

Thir or these: A corpus study of variation and change in Older Scots demonstratives

Ian Clayton, January 17, 2007

1. Introduction

Scots, the distinctive sibling of English which evolved in Scotland in the course of the fifteenth century, has attracted a considerable amount of scholarly attention in the last decade or so. This development is a felicitous one, because Scots has heretofore occupied a shadowed but entirely undeserved place in the linguistic literature in comparison with its more prominent southern neighbor, even though there is much to be learned from the study of Scots about, among other things, the development of standards and prestige dialects, and the interaction of dialectal and historical forces.

The present paper explores the effect of certain extra-linguistic forces on the linguistic development of Scots during the period of anglicization extending from the mid-16th to the late 17th centuries, in the spirit of recent quantitative diachronic research programs carried out on the subject of Scots (see e.g. Meurman-Solin 2000b and 1993b), and southern “English” English (van der Wurff 2000, and a variety of productive studies in the Helsinki and Edinburgh schools). The subject of my research has been the Scots form of the demonstrative article/pronoun *thir* and its ultimate replacement by the southern English form *these*. I have explored how the competition between these two forms is reflected in their relative frequency, as well as the important roles played in that competition by the extralinguistic forces of prestige and geographic distance.

This paper assumes that a quantitative analysis of variation can effectively illustrate the impact of extralinguistic forces such as social class, nationality, political context, and literary genre on the adoption and diffusion of linguistic innovations. Further, it assumes that these

extralinguistic forces are in fact indispensable to a full explanation of language change: inasmuch as a given innovation is not—cannot be—instituted simultaneously by each member of a broader linguistic community, its introduction must instead have an identifiable locus, and any subsequent dissemination through the linguistic community should likewise follow identifiable pathways, be these pathways through geographic or social space.

Study of these extralinguistic forces is crucial to solving the so-called *actuation problem* (Weinreich, Labov, and Herzog 1968): as Peter Trudgill points out in his preface to Milroy (1992), linguistics has made notable strides in identifying what kinds of changes are most likely to occur, and which least, and even why; but we are as yet inexpert at accounting for the manifestation—the actuation—of these changes at particular places and times. This paper does not solve the actuation problem, but it makes an important contribution toward such a solution, by illustrating how a quantitative examination of textual sources can identify the point in geographic and social space where a given innovation first appears, trace its diffusion (or actuation) within and between communities of practice (Meyerhoff 2002), and even identify the probable sociopolitical motivation for its adoption in the first place.

In this respect, this study follows in the footsteps of such pioneering works of historical sociolinguistics as Romaine (1982). Romaine's study was notable for demonstrating that the variationist paradigm can be applied within a broad diachronic scope, and what is more, that the paradigm would not suffer through use of textual rather than spoken media as its principal informants. Rather, the written medium has a value on a par with the spoken: "once we accept that the basic dichotomy exists between language and medium (and not between language and writing, or speech and writing), it can no longer be argued that all forms of the written language . . . are outside the field of linguistics" (Romaine 1982, p. 15).

However, to ensure their value, principled criteria should be employed in the selection of written texts, just as with living informants, for example those offered by Schneider (2002):

1. texts should be as close to the vernacular as possible, to the exclusion of literary and formal texts; the value of the latter is obviated by their adherence to “categorical, invariant usage” (p. 71);
2. authors should represent a range of social categories (age, class, sex);
3. texts must contain variation in the feature to be investigated;
4. texts must be large enough to offer some assurance of statistical validity.

Schneider’s first criterion will be challenged in this paper: it is not always the case that formal or literary texts are of little value. While relatively invariant usage is indeed the norm in such texts, any departure from such invariance is itself interesting: in the present study, variation in formal Scots usage was symptomatic of a shifting standard, the product of a sea change in the balance of political and social power which was to have enormous linguistic ramifications.

The balance of the paper is organized as follows. In section 2, I present a brief overview of the historical development of Scots, making reference to landmark events in its history, and mentioning a number of features that distinguish it from the southern English standard. Section 3 is a description of the Helsinki Corpus of Older Scots, the principal research tool upon which this study is based. Finally, section 4 describes the study and summarizes its results.

2. Divergence and convergence: historical trends in the development of Scots

The historical development of Scots is almost impossible to discuss without reference to English. Both speech forms evolved from the same linguistic stock, the so-called Anglo-Saxon language brought to Britain in the fifth century AD by Germanic tribes. Specifically, Scots is a

daughter of the northern or Northumbrian variety of Middle English, from which it was virtually indistinguishable until approximately the beginning of the fifteenth century; the inscription on the Ruthwell Cross (ca. 800 AD) in Dumfriesshire, Scotland, is geographically the earliest extant example of Scots, but "might have been carved, as far as linguistic propriety is concerned, at Edinburgh or at York" (Smith 1908: xii).

The Scots with which this study is concerned first arose, by a process of incremental differentiation from northern Middle English, in the early part of the fourteenth century. It is variously termed Old, Older, or Early Scots; "Early Scots" is perhaps the best term, since it reflects historical fact without prompting undue comparisons with such other "Old" Germanic languages as Old English, Old Norse, or Old High German, which are considerably more archaic in nature. Nevertheless, because the bulk of literature to which this study refers uses it, and because it is part of the name of the corpus my own research is principally based on, I will use the term "Older Scots" throughout this paper.

This Older Scots of which we speak is, of course, accessible only the form of texts written by authors who were, for the most part, cultured, highly educated, and literate. While Smith perhaps overstates the case when he says that "Middle Scots was, more exclusively than any companion phase in the languages of north-west Europe, the special affair of literary habit, as distinguished from spoken dialect" (Smith 1908: xi), it is important to remember that this literary language was indeed largely the product of conscious and active linguistic accommodation and adjustment on the part of its writers; this fact was to have deeply significant consequences for the long-term development of written Scots. Aspects of this conscious accommodation will be addressed in section 4.

What propelled this early differentiation of Scots from English? There is no simple answer to this simple question. Of course, the force of time and distance would have done its work. But we must also consider the effects of national sentiment and historical events. Popular, anti-English sentiment during and after the wars of independence between Scotland and England certainly encouraged the development of a divergent standard by the first half of the fifteenth, a standard then to see good use in the developing political institutions of an independent Scotland (Meurman-Solin 1993). A century and a half later, growing political and social connections between the two kingdoms, in which England played a dominant role, eroded the Scottish standard. All of the characteristic Scots forms described in the following paragraphs, and many more, underwent a prolonged period of attrition beginning in the mid-16th century, a process steeply accelerated following the union of the two kingdoms under James VI/I in 1603. By the end of the seventeenth century, many if not most distinctive Scots forms were rare to the point of vanishing in written documents.

The Older Scots literary standard is distinguished from the southern standard by a number of genuinely linguistic features—lexical, syntactic, morphological, and phonological—as well as other features that are more graphematic than linguistic. Lexical items particular to Older Scots included *wrangwis* "unrighteous," *gudeserr* "grandfather," *barnis* "children," *fretis* "prophecy," and *soldatis* "soldiers" (examples from Görlach 2002: 113). Syntactic divergence initially included a lack of *do*-support (later acquired through anglicization; see Meurman-Solin 1993b, and Moessner 1997). Morphological features included the invariant indefinite article *ane*; the archaic present participial ending *-and* beside gerundial *-ing*, as *quhare Johnne was baptysand* beside *This man com into witnessing*; and retention of the Northern plural demonstrative *thir* rather than the Southern *these*—the subject of the present study.

Phonologically, Older Scots differed from Southern English in the absence of rounding in long /a:/ (yielding *stane*, *wrang* "stone," "wrong"); the extension of open-syllable lengthening to the high vowels /i, u/, in contrast to southern English which applied the rule only to mid and low vowels; and the retention of the archaic voiceless velar fricative /x/ (Johnstone 1997).

Graphematic peculiarities of Older Scots included past tense indicative *-it*, *-et* in lieu of southern *-ed*, as in *defamet*, *deformet*, *turnit* 'defamed,' 'deformed,' 'turned'; plural *-is* rather than *-es*, as in *schouris*, *towris* 'showers,' 'towers'; and the use of *quh/qwh* to indicate /w/ or voiceless /ɹ/, as in *quhip*, *quhen* 'whip,' 'when.'

3. The Corpus

The Helsinki Corpus of Older Scots (HCOS, Meurman-Solin 1995a, 1995b, 1993a) is designed to supplement the Helsinki Corpus of English Texts, by providing a substantial body of material representing the language form that arose in Scotland over the fifteenth and sixteenth centuries, and which characterized the bulk of Scottish literary production at that time. The purpose of the Corpus is to allow for study of the linguistic developments of Scots over the period 1450-1700, through the quantitative investigation of sociolinguistic variables in light of relevant and consistent extralinguistic factors such as genre, function, audience, and geography.

The corpus consists entirely of prose texts, taken from sources available in printed, edited form, and specifically from those editions which have made no regularization or redaction of the original material (M-S 1993a). Only texts original to Scots have been admitted, with the exception of a small number of translations among the very earliest samples, which have been included in order to provide the desired distribution of samples across type. Fourteen genres are included in the corpus: these include law (from burgh records), (auto)biographies, Biblical texts,

travelogues, sermons, diaries, educational texts (chiefly for religious instruction), private letters, official letters, handbooks, trial proceedings, histories, and pamphlets. A greater emphasis has been placed on the period 1570-1640, since this period, marked by the Scottish Reformation and Unification, features the greatest density of linguistic changes (Meurman-Solin 1995c).

Eighty texts have been included in all, distributed as follows:

Period:	No. of texts:	Word count:
1450-1500	9	95,900
1500-1570	18	200,500
1570-1640	29	302,300
1640-1700	24	247,000
<i>Total:</i>	<i>80</i>	<i>Total: 845,700</i>

At present, the HCOS contains no syntactic or other structural tagging, but each entry does contain tagging for a variety of extralinguistic variables; these include the mode of the text (i.e. speech-based, as in trial records or sermons, or written, as in most other genres), formality (formal, as in official letters, or informal, as in private correspondence), the writer's age and gender, and the status of the audience relative to that of the writer.

While the compiler's goal of diversity has largely meant that multiple works by the same author have not been included within the corpus, there is a small number of fortuitous exceptions, which prove valuable in comparing the relative abundance of certain linguistic forms in a given author's works, depending on e.g. the function of the text in question, and the sociopolitical climate in which it was produced.

Finally, I emphasize the point that a corpus is best thought of as a collection of individual informants, each of which can and will behave in a highly idiosyncratic manner, and display great individual variation. Therefore, any statements regarding general, corpus-wide trends must be interpreted with that fact in mind. On the other hand, it is this very idiosyncratic nature of

individual texts that provides insight into the extralinguistic factors that may condition linguistic variation and change.

4. Prestige, distance, grammatical markedness: three sources of variation

4.1 The relative abundance of thir and these across time and genre

In this section of the paper, I illustrate variation between the Scots form *thir* and the English form *these* in three different contexts. First, I examine variation in the writings of individual authors: in the private diary of John Lesley, Bishop of Ross, excerpted from 1571, and in Lesley's *History of Scotland*, published in 1570; and in two writings of James VI/I: the *Basilicon Doron* of 1598, and *A Counterblaste to Tobacco*, written in 1604. The examination reveals that both authors were capable of making deliberate adjustments to their use of Scots or non-Scots forms, according to the demands of audience and historical events. Second, I investigate variation in the use of *thir* and *these* in the language of contemporaneous Parliamentary Acts and Burgh Records, illustrating how greater distance from the socio-political center of Scotland correlates with higher frequency of the conservative *thir*. Finally, I investigate the interaction of grammatical function with the respective frequencies of *thir* and *these*, showing that a greater degree of functional markedness correlates with a greater degree of regularity in use.

As part of my preliminary investigation, I tallied all the examples of *thir* and *these* in each of the eighty texts of the HCOS. The percentage of instances composed by *these* has been charted in Figure 1. There it is plainly to be seen that throughout the first century and a half of material represented in the corpus, the southern form *these* was completely absent, not making an appearance until 1570. Once it appears on the scene, *these* is able to achieve a virtually complete

victory within a century, having displaced *thir* from the corpus altogether by 1689. Not only is the speed of displacement remarkable: notable also is the great degree of variability between texts. While the first text to display the southern form (in 1570) does so in only 29% of instances, the second text featuring *these* shows this form in 100% of instances. On the other hand, as late as 1680, some texts continue to show a preponderance of the conservative Scots *thir*, as for instance the diary of 1680, with only 33% *these*.

Figure 1.

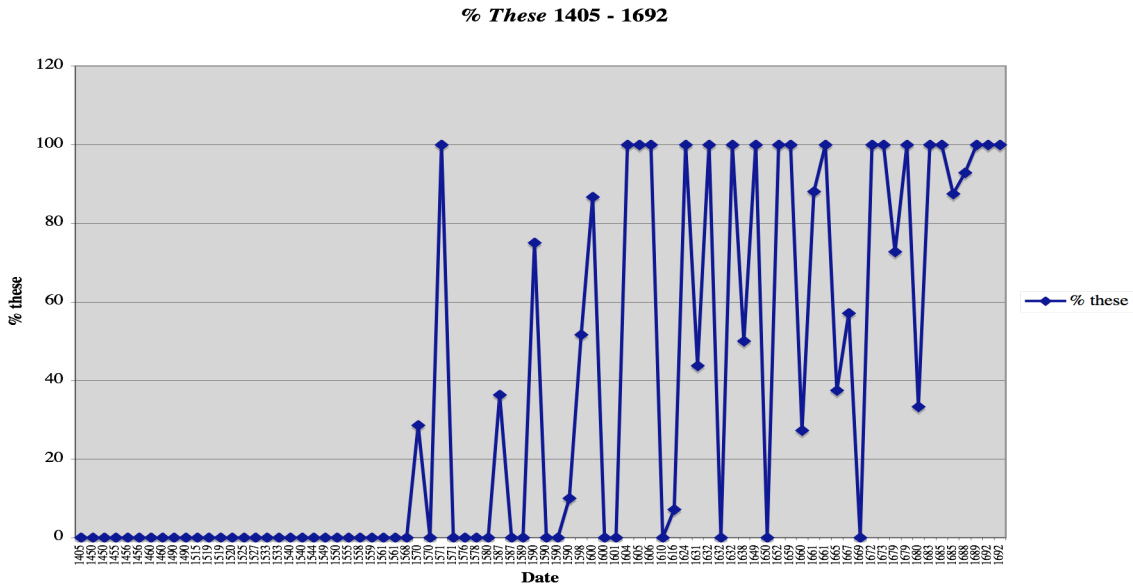
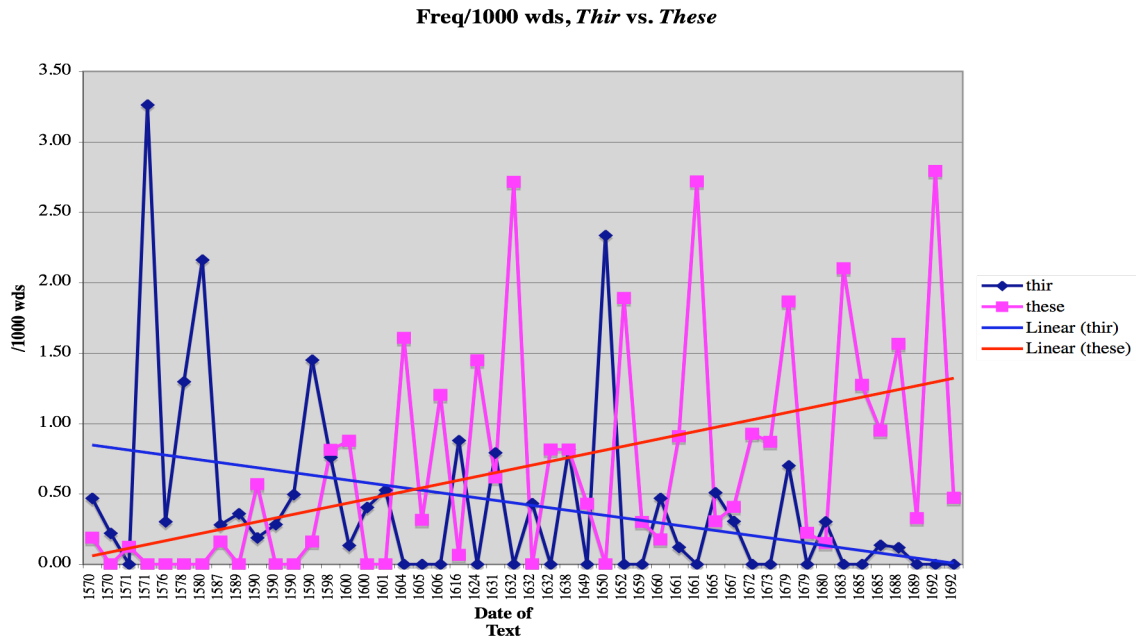


Figure 2 provides similar information, this time ignoring all texts prior to 1570, and representing the occurrence of *thir* and *these* normalized as instances per 1000 words of text. Once again, the extreme variability is apparent.

Figure 2.



4.2 *Thir* and *these* as prestige forms

Comparison of multiple texts within the HCOS produced by the same authors indicates that both *thir* and *these* were highly marked forms, that their users were conscious of this fact, and that speakers made active use or disuse of either form as the circumstances warranted. What is more, in the case of one author, James VI/I, the changing fortunes brought by the winds of historical events to each of the competing forms are easily seen.

That a writer could easily alternate between *thir* and *these* according to the function and audience of the text is illustrated by the writings of John Lesley, Bishop of Ross. The HCOS includes two texts by this author: an excerpt from his diary, written in 1571, and a portion of his *History of Scotland*, published in 1570. In the approximately 8000 words of the diary excerpt, *thir* appears not at all. This indicates that, for whatever reason—be it education, childhood environment, or other—Lesley did not include *thir* as a component of his basic idiolect. Had he done so, *thir* would surely have surfaced in such an informal, minimally redacted, and private

context as his diary. By contrast, at virtually the same period in his life, Lesley used *thir* as frequently (71% of instances) as he did *these* in the *History* that, given the historical context, must have been intended principally for Scottish eyes. In short, Lesley made use of *thir* when motivated to do so by audience considerations, but apparently not otherwise.

The HCOS also contains two texts written by James VI, King of Scotland, and as of 1603 King James I of England, which show that writers were able to equally favor *these* once the political winds had changed following the unification of the two kingdoms. In the *Basilicon Doron* of 1598, a kingship manual intended for his son Henry, James actively employs Scots forms. Addressing this narrow, private, albeit royal audience, James used *thir* in approximately 48% of instances. The picture is entirely different in *A Counterblaste to Tobacco*, published in 1604, only 6 years later. This pamphlet, disparaging the use of tobacco and addressed to the King's "good Countrey men," is "programmatically [Southern] English," according to Görlach (2002: 146), and features Southern forms practically across the board: spelling, morphology, and syntax (e.g. *do*-support) are Southern in character. *Thir* is no exception: James made no use at all of this form in *A Counterblaste*. Following the Unification of 1603, the King was now powerfully motivated to adapt his language to the new political climate.

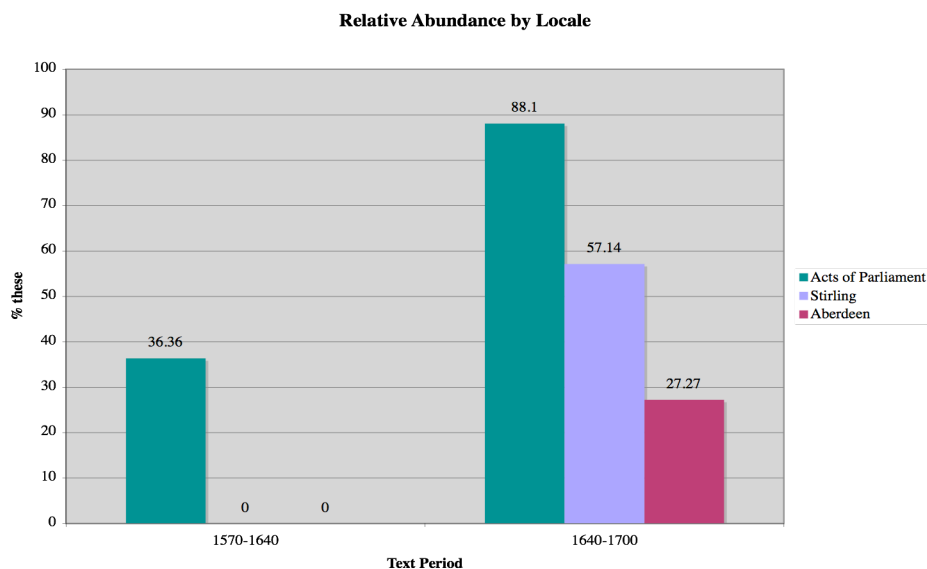
4.3 Variation across geographic space

One expects to see a linguistic innovation imposed from above by a superstrate to first take effect among those closest to sources of prestige, either geographically (e.g. proximity to sociopolitical centers), or socially (among those in greatest contact with the superstrate, e.g. government officials or other representatives of political power). In this scenario, once established at a sociopolitical center of gravity, the novel form spreads towards the periphery

(see for example Britain 2002), either through inter-speaker contact, or through the dissemination of textual materials.

This pattern of diffusion seems to be confirmed by certain texts in the HCOS. While the corpus is limited in its ability to display trends dependent on physical geography because of its emphasis on Central Scots, the inclusion of burgh records from both Stirling (a socially and politically central region) and Aberdeen (a peripheral region) (M-S 1995c) allows us to examine the diffusion of *these* across geographic space. If burgh records and legislative material are accepted as having similar function and audience, we may also include Acts of the Scottish Parliament (located in Edinburgh) into our comparison. As figure 3 shows, a comparison of texts from these three sources demonstrates the early intrusion of *these* in the Parliamentary Acts in the period 1570-1640, when the Southern form is still absent in the somewhat less central Stirling burgh records and the peripheral burgh records of Aberdeen. By the period 1640-1700, *these* has nearly supplanted *thir* in the language of the Acts, and is beginning to dominate the records of Stirling, but shows a distinct lag in the records of Aberdeen.

Fig. 3



These findings largely agree with those of M-S (1995c), which investigates the relative abundance of the native Scottish past tense grapheme *-it* versus the Southern English *-ed* in the same three groups of texts. In the period 1570-1640, *-it* was highly favored, at 88%, 84%, and 94% for Acts of Parliament, the Stirling Records, and the Aberdeen Records respectively. By the period 1640-1700, the use of *-it* in the Acts of Parliament and the Stirling Records had dropped to 2% and 12%, respectively, but remained at 75% for the Aberdeen Records, again illustrating a lag in the adoption of novel forms in this peripheral region.

More broadly, these findings are partially consistent with Labov's (2003) generalization that "the crucial division in the society from the point of view of language change was not middle class vs. working class, but rather [socially] centrally located groups as against peripherally located groups," what he terms the Curvilinear Model (p. 32). The HCOS is not sufficiently representative across social class to permit a fine-grained application of Labov's model, but it certainly appears that the novel form *these* was introduced first among the upper and middle classes and disseminated thence to the balance of society. On the other hand, there does not appear to be confirmation of the model's generalization that the uppermost classes are resistant to innovation. I do not take this as a direct challenge to Labov's model; rather, I assume that it is a product of the unique sociopolitical situation that obtained in Scotland in the late sixteenth and early seventeenth century: the innovative form was introduced from outside, and as a token of a new power balance, was naturally gravitated to by those individuals who had both the greatest motivation and the greatest degree of contact with users of the southern English standard.

The pattern of geographic diffusion observed in this study, on the other hand, is fully consistent with the wave model, as elucidated in Hock (1991): linguistic innovations are

disseminated gradually from a focal area (Edinburgh), where they were first adopted, to a transition area (Stirling), and only then to a relic area (Aberdeen).

5. Conclusions

This study has shown that a corpus-based, quantitative investigation of linguistic variables may be a profitable means to discover the nature and mechanism of superstratal or convergent forces in language change. The study reveals that language users are capable of making conscious use or disuse of novel linguistic forms, varying in such use according to audience and political context, among other extralinguistic factors. The study further demonstrates that, in the case of convergence toward a recently imposed superstratal linguistic form, it is indeed at the apex of the sociopolitical body that novel forms are introduced into the local language from the superstrate, and from which these innovations are diffused through the linguistic community.

While there are a great number of further directions such investigations may be taken, those which seem to me to be most immediately desirable include determining whether other forms introduced into Scots from southern English follow a timeline similar to that I have outlined for these, or whether there are significant differences in the date and speed at which such introductions take place. It would also be of interest to see if such innovative forms follow the same pattern of geographic distribution, i.e. from the sociopolitical center toward the periphery, as *these* and *-ed*. Finally, it would be of interest to learn whether there are distinct correlations between the time and rate of *these* adoption and extralinguistic factors not examined here, such as sex, degree of education, or text genre.

References

- Britain, D. 2002. Space and spatial diffusion. In J.K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, eds. *The Handbook of Language Variation and Change*. Malden, MA: Blackwell. 603-637.
- Görlach, M. 2002. *A Textual History of Scots*. Heidelberg: Universitätsverlag C. Winter.
- Hock, H.H. 1991. *Principles of Historical Linguistics*. New York: Mouton de Gruyter.
- Kay, B. 2006. *Scots: The Mither Tongue*. Edinburgh: Mainstream Publishing.
- Johnstone, P. 1997. Older Scots phonology and its regional variation. In Jones 1997, 47-111.
- Jones, C., ed. 1997. *The Edinburgh History of the Scots Language*. Edinburgh: Edinburgh UP.
- Labov, W. 2001. *Principles of Linguistic Change. Volume 2: Social Factors*. Oxford: Blackwell Publishing.
- Meurman-Solin, A. 2003. Corpus-based study of Older Scots grammar and lexis. In John Corbett, J. Derrick McClure, Jane Stuart-Smith, eds. *The Edinburgh Companion to Scots*. Edinburgh: Edinburgh UP. 170-196.
- 2000a. Change from above or below? Mapping the *loci* of linguistic change in the history of Scottish English. In Laura Wright, ed. *The Development of Standard English 1300-1800: Theories, Descriptions, Conflicts*. Cambridge: Cambridge UP. 155-170.
- 2000b. The progressive in Middle Scots. In Teresa Fanego, María José López-Couso, Javier Pérez-Guerra, eds. *English Historical Syntax and Morphology: Selected Papers from 11 ICEHL, Santiago de Compostela, 7-11 September 2000*. Philadelphia: John Benjamins. 203-229.
1997. Differentiation and standardization in Early Scots. In Jones 1997, pp. 3-23.
- 1995a. Ed. *The Helsinki Corpus of Older Scots, 1450-1700*. Distributors: Oxford Text Archive (Oxford) and Norwegian Computing Centre for the Humanities (Bergen).
- 1995b. A new tool: the Helsinki Corpus of Older Scots (1450-1700)." *ICAME Journal* 19. 49-62.
- 1995c. Text profiles in the study of language variation and change. In Raymond Hickey, Merja Kytö, Ian Lancashire and Matti Rissanen, eds. *Tracing the Trail of Time: Proceedings from the Second Diachronic Corpora Workshop, New College, University of Toronto, May 1995*. Atlanta: Rodopi. 199-214.
- 1993a. "Older Scots." In Rissanen et al. 1993, 75-82.

- 1993b. Periphrastic *do* in sixteenth- and seventeenth-century Scots. In Rissanen et al. 1993, 235-251.
- 1993c. *Variation and Change in Early Scottish Prose: Studies based on the Helsinki Corpus of Older Scots*. Helsinki: Suomalainen Tiedeakatemia.
- Meyerhoff, M. 2004. Communities of Practice. In J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, eds. *The Handbook of Language Variation and Change*. Blackwell Publishing. 526-548
- Milroy, J. 1992. *Linguistic Variation and Change. On the Historical Sociolinguistics of English*. Oxford: Blackwell.
- Moessner, L. 1997. The syntax of Older Scots. In Jones 1997, 112-155.
- Rissanen, M., Merja K., and Palander-Collin, M., eds. 1993. *Early English in the Computer Age: Explorations through the Helsinki Corpus*. New York: Mouton de Gruyter.
- Romaine, S. 1982. *Socio-historical Linguistics: Its Status and Methodology*. Cambridge University Press.
- Schneider, E. 2004. Investigating Variation and Change in Written Documents. In J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, eds. *The Handbook of Language Variation and Change*. Blackwell Publishing. 67-96.
- Smith, G. 1903. *Specimens of Middle Scots*. New York: Dodd, Mead, and Co.
- Taavitsainen, I. 1993. Genre/subgenre styles in Late Middle English? In Rissanen et al. 1993, 173-200.
- van der Wurff, W. 2000. Variation and change: Text types and the modelling of syntactic change. In Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg, and C.B. McCully, eds. *Generative Theory and Corpus Studies: A Dialogue from 10 ICEHL*. New York: Mouton de Gruyter. 261-282.
- Weinreich, U., Labov, W. and Herzog, M. 1968. Empirical foundations for a theory of language change. In W.P. Lehman and Y. Malkiel, eds. *Directions for historical linguistics*. Austin: University of Texas Press. 95-189.