from speakers ages 10 to 16 (N = 17 in this case). As the pattern is not curvilinear, it is more suggestive of change-in-progress than of age-grading. It can be interpreted as a second level of change on top of the Chapter 5 change towards *like* within the finite subordination system. There are two changes occurring at once, one metaphorically on top of the other.

This provides part of an answer as to why the numbers of finite tokens were not evenly distributed in apparent time in Figure 6.1. Caution is necessary since those were raw token numbers and these are proportions, but given that the two trends look similar, it is reasonably to conclude that the reason why younger speakers on average were using so many more tokens of finite subordination was that it was a description of approximately half of a change. The other half is represented by a decline of Subject-to-Subject raising.

The reason why this does not account for *all* of the pattern in Figure 6.1 is that *seem*, observed here, is only one of the four verbs in the results in Chapter 5. *Seem* *like* is taking over from *seems to*, but the pattern in Figure 6.1 may also require higher numbers of *look* *like*, *sound*
like, and/or feel* like among young people. It is possible that this is where grammaticalization of these forms (López-Couso and Méndez-Naya 2010, 2014) begins to play a major role.

The finite context catching on in this way challenges the prediction of Hopper and Traugott (2003:175) that “(It) seems that he is right [will give way to] He seems to be right.” They defend this notion by appealing to the idea that over time, as grammaticalization progresses, separate clauses fuse into one. In other words, the clausal boundary breaks down: in It seems that he is right, there is a CP that could serve as a stand-alone sentence, but in He seems to be right, the subordinate clause lacks both tense and an overt pronoun, meaning that the clausal boundary has been eroded.

This is not what has been observed in the Ontario data. With He seems to be right yielding to It/he seems like he’s right in Ontario, the clausal boundary may be, if anything, being strengthened – though the possibility that the subordinate clause is a PP (Heycock 1994:295; see also Potsdam and Runner 2001:465-466, Asudeh 2002:6-8) and the link between a copy-raised NP and the coreferential pronoun that is usually in the subordinate clause may mitigate this somewhat. 95

Now that it has been established that there is a trade-off in apparent time between the finite and non-finite tokens, the next step is to determine which of the two ways of dividing up the finite tokens yields more insight. A comparison will indicate whether one approach does a better job than the other of revealing underlying reasons for the change.

6.3.3 Comparative study of two approaches to the same data

Figure 6.7 shows the results of splitting the finite subordination tokens according to envelope-of-variation. The Subject-to-Subject raising tokens are exactly as they were in Figure 6.6, but the finite tokens have been divided into [(it) seem* + finite] and [NP seem* + finite].

95 I am grateful to Belén Méndez-Naya for pointing this out to me.
Figure 6.7: Tokens of seem in apparent time in Ontario English, split into types [it seem* finite], [NP seem finite*], and [Subject-to-Subject raising] (N = 735).

Figure 6.7 reveals that both of these subtypes of finite tokens – expletive and matrix subjects – are increasing relative to Subject-to-Subject raising. In particular, [(it) seem* + finite] is consistently more prevalent than [NP seem* + finite]; the latter is only becoming non-negligible among the youngest speakers. There are at least two separate things going on here. One is that like is the only extant complementizer in vernacular speech that permits copy-raising (NP subjects). The other is that the copy-raising operation itself seems to be catching on even within like tokens.

Figure 6.8 shows the alternative approach: Subject-to-Subject raising as before, but with the finite tokens divided into [blocked copy-raising] – i.e. that and Ø – and [permitted copy-raising] – like, as if, and as though.

The results are not entirely surprising given Chapter 5. The one complementizer in vernacular speech that permits copy-raising is like, and the two that block it are that and Ø (Huddleston and Pullum 2002:962, Gisborne 2010:275); the previous chapter established that like is the single incoming form among the complementizers. However, Figure 6.8 recasts that result.
(at least with *seem*) in terms of the syntactic properties, and with respect to Subject-to-Subject raising.

This is a more informative result relative to Figure 6.7 in that it squarely identifies what is coming in to replace Subject-to-Subject raising: contexts (with *like*) where copy-raising is permitted, whether or not it is actually realized. The tokens of blocked copy-raising (S1) – *it seems that* and *it seems Ø* – are infrequent to begin with and are being lost along with Subject-to-Subject raising. Both of the contexts that Hopper and Traugott (2003:175) suspect of being in a trade-off are in fact being lost in Ontario, with the *like* option coming in at their expense.

Given that the set of tokens where copy-raising is permitted is almost identical to the set of finite tokens with *like* as the complementizer, it could be argued that Figure 6.8 simply restates the results from Ontario in Chapter 5. What makes this potential objection not the case is that the results here are placed opposite Subject-to-Subject raising. That is, while a decline in [blocked copy-raising] *only* against [permitted copy-raising] would fall straightforwardly out of Chapter 5, there was no guarantee that [permitted copy-raising] – essentially the same as the *like* tokens – would *still* show any kind of increase when plotted on the same graph as Subject-to-Subject raising. Figure 6.8 is therefore not simply a redundancy. It does not speak to *like* catching on among the complementizers in apparent time. Rather, it shows the entire finite structure – mostly fuelled by *like* – breaking out of its envelope-of-variation and beginning to detract from a more distantly related syntactic structure.

---

96 Note, for instance, that there is very little increase in the [permitted copy-raising] variant in speakers under the age of 45 in spite of the fact that within the finite variation, *like* after the verb *seem* is clearly on the increase in the same data-set for all of the older age cohorts (Figure 5.4).
The findings in Figure 6.8 are very intriguing from the perspective of envelope-of-variation. On the surface, there is no reason to distinguish *It seems that something is wrong* and *It seems like something is wrong* simply because the latter happens to have a syntactic property that *could* have allowed the speaker, if they had so chosen, to raise the embedded subject to matrix subject position. Disregarding the tokens with copy-raised NPs, the variable contexts for *It seems that something is wrong* and *It seems like something is wrong* are identical, just with the complementizer swapped. But the context that holds the *that* complementizer with *seem* is being lost. As above, attributing this shift simply to *like* being the incoming form and happening to have this property – while not incorrect – would be insufficient. It would overlook the fact that *seem* *like* is taking over not only from *seem* *that*, *seem Ø*, *seem* *as if*, and *seem as though*, but also *seem to* – a different construction altogether. The new form is ushering in a context in which there is always the option to raise the subject of the embedded clause to the matrix subject position – whether or not speakers take advantage of this. As for [blocked copy-raising], it seems to be patterning on its own.
There is additional explanatory power in this, and implications for the narrower level of variation between the finite comparative complementizers. The result in Figure 6.8 suggests that [blocked copy-raising] – *that* and Ø – is its own variant on the level of whole syntactic structures, and set apart from [permitted copy-raising], i.e. *like*, *as if*, and *as though*. If this is the case, then the expectation would be for *like* to be able to compete directly with *as if* and *as though* (within its variant) but not directly with *that*/Ø (separate variant). *That* and Ø are buffered, in a sense, from the influx of *like* as a comparative complementizer.

This is precisely what has been found (see Chapter 5, and also Brook 2011a, 2014): *like* appears to have taken over from *as if* and *as though* readily in Ontario English, leaving *that* and Ø as the only non-negligible complementizers when it comes to vernacular speech; *like* has competed with *that* and Ø more recently, but the Chapter 6 results suggest that only the entire finite structure (now unified by *like*) can compete with *that*/Ø, which is consistent with previous observations. They also suggest that the initial division at the beginning of this chapter – between [Subject-to-Subject raising] and finite tokens all together, as in Table 6.2 and Figure 6.5 – conflated two very different things: [blocked copy-raising] and [permitted copy-raising]. The former is low-frequency, but both its properties and its apparent time trajectory are very different from those of [permitted copy-raising].

At least within the present work, it may not be possible to ascertain exactly what, on the broader syntactic level, is competing directly with what. In other words, are the non-finite tokens competing with the finite ones, which have internal variation between blocked and allowed copy-raising? Or are all three covarying directly? The question is particularly tricky given the small number of finite tokens with *seem* in which copy-raising is blocked. For now, the subsequent analysis in this chapter will focus on [Subject-to-Subject raising] versus the incoming [permitted copy-raising], but I acknowledge that how many sub-levels of variation there are within this single broader level is not clear.
The next step, given the hints of a sex effect (see section 6.1.1), is to establish whether there is a female lead to this change.

### 6.3.4 Sex

To begin with, women do favour [permitted copy-raising] overall, and this effect of sex is highly significant (Pearson’s χ²: df = 1, p < 0.0001) (N = 694) shows this cross-tabulation in terms of proportions, revealing a female lead in terms of the incipient change.

![Graph](image)

*Figure 6.9: Cross-tabulation of sex and type-of-complementizer construction variants in Ontario English, excluding the cases of blocked copy-raising (N = 694).*

Figure 6.10 shows the sex effect in apparent time. There is a consistent female lead in every age-group except the oldest, who hardly use the [blocked copy-raising] construction at all.
The male speakers are consistently behind by about fifteen years; the expected female lead in apparent time (Labov 2001:501) is present. In terms of both apparent time in general and sex, this change on the level of entire syntactic constructions is acting in accordance with general expectations. Classic variationist ideas about gender amidst variation and change are seen to apply as readily to whole syntactic structures as to traditional narrower variants.  

6.3.5 Multivariate analysis

It is possible to run a multivariate analysis on the variation between [Subject-to-Subject raising] and [permitted copy-raising], but only with a very simple model due to the small number of factors being analyzed. This is shown in Table 6.5. Even with individual taken into account as a random effect, both age and sex turn out to condition the variation at the significant level. Both of these factors are in the directions expected for a change-in-progress.

---

97 Two younger twenty-something linguistics students have spontaneously reported to me that the entire [Subject-to-Subject raising] structure sounds formal or old-fashioned to them (Ernest Leung, p.c., 23 January 2016; Brianne Süss, p.c., 29 February 2016).
Table 6.5: Factors selected as significant to the probability of the [permitted copy-raising] structure (versus [Subject-to-Subject raising]) for the verb in Ontario English as produced by Rbrul (Johnson 2009), with individual run as a random effect.

6.3.6 Belleville 1975

A hypothesis about the distribution of variant structures in the Belleville 1975 data can be made on the basis of the apparent time trajectory in Figure 6.7. Among the tokens of *seem* plus subordination should be mostly tokens of [Subject-to-Subject raising] and possibly a few instances of [blocked copy-raising], but very few cases of [permitted copy-raising].

This is precisely what is observed. Out of 23 tokens of *seem*, 21 are [Subject-to-Subject raising], 2 are [blocked copy-raising], and *none* are [permitted copy-raising] in spite of this being an earlier stage of Canadian English with presumably more *as if* and *as though* available (though Chapter 5 established that there were no tokens of either of these in Belleville 1975).

6.3.7 Clara

As in Chapter 5, another source of insights into an apparent change-in-progress in Ontario would come from Clara Felipe, a Toronto speaker who has been interviewed yearly since she was 16 (Tagliamonte 2012:274-276; Wagner and Tagliamonte to appear).

Over ten interviews, Clara only uses 20 tokens of *seem* with either type of subordinate clause (finite or non-finite), so the numbers are necessarily small and unreliable across the board. What
they show, displayed in Figure 6.11, is that Clara does not use Subject-to-Subject raising before the age of 21, but she has used both constructions since.

![Figure 6.11: Proportions of finite (both blocked and permitted copy-raising) and of Subject-to-Subject raising tokens in the interviews with Clara Felipe (N = 20).](image)

Although there is also the evidence for change (linear rather than curvilinear pattern in apparent time; female lead), these results are consistent with some degree of age-grading occurring in the community. Clara appears to have adopted some use of Subject-to-Subject raising as she has moved into the workforce. It is intriguing that in Chapter 5, Clara does not show age-grading away from an innovative variant (Figure 5.15), but here, she does (Figure 6.11).

The difference may come down to how far along the community is in terms of the change. The like complementizer is well-established in Ontario English; even the oldest age-cohort uses it at a rate of more than 30 percent among the other comparative complementizers (Figure 5.2). The change towards a whole syntactic structure, as Figures 6.6 and 6.8 show, is newer, which might make the [permitted copy-raising] structure something that is more saliently associated
with young people – which in turn would open up a subtle, possibly even unconscious need for age-grading once adolescents reach the workplace.

6.3.8 Summary: Ontario

The findings from Ontario English show a change-in-progress entirely consistent with general principles of variationist sociolinguistics, albeit possibly contra Hopper and Traugott (2003:175). The [Subject-to-Subject] construction is being lost in apparent time to [permitted copy-raising] structures – almost all it/NP seem* like before a finite subordinate clause. An evaluation of two ways of partitioning the finite tokens opposite Subject-to-Subject raising shows that [blocked copy-raising] is low-frequency and declining. The innovative form is [permitted copy-raising] rather than the finite structures more generally; [permitted copy-raising] and [blocked copy-raising] have very different trajectories in apparent time. Clara’s results suggest a role for age-grading, but the apparent time trajectory and female lead point towards a change-in-progress underway in the community on top of the like innovation described in Chapter 5. The key to the change may be that the syntactic and semantic flexibility afforded by [permitted copy-raising] – the option to raise an NP to the matrix subject position – puts this type of structure at an advantage relative to both [Subject-to-Subject raising] and [blocked copy-raising], neither of which allows more than one type of matrix subject. This idea will be explored further in Chapter 7.

One of the implications for the analysis in Chapter 5 is that a study of comparative complementizer variation that combines that/Ø with like/as if/as though is likely mixing two types of structure that are not necessarily competing directly with each other in terms of variation and change. (Arguably the Chapter 5 results for York and even the original historical split imply the same; however, this is evidence.) If [blocked copy-raising] is a separate variant on the level of broader variation between syntactic structures, then that and Ø may not be covarying directly with like as complementizers in spite of surface similarities in terms of the
expletive constructions that the complementizers can all appear in. The extra advantage of this interpretation is that it provides an account of why like has replaced as if and as though early on in Canadian English but that and Ø only more recently. This result sets up the prediction that similar findings should be identifiable in other parts of the English-speaking world. For now, the results have shown that even on this broader syntactic level, there is familiar variation and change to be found; at least in this case, stratification according to age and sex seems to be straightforwardly applicable to fairly broad grammatical features.

6.4 Results: United Kingdom

6.4.1 Overall: Finite versus non-finite

As Table 6.6 shows, there are 200 tokens of seem plus some kind of subordinate clause in York, but very few of them (N = 24) are finite.

<table>
<thead>
<tr>
<th></th>
<th>Finite</th>
<th>Subject-to-Subject raising</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>24</td>
<td>176</td>
<td>200</td>
</tr>
<tr>
<td>%</td>
<td>12.0%</td>
<td>88.0%</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.6: Overall tokens of non-finite (seem* to) and finite (seem + comparative complementizer) in York.

This is in spite of the fact that York’s complementizer system, as Chapter 5 showed, is divided between five separate active variants, all of them theoretically compatible with seem (though as noted in Figure 5.21 there are no tokens of seem* that in the York data). There are more than seven times as many non-finite tokens as finite ones; in York, [Subject-to-Subject raising] is by far the dominant structure on the broader syntactic level that is the focus of this chapter.

6.4.2 Apparent time: Finite versus non-finite

An apparent time view of the finite versus non-finite tokens in York, shown in Figure 6.12, shows stability; the non-finite structures are dominant for everyone across the board.
For every age-cohort, the rates of the finite structures are under 20 percent, and there is no particular sign of change in apparent time.

6.4.3 **Comparative study of two approaches to the same data**

Given that the finite tokens are infrequent and not changing very much, it is unsurprising that neither approach to dividing them into two subgroups is especially revealing. The finite tokens of all sorts stay at low frequencies throughout, as Figures 6.13 and 6.14 show.
Figure 6.14: Tokens of seem* in apparent time in the York English Archive, split into [Subject-to-Subject raising], [permitted copy-raising], and [blocked copy-raising] (N = 200).

There is a small sex effect whereby female speakers in York use a higher proportion of the [permitted copy-raising] than male speakers do – see the cross-tabulation in Figure 6.15 – and it is statistically significant (Fisher’s exact test: df = 1, p < 0.05).

Figure 6.15: Cross-tabulation of sex and two syntactic covariants in York, excluding [blocked copy-raising] (N = 198).
However, given the lack of evidence of change-in-progress, the female lead cannot be reasonably attributed to [permitted copy-raising] coming in as it is in Canada. It might reflect local prestige of the *as though* and *as if* variants, but that is speculation.

### 6.4.4 Summary: York

The broader-level syntactic change evident in Ontario English whereby finite structures with permitted copy-raising take over from Subject-to-Subject raising is not apparent in York. There are no signs that any such change is underway; the finite tokens with *seem* are few in number for all age cohorts. While the lack of an effect in York is inconclusive, it does suggest that the second level of change observed in Canada necessarily follows *like* saturating the comparative-complementizer system. Chapter 7 will examine this possibility in more detail.

### 6.5 Conclusion

The findings in this chapter demonstrate that consistent variation and change with the expected age and sex effects given a change-in-progress can be found among covarying syntactic structures. These results show that the central variationist notion of orderly heterogeneity (Weinreich et al. 1968:100-101) can apply readily to categories of generative grammar as they are employed in vernacular speech. Additionally, they serve as a demonstration that morphosyntactic change can be operating on two levels in tandem. Younger speakers in Canada are not only saying *seem* *like* instead of *seem* *as if/as though/that/Ø*; they are also using *seem* *like* instead of *seems to*. Both changes are subject to a straightforward age effect and a female lead as occurs in most changes-in-progress (Labov 2001:501).

Dividing up the tokens in multiple ways as per Tagliamonte and Poplack (1993) showed that the structure replacing Subject-to-Subject raising in Canadian English is the one in which copy-raising is permitted – mostly with *like*. This result is not an automatic consequence of the findings in Chapter 5; it shows that *like* has broken through its original envelope of variation and caused the structure in which it appears to start drawing off another type of subordination.
Why this has occurred is a question left for Chapter 7. On the surface, there are no differences between \textit{it seems that} and \textit{it seems like}; if even among the expletives \textit{like} is catching on, there are hints that underlying properties might be involved. Chapter 7 will investigate this idea in more detail.

Inspired by comparative sociolinguistics, these findings have established that variationist methodology can incorporate abstract syntactic concepts. They also suggests that the categories of raising and copy-raising are distinct but closely enough related to be susceptible to covariation.\footnote{They may also imply that copy-raising does indeed involve movement of some kind, though I leave this suggestion for future work.} They show that \textit{that} and \textit{Ø} are distinct enough to be a separate syntactic variant from \textit{as if}, \textit{as though}, and \textit{like} – thus accounting for why \textit{like} has been observed to overtake \textit{as if} and \textit{as though} more easily, and also implying that \textit{that}/\textit{Ø} and \textit{like} are different enough that including them in the same analysis (along the lines of Chapter 5) is incomplete without a broader look at the syntax.

Aaron (2010)'s analysis of Spanish future variation is centred on the insight that when it comes to variationist approaches, “sensitivity to quantitative patterns outside the variable context may be the key to understanding otherwise inexplicable changes in distribution within the variable context” (2010:30). Her interpretation of the alternation between Peninsular Spanish periphrastic and synthetic future hinges on what is happening when it comes to epistemic markers that share a form – but not a function – with the synthetic future. Even though these are outside the conventional envelope of variation, they exert an influence. That is, the emergence of the periphrastic future as the dominant expression of futurity in Peninsular Spanish makes sense in light of the synthetic form drifting towards a different function, in a manner that is reminiscent of a chain shift (see Dinkin \textit{to appear}).
The phenomena under study in this chapter are not directly comparable to the shift uncovered by Aaron (2010). However, they share the property of being illuminated to a substantial degree by a wider look at the syntactic neighbourhood surrounding a case of morphosyntactic variation. Here, as for Aaron (2010), outside constructions that are related to the ones inside the traditional variable context have proven to be crucially implicated. The findings in this chapter suggest that, in general when it comes to morphosyntactic change, it is advisable to check on the behaviour of grammatical phenomena closely related to the main variable under study – even if the related contexts are internally non-variable. In other words, if there are related syntactic structures that can be substituted with little or no appreciable change in meaning, they ought to be studied to see whether they can help to explain the overall distributional patterns of the primary object(s) of study. This is especially the case when, as here, the number of tokens of the phenomenon under investigation is not stable in apparent time. This is a sign that the entire construction might be varying with something else.
Chapter 7
Implications and conclusions

7.1 Introduction

Chapters 5 and 6 established that two changes involving ostensibility and subordination are occurring in Canadian English. The present chapter represents an attempt at answering larger questions about what it means for there to be two linked changes (with one essentially nested inside the other). It discusses the findings in Chapters 5 and 6 with respect to some major ideas in variationist sociolinguistics and/or historical linguistics.

From the standpoint of grammatical change, “in accordance with standard theory that syntax is projected from the lexicon, it is not surprising that lexical changes can entail syntactic changes” (Miller 2002:3). Likewise, Whitman (2000:238) argues for “syntactic reanalysis formulated in terms of changes in the features of lexical items... a lexical conception of a type of change that has been thought to be quintessentially syntactic”. Words, especially heads, change their properties, and this affects the syntactic structures in which they appear (see also Heine and Reh 1984). Chapters 5 and 6 together might represent a case of this: of like displacing its covariants through lexical replacement, changing its properties, and setting off broader syntactic change. This chapter will evaluate, among other major questions, what the Chapter 6 pattern represents and whether it is a natural consequence of the Chapter 5 change.

---

99 As a useful shorthand for the two levels of change, I refer to the inter-complementizer variation as the Chapter 5 change, and the broader level of change with seem-subordination as the Chapter 6 change.
7.2 What type of change is the pattern from Chapter 6?

7.2.1 Classification

The change-in-progress from Subject-to-Subject raising involves syntactic structures and could straightforwardly be labelled a syntactic change, but beyond that its categorization is not immediately apparent. It does not qualify as a large, parameter-level change along the lines of word-order (see e.g. Faarlund 1990:Chapter 3; Kroch 2001:700), and there is no obvious place for it among Campbell’s (2004:283) exclusive “three mechanisms of syntactic change: reanalysis, extension, and borrowing” (see also Harris and Campbell 1995:3). A satisfactory resolution is provided by Campbell himself (2004:296), referring to “changes from one syntactic structure to another” as one of several types of additional, miscellaneous syntactic shift.

If this sort of change is understudied, this is probably due to it being less than conspicuous, especially while it is underway. According to Heine and Kuteva (2005:75), “a change where an existing use pattern simply acquires a higher frequency of use and comes to be used in new contexts...is hard to identify, especially since it does not affect the conventionalized structure of grammatical categorization; it therefore comes as no surprise that – more often than not – this process tends to go unnoticed in the relevant literature.” In the case of Chapter 6, a trade-off between two syntactic structures and a link to a narrower level of change make it more noteworthy. Neither structure is new to English, and neither has suddenly become dramatically different – which is as expected. Chapter 6 revealed variation and orderly heterogeneity (Weinreich et al. 1968:100,188) operating even on the level of whole syntactic units. For a

---

100 This is not meant to deny that discourse-pragmatic factors might be involved.
101 Campbell considers this to be a subset of reanalysis (2004:296). I am not convinced that this necessarily applies in the case of Chapter 6, or that it is true more generally. Bácskai-Atkári and Dekány (2014:166) agree with me in that while some cases of change from one syntactic structure to another can clearly be linked to reanalysis, even then the reanalysis may well be “optional”.
number of reasons, this is not particularly surprising. One of these reasons is the presence of partially analogous cases in the historical linguistics literature, as follows.

7.2.2 Precedents

It is not unprecedented for non-finite and finite subordination to engage in covariation or change. *Seem* is not even necessarily the only case of this in present-day English. Along with a brief mention of *looks to be* in British English, Algeo (1988:22) refers to variable complements of *recommend* and suggests that British English accepts an infinitival complement that is “nonoccurring or less common” in American English:

(100a) The Scottish teachers’ union has recommended its members to reject the offer.
(adapted from Algeo 1988:22)

(100b) The Scottish teachers’ union has recommend that its members reject the offer.

Cases of finite subordination broadly taking over from non-finite subordination within a language are also attested. Most notably, a geographic cluster of languages, not necessarily related to each other, have lost an infinitive complementation structure to the gradual encroachment of a finite one. These include Greek and a range of other Balkan languages.

**Ancient Greek** (Meillet 1948[1912]; Joseph 1983; Campbell 2004:293): Campbell (2004:293), summarizing the study of Meillet (1948[1912]), outlines how “Greek lost its original infinitive construction and replaced it with a subordinate clause construction”. Infinitive subordination is no longer used in Modern Greek; instead, the language relies on finite control structures for subjunctive constructions and the like (see e.g. Terzi 1992, 1997; Sampanis 2011; also brief discussion by Bácskai-Atkári and Dekány 2014:179). For instance, the example in (101) is literally ‘Mary wants that John buys the newspaper’; the complement is tensed.

(101) I Maria theli na agorasi o Yiannis tin efimerida.
Mary want-3sg COMP buy-3sg John the newspaper
‘Mary wants John to buy the newspaper.’
(adapted from Terzi 1997:355).
Balkan Turkish, and Balkan languages more generally (Joseph 1983; Banfi 1990; Terzi 1992, 1997; Matras 1998): A diachronic shift from non-finite to finite complementation in the case of subjunctives, as in Greek, has been found across the Balkan languages; it has spread through language contact (Joseph 1983; Banfi 1990; Matras 1998:93). For instance, according to Matras (1998:91), Balkan Turkish is distinguished from standard Turkish by “the total replacement of infinitives by finite forms” in subordinate clauses. While finite subordinate clauses were always a possibility in this context in Balkan Turkish, they have become mandatory and the non-finite subordinate clauses have been lost (Matras 1998). The process of the finite structure catching on is simply “the extension of an option already available in the language” Matras (1998:95).102

Hungarian: This is an independent historical case of non-finite subordination giving way to finite subordination (Bácskai-Atkári and Dekány 2014). Old Hungarian started out with several types of non-finite clause (infinitives, gerunds, and two types of participles), but gradually these types waned in frequency and in productivity as finite subordination caught on (Bácskai-Atkári and Dekány 2014:149). The participles have almost completely vanished from most dialects (Bácskai-Atkári and Dekány 2014:159-160); infinitives remain grammatical but have become restricted in their distribution (Bácskai-Atkári and Dekány 2014:163-166). The combined set of extant non-finite subordination strategies in modern Hungarian have become less and less like independent stand-alone clauses; finite subordination has caught on in the meantime (Bácskai-Atkári and Dekány 2014:179).

Other cases: Bácskai-Atkári and Dekány (2014:179) refer to two other cases of non-finite complementation losing ground to finite: in Khanty (Uralic) (Csepregi 2012) and Ewenki (Tungusic) (Comrie 1998), both apparently under the influence of Russian. Romani, as well, appears to have undergone the same shift (Matras 1998). Given the geographic span of the

102 Similar discussion of this sort of extension/generalization is provided by Harris and Campbell (1995: 107-108).
convergent Balkan change, it is unclear whether the change in Romani is actually the same as
the one found in Classical Greek, but Matras touches on the possibility, citing early work
investigating Greek influence on Romani (Miklosich 1880:50, as cited by Matras 1998:96).

Regardless of whether the Balkan area really represents only a single
case, this section shows that processes by which finite subordination strategies take over from non-finite ones are
attested cross-linguistically.

7.2.3 A challenge for grammaticalization theory?

As mentioned throughout Chapter 6, Hopper and Traugott (2003:175) refer directly to the
verb seem with various types of complements, exemplifying the possibility of

syntactic changes whereby initially separate clauses may become totally interlaced such that the boundaries
between clauses may become obscured at the surface level (e.g. (It) seems that he is right > He seems to be right).

On one hand, this suggests that changes on the level of entire syntactic structures, as per
Chapter 6, are to be expected. However, the problem is the direction. Hopper and Traugott refer
to almost the opposite of what was found in Chapter 6 (a case of non-finite subordination
giving way to finite subordination). Indeed, grammaticalization theory generally predicts that
“complex clause structure can be modified over time. Typically the shift...is from less to more
bonded. The final outcome of this increase in unification may be the development of simple
clauses out of complex structures” (Hopper and Traugott 2003:211). He seems to be right counts
as a ‘simple clause’ because the subordinate clause is not a self-contained CP; it is an IP that is
dependent on the matrix subject and verb.

As mentioned in Chapter 6, it is possible that Hopper and Traugott were talking about
that/Ø rather than like; and I found evidence that [blocked copy-raising], e.g. It seems that he is
right, was low-frequency and needed to be considered apart from [permitted copy-raising].
However, the larger question here is whether the change that I have observed is a problem for
the more general principle that Hopper and Traugott propose: that of clausal boundaries
eroding and subordinate clauses becoming increasingly dependent on the matrix structure over time. A shift from He seems to be right to It/he seems like he is right, as observed in Chapter 6, is not clearly one of “boundaries...becom[ing] obscured at the surface level” as per Hopper and Traugott (2003:175). It is very close to being the opposite, considering that finite subordination reintroduces tense and an overt subject pronoun – though a copy-raised nominal remains linked to its subordinate clause, which may even be a PP rather than a CP (Heycock 1994:295, Potsdam and Runner 2001:465-466, Asudeh 2002:6-8).

In and of itself, the present study is not necessarily a major problem for this aspect of grammaticalization theory. This is especially true if it is simply regarded as atypical and an exception to the notion of “less to more bonded” clauses over time (Hopper and Traugott 2003:211). The issue is that the Chapter 6 change is not easily described as an anomaly. As noted, non-finite complementation giving way to finite complementation has been observed more broadly in many languages (Meillet 1948[1912], Joseph 1983, Banfi 1990, Matras 1998; see 7.2.2); the reverse is attested but appears to be less common (Koptjevskaja Tamm 1994, as cited by Bácskai-Atkári and Dekány 2014), and Bácskai-Atkári and Dekány (2014:182) even argue that this direction is more likely across languages. The opposite notion – that of clauses tending to break down over time, e.g. finite subordination giving way to non-finite subordination – does not appear to hold.

7.2.4 Conclusion to section 7.2

The change observed in Chapter 6 can be described as a syntactic change: a shift from one syntactic structure to a second, related syntactic structure in terms of proportions in accordance with the notion of orderly heterogeneity (Weinreich et al. 1968, etc.). Like retains its old insensitivity to metaphoricality (Brook 2011a:63-64) and its ability to support but not require copy-raising.
Cases in which non-finite subordination yields to finite subordination have been reported from a range of languages, albeit often as the result of pressure from language contact, e.g. in the Balkans (Matras 1998) and in Russia (Comrie 1998; Csepegi 2012; Bácskai-Atkári and Dekány 2014). The opposite direction of change is possible – Bácskai-Atkári and Dekány (2014:179) refer to the example of Amharic (Koptjevskaja Tamm 1994) – but less frequent. It is possible that the more common direction of change represents a cross-linguistic tendency (Bácskai-Atkári and Dekány 2014:182); however, it contradicts, at least to some extent, the prediction of grammaticalization theory that clausal boundaries will tend to fuse or erode over time (Hopper and Traugott 2003:175).

The case-studies referred to here have each occurred more generally across a language than the change in Chapter 6. The shifts that occurred in Greek, Balkan Turkish, Hungarian, and neighbouring languages have involved the obsolescence of non-finite subordination. English is unlike these in that it is not, to my knowledge, losing its infinitive constructions in general. However, what the cross-linguistic examples establish is that there are precedents for a finite subordinate structure taking over from a non-finite one. If it can happen on a larger scale, there is no reason to suspect it could not within a limited syntactic domain – such as a single verb in English that has both non-finite and finite complementation strategies available. In sum, the change from one subordinate structure after seem to another in Canadian English is a lexically restricted version of a change that is cross-linguistically not unusual. It is also in the same direction that has more often been observed across languages in the historical linguistics literature. Given this, there is nothing inherently strange about it.

However, Chapter 6 described just one of two changes in the present work. Still unaddressed is the matter of whether the timing of the two changes is coincidental or whether

---

103 If any dialects of present-day English exist in which finite alternatives (e.g. I want that you take the subway) are grammatical, it would be worthwhile to watch for signs of change.
the Chapter 6 change is a consequence of the comparative complementizer *like* having reached
the saturation point among its four covariants, as observed in Chapter 5. The next section
examines this question.

7.3 **Is there a causal link between the two levels of change?**

7.3.1 **Introduction**

The Chapter 5 change describes the influx of *like*; the change in Chapter 6 involves the entire
[permitted copy-raising] structure, now mostly with *like*, increasing at the expense of Subject-to-Subject raising. There are two levels of change; is the second a result of the first?

This possibility is not intrinsically far-fetched. It is reasonable to propose that one change
might affect another – even aside from the easy example of chain-shifts in the phonological
domain. Martins and Nunes (2009), for instance, rely on the notion of syntactic “chain
reactions” to capture a multiple-step change in Brazilian Portuguese. This is not to say that
domino effects are the only possibility. One change may be able to affect another simply on the
basis of an overlap between the forms of the variants. For instance, Aaron (2010) finds mutual
influence of the epistemic synthetic future and periphrastic future in Spanish in spite of
divergent (and changing) functions; D’Arcy (2015) argues that the emergence of *have* as the
preferred variant for stative possession in Victoria, British Columbia can be attributed to the fact
that unlike *have got*, it is compatible with *do*-support, which itself has been an incoming form in
early Canadian English. Dinkin (2015, forthcoming) directly investigates whether several
different syntactic changes with *like* in North American English can be treated in the same
manner as a ‘conspiracy’-type change in phonology. Do the signs of a broader-level change owe
themselves to the change towards the complementizer *like*?
7.3.2 Canada versus the UK

As seen in Chapter 6, in both Ontario and York in apparent time, the relative proportions of the two syntactic constructions show stability at first – until in Ontario the [Subject-to-Subject raising] structure begins to decline in apparent time. This change is underway in Ontario (Figure 6.8), but there is no sign of it occurring in York (Figure 6.12). It is not a change that is occurring across dialects. Something has led to it beginning in Canada, but not in York.

Without evidence of the change beginning in York (yet), there is no way to be certain what the difference is. However, the findings are at least consistent with the hypothesis that the comparative complementizer like needs to have reached a certain threshold in the community grammar before its whole finite structure begins overtaking Subject-to-Subject raising. As an illustration of this possibility, Figure 7.1 superimposes the two innovations in Ontario English, each as proportions within its own change, onto a single graph in apparent time. It suggests that, if the hypothesis holds, the proportion of like among the complementizers needs to reach about 70 percent before the whole structure begins overtaking Subject-to-Subject raising.

Figure 7.1: Superimposition of proportion of like (out of the comparative complementizers) and proportion of [permitted copy-raising] (out of the seem-complements) in Ontario English.
The Chapter 5 results for York showed that the local comparative-complementizer system is still split between five variants. Further, although *like* has recently skyrocketed in apparent time against the other complementizers, the change has been abrupt. For older speakers, *like* is a marginal variant subject to considerable prescriptive disapproval (Huddleston and Pullum 2002:1158, Mair and Leech 2006:318). It is therefore unclear whether a similar change could be expected to occur in York someday. Nonetheless, the fact that it has not yet happened there is logical if the existence of the Chapter 6 change in Canada depends on the saturation of the comparative-complementizer system by *like*. If this is the case, then there are too few speakers in York across too narrow an age-range for whom there is active competition between *seem to* and a unified *seem*/*like* as the main finite form for this to be the case.

### 7.3.3 Hypothesized causality

In the rest of this section 7.3, I will set up this hypothesis: that once the *like* complementizer nears saturation of the comparative-complementizer system, the [permitted copy-raising] structure in which it appears will begin to take over from Subject-to-Subject raising. I will support this position with arguments from a) the Hungarian case, which proves to be fairly similar; b) an examination of the role of levelling among the complementizers; and c) a look at what this means for the expression of epistemicity and evidentiality after *seem* and the other ostensibility verbs.

Bácskai-Atkári and Dekány (2014) attribute the Hungarian change from non-finite to finite subordination to two factors. First, the Hungarian equivalent of *that* took on a wider range of uses and its versatility gave it an advantage over non-finite subordination. Second, finite subordination itself became more multi-purpose:

> On the one hand, the general finite subordinator *hogy* ‘that’ was extended in its functions and came to be preferred over non-finite structures. On the other hand, the appearance and the strengthening of specific complementizers also meant that finite subordinate clauses could be used for several functions (Bácskai-Atkári and Dekány 2014:187).
The comparative complementizer *like* is relatively new. To my knowledge, the earliest attestation dates to 1864, found by López-Couso and Méndez-Naya (2012b:316) in the Corpus of Historical American English (Davies 2010–). The oldest examples that I have found in the Oxford English Dictionary are from 1866 and 1890, both with *look* – see (35) and (36) (“*look, v. P2.c.ii*”). In the corpus of Canadian writing from 1830 to 1960 that I compiled from Project Gutenberg (Brook 2011a), the oldest comparative complementizer *like* is from 1896 – see (37).

It is thought that this *like* was previously a conjunction introducing (adjunctive, optional) clauses of manner across a range of verbs (e.g. *He sat down as if/as though/like he’s never seen a chair before*). In these contexts, *like* introduces adjuncts rather than complements (see e.g. Quirk et al 1985:1183, McCawley 1998:149, Bender and Flickinger 1999, Asudeh 2002; López-Couso and Méndez-Naya 2012a:174-175, 2012b:316; Kim 2014:176-178). What likely happened is that *like* underwent reanalysis and became able to serve as a complementizer as well as an adjunctive marker (Bender and Flickinger 1999, López-Couso and Méndez-Naya 2012b).

Relevant to this is the observation of Whitman (2000:223) that “reanalysis always proceeds from relabelling, in the sense of change in a categorial feature without any change in surrounding syntactic structure”. The lexical item *like* needed to change subtly *in situ* before other types of grammatical change could proceed. The 19th-century reanalysis of *like* as a comparative complementizer can be considered an earlier stage in the process, with stages perhaps along these lines:

---

104 This does not mean that *like* complements are necessarily CPs – the fact that they allow copy-raising may indicate that they are prepositional phrases – but that is a different matter (see section 2.1.6, as well as Potsdam and Runner 2001:466; Asudeh and Toivonen 2007, 2012; and Asudeh 2012:331).
Reanalysis of *like* as a comparative complementizer after ostensibility verbs (early 19th c.) (Bender and Flickinger 1999, López-Couso and Méndez-Naya 2012b).

Lexical replacement: *like* increasing in frequency as it overtakes other variants (20th c.) (Figure 5.2)

Robust variation between *seems to VP* and *seems like CP* (later 20th c.) (Figure 6.12)

Seems like CP begins overtaking *seems to VP* (end of 20th c.) (Figure 6.12)

If this is the case, then the Canadian English *like* is strikingly similar to the Hungarian *hogy*: *like* became able to take complements of *seem, appear, look, sound*, and *feel*; via lexical replacement it took over all of the functions of its competing variants (metaphorical subordinate clauses and concrete ones; expletive subjects and copy-raised subjects; *look* and *sound*, then *seem* and *feel* alike). It has become the default complementizer in spoken Canadian English, and it accommodates all sorts of subordinate clauses.105

**7.3.4 Conclusion to section 7.3**

The case of the complementizer *like* in Canadian English resembles the Hungarian *hogy* in that it developed considerable multifunctionality, and that this stage preceded syntactic change whereby finite subordinate structures took over from non-finite ones. In York, where *like* is very low-frequency for speakers under 30, there is no evidence of the Chapter 6 change. Given both of these findings, my answer to the question of whether there is a causal connection between the levels of change is ‘most likely’.

In the subsequent section, I assume that the answer is ‘yes’ – that the change towards *like* as the dominant comparative complementizer did cause the later change toward the permitted

---

105 There is a parallel here to the *be like* quotative, with Buchstaller (2014:9) attributing at least some of its appeal to its being a “wild card” of a variant in terms of what it can introduce (speech, thoughts, gestures, etc.). See also Dinkin (2015, forthcoming) on whether various uses of *like* do this in general, across variables. The idea that reanalysis is only possible given ambiguity (see e.g. Harris and Campbell 1995:51, Brinton and Traugott 2005:26-27) may be relevant.
copy-raising] structure – and explore what might have given *like* the ability to instigate this broader-level change.

7.4 Why has *like* been able to do this?

7.4.1 Introduction

*Like* allows, but does not demand, copy-raising: the operation is optional. As mentioned, this is one of the things that gives *like* its flexibility, along with its comparative insensitivity to verb/metaphoricality/etc. However, support for optional copy-raising is not a new property among the comparative complementizers. *As if* and *as though* allow copy-raising as well (see Chapter 2), and aside from their preference for metaphorical subordinate clauses, function fairly similarly to *like*. A crucial question, then, is why the comparative complementizer *like* could plausibly be said to have set off the Chapter 6 change if *as if* and *as though* never did so.

I suggest that the answer is primarily a combination of two things. One is that *like* has levelled the finite paradigm, helping to position finite subordination after *seem* as a unified structure. No comparative complementizer has done this before. The other, following from the first, is the redefined distinction between levels of evidentiality in the choice of matrix subject. I examine each of these in turn.

7.4.2 Two stages of levelling, and the consequences thereof

Both the Chapter 5 and the Chapter 6 changes could be described as cases of levelling. Subject-to-Subject raising is a strategy only available with *seem* and *appear* to begin with (see section 6.2.1). If *like* has unified the set of comparative complementizers by taking over from all of its covariants, then the entire set of verbs shares a finite subordination strategy – *like* – which is very flexible in terms of the arguments that it accepts. In the Chapter 5 change, *like* has levelled the set of them and ensured that all of the verbs can support copy-raising (though *feel* is usually an experiencer verb instead – see section 2.1.7).
Raising, in comparison, is a structure available for a subset of these verbs – only *seem*, if the rarity of *appear* in spoken discourse (see sections 3.2.4 and 5.1.1) is taken into account. With a finite paradigm all acting similarly in terms of the complementizer, the *seems to VP* raising structure is reduced to being a quaint alternative to one of them. If *seems like*, *looks like*, and *sounds like* have all been acting increasingly alike thanks to a frequent complementizer that behaves the same way for each of the verbs (with possible influence from new parenthetical usages as per López-Couso and Méndez-Naya 2010, 2014), then the subsequent loss of *seem to VP* may reflect a second form of levelling: the loss of an idiosyncratic alternative structure available to only one of the verbs in vernacular Canadian English. This is especially plausible given that the raising structure does not offer anything in terms of syntactic functionality that *seem* like plus a finite subordinate clause cannot, and it is more limited. Raising is obligatory, for one thing (102a), and the coindexed PRO in the lower clause has to be in the subject position (102b):

(102a) * It seems to he want to go to the bank.
(102b) * He seems to they chased.

*Like* can handle both of these:

(103a) It/he seems like he wants to go to the bank.
(103b) He seems like they chased him.

While coindexed nominals in the lower clause that are not in subject position, as in (103b), are non-canonical, they do exist (Rogers 1974a; Heycock 1994; Bender and Flickinger 1999:17; Potsdam and Runner 2001; Asudeh 2012; Asudeh and Toivonen 2007, 2012; Kim 2014). Such examples may be subject to dialectical variation in terms of their acceptability (Asudeh and Toivonen 2007, 2012; Asudeh 2012), but they are not nearly as problematic as attempting to force an object into Subject-to-Subject raising, as in (102b). The finite structure is much more versatile, and although non-canonical cases of copy-raising are unusual in the present data, the fact that they are attested means that the finite structure has additional syntactic functionality that ordinary raising determinedly lacks.
Given that *like* is the default comparative complementizer spoken Canadian English by this point and that it has usurped both the other complementizers and Subject-to-Subject raising after *seem*, the template for finite subordination constructions after *seem* (and the other verbs) has been simplified considerably:

| expletive (it/Ø) NP | seem* look* sound* feel* | like | [subordinate clause] |

To borrow a concept from statistics and/or mechanics: it is possible to talk about the number of ‘degrees of freedom’ of this template in terms of what is permitted to vary. If we oversimply slightly and assume that the (vernacular) template for speakers under 30 in Ontario no longer has any meaningful proportion of the other variants from Chapter 5, then the levelling effect of *like* has reduced the number of degrees of freedom in this template by one, and also radically decreased the number of output combinations that can be produced.\(^{106}\) With this, the other ‘degrees of freedom’, inasmuch as they show ‘variation’, no longer depend on the choice of comparative complementizer. I cover these individually:

**Variable expletive ‘it’ realization:** Now that comparative complementizer is controlled for, I assume that this ‘variation’ is modulated by style/register or degree to which the *seems like* collocation is acting as a clause-initial parenthetical. It probably reflects subtle layering, since null expletives “could indicate that [these finite structures] have already taken the first step down the path of grammaticalisation” (López-Couso and Méndez-Naya 2014:57).

**Verbs:** The four verbs in the template are subtly different in terms of their syntactic/semantic properties – especially *feel* (see 2.1.7) – but with the comparative complementizer system

\(^{106}\) This is true even if ungrammatical ones are manually excluded first (e.g. *looks that*).
levelled, as far as the *template* is concerned they are interchangeable. \(^{107}\) In other words, with *that* and Ø out of the picture, there are no combinations of verb and complementizer left that are inherently ungrammatical (e.g. *looks that*) (Huddleston and Pullum 2002:962, Gisborne 2010:276, López-Couso and Méndez-Naya 2014:39). The speaker no longer has to dedicate as much cognitive processing power to the matter of *which* complementizer to select – and whether the subject or the verb have restricted the set.

**Matrix subject type:** What conditions the ‘variation’ between a matrix expletive (whether realized or not) and a matrix noun phrase? Disregarding *feel* for being divergent in terms of how it selects its matrix subjects, the answer must lie in epistemic/evidential stance. Given the amount of detail in this argument, I assign it to its own section, focusing on evidentiality.

### 7.4.3 The role of evidentiality (and maybe epistemicity) in the Chapter 6 change

The literature points towards the respective evidential values of the expletive matrix subject and the NP subject being very different. With *like* having levelled the complementizer system, my argument is that this straightforward distinction between an expletive and an NP in terms of pragmatic value is an central advantage of the finite structure over the non-finite (raising) one, to the point that it helps account for *why* the finite pattern with *like* is taking over from Subject-to-Subject raising.

It is not contentious that a matrix NP subject before a perception verb is normally interpreted as the source of perception (Rogers 1974:76-78, Asudeh and Toivonen 2012, Rett et al. 2013, Rett and Hyams 2014:177). For instance:

\(^{107}\) A potentially interesting direction for future work, especially once *like* has saturated the system over several sequential age-cohorts, would be to hold the presence of the complementizer *like* and expletive subjects constant, treat the *verbs* as the dependent variable, and investigate whether they develop new and/or stronger systematic grammatical conditioning of their own. Another way of phrasing this – revisiting Chapter 2 – is whether *seem* will continue to be something semantically distinct from *look, sound, and feel.*
Rachel looks like she won the game.

= the speaker has seen Rachel and encountered visual evidence of her having won the game.\(^{108}\)

However, the sentence *It looks like Rachel won the game* makes no commitment to whether Rachel herself is the source of the information; this impression *could* be based on direct perception of Rachel but could also be based on a report, indirect evidence, deduction, etc. There is no commitment to information source; it is unmarked for evidentiality (Asudeh and Toivonen 2012, Rett et al. 2013, Rett and Hyams 2014).

Experimental and acquisitional research (Rett et al. 2013, Rett and Hyams 2014) supports these two conclusions. Results from an online grammaticality-judgment task with adults show that across *seem, look, and sound*, matrix NP subjects as in (104) were consistently judged problematic in the context of only *indirect* evidence of the assertion in the finite subordinate clause (Rett and Hyams 2014:182).

Rett and Hyams (2014) also perform an acquisitional study based on 45 children in the CHILDES database (McWhinney and Snow 1985). They used the conversational context to determine perceptual source and coded 55 unambiguous finite subordination tokens of *seem, look, and sound* (Rett and Hyams 2014:189). Some had matrix NP subjects and some did not.

When they had direct evidence, the children used a matrix NP subject as per (104) about half of the time (21 out of 41 tokens), otherwise opting to use an expletive *it*. With indirect evidence, there were 14 tokens, *none* of which had a matrix NP subject (Rett and Hyams 2014:189).

\(^{108}\) Several analyses (Heycock 1994:289, Gisborne 2010:Chapter 7, Mack 2010:169, Landau 2011; Kim 2014) defend scenarios in which there is a matrix NP subject but “the speaker has no direct perceptual contact with the referent of the subject” (Kim 2014:182). That is, perhaps *Rachel looks like she won the game* could be uttered if someone is reviewing the sheet of scores long over the players have left (adapted from Kim 2014:182). However, given that experimental evidence (Rett et al. 2013; Rett and Hyams 2014) clearly indicates the *unacceptability of copy-raised subjects for such indirect evidentiality, I assume that cases along the lines of the above are highly marked under most circumstances and that such scenarios can be constructed but are unlikely in spontaneous language. Mack (2010:172) concludes that although “[copy-raised] subjects need not be evidential sources, it is undeniable that they are interpreted this way most of the time.” Contexts that allow for exceptions to the general P-source rule (Asudeh and Toivonen 2012; Kim 2014) – e.g. reading Rachel’s score-sheet long after the game – also necessarily make it *obvious* that Rachel herself is not the source of perception in spite of the speaker saying *Rachel looks like she’s won the game.*
Likewise, a corpus of adult speech (Demuth et al. 2006) yielded 112 tokens. Of the 45 examples of raised NPs, all 45 were in the context of direct evidence (Rett and Hyams 2014:191).

What this all means is fairly unequivocal: that NP matrix subjects with finite subordinate clauses e.g. (104), in the speech of both adults and children, “are absolutely unacceptable [when used to describe] indirect scenarios” (Rett et al. 2013). That is, matrix NP subjects necessarily indicate direct evidence. They are inexorably marked for evidentiality, whereas expletive matrix subjects are unmarked and can refer to either direct perceptual evidence or other sources of information. If speakers want to indicate direct perception, they must use an NP matrix subject.

Using a template with only like as a complementizer means that copy-raising is always permitted. There is variation between NP matrix subjects and expletive matrix subjects. If the NP matrix subjects necessarily denote evidentiality (direct sensory perception), and the expletive matrix subjects are grammaticalizing towards becoming speaker-centered epistemic markers representing distance from a statement (López-Cuso and Méndez-Naya 2010, 2012a:180, 2014:57), then the variation between the two types of subject is easily accounted for. Rachel seems like she won the game is reserved for direct perception of evidence to that effect; it/Ø seems like Rachel won the game is the unmarked, multi-purpose statement meaning that for whatever reason(s), the speaker wants to express a mitigated level of commitment to the observation or impression. The verbs seem, look, and appear – which “combine the notions of epistemicity and evidentiality” to begin with (López-Cuso and Méndez-Naya 2014:38) – now express these two functions in a streamlined binary system.

This was not always the case. The complementizers that and Ø block copy-raising (Huddleston and Pullum 2002:962, Gisborne 2010:276, López-Cuso and Méndez-Naya 2014:39), so before like, the only ones that permitted copy-raising were as if and as though. These were not guaranteed to be equal in epistemic/evidential value. For instance, López-Cuso and Méndez-Naya (2012a:188) suspect that the choice of complementizer, especially after seem, “may
reflect the degree of the speaker’s endorsement of the embedded proposition”. With like, the system has been levelled, and copy-raising is always allowed but not obligatory. The epistemic/evidential value of the sentence now depends mainly on the type of subject.\textsuperscript{109}

This is a much simpler system than before: all four ‘degrees of freedom’ affected the epistemic stance of a template sentence before (see Mack 2010:173-174 on null versus overt expletives, plus the references above). Now there are only three ‘degrees of freedom’, the system has two fairly divergent levels of epistemicity/evidentiality, and on top of that there are fewer syntactic complications.

Evidentiality could not have been as straightforward before copy-raising was always a possibility. It seems that/Ø Rachel won the game has an expletive subject, but copy-raising is blocked; only one type of matrix subject is available for this structure with that and Ø. What does this mean for its evidential value?\textsuperscript{110}

The exact same problem affects the evidential status of Subject-to-Subject raising; there is only one option for the matrix subject as a result of the obligatory raising. I argue that, now with like having saturated the comparative-complementizer system, optionality is crucial to why the streamlined binary epistemic/evidential system as described above is coming in at the expense of Rachel seems to have won the game.\textsuperscript{111}

According to Rett and Hyams (2014:177), Subject-to-Subject raising is unmarked for evidentiality, the same way the matrix expletive structure is. That is, the raising is mandatory

\textsuperscript{109} It is also possible that the epistemic value of seem differs subtly from that of look and sound (López-Cuso and Méndez-Naya 2012b:327).

\textsuperscript{110} This, in my view, is the reason why the conclusions Grimm (2010) do not quite agree with the findings of Rett and Hyams (2014). Grimm (2010) defends the position that expletive subjects are marked for indirect evidentiality (see also Matushansky 2002:225). However, the only complementizer that Grimm examines is that. With like and the other complementizers that permit copy-raising, I am inclined to agree with Rett and Hyams (2014:192) that the expletive structure is unmarked for evidentiality.

\textsuperscript{111} See Gisborne and Holmes (2007) on the history of appear: “Pragmatic effects can drive semantic change, and semantic change, in turn, can drive syntactic change.”
but *Rachel seems to have won the game* can refer to direct or indirect evidence. If it is the case that
with a whole range of other complements – finite subordinate clauses as above, adjectives, etc. –
*seem, appear, look, and sound* are functioning as evidentials (see e.g. Gisborne 2010:Chapter 7),
then the raising structure is an anomaly. It is the only type of structure with an NP matrix
subject for which a connotation of direct evidentiality is *not* guaranteed:

(105a) This boy seems lost.  
(105b) This boy seems like he is lost.  
(105c) It seems like this boy is lost.  
(105d) This boy seems to be lost.

My argument is that there is no place for an extra ambiguous structure with rigid matrix-
subject restrictions, especially one whose evidential value is not in-line with the other options.
The two finite options are now streamlined with *like*, and free from the confusion of what the
evidential status of *It seems that* is given that it blocks raising (Grimm 2010). The way the binary
system now works with *like* is presumably as follows: if the speaker wants to indicate direct
evidentiality, they use a matrix NP subject. If they do not want to commit to this, for any
reason(s) (indirect evidence only, uncertainty, hedging, etc.), the epistemic *It seems like* is
available.112 There is no place for an inflexible matrix NP subject structure that does not clearly
indicate either of these options – unmarked for evidentiality, but also not an epistemic structure.

Figure 7.2, though it is a slight oversimplification, captures this:

![Figure 7.2: Schematic of evidential/epistemic stance of two finite strategies with like and Subject-to-Subject raising. The Subject-to-Subject raising option offers no syntactic flexibility and does not clearly correspond to either option.](image)

112 This may also account for why comparative complementizers also seem to be increasing with the other verbs
besides *seem* (Figure 6.1) – if the entire epistemic/evidentiality system with this class of verbs has been reorganized
and unified.
The system has become organized such that there is essentially a one-to-one mapping between semantic/pragmatic function and syntactic structure. Subject-to-Subject raising does not have a place in it.

7.4.4 Conclusion to section 7.4

I have argued that two properties associated with like have given it the ability to recalibrate the finite subordinate system (It seems like he is ill and He seems like he is ill) such that Subject-to-Subject raising is no longer of much use in comparison. The first is that it has saturated the comparative complementizer system and streamlined the template of [ostensibility verb] + like + [finite subordinate clause]. The second is that this has allowed for the reorganization of the finite subordination options such that they line up with evidential/epistemic values. Copy-raising is always a possibility now; speakers can select an NP matrix subject to convey direct evidentiality and the matrix expletive for a lessened degree of commitment for reasons having to do with indirect evidentiality and/or hedging. Not only does Subject-to-Subject raising offer no functionality or meaning that these two options do not, but in comparison it is both ambiguous (it conveys neither evidentiality nor epistemicity in any sort of straightforward way) and confusing (given that matrix subjects with seem-type verbs are otherwise evidentials). This is the explanation that I propose for the change in Chapter 6.113

7.5 What about feel?

Most of this chapter has focused on seem, look, and sound. While feel is normally different in terms of its argument types (see section 2.1.7), it has followed the same change towards like in Chapter 5, and is barely behind seem in this regard.

113 It is not necessarily the only factor. There may also be a cross-linguistic tendency for finite subordination to overtake non-finite subordination (Bácskai-Atkári and Dekány 2014:182).
While *feel* is capable of acting as a copy-raising verb, as in (106a), it more often simply functions as an experiential verb referring to the internal emotions/opinions of the speaker, as in (106b) (see section 2.1.7 on disambiguating tokens of *feel* followed by a finite subordinate clause; see also Rett and Hyams 2014:176).

(106a) This chair feels like it was made by extraterrestrials with no knowledge of human anatomy.
(106b) I feel like this chair was made by extraterrestrials with no knowledge of human anatomy.

In spite of how different *feel* usually is from the other verbs, it is going through the change towards the comparative complementizer *like* just as *seem* is (Figure 5.4), and the difference is not significant (Table 6.5). This speaks to the form being more important than the function, at least in the case of this verb in this change. I conclude that surface similarity of the syntactic templates (and the overlapping ability to support copy-raising, at least in theory) enabled *it feels like/I feel like* to follow *it seems like/she seems like*, in spite of the multiple ways in which the typical argument structure and theta-role assignment of these two verbs diverge (see Chapter 2).

It is also possible that *I feel like* is already competing with *I think* as an epistemic parenthetical (Thompson and Mulac 1991b, etc.). There has been some attention paid to this issue in several online articles. Liberman (2013) refers to *I think, I feel like*, and *seems* as just a few of many colloquial “words and constructions that can be used to hedge statements” – a function closely linked to epistemicity (Kärkkäinen 2003:22).114 Baker (2013) wonders why “over a dozen” of her female friends report “us[ing] ‘I feel like’ liberally” – the phrase strikes her as a “verbal tic” particularly associated with women, and Liberman (2013) confirms a female lead,115 which hints at change-in-progress (Labov 2001:501) affecting *I feel like*. Two scholars independently use the Google Ngram Viewer (Michel et al. 2011) to show evidence of its

---

114 Thompson and Mulac (1991b:322) cite Benaviste (1971:228-229) saying something similar: that verbs of “mental operations [with first-person NP subjects] convert into a subjective utterance the fact asserted impersonally”.

115 I would like to thank Rich Smith of Seattle’s The Stranger for bringing these two articles to my attention in our email discussion that formed part of his research for an article on *I feel like* (Smith 2015).
increase in raw frequency: Liberman (2013) with *I feel like* (Liberman 2013), and Squires (2013) with *feel like* on its own. Both of these are replicated below.

Figure 7.3: Replication of Liberman (2013), showing a sharp increase in the frequency of the collocation *I feel like* across English written sources via the Google Ngram Viewer.

Figure 7.4: Replication of Squires (2013), showing a sharp increase in the frequency of the collocation *feel like* across English written sources via the Google Ngram Viewer.

Figure 7.5: Synthesis of the above two graphs. Note that *feel like* includes all of the examples of *I feel like*. 
There is a need for caution here, and for multiple reasons. One is that *feel like* in the Google Books database is not unambiguously followed by a finite subordinate clause – *feel like NP* and related structures are all included in the Google Ngram Viewer search, including a few pop-culture phrases that may well be disproportionately common, for instance:

(107a) Man! I feel like a woman! (Shania Twain, *Come On Over*, 1997).
(107a) I feel like such a turnip! (*The IT Crowd*, “Moss and the German”, Linehan 2007).

The other reason to be careful with the Ngram results is that *like* itself is an incoming form. That is, an increase in *feel like* is to be expected given that *feel like* is simply taking over from *feel as if*, *feel as though*, and even *feel that/Ø* (see Chapter 5). The below figure shows that because of the rise of the *like* complementizer in general, *feel like* and *I feel like* are not necessarily set apart from *seem(s) like*, *sound(s) like*, and *look(s) like* from the standpoint of the Google Ngram Viewer.

![Graph showing the usage of various like-structures](image)

*Figure 7.6 Look like, looks like, feel like, sound like, sounds like, seem like, seems like, and I feel like across English-language sources via the Google Ngram Viewer (Michel et al. 2011).*

Evidence beyond the above charts, therefore, is needed in order to make a strong case for grammaticalization of *I feel like* (and possibly *it feels like* as well). While it may be true that *I feel like* is taking on “a usage that’s mostly a somewhat more involved way to say ‘I think that’” (Liberman 2013), future work will need to establish whether *I feel like* is doing anything that *I feel as if*, *I feel as though*, *I feel that*, and *I feel Ø* have never done.
*Feel* is a hybrid. It differs from the epistemic markers in the *I think* class in that it accepts *like, as if, and as though* complementizers on top of *that* and Ø. Its trajectory of change has followed that of *seem*. But it is not (ordinarily) a copy-raising verb, either. *I feel* was already an epistemic marker; whether grammaticalization is necessary to account for the insertion of *like* under the influence of *seem* and the over verbs is unclear. Also, Denis (2015) offers a cautionary study in why assuming prototypical grammaticalization in the case of pragmatic markers may not be advisable in the first place.

Nonetheless, if *I feel like* is taking on greater grammatical functions, even if quite abruptly as per Denis (2015), several predictions can be made – among them, that the parenthetical usage (*This is pretty good, I feel like*) will begin to appear and then increase in frequency (Rodríguez Louro and Harris 2013:419; see also Thompson and Mulac 1991b:326, López-Couso and Méndez-Naya 2014). The present study cannot speak directly to whether this is occurring already, since *like*-parentheticals were excluded from the variable context in favour of examining the finite subordinate clause template, but *I feel like* is worth watching since it may be moving into at least one new envelope of variation.

### 7.6 Is there a link to the *be like* quotative?

There are a number of reasons to draw parallels between the *seems like* type changes and the *be like* quotative. Most saliently, both involve *like* as a subordinator and are setting off major grammatical change. It is even possible that they are related constructions: the *it-is-as-if* construction (Declerck 1992, Calude and Delahunty 2011) appears to have given way to *it-is-like* (Brook 2011:69-71; see also López-Couso and Méndez-Naya 2012a:181). *It-is-like* or *it’s like* is also a construction able to introduce quotatives (see e.g. Romaine and Lange 1991:230,

---

116 The *it-is-as-if* pattern is closely enough related to *seems like, looks like, sounds like*, etc. that López-Couso and Méndez-Naya (2012a) combine it in their analysis.
Given that this use of *it’s like* is not compatible with older quotatives (*it says, *it goes), it is reasonable to propose that it was associated with the introduction of *be like* (as well as *be all* and *be just* and other *be-* centered quotatives).

*Seem* is a copula just like *be* (Quirk et al. 1985:1174), so it is not surprising that phrases along the lines of *it seems like* are occasionally used as quotatives themselves. Blondeau and Nagy (2008) cite several examples of this in the process of investigating complementizer functions of *like*. López-Couso and Méndez-Naya (2014:58, note 11) point out a small number of examples from the Corpus of Contemporary American English (COCA) (Davies 2008–). Davydova and Buchstaller (2015:452) even count *feels* as a copula that can appear in their quotative template. Indeed, there are several *feel* quotatives in the Ontario data. For example:

(108) He just feels like, "Oh, I have to say anything and do anything t-- sk-- like keep her."
(Clara Felipe, F 22, Toronto)

The *it’s like* construction might therefore be a sibling of *it seems like* and an ancestor of *I’m like*. Certainly both the *seems-like* class of changes and the introduction of the *be like* quotative have been fairly rapid changes (see e.g. Tagliamonte 2012:248). It is interesting and possibly telling that as with *seems like* structures, the *be like* quotative appears to have caught on very abruptly in York (Gardner et al. 2013), likely due to transmission/incrementation in Canada and diffusion in the UK (see e.g. Labov 2007, Tagliamonte and Denis 2014). This could be coincidental, but it could also reflect underlying similarities or mutually-reinforcing influence of the change towards *seems like* and the change towards *be like*.

---

117 When it comes to the Chapter 5 change, *like* shares an advantage with the *be like* quotative: a lack of sensitivity to a grammatical factor that majorly constraints its competitors. These are metaphoricality in the case of *seems like* and content-of-the-quotation for *be like*. See Dinkin (2015, forthcoming) for further analysis of *like’s* versatility across variables.
There are other commonalities. Dinkin (2015, forthcoming) points out that seems like structures and be like quotatives, as well as the discourse like, all share “vague literality”: that is, a lack of sensitivity to how precise their arguments are. This makes them flexible: for instance, the like comparative complementizer can take either metaphorical or literal subordinate clauses, and the be like quotative can introduce either reported speech or reported thought (Romaine and Lange 1991, Ferrara and Bell 1995). Given the commonalities and the possibility of a shared relative in the it’s like structure, while the Chapter 5 change is primarily lexical replacement and be like is a case of the quotative system completely reorganizing itself, my sense is that there has been some degree of reciprocal influence.

7.7 Conclusions

The present work has argued that the change towards the complementizer like in Ontario English is only one part of a much larger picture. To begin, it is the smaller of two levels of change. While like has overtaken as if and as though in vernacular speech and is now also driving out that and Ø, the story does not stop at the saturation point. I have defended the hypothesis that when like has reached a certain threshold in the community’s comparative-complementizer system on the vernacular level – once it is the central variant and its covariants are low-frequency and/or restricted to different registers – it sets off a change on a broader syntactic level. With like prevalent enough, the finite subordinate structure that it appears in will consistently allow for two types of matrix subject, which is twice as many as any of the individual competing structures allow. That and Ø only permit expletive matrix subjects; Subject-to-Subject raising requires NP raising. In both of these cases, there is no grey area. As section 2.1.6 established, there are no unambiguous exceptions to the former in the Corpus of Contemporary American English (Davies 2008–); syntactic rules prohibit any attempts at

forcing Subject-to-Subject raising to act differently. *Like* permits either, in what is otherwise the same syntactic template. The acquisitional/experimental literature strongly suggests that what governs the ‘variation’ between matrix NP subjects and expletive subjects before copy-raising verbs is evidential/epistemic status (Rett et al. 2013, Rett and Hyams 2014). A copy-raised (matrix NP) subject is an unequivocal way of indicating direct perceptual evidence for the assertion in the subordinate clause, whereas an expletive matrix subject is a multifunctional epistemic phrase that can represent uncertainty, ambiguity, hedging, or more. With this distinction emerging and always possible with *like* as the complementizer, structures that do not fit into the paradigm for either syntactic or semantic reasons are suddenly less useful. I have argued that this accounts for why with a large enough proportion of *like*, Subject-to-Subject raising has begun to decline (along with the *that*/Ø complementizers). Once *like* has reached a critical point that seems to be around 70 percent (Figure 7.2), there is enough of a streamlined evidential system in place (*it seems like* versus *NP seems like*) that Subject-to-Subject raising is no longer necessary. When there is robust variation between the permitted-copy-raising structure and Subject-to-Subject raising, the former is likely to catch on assuming that it has reorganized itself to align with evidentiality and epistemicity as I have hypothesized. Note, for instance, in the variable-context-division approach to dividing the finite tokens (Figure 6.7), the [NP seem* finite] and [it seem* finite] structures are increasing. An evidentiality system with two options requires both of these (although the copy-raising function seems to be on the increase relative to the expletive one with *seem*, at least, in Chapter 6 – perhaps *as if* and *as though* were absent from the vernacular complementizer inventory for long enough that this is simply the verb *seem* reclaiming its ability to encode direct evidentiality with copy-raised subjects).

Subject-to-Subject raising – mandatory raising – is ambiguous in terms of evidential/epistemic stance, and it is unlike other matrix NP structures with *seem* in that it is not a direct evidential. Both of these properties put it at a distinct disadvantage opposite a binary
finite system in which the type of subject necessarily corresponds to the information source and level of commitment.

Whether this process could account for cases of finite subordination taking over from non-finite subordination in other languages – Greek, the Balkans in general, and Hungarian, among others (Meillet 1948[1912]; Joseph 1983; Banfi 1990; Matras 1998; Bácskai-Atkári and Dekány 2014) – is not obvious. It might well be idiosyncratic to some degree given its lexical restriction. However, there are major similarities between *like* and the Hungarian complementizer *hogy*: they both acquired a new range of functions, and consequently a finite subordinate structure increased in rate of use opposite a non-finite one. While these cases contradict the hypothesis from grammaticalization theory about clausal boundaries gradually eroding (Hopper and Traugott 2003:175), they appear to represent a tendency across languages (Bácskai-Atkári and Dekány 2014:182). Further work on diachronic approaches to clause boundaries is merited.

This study suggests that looking beyond the conventional variable context in a morphosyntactic change – as per Aaron (2010) – is potentially highly valuable. Getting to know the syntactic ‘neighbourhood’ around a particular variable and/or phenomenon may produce considerable additional insights. This is especially advisable in the case of token *numbers* that, as in Figure 6.1, are correlated with speaker age and/or year-of-birth rather than evenly distributed. It is a sign that the variation under question may represent only half of a broader-level change.

The Chapter 6 case can be described as a change from one syntactic construction to a second (Campbell 2004:296), making both the variable and the variants more abstract than usual for variationist studies. However, the notion of orderly heterogeneity (Weinreich et al. 1968) applies just as well as it usually does, and the results show the same effects of age and sex that are typically observed for changes-in-progress.
This work is fundamentally variationist, driven by cautious delineation of both variable contexts and variants on two levels and comparative methodology that has allowed me to evaluate multiple ways of dividing up the same data (Tagliamonte and Poplack 1993). However, the study has drawn extensively on theoretical background from syntax, semantics, and pragmatics.

Copy-raising verbs have occasionally been called an encoded system of evidentiality for present-day English (Rett and Hyams 2014), but this is far from agreed-upon; others argue that “evidentiality is neither [embedded in the grammar] nor obligatory” in modern-day English (Rett et al. 2013). While English is certainly not a language in which every clause and/or assertion requires a statement as to the speaker’s source of information, the influx of like is giving Ontario English a more consistent and reliable evidential system than it had. Operating in tandem on two levels of change in accordance with classic behaviour of variants in the midst of change, an innovative comparative complementizer variant and a moderately offbeat syntactic transformation have not only levelled a paradigm in two ways but harnessed a small set of perception verbs to create a binary epistemic/evidential system for subordinate clauses. This discovery owes itself to variationist methodology and syntactic/semantic theory combining forces. It can be concluded that there is ample room for abstract syntactic properties and categories in the rigorously quantitative realm of variationist research – and vice versa.
References


ARCHER: *A representative corpus of historical English registers* (2006). Northern Arizona University, University of Southern California, University of Freiburg, University of Helsinki, and Uppsala University.


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


Butters, Ronald (1982). Editor’s note [on be + like]. American Speech, 57, 149.


Hothorn, Torsten; Peter Buehlmann; Sandrine Dudoit; Annette Molinaro; and Mark Van Der Laan (2006). Survival ensembles. *Biostatistics, 7*(3), 355–373.


López-Couso, María José; and Belén Méndez-Naya (2012a). On the use of *as if, as though, and like* in present–day English complementation structures. *Journal of English Linguistics, 40*(2), 172–195.


Matras, Yaron (1998). Convergent development, grammaticalization, and the problem of ‘mutual isomorphism’. In Winifred Boeder; Christoph Schroeder; Karl Heinz Wagner; and Wolfgang Wildgen (eds.), *Sprache in Raum und Zeit: Beiträge zur empirischen Sprachwissenschaft*, 891–103. Tübingen: Gunter Narr Verlag.


Michel, Jean-Baptiste; Yuan Kui Shen; Aviva Presser Aiden; Adrian Veres; Matthew K. Gray; William Brockman; The Google Books Team; Joseph P. Pickett; Dale Hoiberg; Dan Clancy; Peter Norvig; Jon Orwant; Steven Pinker; Martin A. Nowak; and Erez Lieberman Aiden (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014): 176–182.

Miklosich, Franz (1880). Über die Mundarten und die Wanderungen die Zigeuner Europas XII. Vienna: Karl Gerhold’s Son.


Quirk, Randolph; Sidney Greenbaum; Geoffrey Leech; and Jan Svartvik (1985). *A comprehensive grammar of the English language*. Harlow, UK: Pearson Education.


Rodríguez Louro, Celeste; Alexandra D’Arcy; and Sali A. Tagliamonte (2013). Before *be like*: Digging into earlier global systems to understand a 21st-century innovation. Paper presented at NWAV 42, Pittsburgh, PA, October 17–20.


Squires, Lauren (2013, October 19). [Chart showing the rise in frequency of the collocation ‘feel like’ from Google N-Gram Viewer] [Infographic]. Retrieved from <https://twitter.com/prof_squires/status/391661379247013889>.


Tagliamonte, Sali A.; and Marisa Brook (*submitted*). Why does North American English use *try to* but British English use *try and*? Let’s try *and/to* figure it out.


