

## From inventory to typology in English historical dialectology\*

Anneli Meurman-Solin

Helsinki Collegium for Advanced Studies, University of Helsinki

This research is part of a project examining the theoretical and methodological implications of corpus-based inventories in the production of typologically relevant new knowledge. The focus on Scots allows me to highlight the importance of full-scale inventories of local and regional variation in varieties that failed to become national standards, diatopic variation often being downplayed in typological research. The main area of linguistic interest is clause combining systems in Older Scots.

### 1. General goals

In the reconstruction of language varieties of the past, the following three general goals seem to me particularly relevant. Firstly, the creation of a wider range of diplomatically transcribed manuscript sources, preferably as part of large electronic databases; secondly, the development of sophisticated tagging systems and software for data retrieval and presentation tailored for diachronic text corpora; and thirdly, the creation of metalanguage and data presentation conventions which are more appropriate for reporting on continued variation and variability attested in corpus-based descriptions of language than those applicable to describing relatively unidirectional processes of change such as language standardization, or construed realities such as language standards. The relevance of this third goal is obvious in my own field of study, that of Scottish historical linguistics. The comparative approach and its metalanguage, frequently

---

\* This research was supported by the Academy of Finland Centre of Excellence funding for the Research Unit for Variation and Change in English at the Department of English, University of Helsinki.

adopted in the field, have not always enhanced our understanding of developments in Scots. Comparability across categories and the typological relevance of findings across languages and language varieties has sometimes been assumed, rather than shown to be valid, as sufficiently detailed corpus-based descriptions have not been available. The risk of assuming direct comparability has become more obvious with the improvement in the quality and quantity of data.

In my work I focus on reconstructing variation and variety in the history of English as used in the various regions and localities of Scotland. This reconstruction takes place by producing detailed inventories of idiolect- or text-specific practices. When such things as careful descriptions of micro-level language use based on family letters are available, these micro-level practices are examined to find out the extent to which they converge with or diverge from those attested in other idiolects, texts or groups of texts in databases tailored to be diachronically, diatopically and diastratically representative. Manuscript-based data representing chiefly non-literary texts are made available in large quantities while trying not to lose any of the quality achieved by applying good philological computing practices in the digitization process. My approach is similar to that adopted at the Institute for Historical Dialectology, University of Edinburgh (Laing & Williamson 2004). The two Edinburgh projects working towards *A Linguistic Atlas of Early Middle English* (Laing 1993) and *A Linguistic Atlas of Older Scots* (Williamson 1992/1993) were launched from the tradition of linguistic atlases based on questionnaires, but the methodology has been developed further so that the databases of linguistic material are lexico-grammatically tagged corpora of full texts, diplomatically edited, rather than questionnaire-delimited sets of isolated word-forms. Further, the 'fit-technique', a method of interpolating texts of unknown provenance into a dialect continuum, has been computerized (Williamson 2000). To guarantee the validity of linguistic description, the analysis of language variation over time and space is based on microscopically precise text profiles containing each linguistic item in the shape attested in the original manuscript, with no preselection or hierarchization involved. Manuscript-based, lexico-grammatically tagged corpora of this kind and the software developed for extracting and presenting data will give us reliable descriptions, or valid inventories, if you like. The full scale of variation can be detected, and the patterns of change triggered by variety-internal or -external factors can be identified as complex processes rather than as properties of objectified systems. In the case of Scots, the tradition of focusing on the rise of the so-called Scottish English Standard, and especially the anglicization of this standard, often restricting the focus to texts originating from the south-east, has

sometimes imposed a perspective that may prevent the researcher from seeing the multidirectionality of developments over time, place and social milieu.

My third general point, the inadequacies of metalanguage, will allow me to contrast our awareness of language variation and variability and the tendency to objectify or reify language varieties and historicize them.

## 2. Metalanguage: Language variability and the objectification and historicization of varieties

### 2.1 Reification

The concepts of 'objectification', 'hierarchization', and 'historicization' summarize the essence of trends that, in my view, have perhaps slowed down the process of increasing our knowledge of varieties of English, both those varieties which developed into the so-called standardized varieties and those that have remained local or regional. I would like to claim that these three trends have greatly influenced both the theoretical and methodological approach widely adopted in the study of varieties of English and the metalanguage used in the description of linguistic findings.

In his chapter entitled "The consequences of standardization in descriptive linguistics" James Milroy (1999: 17) discusses "the interaction between scholarly linguistic attitudes to language and the publicly expressed attitudes of non-linguists and critics of linguistics" and claims that "the consequences of standardisation are discernible in the attitudes of linguistic scholars themselves: their judgements as to what the object of description consists of have been influenced by their knowledge that a standard form exists in some abstract dimension and by some consequences of the ideology of standardisation." Crowley (1990) and Benson (2001) have also drawn the conclusion that despite our awareness of the inherent variability of language we tend to provide descriptions suggesting that we have in fact identified a language variety. With reference to Crowley (1990), Benson (*ibid.* 20) points out that "[i]n the Saussurean scheme ... language variation can only be explained if each non-standard form is treated as a system in its own right" and, as a consequence, "[t]he inherent variability of language, which justifies the distinction between *langue* and *parole*, is thus reduced to a proliferation of distinct language 'varieties', each having its own invariant systematicity".

Benson (*ibid.* 21) has also identified the tendency to objectify in the

lexicographical practice of using national and regional labels, which is his main concern:

Anthropological and sociological approaches to language variation both tend to reduce the inherent *variability* of language to language *variety*. Variety imposes order on variability and produces innumerable independent objects for scientific investigation within a single language. This process is especially apparent in the literature on World Englishes, which has tended to emphasise the identification, description and classification of regional and national varieties of English over the description of the fluidity of English as a language functioning across geographical and political boundaries. It is also apparent in the broad approach of twentieth-century English dictionaries to the international dimensions of English, which tends to adopt the concept of language variety uncritically in the form of national and regional labels. While language variation is now generally seen as 'good', there is at the same time a tendency to systematically reduce it to objectified and describable forms. Twentieth-century dictionaries are implicated in the objectification and homogenisation of language varieties in much the same way as their predecessors were implicated in the objectification and homogenisation of the English language as a whole. (Benson 2001:21)

Objectification or reification may be an innate characteristic reflecting the functioning of the human brain, but mental frames or schemata of this kind are problematic as they may regulate processes of analysis in which linguists are trying to identify some order in heterogeneity, that is, some relatively consistently preferred practices in data that otherwise chiefly give evidence of heterogeneity and continued variation. Demarcating areas as the territories of specific varieties may divert our attention from examining ordered heterogeneity recordable only by crossing such imposed boundaries. The following questions seem relevant. To what extent do the methods applied to the study of varieties of English reflect this tendency to insist on the existence of language entities, or systems, even at a stage when there may not yet be sufficient, i.e. statistically significant, evidence of uniformity or systematicity within such an entity? How has the tradition of giving legitimacy to a description of language use in a specific place by identifying it as a distinctive entity affected our ability to develop appropriate discourses for discussing linguistic variability? How has reification affected the way in which we conceptualize Scots or Scottish English?

One of the main points of my paper is that the negative influence of reification can be avoided by improving both the quantity and quality of text corpora containing digitized and preferably tagged data produced by as wide a

range of text and discourse communities as possible. A particularly important consequence of de-reification in describing Scots is that variation and variety resulting from varieties and languages having come into contact on Scottish soil will be given due attention.

## 2.2 Hierarchization

A further problem arises from this tendency to reify varieties of English: the entities, or objectified systems, tend to be hierarchically ordered. This seems to lead to the widespread practice of adopting a comparative viewpoint especially when examining the non-standardized or otherwise less prestigious varieties. From the centrist perspective, regional and local varieties are satisfactorily described in terms of how they relate to standardized varieties. Moreover, descriptions of this kind are often expected to focus on how regional or local varieties diverge from standardized varieties, so that the often considerable proportion of shared features between them will not be discussed with sufficient thoroughness. The tendency to adopt the comparative approach to the study of regional and local varieties can be seen as a reflection of the views in Milroy (1999) quoted above.

Judging by earlier research, it seems to have been difficult to resist the temptation of approaching the study of the Scots language from this comparative angle. For a number of political and socio-cultural reasons, the tradition of objectifying the Scottish English variety may be especially powerful, and in the non-corpus-based studies in particular the tendency to highlight divergence and downplay convergence has not taken us closer to understanding the history of Scots. No comprehensive account of the Scots variety has thus yet been made available. The most important achievements are in the field of lexicography — the *Dictionary of the Older Scottish Tongue* and *The Scottish National Dictionary*, in particular (Dareau 2004).

## 2.3 Historicization

The third trend, 'historicization', has had similar consequences in the reconstruction of text languages as reification. Communities using regional varieties may have a tendency to historicize the variety for it to gain legitimacy (cf. Milroy 1999:28). As Milroy points out, "[m]uch of the legitimisation of the standard variety is in fact achieved through creating a history for the standard language — a process that we can reasonably call *historicisation* (involving the

creation of a legitimate historical canon)". While the legitimacy of the standard variety has influenced the ways in which we conceptualize regional and local varieties, the historicization of the Scots language, for instance, has maintained attitudes that have justified the use of an allegedly independent language variety in the pursuit of nationalistic goals. Having been educated in a canon-saturated tradition makes it difficult not to be influenced by the historical canon. More work at the record offices browsing through bundles of documents is needed to improve the range of data (Meurman-Solin 2001a). When a text community is defined in terms of what written texts verifiably had a social and communicative function among the literate members of that community, it may be possible to deobjectify and dehistoricize.

### 3. Tagged corpora

Four corpus projects are relevant to providing better inventories of Scots. The Helsinki Corpus of Older Scots (HCOS) is available on the ICAME CDROM, but a new version is being produced by checking the texts against manuscript and annotating them using the Institute for Historical Dialectology software developed for lexico-grammatical tagging. Cooperation between research units in Edinburgh and Helsinki will allow us to create a diatopically, diachronically and diastratically representative corpus of Scots which covers the time-span from the earliest vernacular texts around the mid-fourteenth century up to the end of the eighteenth. In addition to the Edinburgh-Helsinki corpus (E/HCOS), I am compiling a Corpus of Scottish Correspondence (CSC), based on a careful search of the family archives, in order to ensure diversity as regards dialect, the writers' social and geographical mobility and linguistic and stylistic competence. This corpus will also be tagged. My fourth database (CESWW) will consist of early Scottish women's writings. In all of these, the texts will be diplomatically transcribed and then keyed in using principles of philological computing (Meurman-Solin 2001a: 18–22). No normalization or emendation is allowed.

At this point some comments on tagging will shed light on tools that facilitate the use of detailed inventories in typology-oriented work. Not all tagging is useful. It is possible to distort evidence by applying overly rigid rules for category membership or ignoring the innate fuzziness between categories, and by simplifying complex patterns of variation or using tags that fail to reflect processes of change over a long time-span, for instance. Tagging is not just a tool that facilitates searching. To be of general relevance it cannot apply

definitions created by a specific theoretical approach. We focus on giving information about structural, i.e. formal, and semantic features rather than category-defined or syntactic properties. In cases in which it has not been possible to avoid theory-specific practices, these must be made as transparent as possible. Since tagging must allow a sufficient degree of flexibility, our tagging system allows different degrees of refinement in the choice of type of tag by introducing a hierarchical way of ordering information in the tags. The lexeme *consider* usefully illustrates the practice of refining tags to indicate categorical fuzziness or polyfunctionality, so that the grammaticalization of the present participle can be traced:

```
$consider/vpsp_CONSIDER+ING $/vpsp_+ING
$consider/vpsp-cj_CONSIDER+ING $/vpsp-cj_+ING
$consider/vpsp-pr_CONSIDER+ING $/vpsp-pr_+ING
$consider/vn_CONSIDER+ING $/vn_+ING
```

The practice is similar when tagging conjunctive phrases incorporating nouns such as *on the consideration that*:

```
$on/pr_ON $/T_THE $consideration/n-cj_CONSIDERATION
$that/cj<_THAT
```

Even though the tag also provides a basis for the analysis of syntactic properties by indicating the connective function, the main criterion in the identification of candidates for a full inventory of realisations is semantic. The tagging principles applied to clause-combining devices at this experimental stage have been influenced by the discussion of notional or conceptual properties of word-classes (Anderson 1997, Jackendoff 2002). This can be illustrated by the following set of conjunctive phrases implying purpose:

- (1) \$for/pr\_FOR \$fear/n-cj\_FEAR \$that/cj<\_THAT  
 \$to/pr\_TO \$/T\_THE \$intent/n-cj\_INTENT \$that/cj<\_THAT  
 \$to/pr\_TO \$/T\_THE \$end/n-cj\_END \$that/cj<\_THAT  
 \$to/pr\_TO \$/T\_THE \$effect/n-cj\_EFFECT \$that/cj<\_THAT  
 \$in/pr\_IN \$intent/n-cj>vi\_INTENT \$to/im+C\_TO \$live/vi\_LIVE \$long/  
 av\_LONG  
 \$in/pr\_IN \$intent/n-cj>pr\_INTENT \$of/pr\_OF \$live/vn\_LIVING  
 \$long/av\_LONG

We will find that in the case of some candidates a grammaticalization process may not actualize; if it does, the spread may take place at a varying rate or only

affect restricted domains (on the grammaticalization of *seeing* and *considering*, see Meurman-Solin and Pahta forthcoming).

These practices allow us to mark topic areas and specific features for a later more detailed analysis, and are therefore appropriate for studying language change over a long time-span. In other words, this method ensures that category changes, polyfunctionality or fuzziness requiring closer analysis can be examined with all the necessary rigour, and items that may play a role in a pattern or taxonomy can be identified and defined.

The order of information in the tag is from general to particular, so that searches can be modified according to the kind of information considered relevant in a particular piece of research. Arrows are used to indicate that an item should be examined as being possibly part of a multi-unit structure or collocate. Thus I have indicated the presence of an adverb in the immediate context whenever that adverb can be viewed as having semantic properties enriching those of the linking device in some systematic way (e.g. *providing always, if ever, or ever, ay quhill* (cf. *ay and quhill*), *as soon as ever, how soon that ever, how soon and incontinent, how soon... incontinent*). An interesting cluster of a conjunctive participle semantically enriched or refined by the frequently co-occurring adverb is *providing always*, as illustrated in example (1):

- (1) The inquest ordains to ansuer Robert Atzin, and ilk ane of the officaris, of ane ferlot of meill in this storme to help thair wiffis and barnis, providing allwayis that thai clame na possessioun thairof in tyme cuming. (HCOS 1555 Peebles records, 225; ed. W. Chambers 1872)

The tagging is as follows:

```
$provide/vpsp-cj_PROVYD+ING $/vpsp-cj_+ING $always/av<_ALWAYS
$that/cj<_THAT
```

This cluster occurs in 62 per cent of the occurrences of *providing* as a subordinator in the HCOS data, the proportion in statutory and administrative texts being as high as 73 per cent. This feature is related to the dominance of generic time reference in such texts.

Corpus compilers are improving the quantitative and qualitative validity and relevance of electronic databases but, for reasons of research economy they may make compromises in the tagging by relying on pre-corpus-linguistics descriptions, resorting to automatic (i.e. non-interactive) tagging, or imposing neat category labels on the data. Yet the high degree of scholarly rigour in ensuring the authenticity of data should also be applied in developing methods

for annotating texts. In our view, tags suggesting membership of predefined categories based on theory-bound syntactic properties must be avoided in the study of regional and local varieties of English in particular, since the grammar of such varieties may significantly differ from that of standardized varieties. In other words, great caution is necessary as decision-making about individual tags may reflect an implicit application of grammatical rules written on the basis of data not valid or relevant in the context.

#### 4. Inventories and typologies: Clause-combining

The typology of clause combining-devices in English as well as other European languages has been extensively discussed in recent literature, diachronic developments of adverbial subordinators in Kortmann (1997) being particularly relevant as regards points discussed in this section; for synchronic descriptions, see, for example, Devriendt et al. 1996 and van der Auwera 1998. As the ongoing corpus-based work has not yet produced detailed inventories that draw on data extracted from qualitatively and quantitatively representative text corpora, typologies have been construed chiefly by using secondary sources such as dictionaries and grammars (Kortmann 1997: 53). Regional or dialectal variation tends to be ignored or marginalized, which may lead to an excessive degree of streamlining and simplification in the selection and presentation of information to ensure the production of valid typologies. Problems related to insisting on typological relevance even though inventories are insufficient are multiplied in the comparative approach since the lack of consistency and coherence in earlier descriptions may flaw comparisons seriously.

The streamlining effect can be illustrated by examining how the non-corpus-based inventory of adverbial subordinators in Present-day English (Kortmann 1997: 292–294) differs from a corpus-based inventory of Present-day Scots (Häcker 1999: 211). Beside information about variant forms (*whenever/finiver/whaneir, frae/fae/from/faem, once/yince/eence*), the following subordinators in Häcker's list are absent in Kortmann's inventory: *as gin, barrin, da wye at, even for as, for aa, gin, the minute/da meenit, onietime, say* and the doubtful *okay*. Häcker's study is based on both written and spoken texts dating from the period 1980–1994, but, as regards genres, only narrative prose is represented (ibid. 18–19). As numerous written genres in present-day Scotland are not assumed to contain diatopically relevant data as a result of the widespread preference for Standard English, the differences can be assumed to be more salient earlier in

the history of Scots. Section 5 will highlight changes in the inventories of some adverbial subordinators in Scots in the 1450–1700 period, focusing on the chronology of changes, the identification of periods when the most dramatic changes take place, and the developments in the general frequency of items in specific semantic roles. I will also illustrate developments in the morphological complexity of adverbial subordinators and their preferred categorial sources.

As the present study is chiefly based on the HCOS, a corpus of not more than 850,000 words, a full inventory cannot yet be provided even though the wide range of genres in this database, fifteen altogether, increases the significance of the findings. Comments on the typological relevance of the inventory must thus also be preliminary. Beside the expansion of databases, my earlier research has suggested a number of areas where further research is necessary. Meurman-Solin (2001a: 20–22) highlights the importance of using diplomatically transcribed manuscript sources as data because patterns of complex sentence structures in Present-day English may have been imposed as part of the editorial practice of introducing punctuation. In assessing the representativeness of corpora it also seems important to use stylistic competence as a criterion. Meurman-Solin (2001b: 42–45) contrasts letter-writers whose repertoire of connectives only includes items such as *and* and *for* with those who manage to indicate logical relationships between propositions more explicitly. Leaving aside regional and dialectal variation, the question of what can be claimed to provide a representative range of data will have to be considered in the selection of genres and discourses as well as the varying stylistic competence of the informants. A thorough knowledge of the evolution of discourse- and genre-specific practices over time and space (see Brinton 2000, for instance) will permit the assessment of the performance of individual informants in a wider context.

Moreover, Meurman-Solin (2002: 189–193) discusses the problems related to categorization, overlapping and intersection between categories such as conjunctions, prepositions and adverbs, suggesting that in order to define the concepts of ‘subordination’ and ‘subordinating clausal link’ the whole repertoire of connectives will have to be analysed. Degrees of fuzziness have important typological implications. Instead of seeing fuzziness as ambiguity, it is useful to apply the concept of polyfunctionality here (cf. Kortmann 1997: 89–94). In addition, in considering the hierarchical downgrading of subordinate syntagms from parataxis to adjoined clauses, correlative diptychs, clause chaining (including co-subordination), and embedding (Lehmann 1988: 183–186), the identification of subordinating conjunctions that should be included in

inventories of the kind illustrated above complicates matters further.<sup>1</sup> As suggested in Section 3, identification can be facilitated by a tagging system which is protean in the sense that it can flexibly record information about structural features — single features and collocates — as well as arrange that information in various ways without preconceiving their syntactic properties.

Within the scope of the present study, the focus has been restricted to what Lehmann calls ‘adjoined clauses’, where “[o]ne of the two clauses constituting the complex sentence contains a subordinative conjunction and may thus be identified as the subordinate clause” (ibid. 185). It has also been necessary to exclude non-finite and verbless adverbial clauses, as their inclusion may also have required including a wide range of structural realisations on the continuum from non-finite clauses to nominalizations.

##### 5. Adverbial subordinators in English

Kortmann offers a valuable description of the typology and history of adverbial subordinators in European languages as well as “a macro-analysis of the history of adverbial subordinators in English” (1997: 346–348). He draws our attention to a number of topics where further research is necessary (ibid. 291). Fluctuations over time in the composition of the inventories and the relative proportions of adverbial subordinators signalling temporal relations compared with causal, conditional, concessive and related interclausal relations should be traced. Secondly, he stresses the need to examine changes in “the morphological make-up of adverbial subordinators” and to assess the varying importance of the origins of subordinators in English. Other interesting questions are whether adverbial subordinators in English developed “towards an ever higher degree of semantic differentiation” or whether “a decrease in (semantic and syntactic) polyfunctionality” can be attested. In the summary of developments from Old English to Present-day English, Kortmann (ibid. 291–335) compares the donor rates of the different periods (ibid. 301), and lists the major results as regards diachronic developments (ibid. 345–348), also formulating further specific research questions (ibid. 348–351).

1. In ‘correlative diptychs’... “the relative clauses ... are subordinate, but not embedded”; some types of ‘clause chaining’ are referred to by Foley & Van Valin (1984: 238–263) as ‘co-subordination’ and involve the use of adverbs defining the semantic link (Meurman-Solin 2002: 188, 206, fn 4).

According to Kortmann, Old English “provided the majority of those adverbial subordinators in Present-Day English with the highest frequency of use”, including *after*, *as*, *as long as*, *as soon as*, *if*, *since*, *so*, *that*, *though*, and *while*, so that “about 70% of those Old English subordinators which made their way into the 20th century belong to this set of high-frequency items”. He writes that “[t]he adverbial subordinators with Old English ancestors can therefore justly be viewed as forming the backbone of the subordinator inventory in Present-Day English, both with respect to their frequency of use and the range and nature (i.e. basicness) of the interclausal relations they cover” (ibid. 296). The innovativeness of Middle English is obvious since “roughly 75% (!) of its adverbial subordinators had no predecessors or were not used in this function in Old English” (ibid. 299). Kortmann (ibid. 302) explains the fact that “Middle English, especially Late Middle English... crucially shaped the inventory of adverbial subordinators” referring to the emergence of “a fairly stabilized written language in the period from roughly 1350 onwards”, both “as a literary language” and “the language of official use”. He considers the low donor rate and high drop-out rate (ibid. 301, Tables 10 and 11), and the sheer size of the Early Modern English inventory of adverbial subordinators (“larger by a third compared with Middle English and by some 50% compared with Present-Day English”), and draws the conclusion that “Early Modern English was essentially a period of experiment and transition” (ibid. 302).

Section 6 provides some answers to Kortmann’s questions, illustrating the quite dramatic changes in the inventories based on Scots data.

## 6. Changes over time in the inventories of some adverbial subordinators in Scots

The data in the HCOS have been structured by the variable of time into the following four sub-periods: 1450–1500, 1500–1570, 1570–1640 and 1640–1700. Information in tables and graphics in this section has been arranged accordingly. In the first sub-period there are fewer texts (a third of those dating from the third period) as the range of prose genres was smaller at that time. The selection of topics concentrates on the chronology of changes, that is, the identification of periods when the most dramatic changes take place. I will examine developments in the general frequency of semantically more or less closely related subordinators that compete with each other as well as illustrate developments in the morphological complexity of adverbial subordinators and their preferred categorial sources.

### 6.1 Relative frequencies of semantically related subordinators

Can we identify varying degrees of dynamism in processes of change in the history of Scots? Are there periods of transition reflected in patterns of co-variation, variants being used interchangeably, or can specific types of intermediary variant be attested? I would like to illustrate some of the major changes in the relative frequencies of semantically-related linking devices in Scots by first examining the semantic fields of “before a point of time” and “up to a point of time”. All variant forms are included in the statistics. In inventories of the kind listed here syntactic, semantic and pragmatic differences between the items have been ignored.

We can see the subordinator *before* taking the status of a prevailing subordinator in the third sub-period, thus replacing *or* as an abrupt change in the percentage and the mean frequency takes place: *before* from 14 (0.03/1,000) to 72 per cent (0.19/1,000) and *or* from 70 (0.17/1,000) to 23 per cent (0.06/1,000). It is significant that the other two semantically closely related subordinators are shown to occur only in a few texts (*afore*) or quite late (*ere*), both remaining quite infrequent (0.04 and 0.01–0.04/1,000 respectively) (Table 1 and Figure 1).

As regards subordinators sharing the sense “up to a point of time”, the most significant change in relative frequencies takes place later than in the cases of *before* and *or* (Table 2).

A change in the relative frequencies of the two main subordinators takes

Table 1. Percentages (absolute numbers in brackets) and mean frequencies (M per 1,000) of *or*, *before*, *afore* and *ere* in the HCOS. TM = mean frequency of the total.

Period	<i>or</i>	M	<i>before</i>	M	<i>afore</i>	M	<i>ere</i>	M	Total	TM
1450–1500	94 (17)	0.18	6 (1)	0.01	(0)		(0)		(18)	0.19
1500–1570	70 (35)	0.17	14 (7)	0.03	16 (8)	0.04	(0)		(50)	0.25
1570–1640	23 (18)	0.06	72 (56)	0.19	(0)		5 (4)	0.01	(78)	0.26
1640–1700	5 (5)	0.02	84 (79)	0.32	(0)		11 (10)	0.04	(94)	0.38

Table 2. Percentages (absolute numbers in brackets) and mean frequencies (M per 1,000) of *quhill/while*, *till* and *until* in the HCOS. TM = mean frequency of the total.

Period	<i>quhill</i>	M	<i>till</i>	M	<i>until</i>	M	Total	TM
1450–1500	97 (38)	0.40	3 (1)	0.01	(0)		(39)	0.41
1500–1570	94 (103)	0.51	6 (7)	0.03	(0)		(110)	0.55
1570–1640	65 (107)	0.35	25 (41)	0.14	10 (16)	0.05	(164)	0.54
1640–1700	18 (28)	0.11	61 (92)	0.37	21 (32)	0.13	(152)	0.62

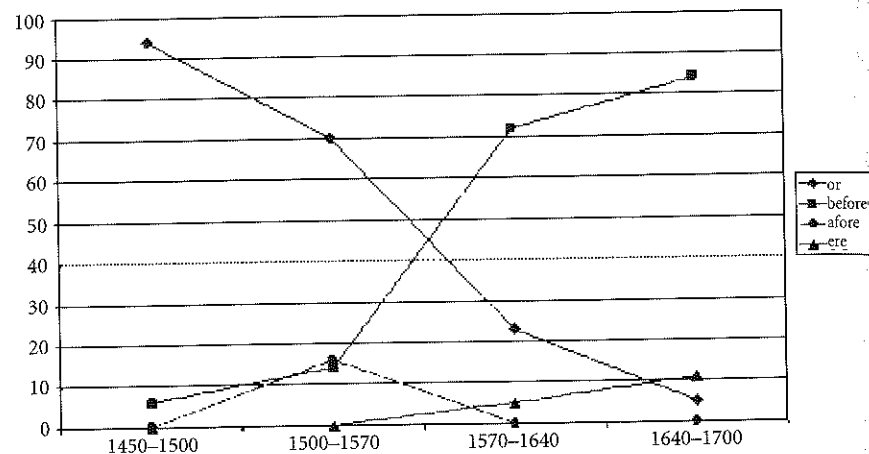


Figure 1. Percentages of *or*, *before*, *afore* and *ere* in the HCOS.

place in the latter half of the seventeenth century, the proportion of *till* increasing from 25 to 61 per cent and, as a contemporary development, that of *quhill/while* decreasing from 65 to 18 per cent. These two processes are accompanied by a significantly smaller increase in the frequency of *until*. Figure 2 shows that *until* is relatively late and only slightly more frequent than the receding variant *quhill*, its proportion increasing from 10 to 21 per cent in the latter half of the seventeenth century.

Statistical information about other sets of semantically related subordinators identifies the third sub-period as one of dramatic change.

An increase in the relative proportion of *unless* from 6 to 80 per cent in the third sub-period has been recorded, while *but if* is very rare in the fourth sub-period (Table 3 and Figure 3).

The shape of the curve in Figure 4 is somewhat different from the previous ones, the change towards the preference for *as soon as* being less abrupt (Table 4).

As in the case of *unless*, the proportion of the spreading variant, *as soon as*, is very high in the last sub-period, a development also reflected in the mean frequencies.

Inevitably, graphs of this kind do not give the full picture, but they do show that we are dealing with major changes in a relatively short time. In Figure 5, Figures 1-4 have been superimposed.

Figure 5 shows that the difference between the prevailing and quite infrequent variants in the 1450-1500 period is quite striking. The proportions do

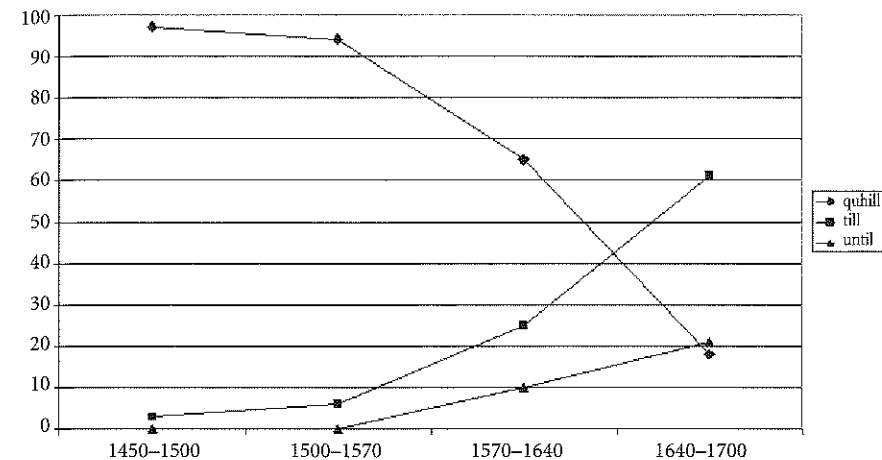


Figure 2. Percentages of *quhill/while*, *till* and *until*.

Table 3. Percentages (absolute numbers in brackets) and mean frequencies (M per 1,000) of *but if* and *unless* in the HCOS. TM = mean frequency of the total.

Period	<i>but if</i>	Mean	<i>unless</i>	Mean	Total	TM
1450-1500	100 (23)	0.24	(0)		(23)	0.24
1500-1570	94 (17)	0.08	6 (1)	0.00	(18)	0.09
1570-1640	20 (3)	0.01	80 (12)	0.04	(15)	0.05
1640-1700	3 (1)	0.00	97 (38)	0.15	(39)	0.16

Table 4. Percentages (absolute numbers in brackets) and mean frequencies (M per 1,000) of *how soon as* and *as soon as* in the HCOS. TM = mean frequency of the total.

Period	<i>how soon as</i>	Mean	<i>as soon as</i>	Mean	Total	Mean
1450-1500	(0)		(0)		(0)	
1500-1570	65 (17)	0.08	35 (9)	0.04	(26)	0.13
1570-1640	45 (9)	0.03	55 (11)	0.04	(20)	0.07
1640-1700	12 (5)	0.02	88 (35)	0.14	(40)	0.16

not change significantly in the 1500-1570 period, the decrease in *or* being somewhat greater because of the competing *before* and *afore*, which together represent 30 per cent. The previously less frequent item becomes the majority variant in the 1570-1640 period. Interestingly, *quhill* is a more resistant feature, *till* beginning to gain more ground in the post-1640 texts.

Inventories based on purely quantitative data can sometimes be misleading



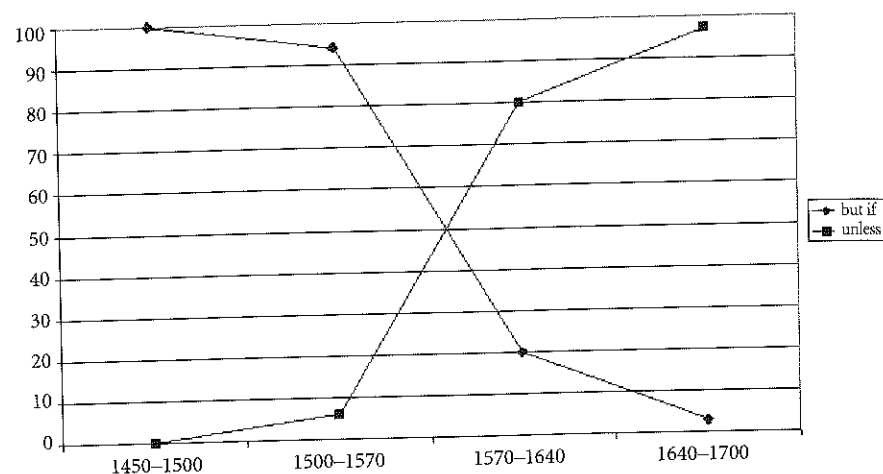


Figure 3. Percentages of *but if* and *unless* in the HCOS.

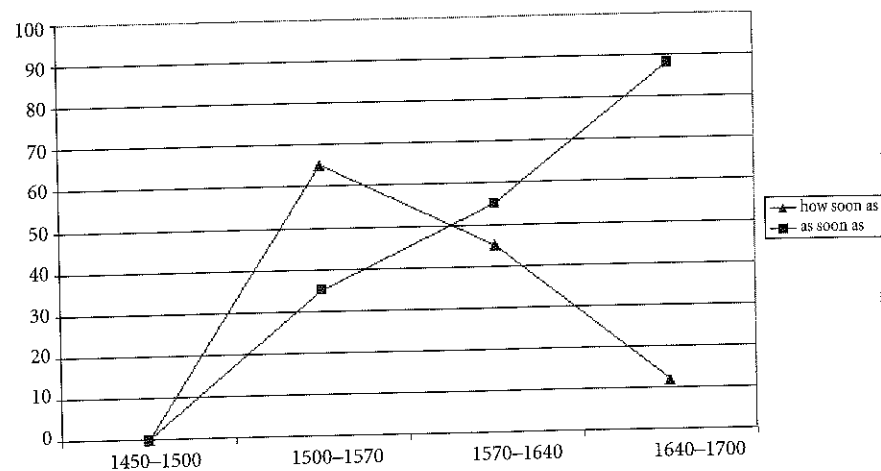


Figure 4. Percentages of *how soon as* and *as soon as* in the HCOS.

as examination of the subordinators *though* and *although* in Older Scots illustrates.

The peak in the use of *although* cannot be related to general trends in Older Scots.<sup>2</sup> The high proportion of this subordinator in the 1570-1640 period can be explained by the idiolectal preference of James VI, since almost 70 per cent

2. The subordinator *although* also remains much less frequent than *though* in Present-day written Scots (Häcker 1999: 140).

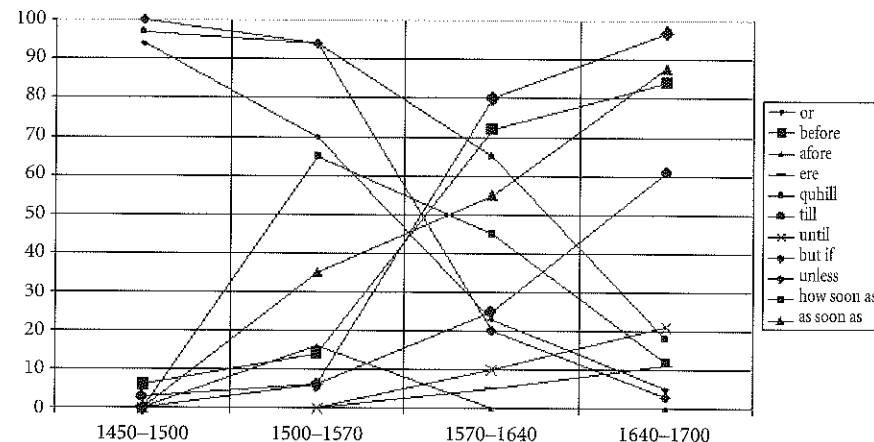


Figure 5. Change over time in the relative frequencies of some adverbial subordinators in the HCOS.

Table 5. Percentages (absolute numbers in brackets) and mean frequencies (M per 1,000) of *though* and *although* in the HCOS. TM = mean frequency of the total.

Period	<i>though</i>	Mean	<i>although</i>	Mean	Total	TM
1450-1500	100 (7)	0.07	(0)		(7)	0.07
1500-1570	100 (9)	0.04	(0)		(9)	0.04
1570-1640	39 (27)	0.09	61 (42)	0.14	(69)	0.23
1640-1700	93 (154)	0.62	7 (11)	0.04	(165)	0.67

of the occurrences of *although* in the period 1570-1640 have been attested in his writings (Table 5 and Figure 6).

The total mean frequencies in the tables show a significant difference between the various subordinators. Listed in the order of mean frequency in the period 1640-1700, *though* is by far the most frequent, *till* and *before* being the second and third most frequent subordinator among those analysed here (Table 6).

## 6.2 Changes in the morphological complexity of adverbial subordinators

Since a full account of changes in the morphological complexity of adverbial subordinators in Older Scots cannot be provided within the scope of the present study, the data are restricted to subordinators of concession. The number of types varies between nine in the first sub-period and thirteen in the third.

The prevailing concessive subordinator is *whatever* or *whatsoever* in the first sub-period (28 per cent), while *howbeit* takes this position in the second (43 per

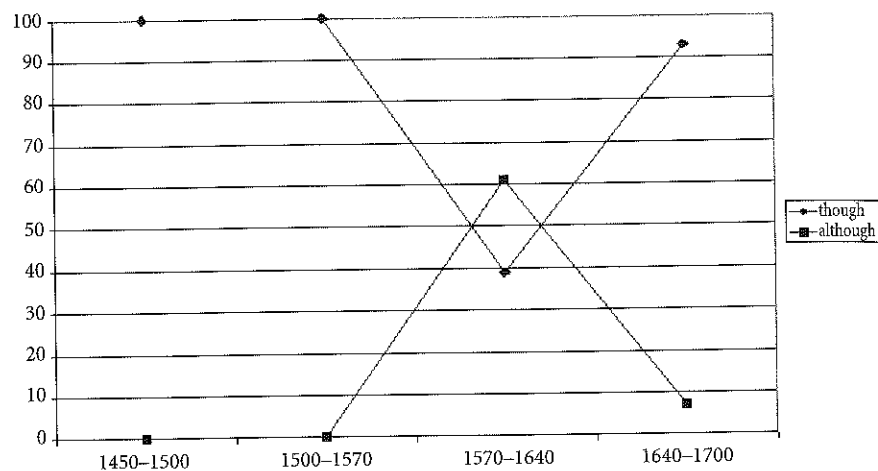
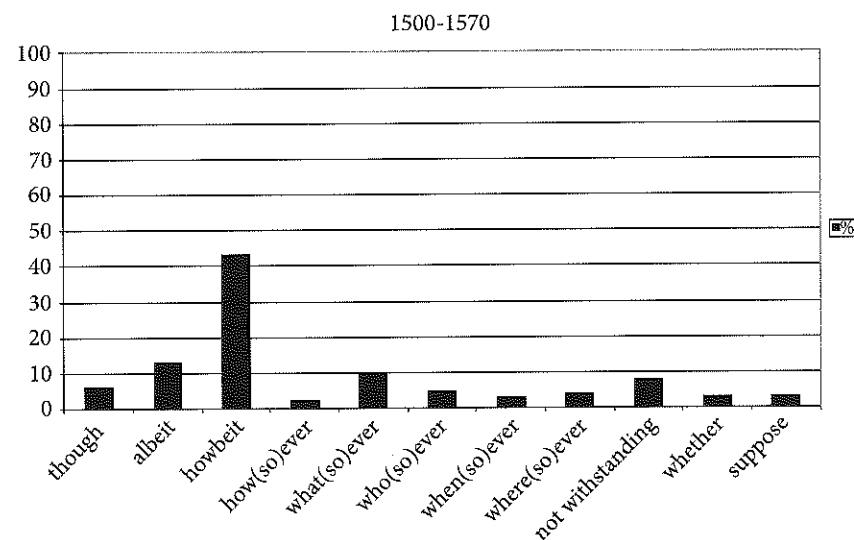
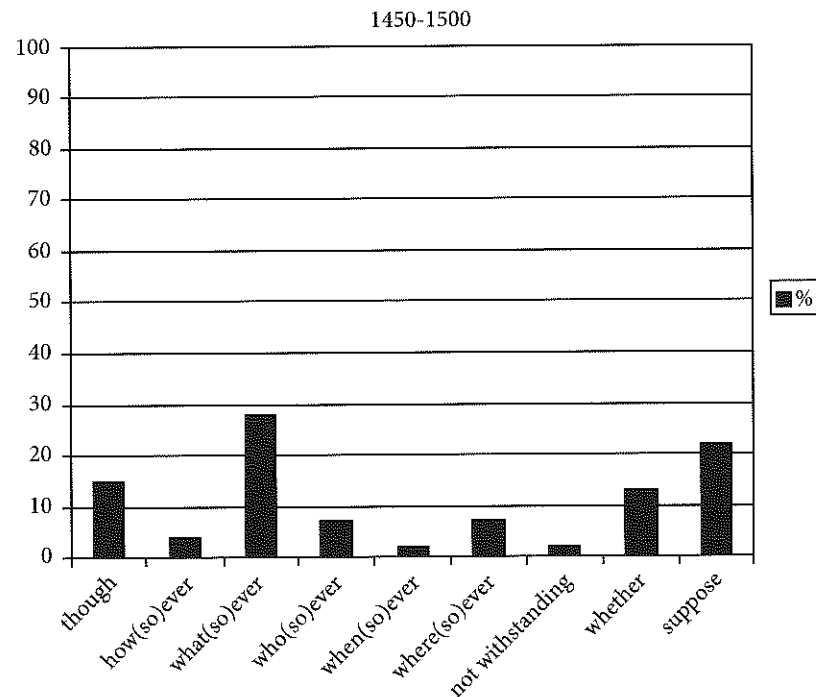
Figure 6. Percentages of *though* and *although* in the HCOS.

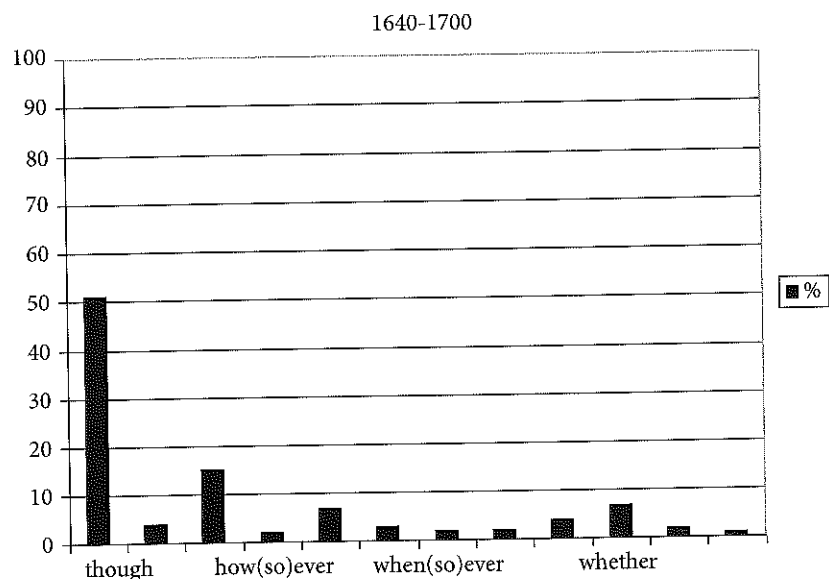
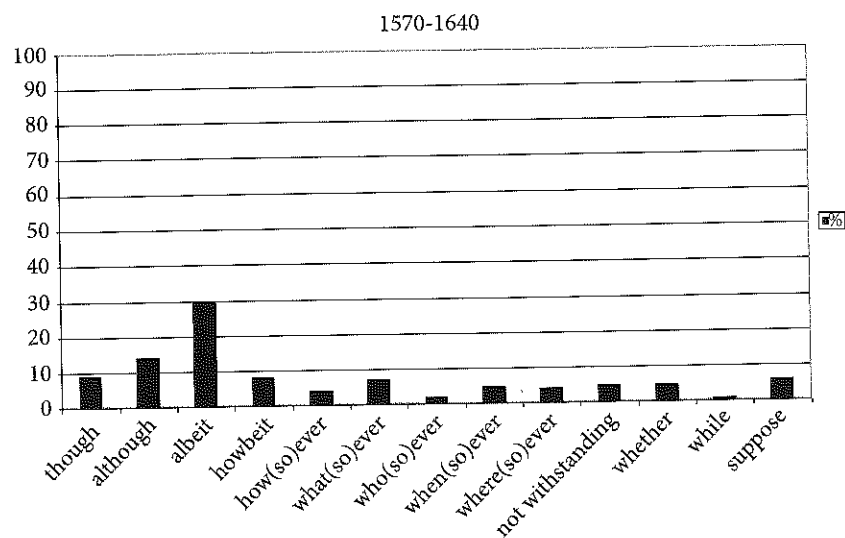
Table 6. Change over time (1450-1700) in the mean frequencies of some adverbial subordinators in Older Scots. M/10,000.

Subordinator	1450-1700
<i>though</i>	0.7-6.2
<i>till</i>	0.1-3.7
<i>before</i>	0.1-3.2
<i>as soon as</i>	0.0-1.4
<i>unless</i>	0.0-1.5
<i>quhill</i>	4.0-1.1
<i>until</i>	0.0-1.3
<i>although</i>	0.0-0.4
<i>or</i>	1.8-0.2
<i>how soon as</i>	0.0-0.2
<i>but if</i>	2.4-0.0

cent), *albeit* in the third (30 per cent), and *though* in the fourth (51 per cent). In addition to being used for condition, *suppose* also implies concession in Scots. We notice, however, that, after the first sub-period, the concessive *suppose* becomes one of the many relatively infrequent subordinating devices in this semantic group. I would also like to draw attention to the very late use of *while* as a subordinator of concession in Scots. There are only eight examples in the HCOS, all dating from the seventeenth century. Overall, the difference between high- and low-frequency items increases over time (Figures 7a-d).



When we look at the developments in the use of the *-ever* type (*what(so)ever*, *when(so)ever*, *how(so)ever*, etc.) and the syntactic phrase type (*albeit* and



Figures 7a-d. Percentages of subordinators of concession in the HCOS.

*howbeit*), we see an overall decrease in their use over time. I find it interesting that the syntactic phrase type should have as high a proportion as 56 per cent in the second sub-period, the percentage (38) also being quite high in the next sub-period (Table 7).

Table 7. Percentages of the *-ever* type and the syntactic phrase type for concession in the HCOS.

Period	<i>-ever</i> type	syntactic phrase
1450-1500	48	—
1500-1570	24	56
1570-1640	22	38
1640-1700	17	15

### 6.3 Changes in the preferred categorial sources of adverbial subordinators

To illustrate changes in the preferred categorial sources of adverbial subordinators in Scots, we might consider the following charts in which nouns and verbs have been compared as sources of subordinators. Figure 8a tells us about changes over time in the mean frequencies: firstly, of subordinators incorporating nouns such as *by reason (that)*, *in case (that)*, *to the end (that)* and, secondly, those that have been derived from verb forms, in Scots mostly present participles, *considering (that)*, for instance.

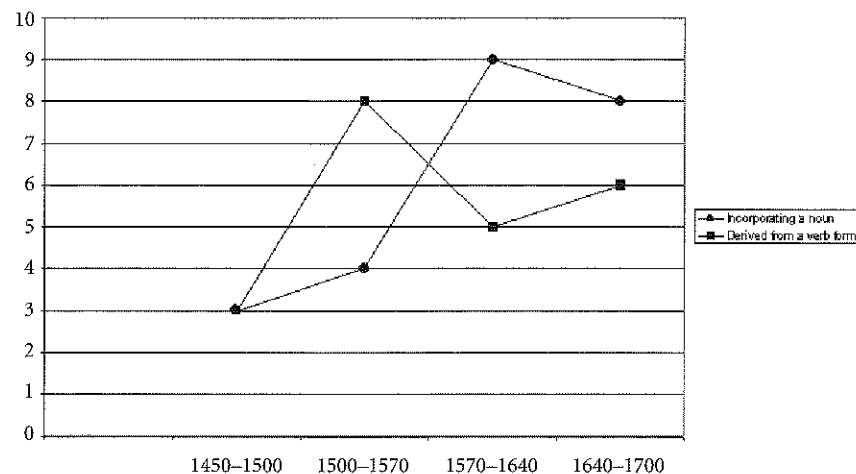


Figure 8a. Number of subordinator types incorporating nouns and derived from verbs in the HCOS.

We see a significant increase in both types, which is earlier in the group having verbs as a source. We also notice that the mean frequency of connective

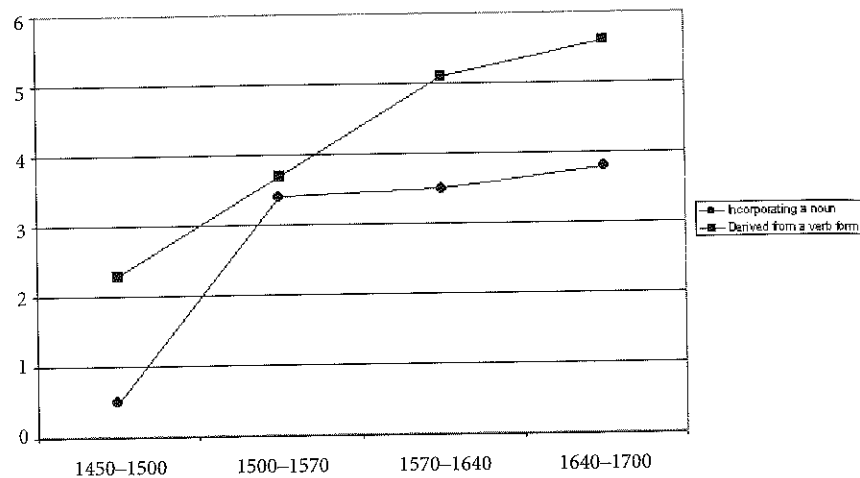


Figure 8b. Mean frequencies of subordinator types incorporating nouns and derived from verbs in the HCOS.

Table 8. Occurrences of some conjunctive phrases incorporating nouns in the HCOS. TM = mean frequency of the total.

Conjunctive phrase	1450-1500	1500-1570	1570-1640	1640-1700	Total	TM
<i>to the intent (that)</i>	1	—	1	—	2	0.00
<i>to the end (that)</i>	2	—	2	8	12	0.01
<i>to the effect (that)</i>	—	27	42	31	100	0.12
<i>for fear (that)</i>	—	—	5	—	5	0.01
<i>by reason (that)</i>	—	18	13	4	35	0.04
<i>in case (that)</i>	2	9	17	21	49	0.06
<i>upon condition (that)</i>	—	—	1	—	1	0.00
<i>in regard (that)</i>	—	—	3	8	11	0.01
<i>in respect (that)</i>	—	3	19	7	29	0.03
<i>on the consideration (that)</i>	—	—	—	1	1	0.00
Total	5	57	103	81	246	0.29
TM	0.05	0.28	0.34	0.33	0.29	

phrases incorporating nouns remains relatively stable after the increase in the first half of the sixteenth century, while the mean frequency of those derived from verbs increases considerably. Thus, assessed by number of types, the use of nouns as a source becomes particularly productive towards the end of the sixteenth century, whereas the number of verbs which have adopted the connective function reaches its peak in the 1500-1570 period. However, the

mean frequencies of connectives incorporating nouns increase considerably as early as this second sub-period, remaining relatively stable after that. Verbs as a source develop differently, as reflected in the continuously increasing mean frequencies, which are almost 2.5 times more frequent in the last sub-period than in pre-1500 texts (Figure 8b).

The data reflect an increase in conjunctive phrases incorporating nouns in the sixteenth and seventeenth centuries, their frequency being very low in pre-1500 texts. Moreover, the number of types is smaller in the second sub-period than the considerably enriched inventory in post-1570 texts. Some of the conjunctive phrases remain relatively rare (*to the end (that)* 0.02-0-0.02-0.03), while some show an increase over time (*in case (that)* 0.02-0.04-0.06-0.09), but there are also phrases that decrease in frequency (*by reason (that)* 0-0.09-0.04-0.02) after a peak in the first half of the sixteenth century. It is not possible for me to go into details here, but I would like to stress that the spread of items of this type does not suggest that their use is conditioned by genre, so that discursal preferences may be related to a formal communicative setting. In the Corpus of Scottish Correspondence *in respect (that)* is used in the following way:

- (2) I think he was asshamed to enter with you in respect \ he did not creue your opunione nor your \ housbands in his vp coming (CSC 1616 Lady Anna Livingston, Countess of Eglinton, NLS Adv. MS 33.1.1., vol. 10:120)
- (3) I doutt not bot our brother \ Jhon cann gei? {torn} zour lo gritter assurance thairof, in respect he hes seinne, whatt hes \ beine my estaitte thir dayes bygaine (CSC 1613 George, 3rd Earl of Winton, NAS GD 3/5/50)
- (4) I am sorie \ sister that ze sould geif me thanks whar ther is \ non deseruid be me Inrespek that I deid nothing \ bot that vhilck vas my bound deutie to do (CSC 1617 Lady Isabella Seton, Countess of Perth, NAS GD 3/5/133)
- (5) pleis zour \ la that this zung gentilman, my sister the Lady carnebeis \ sone being desyrus to be placit in service v<sup>t</sup> sum nobilman, and \ hes villit me to mak my moyen for the saming, in respect his \ father hes sumquhat outschot him self in misgoverning his \ rent, and I accounting zour la and {torn} husband \ as the most speciall freinds that I heaue to imploy vill most \ affectuiesly requeyst zour la that ze vill place this zung man \ ather in my lord zour husbands chamber to put on or af his \ cloths {ins} and vait on his lo {ins} or in qt vther office zour la thinks meit for such a \ ane, and I vilbe ansuerable to zour la that he sal {damaged} be faith\ful and obediand to my lord or zour la in qtsoeuer ze imploy \ him (CSC 1612 Dame Grissall Roos, Lady Keir, NAS GD 3/5/32)

Table 9 provides a more detailed account of verbs as a source of adverbial subordinators and information about the chronology of their spread in Older Scots.

Table 9. Occurrences of some conjunctive phrases using verbs as source in the HCOS. TM = mean frequency of the total.

Conjunctive phrase	1450–1500	1500–1570	1570–1640	1640–1700	Total	TM
<i>notwithstanding (that)</i>	1	13	14	12	40	0.05
<i>suppose (that)</i>	18	9	18	5	50	0.06
<i>conditioned (that)</i>	–	1	–	–	1	0.00
<i>giving (that)</i>	–	3	–	–	3	0.00
<i>given (that)</i>	–	1	–	–	1	0.00
<i>providing (that)</i>	–	31	22	24	77	0.09
<i>provided (that)</i>	–	–	–	5	5	0.01
<i>seeing (that)</i>	–	3	80	50	133	0.16
<i>considering (that)</i>	3	14	20	47	84	0.10
Total	22	75	154	143	394	0.47
TM	0.23	0.37	0.51	0.58	0.47	

While the general frequency of subordinators derived from verbs is greater than that of incorporating nouns, we can see a similar chronology in their spread. This is even more obvious if we exclude the non-participial *suppose* (conjunctive participles: 0.04–0.33–0.45–0.56). The table also shows that past participles are very rare. There are no occurrences of *granted (that)*, *outtaken (that)*, *during (that)*, *excepting (that)* or *supposing (that)*, listed in Kortmann's inventory of Early Modern English adverbial subordinators (1997:293). The high-frequency items *providing* and *considering* rarely occur in letters, *seeing (that)* being used instead. In fact, in Scots *seeing (that)* was introduced into mid-sixteenth-century official letters and has not been attested in other texts in this period in the HCOS. In the 1570–1640 period its mean frequency is 1.2/1,000 in letters, that of all texts representing this period being as low as 0.3 (for further information, see Meurman-Solin and Pahta forthcoming).

## 7. Concluding remarks

I would like to conclude by relating this brief summary to the three general points I made at the beginning. I have only briefly referred to problems related to categories in the analysis of connectives in complex sentence structures,

pointing out that a full-scale manuscript-based electronic corpus is required to deal with such problems. At the moment, my corpus based on diplomatically transcribed manuscripts is too small to allow such an endeavour. When the tagged databases being compiled are available we will have better ways of analysing developments over time in the system of clause linkage. The introduction of modern punctuation into text editions has of course fatally distorted evidence in edition-based corpora, however representative these may be by other criteria. It is also necessary to examine texts in manuscript to understand how each writer structures the text, whether the clause boundaries have been left unmarked or devices such as capitals or spaces have been used to mark them (for illustrations see Meurman-Solin 2004). The complexities of clause linkage systems may be difficult to deal with appropriately without annotating the manuscript-based texts with tags providing information about the lexicogrammatical properties of linguistic items. We may not succeed in identifying all the items entering patterns of syntactic variation in any other way. The reanalysis of category membership may necessitate redefining conjunction plus adverb collocates, and positioning them in our subordinator inventories in a semantically appropriate way. Polyfunctionality will have to be paid due attention. I have referred to Kortmann's typology of adverbial subordinators in European languages to arouse our interest in questions that remain unanswered, not to suggest that I find the comparative perspective timely. We cannot compare before we have managed to describe what we want to compare fully. The relevance of a typology depends on the quality of inventories available for creating it.

## References

- Anderson, John M. 1997. *A Notional Theory of Syntactic Categories*. Cambridge: Cambridge University Press.
- Auwera, Johan van der, ed. 1998. *Adverbial Constructions in the Languages of Europe*. Berlin & New York: Mouton de Gruyter.
- Benson, Phil. 2001. *Ethnocentrism and the English Dictionary*. London & New York: Routledge.
- Brinton, Laurel J. 2000. "The Importance of Discourse Types in Grammaticalization: The Case of Anon". *Textual Parameters in Older Languages* ed. by Susan C. Herring, Pieter van Reenen & Lene Schøsler, 139–162. Amsterdam & Philadelphia: John Benjamins.
- Croft, William. 1990. *Typology and universals*. Cambridge: Cambridge University Press.
- Crowley, T. 1990. "That obscure object of desire: a science of language". *Ideologies of Language* ed. by J. E. Joseph & T. J. Taylor, 27–50. London: Routledge.

- Dareau, Marace. 2004. "DOST: a significant instance of historical lexicography". *New Perspectives on English Historical Linguistics (2): Lexis and Transmission. Selected Papers from 12ICEHL (Glasgow, 21–26 August, 2002)* ed. by Christian J. Kay, Carole A. Hough & Iréné A. W. Wotherspoon, 49–64. Amsterdam & Philadelphia: John Benjamins.
- Devriendt, Betty, Louis Goossens & Johan van der Auwera, eds. 1996. *Complex Structures. A Functionalist Perspective*. Berlin & New York: Mouton de Gruyter.
- DOST = *A Dictionary of the Older Scottish Tongue*, ed. by William A. Craigie, A. J. Aitken et al. Aberdeen: Aberdeen University Press & Oxford: Oxford University Press.
- Foley, William A. & Robert D. Van Valin Jr. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Haiman, John & Sandra A. Thomson, eds. 1988. *Clause combining in grammar and discourse*. Amsterdam & Philadelphia: John Benjamins.
- Haumann, Dagmar. 1997. *The Syntax of Subordination*. Tübingen: Max Niemeyer.
- Häcker, Martina. 1999. *Adverbial Clauses in Scots. A Semantic-Syntactic Study*. Berlin & New York: Mouton de Gruyter.
- Jackendoff, Ray. 2002. *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Jones, Charles, ed. 1997. *The Edinburgh History of the Scots Language*. Edinburgh: Edinburgh University Press.
- Kortmann, Bernd. 1997. *Adverbial Subordination: A Typology and History of Adverbial Subordinators Based on European Languages*. Berlin & New York: Mouton de Gruyter.
- Laing, Margaret. 1993. *Catalogue of Sources for a Linguistic Atlas of Early Medieval English*. Cambridge: D. S. Brewer.
- . 2000. "Never the twain shall meet. Early Middle English — the East-West divide". Taavitsainen et al. 2000. 97–124.
- & Keith Williamson. 2004. "The Archaeology of Medieval Texts". *Categorization in the History of English* ed. by Christian Kay & Jeremy Smith. Amsterdam: John Benjamins.
- Lehmann, Christian. 1988. "Towards a typology of clause linkage". Haiman & Thomson 1988. 181–225.
- McIntosh, A., M. L. Samuels & Michael Benskin. 1986. *A Linguistic Atlas of Late Mediaeval English*. Aberdeen: Aberdeen University Press/Edinburgh: Mercat Press.
- Matthiessen, Christian & Sandra A. Thompson. 1988. "The structure of discourse and 'subordination'". Haiman & Thomson 1988. 275–329.
- Meurman-Solin, Anneli. 1995a. *The Helsinki Corpus of Older Scots (HCOS)*. Distributors: Oxford Text Archive, Norwegian Computing Centre for the Humanities. <http://www.hd.uib.no/corpora.html>
- . 2001a. "Structured Text Corpora in the Study of Language Variation and Change". *Literary and Linguistic Computing* 16:1.5–27.
- . 2001b. "Women as informants in the reconstruction of geographically and socio-culturally conditioned language variation and change in 16th and 17th century Scots". *Scottish Language* 20.20–46.
- . 2002. "Simple and complex grammars: The Case of Temporal Subordinators in the History of Scots". *Variation Past and Present. VARIENG Studies on English for Terttu Nevalainen* (= Mémoires de la Société Néophilologique de Helsinki, 61) ed. by Helena Raumolin-Brunberg, Minna Nevala, Arja Nurmi & Matti Rissanen, 187–210. Helsinki: Société Néophilologique.
- . 2004. "Towards a Variationist Typology of Clausal Connectives. Methodological Considerations Based on the Corpus of Scottish Correspondence". *Methods and Data in English Historical Dialectology* ed. by Marina Dossena & Roger Lass. Bern: Peter Lang.
- . forthcoming a. *Corpus of Scottish Correspondence (CSC)*, 1500–1800.
- . forthcoming b. *Corpus of Early Scottish Women's Writings (CESWW)*.
- & Päivi Pahta. 2004. "Circumstantial adverbials in discourse: A synchronic and a diachronic perspective". *Language and Computers: Proceedings of the ICAME conference, Guernsey 2003* ed. by Antoinette Renouf. Amsterdam: Rodopi.
- Milroy, James. 1999. "The consequences of standardization in descriptive linguistics". *Standard English. The Widening Debate* ed. by Tony Bex & Richard J. Watts, 16–39. London: Routledge.
- Moessner, Lilo. 1997. "The Syntax of Older Scots". Jones 1997. 112–155.
- Rissanen, Matti. 1999. "Syntax". *The Cambridge History of the English Language*, Vol. 3, *Early Modern English 1476–1776* ed. by Roger Lass, 187–331. Cambridge: Cambridge University Press.
- Taavitsainen, Irma, Terttu Nevalainen, Päivi Pahta & Matti Rissanen, eds. 2000. *Placing Middle English in Context*. Berlin & New York: Mouton de Gruyter.
- Thompson, Sandra A. & Robert E. Longacre. 1985. "Adverbial Clauses". *Language and typology and syntactic description*, vol. II: *Complex constructions* ed. by Timothy Shopen, 171–234. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs & Bernd Heine, eds. 1991. *Approaches to Grammaticalization*, vol. 1: *Focus on theoretical and methodological issues*. Amsterdam & Philadelphia: John Benjamins.
- Williamson, Keith. 1992/93. "A Computer-aided Method for Making a Linguistic Atlas of Older Scots". *Scottish Language* 11/12.138–73.
- . 2000. "Changing spaces: Linguistic relationships and the dialect continuum". Taavitsainen et al. 2000. 141–179.
- . 2001. "Spatio-Temporal Aspects of Older Scots Texts". *Scottish Language* 20.1–19.