Defining relatives

6 Mar 2013 Final authors' revision before copy-editing (to include renumbering of notes from 13 onwards).

To be published in *Journal of English Linguistics* 41.2

David Denison

University of Manchester

Marianne Hundt

University of Zurich

Abstract

This paper uses a case study of scientific English from the seventeenth to the nineteenth centuries to test the traditional demarcation between restrictive and non-restrictive adnominal relatives and to reconsider appropriate methods for analyzing such historical data.

After an initial classification of some 1160 clauses from the ARCHER corpus using the traditional dichotomy, alternative proposals for analyzing relative clauses are reviewed in the light of problematic examples. The concept of "aspective" relatives (Wood 1952; Sigley 1997) is adopted: those having most of the formal and pragmatic hallmarks of restrictive relatives while not strictly restrictive in the set-theoretic sense. We identify these with the clauses in present-day English that are not restrictive but which Huddleston, Pullum, and Peterson (2002) include under the heading "integrated." We add a fourth, minor type, "continuative." Nevertheless our data present problems even for a four-way classification. We demonstrate how some analytic difficulties are due to changes in text-type-specific style and conventions over time, plus general diachronic change, but also that some examples genuinely resist hard-and-fast classification. We therefore treat our classes as overlapping bands on a one-dimensional gradient, testing the revised classification in detail on our seventeenth-century data. The overlaps are less numerous than any of the three main types (restrictive, aspective, and non-restrictive), but they help to reduce the number of unclear cases from over a quarter of the total in our initial classification to a mere one percent. We

distinguish carefully between vagueness (underspecified examples where the interpretation is in no doubt) and ambiguity (involving a choice between interpretations that depends on missing information). We suggest that our classification might be more widely applicable to the study of relative clauses. By and large, modern *Sprachgefühl* can be used cautiously alongside other, more objective tests to classify relatives in historical scientific texts--a point of more general theoretical relevance.

Keywords

English grammar, stylistic variation, written language, American English, British English, corpus linguistics, historical linguistics, history of English, relative clause, scientific writing

Introduction

Background to Study

All modern studies of adnominal relative clauses (that is, those with a nominal antecedent) base their analysis on a distinction between two main types of relative clause.[1] The terminology and sometimes the precise definition varies--restrictive vs. non-restrictive, defining vs. non-defining, integrated vs. supplementary, tense vs. lax and so on--but the distinction is taken to be crucial for English both by prescriptive and descriptive grammarians. Sigley (1997:129-130) attempted to construct a synoptic diagram to represent the various gradients and polarities proposed in the literature to date, and Huddleston, Pullum, et al. have defended their own distinction at some length (2002:1034-1035, 1058-1066, 1352-1353). Most of these discussions concern present-day English (PDE). In a recent study (Hundt, Denison & Schneider 2012b), we investigate relativization strategies in British and American scientific English in the ARCHER corpus; our focus there is on the effects of prescriptive ideas, of national variety and of register, and on the contribution of relative clauses to noun phrase complexity. The discussion necessarily has to be conducted in the light of the basic dichotomy, particularly given our inclusion in that paper of prescriptive ideas.

However, there are cases in PDE where the distinction is hazy, as is sometimes acknowledged in the literature, and it turned out to be even more difficult to classify clauses in our historical scientific data.

In this paper we intend to revisit that basic distinction. How best are the terms "restrictive" and "non-restrictive" defined? Is the distinction workable, particularly in historical data?[2] And if so, is it a binary choice? All of this raises another crucial methodological problem which is of very wide significance in diachronic linguistics: when, if ever, can present-day translations and intuitions serve as proxies for making analytic decisions about older states of the language? The particular difficulties thrown up by early scientific writing give us the opportunity to develop the classification of relatives and assess its limitations--goals which are necessary for our corpus study but potentially of wider relevance in English grammar and historical linguistics.

The research questions we have are as follows:

- Does the distinction between restrictive and non-restrictive relatives become more problematic as you move further back in time? Our hypothesis is that it does.
- Is it problematic to use the *Sprachgefühl* ('intuition') of speakers of PDE and the criteria appropriate to PDE to classify earlier examples? Our hypotheses are that there are indeed real problems in such procedures, and furthermore, that they become more severe with older texts and in specialized text types.
- What is the most appropriate classification of adnominal relative clauses?

We briefly describe our corpus and discuss some initial exclusions from our dataset. We sketch the familiar restrictive/non-restrictive dichotomy, and then we introduce some alternative classifications and illustrate them with examples from our historical corpus data. A

substantial section details our method of analysis, including the provisional classification we initially adopted. It also looks at groups of examples which proved hard to classify, in the light of which we attempt a modification of the analytic classification. We close by drawing some conclusions from the investigation both for the study of relative clauses and for historical corpus linguistics in general.

Our Corpus and Data Selection

Our database of relative clauses is taken from the British part of the ARCHER 3.2 corpus[3] and is confined to the single genre of science, in which there are 105,410 words in 50 files ranging in date from 1674 to 1899. Science--which in ARCHER covers the natural sciences excluding medicine--is a genre where relative clauses are frequent and sometimes occur in quite complex structures (see Hundt, Denison & Schneider 2012b); early science writing can be particularly challenging to analyze. An approach that manages to deal successfully with such complexities is therefore likely to be of more general utility.

The corpus was annotated with a parser (Pro3Gres) developed by Schneider (2008),[4] by means of which a set of *that*-, *which*- and *who*- relatives was identified and given a preliminary coding in the database (file, sentence number, sentence, antecedent, relativizer). In a previous paper (Hundt, Denison & Schneider 2012a) we describe the success of the parser in identifying adnominal relatives in a similar corpus that also included American data. The parser was adapted after an initial run, and after parser adaptation, the recall for *that*-relatives and *wh*-relatives was around 40 percent overall, though precision was good at 82-86 percent. The low recall figure is largely attributable to missed *whic*h-relatives (Hundt, Denison & Schneider 2012a:8-9), which we subsequently retrieved manually.

The parser treated semicolons as a sentence boundary, yet relative clauses in early texts are occasionally punctuated off by a semicolon--relevant cases will be discussed under the heading “Phonology and Punctuation” below. As a precaution, a string search for <; *that*> was run. There were no relevant examples among the hits. The only possible contenders

involved the *that is* pattern, which we treat as a lexicalized idiom comparable to Latin *videlicet/viz.* or *id est/i.e.* and not as a relative clause:

(1) ...and it will be seen immediately to indicate a polarization of the latter in the directions c d; that is, in a direction perpendicular to the axis of motion... (1825barl.s5b)

Huddleston, Payne, and Peterson include *that is* among a list of what they call "indicators" (2002:1354-1355), and similarly Quirk et al. (1985:635, 1307). However, we did count as an adnominal relative the pragmatically somewhat similar:

(2) I Do herewith send you an $_1$account, $_1$Ø I lately received from New Providence, one of the Bahama Islands, concerning Fish there, $_1$which is as followeth... (1675ai--.s2b)[5]

Intuitively it seemed more likely that a relative clause involving *which* might follow a semicolon, but that eventuality had already been provided for, as we had supplemented the collection by concordance searches for the strings *which, whose,* and *whom,* most of which yielded adnominal relative clauses. Other than the special case noted above, we chose not to search manually for *that*-relatives in the untagged text files--and not merely for the obvious reason that examining all instances of the word *that* is laborious. It turned out that the gain would have been too small to justify the time needed. As we discovered by sampling files from the eighteenth, nineteenth, and twentieth centuries, *that*-relatives are considerably less frequent in the science genre than *wh*-relatives, with just four *that*-relatives missed as against 46 *which/who* in our eighteenth-century sample. In the combined parser/concordance database, the many duplicates for *which*-relatives were then removed. Our collection of

adnominal relatives in the corpus is thus somewhat incomplete for *that*-relatives but should be complete for *wh*-relatives.[6]

In addition to limitations on the relativizers covered, our study is confined to relative clauses which have a nominal antecedent--"adnominal" relative clauses. Both the automatic retrieval and our string-based searches yielded non-adnominal ("sentential") relatives and other false positives that we had to eliminate manually. A clear-cut example of a sentential relative is given under (3):

(3) ...it is now shooting Suckers out of the Root, which proveth that the Branches are as useful to support the Roots, as the Roots the Branches... (1724fair.s3b)

Real data tend to be messy, so it comes as no surprise that some relative clauses defy easy classification, hovering uncertainly between sentential and adnominal:

(4) ...and as it was necessary, for the animal's respiration, that the mouth of the vessel should communicate with the open air, it was made pretty deep, that the cold of the atmosphere round the animal might not be diminished fast by the $_1$warmth of the open air, $_1$which would have spoiled it as a conductor. (1775hunt.s4b)

The relative clause in (4) could either be adnominal (post-modifying *warmth*) or it could refer back to the diminishment of the cold, which would make it sentential. We retained some of these problematic examples in our database but excluded them from the statistics. However, we decided not to exclude (5), having considered but rejected an analysis in which the second *which*-clause is attached loosely as a sentential relative:

(5) That this air is of that exalted nature, I first found by means of nitrous $_1$air, which I constantly apply as a test of the fitness of any kind of air for respiration, and $_1$which I believe to be a most accurate and infallible test for that purpose. (1775prie.s4b)

The following example (6) is of the opposite kind, i.e., it is not straightforwardly adnominal, but has an adnominal feel to it:

(6) ...so that, far from its being a fault, as some ignorant druggists at Rome and Venice believe, it is a mark that the myrrh is fresh gathered, which is the best quality that myrrh of the first sort can have. (1775bruc.s4b)

However, the relative clause headed by *which* does not post-modify *mark* but rather *fresh gathered* or the predicate containing that phrase. It was therefore excluded.

We also removed the examples of relative determiner *which* like (7), whose relative phrase *which opinion* has no nominal antecedent and must therefore be sentential:

(7) ...but those of Jezzo say, that there runs an arm of Sea betwixt them and Tartary: Which opinion may seem to receive some confirmation from… (1675anon.s2b)

Other examples were, however, retained as adnominal, whether the antecedent was lexically identical to the head of the relativized *which*-phrase, (8), contained the same lexical item premodified, (9), or was lexically different, (10):

(8) Every part of the body is to a greater or less degree covered by a kind of down, which seems to be the efficient cause of its $_1$capability of repelling moisture; $_1$which capability is so remarkable, that… (1825kidd.s5b)

(9) ...whereby the magnetic $_1$forces are changed, as you have suggested, from their original direction, parallel to the magnetic axis of the ball into a position oblique to it, $_1$which oblique forces being resolved into two… (1825barl.s5b)

(10) It remain'd in this State about $_1$2 Minutes, during $_1$which time, we saw several Colours… (1720cote.s3b)

Huddleston, Pullum, and Peterson (2002:1043-1044) use the label "upward percolation type VII" and describe the type without preposition before *which*-- our (8) and probably also (9)-- as "quite rare and formal, verging on the archaic" in PDE. A similar characterization is provided in Denison (1998:277).

There are some *that*-clauses that follow a nominal head which is derived from a verb. In these contexts, *that* is a non-relative complementizer rather than a relativizer and we therefore excluded them from our dataset too.[7] The following examples (11 and 12) illustrate this kind of sentence:

(11) *belief that ~ believe that*

First, the all-absorbing interest centred in the bird-remains; and, secondly, the belief that the bones were those of a still-existing gigantic species of Tortoise commonly called Tesdudo indica. (1874gunt.s6b)

(12) *discovery that ~ discover that*

...he in some degree anticipated the discovery…, that in several Mammalia the tooth-germs never pass through any papillary stage... (1874tome.s6b)

Typologies of Relative Clauses

Restrictive and Non-Restrictive Relative Clauses

The dichotomy between restrictive and non-restrictive is the basis of much discussion of relative clauses, both popular and evidence-based. A restrictive relative clause is one which serves to delimit the reference of the antecedent, to restrict it.[8] As a number of writers have pointed out, however, although a restrictive relative clause may be named from this logico-semantic function, the clause type has clear syntactic and phonological correlates which are in many ways more central, such as that a restrictive relative clause forms a constituent with its antecedent, and that it belongs in the same intonation contour as the matrix clause.[9] The phonological property is in turn associated with the orthographic convention in PDE of its not being marked off by commas.

We can illustrate the familiar difference between restrictive and non-restrictive relative clauses with a "minimal pair" of examples from our eighteenth-century data, restrictive (13) vs. non-restrictive (14):

(13) ...whereas the common water, when exposed in a state of tranquillity to $_1$air $_1$<u>that is a few degrees colder than the freezing point</u> may easily be cooled to the degree of such air, and still continue perfectly fluid (1775blac.s4b)

(14) One cause of this variety was plainly a variation of the temperature of the $_1$air, $_1$<u>which became colder in the afternoon</u>, and made the thermometer descend gradually to 25°. (1775blac.s4b)

The first and obvious points to make are that *which* is not, and never has been, confined to non-restrictive use, despite the factitious proscription of restrictive *wh*-relatives which is common in (mainly) American style guides, while *that* is sometimes found in non-restrictive relatives (as in 15 and 16):

(15) One effect of boiling water long, is to expell the $_1$air $_1$which it naturally contains... (1775blac.s4b)

(16) ...and the Sky was ting 'd with a very unusual yellowish Colour, which perhaps might be reflected from a great Quantity of $_1$Snow, $_1$that soon after fell for near a quarter of an Hour... (1721lang.s3b)

Thus (15) is restrictive with relativizer *which*, and (16) is probably non-restrictive and has relativizer *that*. Choice of relativizer can be discounted as a way of discriminating between relative clause types, therefore, especially with restrictive relatives. The other differences summarized in Table 1 seem to be more robust and will come up again in alternative classifications. We will address them in turn, under the headings "logico-semantic function," "phonology and punctuation," even though these are obviously connected properties to some extent. First we briefly introduce some labels for types other than restrictive and non-restrictive which have been offered in the literature.

[TABLE 1 HERE]

As discussed by Lehmann (1984), Geisler and Johansson (2002), and Huddleston, Pullum, and Peterson (2002) among others, there are clauses which bear the distinctive formal signs of being restrictive relatives without being semantically restrictive (see Huddleston, Pullum & Peterson 2002:1064-1065). Conversely, non-restrictive relative clauses, usually regarded as supplying optional additional information, are sometimes in effect semantically obligatory (Geisler & Johansson 2002:96, citing Rydén 1984; see also Sigley 1997:123). The distinction is therefore a problematic one.[10] As Olofsson puts it:

> In practically every account of relative constructions there is a statement to the effect that the binary classification of relative clauses is not a hard and fast

> one, in that some clauses are neither clearly restrictive nor clearly non-restrictive. (Olofsson 1981:46; cf. also 22-23, 27)

One solution, following Lehmann (1984) or Jacobsson (1994), is to regard the distinction as gradient and to reclassify the dichotomy on the basis of the nature of the reference of the antecedent: generic vs. non-generic, and within the non-generic set, non-specific vs. specific vs. unique, etc. We have not followed Lehmann (1984), for reasons noted where we discuss appropriate models below.

Aspective Relative Clauses

Taking up an early observation by Wood (1952:13), Sigley (1997:127) introduces the label "aspective" clause for essential non-restrictive clauses that bear a formal similarity to restrictive clauses in that they (a) occur in the same intonation contour as the matrix clause, and (b) allow for relativizers other than *wh*-pronouns. Sigley's aspective clauses are thus not set-delimiting but add information to the noun in the matrix clause that is essential to the discourse. They are so called because "the relative clause captures an aspect of the antecedent that is extremely relevant to the content of the matrix clause" on logico-semantic grounds (1997:127).

If one did not start from the label "restrictive" and its etymological meaning, aspectives would merely be a subtype of restrictive, since they share most properties with them. Somewhat confusingly, Sigley argues for this (1997:124 and especially 128), *contra* Olofsson (1981) and Wood (1952). However, to strengthen the justification for treating them as a special case, he confines his examples of aspectives to unique antecedents, which safely rules out the possibility of a restrictive clause, *sensu strictu*:[11]

(17) She thanked my $_1$father, $_1$who had saved her life. (example cited by Sigley 1997:125, from Rydén 1974:542)

Since he is dealing with modern data he can test the facts of intonation, substitution, and meaning, and his discussion of the data is persuasive.

In our data we think the same considerations apply, but it is harder to find convincing examples. Consider (18) and (19):

(18) There was then much discourse of the $_1$Gulf of Anian, by $_1$which passage was said to be open into the Tartarian Sea… (1675anon.s2b)

(19) If any person brings an ague to Richmond, he is generally freed from it in a few days; though the $_1$village of Gilling, about a mile and a half distant, $_1$which stands low, and has a large pool of stagnant water adjoining to it, is visited with this complaint every spring and autumn. (1775perc.s4b)

The relative clauses in (18) and (19) are clearly non-restrictive because the antecedents are proper names, but the information given in the relative clauses is central to the discourse and therefore essential. In both instances, a reading with tight juncture (same intonation contour) seems possible, even though there is a phrase in apposition to the antecedent before the relative clause in (19). Juncture is discussed in the section on phonology below. We have coded (18) as aspective but left (19) as non-restrictive, mainly based on our judgment of the juncture.

The antecedent in example (20) is also clearly identified by the non-finite post-modifier “marked p,” but tight juncture is much more likely than loose juncture and the information in the relative clause is essential to the discourse. Example (20) therefore also qualifies as an aspective relative clause:

(20) In figs. 11 & 13 I would draw attention to the $_1$corpuscles marked p, $_1$which appear to be in the act of detaching themselves from the hypoblast, whilst the corpuscle (pd) has the appearance of a hypoblastic cell undergoing quadruple division. (1874lank.s6b)

The following examples are also aspective, but the wider context is needed to discover this. In (21) the author has been describing experiments involving the "golden Ball" that is the antecedent, so it is clearly identified from the previous discourse, and the relative clause cannot therefore be restrictive; however, the information provided in the relative clause is essential to the discourse.

(21) As to the golden $_1$Ball $_1$which had Varnish and Cement upon it to keep the Mercury from sinking into it, I found it to weigh as follows... (1721desa.s3b)

It is quite possible that aspective relative clauses are particularly frequent in scientific texts because the function of relative clauses in this text type is often not of the set-delimiting type but to add essential information, as in (22) and (23):

(22) The two cylinders are connected by small $_1$pieces of thermometer-tubes $_1$which keep them steady with their faces parallel to each other, but turned in opposite directions, and also serve to make the insulation as complete as possible. (1825pond.s5b)

(23) In a little Time appear'd at the same Height with the sun, as near as I could guess, having no Instrument, a luminous Spot, being about four times the largeness of the Sun's Disk, and about 30 D. distant from the Sun to the

> Southward, which was covered with the lively Shades of red and yellow on the Side next the Sun, and encreased in Splendor (so as scarce to be born by the naked Eye) till it exceeded the Brightness of the $_1$Sun, $_1$which was then under a thin Cloud, so as easily to perceive his Disk. (1722cay-.s3b)

The relative clause in (22) does not single out a certain sub-set of thermometer-tube pieces but adds the information on what their role was in the experiment, namely to steady the cylinders. Similarly, the relative clause in (23) is not set-delimiting, in that there is only one sun in our solar system. However, the relative clause adds information that is essential to the topic under discussion.[12]

Finally, we mention a problem that seems to be confined to singular definite antecedents. In some instances it is unclear whether the definite article is anaphoric or cataphoric, as in (24):

> (24) Thirdly, Niter, which is made by the affusion of an Acid Spirit upon an Alcali, may be almost totally distill'd into an Acid Spirit, there appearing not the least footsteps of a Volatil Salt, and scarce any of the $_1$Alcali, out of $_1$which it was chiefly produced. (1676coxe.s2b)

The antecedent, “the Alcali,” is arguably already fully specified, having been previously mentioned (anaphoric *the*). In that case, the relative clause would be aspective. However, a quite natural reading takes the relative clause as specifying the particular alkali concerned (cataphoric *the*), and therefore as restrictive, even if redundantly so. This is an analytic problem, but the choice makes no substantive difference (on ambiguous determiner function as an analytic problem, see also Olofsson 1981:31-34 and Sigley 1997:121-123).

Integrated, Supplementary, and Continuative Relative Clauses

Huddleston, Pullum, and Peterson (2002:1058-1066) show that the set-theoretic definition of restrictive relative clauses does not capture precisely the set of clauses which have the various correlates of restrictiveness listed in Table 1. Instead they divide the spectrum of relative clauses in two at a different point, into "integrated" relatives (of which restrictive relatives are a proper subset) and "supplementary" relatives (2002:1034).[12A] In short, an integrated relative clause is closely linked to its matrix clause on the basis of intonation, syntax, and semantics, all of which--in their account--go together (whereas a supplementary relative clause merely adds extra information about the antecedent which is not fully integrated structurally into the matrix clause). Some relative clauses are integrated in the ways just mentioned but are nevertheless not restrictive in the set-theoretic sense of the term. As far as we can tell, these crucial cases correspond pretty closely to Sigley's aspective category (endorsed by Robert Sigley, p.c. 4 Feb 2012).

Another type is represented by the term "continuative," used by Jespersen (1909-1949:III 105-106) and defined as a type of relative clause that is "always added after what might have been the end of the sentence"; Romaine (1982:83) points out that they "advance the discourse by adding new information." Denison (1998:286-287) treats them as an extreme type of non-restrictive relative which is "in effect coordinated with, rather than subordinated to, what precedes." He tentatively suggests subsuming sentential relatives and relative clauses with determiner relative pronouns--for an example, see (27) below--under the heading of "continuative"; certainly, determiner relatives very frequently share the quality of resuming an apparently complete sentence and sending it off again in a new direction. (Sentential relatives are rather more varied, though some do this too. We have not assumed here that sentential relatives and determiner relatives must all be continuative.) Another characteristic of the semi-independence of continuative relatives is that they allow for non-declarative clauses to be embedded in them (Denison 1998:287, following Jespersen 1909-1949: III 105-106), though such examples do not occur in our data.

We give some examples of continuative relatives from our corpus in (25) to (27):

(25) A $_1$Discourse denying the Pre-existence of Alcalizate or Fixed Salts in any Subject, before it were exposed to the Action of Fire: To $_1$which is added a Confirmation of an Assertion, deliver'd in Numb. 101. p.5. & 6. of these Tracts, viz. That Alcalizate or Fixed Salts extracted out of the Ashes of Vegetables, do not differ from each other: The same likewise affirm'd of Volatil Salts and Vinous Spirits; by the Learned Dr. Daniel Cox. (1676coxe.s2b)

(26) This Sand is $_1$that which is commonly at or near the Sea-shoar, $_1$which to distinguish from what is useless; know, That the wash of the Sea rolls and tumbles stones & shells, &c. one over another, whose grating makes this Sand. If the matter be shelfy (as we call it) that is the grating of stones, it is of small valew. But if it be notably shelly, then it is what we desire. (1675hook.s2b)

(27) The Planks are laid in Sand; the lowest about six or eight Inches above the Iron-Plates, they are well cover'd with the sand, and Boards laid over all, to keep in the heat. The Sand is moistned with warm Water, (for which purpose they have a Cauldron adjoyning to the Stove)... (1722cay-.s3b)

We recognize adnominal continuative relatives as an extreme case of non-restrictive relatives but fully contained within the latter class; although a useful descriptive category, they pose no challenge to the restrictive/non-restrictive dichotomy. Although we treat continuative relatives as a subclass, we have no objection to Huddleston, Pullum, and Peterson's approach, in which they are merely a use to which supplementary relatives can be put in narrative (2002:1064). Example (26) is an infinitival relative of a type no longer found in English as an adnominal continuative relative. Notice, however, that (27) is sentential

rather than adnominal and will not therefore be considered further in the present paper (but see Figure 2 below).

Initial Coding of Corpus Data

We set out initially to test the utility of the restrictive/non-restrictive dichotomy. Three different scholars coded different records in the database and sometimes discussed tricky examples at length. One thing we were interested in was inter-coder variation. The first author is a native speaker. The second author and the third coder (Anja Neukom-Hermann) are both native speakers of German with near-native competence in English. For the field "type of relative" we went for a forced choice between restrictive and non-restrictive unless the example was finely balanced. Later we added the values "aspective" and "continuative," as discussed above.

We coded (or corrected) our examples for the following properties, as shown in Table 2:[13]

[TABLE 2 HERE]

We go into further detail below on some of these properties: the values "questionable," "unclear," and "other" under "type of relative" were labels introduced ad hoc as a first attempt to describe the data without theoretical preconceptions. They were later conflated under "unclear," although only three such examples remained by that stage.

In Table 3 we show a summary of the results for type of relative.[14] In this table, aspectives, ambiguous and uncertain examples are lumped together in the middle row. Note that zero relatives appear nowhere in either Table 2 or Table 3, as such clauses were excluded from the analysis; they are in any case very infrequent in our scientific data (see Hundt, Denison & Schneider 2012a,b). Moreover, zero relatives pose almost no problem with respect to the classification issue at the core of the present paper.

[TABLE 3 HERE]

We have indicated for each century the percentage of relative clauses that are neither clearly restrictive nor clearly non-restrictive. Sigley's twentieth-century figures (from the Written Corpus of New Zealand English) show only 4.5 percent of 1475 examples "ambiguous" between restrictive and non-restrictive, though this category is not precisely comparable with ours (1997:368, Table 10.1.2). Our first research question is answered: the distinction between restrictive and non-restrictive relatives does indeed become more problematic the further we move from the present day, though the differences between the chronological subsets of our data are below the level of statistical significance.

Logico-Semantic Function

As noted in the introduction, logico-semantic function has been considered one of the most important criteria, if not the most important, for the distinction between different types of relative. We started from the assumption that the basic distinction was between restrictives and non-restrictives. Within the non-restrictives, we chose to mark continuative relative clauses as a special case. Later we added the possibility of aspective relatives and recoded accordingly. However, there had been a residue of clauses which could not readily be classified under these four headings. Our initial classification allowed for a variety of problematic types. Coders were allowed to add a mark to signal the questionable status of some examples (e.g. "?restrictive"), and for clauses which could not even be doubtfully assigned to one of the classes we allowed for the problematic types "ambiguous," "unclear," and "other." The intention was always to refine the classification and deal with residual problems more systematically. For the present we give a couple of problematic examples.

In example (28) the referent of the antecedent is already clearly identified by material preceding the relative clause, as a zero relative clause and possibly an additional prepositional phrase delimit the set of possible observations:

(28) Accompanying I beg leave to offer you some $_1$observations, $_1$Ø I made in the year 1767, in the province of Allahabad, on the temperature of the weather, $_1$which will serve to elucidate the extraordinary and sudden changes incident to that part of Asia. (1775bark.s4b)

On those grounds example (28) is perhaps unlikely to be restrictive, though it is of course possible to cumulate restrictive modifiers.[15] Would the relative clause have been spoken in the same intonation contour as what precedes? It makes no difference to the interpretation, which would leave example (28) hovering between aspective and non-restrictive. In our database (28) was finally coded as aspective, but its interpretation is unaffected by the choice of label--which nicely illustrates the sometimes murky distinctions between restrictive, aspective, and non-restrictive.

Somewhat different is (29), where the relative *that*-clause seems to parallel the aspective *which*-relative earlier in the sentence:

(29) ...whereas on the contrary we do not find the lest ['least'] footsteps thereof either in $_1$Blood, Urine, Bones, Horns, & c. $_1$which do all abound with Volatil Salts; nor in some other $_2$parts, Excrements, and Juyces, $_2$that afford store of Acidity, which may frequently by coagulation be brought to a Saline form or consistence. (1676coxe.s2b)

If the clauses are parallel, we have an aspective relative, and if not, a restrictive relative, but the answer matters: the meaning would be different in each case. Example (29), unlike (28), was therefore initially coded as ambiguous--though we later agreed that it was unambiguously restrictive and recoded accordingly (see further the discussion of ambiguity below).[16]

Phonology and Punctuation

Our initial measure of juncture was an impressionistic attempt to judge the likelihood that a relative clause would belong in the same intonation contour as the antecedent (scored as 1) or form its own intonation contour (5). Intermediate scores of 2, 3, and 4 were allowed. Procedurally this is of dubious validity, as we were well aware, but it belonged with our research question about modern judgments on older data, and inter-coder variation was of potential interest. The problems, of course, are many. First, as already noted, we were using the judgments of speakers of PDE (both native and skilled ESL) on texts written long ago. Second, we were taking written texts--and not even written-to-be-spoken ones--and "translating" them into speech. Third, we were having to rewrite many of the examples before carrying out the test! This last point is because in written scientific texts there is often a parenthetic interruption between antecedent and the relevant relative clause, as in (30):

(30) In that paper I described various $_1$pieces of apparatus, chiefly in the form of delicate balances suspended in glass tubes, by means of $_1$which I was enabled to show attraction or repulsion when radiation acted on a mass at one end of the beam, according as the glass tube contained air at the normal pressure, or was perfectly exhausted. (1875croo.s6b)

For some coders, the test involves mentally downplaying or even excising the intervening material. As it happens, the pied piping seen in example (30) may be partly responsible for the juncture being scored as 4 (relatively loose), whether or not the parenthetic interruption is discounted; note that the example was classified as a restrictive relative.

Montgomery (1989:137) points out that punctuation of relative clauses only becomes standardized in the twentieth century.[17] In historical data, therefore, one cannot be sure of a

correlation between speech and punctuation, nor use punctuation as the basis of a distinction between types of relative clause. In this context notice the re-punctuation in (32), supposedly by William Gifford, of Jane Austen's manuscript of *Persuasion* in (31), to remove a comma before a restrictive relative clause (Sutherland 2010):

(31) …it was overwhelmed, buried, lost in those earlier feelings, which I had been smarting under Year after Year. (Manuscript; erasures and line breaks not reproduced)

(32) …it was overwhelmed, buried, lost in those earlier feelings which I had been smarting under year after year. (Printed text of 1818)

We decided that punctuation was not a safe diagnostic, as many writers did not seem to punctuate reliably according to modern conventions. Consider for example the zero relatives included incidentally in (2) and (28) above: both are marked off by commas, even though zero relatives are generally assumed to be restrictive. Some early writers punctuate very heavily in general. Those of (33) and (34), for instance, seem to follow a convention like that of Modern German in which almost all clauses--in one case, even a complex NP--are separated by commas or semicolons:

(33) By the Post-script of Mr. Lucas's Letter, one not acquainted with what has passed, might think, that he quotes the Observation of the R. Society against me; whereas the $_1$relation of their Observation, $_1$<u>which you sent to Liege</u>, contained nothing at all about the just proportion of the Length of the Image to its Breadth according to the angle of the Prism, nor any thing more (so far as I can perceive by your last) than what was pertinent to the things then in dispute, viz. that they found them succeed as I had affirmed. (1676newt.s2b)

(34) The first Experiment, I have to offer to your Observation at present, is made on the New England Cedar, or rather Juniper, grafted on the Virginia; and what is remarkable in it, is, That the $_1$Branch, $_1$<u>which is grafted</u>, is left several Inches below the Grafting, which part continues growing as well as the upper Part above the Grafting. (1724fair.s3b)

Notice that (34) at least is clearly a restrictive relative, despite commas. In both cases the comma that follows the antecedent can at least be counted as punctuation separating off the relative clause. However, as we illustrated in (30) above, early scientific texts often have material intervening between antecedent and relative clauses. In such cases the presence of a comma or other mark often makes it impossible even to test the punctuation status of the relative clause, as the punctuation could be ascribed to the parenthetic material and not necessarily to the relative clause.

Despite these serious caveats, punctuation turned out to be somewhat more consistent than we had expected. Here we confine ourselves to relative clauses which we had marked as unproblematically restrictive or non-restrictive and count instances with no punctuation immediately before the relative clause. As seen in Table 4, it turns out that less than a quarter of restrictives are punctuated, and the chronological trend is towards ever greater conformity with the PDE conventions. Although the great majority of non-restrictives are indeed set off by punctuation, there is no clear chronological trend.

[TABLE 4 HERE]

These findings confirm Montgomery's (1989) claim that punctuation standardizes as we move towards the twentieth century, at least for restrictives; the differences between adjacent centuries both prove significant in a chi-square test at $p \leq 0.001$. Non-restrictive relative clauses have always strongly tended to be preceded by some punctuation mark, whether a comma (including one that marks off a parenthetical clause), a bracket or even a

semicolon; there is no clear diachronic trend. In the section below on modeling the distinction, we will therefore discuss examples that were difficult to classify and consider whether, e.g., a clash of logico-semantic function and punctuation may have given rise to them being classified as unclear.

Measuring the Link between Antecedent and Relative

The relationship between antecedent and relative clause is central. Olofsson (1981:18) follows a Scandinavian tradition in using the term "relative junction," which refers to the constituent formed of an adnominal relative clause and its head: "the noun-like entity that results from the combination of an antecedent and a relative clause."[18] His terms for restrictive and non-restrictive relative clause (+ antecedent) are "tense" and "lax" junction, respectively (1981:18), but he focuses on the "tense relative junction," because its relative clause is part of the NP and not additional information that can be separated from the antecedent without great semantic effect. He claims the binary classification is exhaustive but that some written relative junctions are indeterminate (1981:46; cf. also 22, 27).

Whereas the term "tense junction" makes sense--it does indeed denote a nominal entity that is a syntactic and phonological unit--"lax junction" is an odd label. There we seem to be moving away from the concept of a linguistic unit and more towards a gradable concept that indicates the closeness of relation between antecedent and relative. What exactly is being measured, however? Syntax? Phonology? Semantics? More practically**,** how do we measure closeness in a way that can be operationalized? One could argue that phonology and punctuation are one way of measuring the strength of the link between antecedent and relative clause. In the preceding sections we saw that they are problematic for historical written data. A somewhat less problematic measure might be the distance between antecedent and relativizer, assuming that restrictive relative clauses will, as a general rule, allow for minimal distance between antecedent and relative clause. That assumption is implicit in Montgomery's summary of the situation:

> There has been a continuing tendency since Middle English to reduce the degree of separation of a relative clause from its headnoun, or to put it another way, an increasing tendency for nonrestrictive relative clauses to become more closely attached to their headnouns. (Montgomery 1989:136f.)

On the basis of our coding of the antecedent, the database calculated that distance as number of words intervening before the relative clause. As a general rule, the distance between the antecedent and the relativizer seems to be a relatively reliable criterion for analysis: of the 132 unproblematic restrictive relative clauses in our seventeenth-century data, 129 (97.7 percent) have a maximum of three words between the antecedent and the relativizer, 118 (89.4 percent) even only a maximum of one. The elements separating the relativizer from its antecedent, moreover, are mostly postmodifying prepositional phrases, as in (35), short appositions that do not "define" the antecedent, as in (36), one example of a long intervening apposition, (30) above, or a preposition connecting the relative clause to its place of extraction from the main clause, as in (37):[19]

(35) But then those $_1$motions of the Winds and Seas, $_1$<u>which were favourable to those who sayled Northward</u>, will be contrary to those who stand Southward, and they may long enough expect Northern gales, which seldom blow till towards the latter end of Summer, viz. in the month of August. (1675anon.s2b)

(36) ...unless it be those pieces still remaining in my collection, and a $_1$piece, somewhat smaller than yours, $_1$<u>which I gave to the king of France's cabinet at Paris</u>. (1775bruc.s4b)

(37) It hath been a constant and general perswasion, that many Fixt Salts do retain, some, at least, the Specifical properties of those $_1$Vegetables, out of $_1$whose ashes they were extracted. (1676coxe.s2b)

However, the distance between antecedent and relativizer naturally grows somewhat for the second of two consecutive relative clauses sharing the same antecedent, either joined by a coordinating conjunction, as in (38), or without a conjunction, as in (39):

(38) I may also notice that, notwithstanding the superiority of the more recent observations and the inaccuracy of many of the older ones, there are a certain $_1$number of the latter $_1$which were made with great care, and $_1$which may vie with recent experiments in exactness... (1874pres.s6b)

(39) The process results in the production of a form which I proposed to call the Planula, but which Professor HAECKEL has better termed the Gastrula, reserving the former name for a $_1$condition of the Gastrula $_1$which sometimes presents itself in $_1$which there is no aperture of invagination. (1874lank.s6b)

Example (38) raises a theoretical issue: is the clause in question restrictive in the set-theoretic sense if it does not restrict further the set of possible referents already delimited by the preceding relative clause? We do not have a general answer to this question.

[TABLE 5 HERE]

The results in Table 5 show that restrictive relative clauses have a strong tendency to follow their antecedent quite closely--surprisingly, even more so in our early texts than in the nineteenth-century data. In other words, a distance of more than ten words between antecedent and relativizer makes it very unlikely that the relative clause is restrictive. There is, admittedly, a risk of circularity here, in that our judgment of what constitutes a clear case

of a restrictive or non-restrictive relative might in part be influenced by its distance from the antecedent.

There are two counter-examples in our initial analysis, though. Example (40) below was coded as restrictive on first analysis, probably because the antecedent “Assertion” is preceded by an indefinite pronoun and needs some further specification. However, it could be argued that the relative clause--separated by 21 words from the antecedent--is not restrictive, because the intervening material (non-finite clause and prepositional phrases) sufficiently identifies the antecedent. Moreover, the relative clause is separated by a semicolon and itself contains embedded clauses, i.e., there are a number of formal criteria that suggest we are dealing with a continuative rather than a restrictive relative clause:

(40) Having dispatch’t this, I cannot but take notice, that I am credibly inform’d, that many persons of no ordinary repute for their skill in Chymistry, and other Arts subservient to Experimental Philosophy, have been pleas’d to censure in an unusual measure of severity an $_{1}$Assertion, accidentally dropt from my Pen, in a Discourse concerning the Volatil Salts of Vegetables, in Numb. 101. of the Ph. Transactions; $_{1}$which although circumscribed by a Parenthesis, and an alien to the main design and scope of my undertaking, yet was so far from being thereby protected, that it hath sustain'd the brunt of many unkind Reproaches, and been represented as a Position without foundation in Reason or Experience. (1676coxe.s2b)

Similarly, in (41), which we initially coded as restrictive; the non-finite clause preceding the relative clause could be seen as sufficiently “defining” the antecedent. Given the essential nature of the information conveyed in the clause, the classification of aspective is suggested.

One could even argue that the clause is continuative. We will return to such problematic examples in the section on ambiguity below.

(41) In most Fishes there is a manifest $_1$channel leading from the gullet or upper orifice of the stomach to the said bladder, $_1$which without doubt serves for conveying air thereinto, as may easily be tried by any one that pleases. But though air may be received into the bladder, yet is there a valve or some other contrivance to hinder the egress of it; for you shall sooner break the bladder than force any air out by this channel. (1675ray-.s2b)

Example (41) was in the end coded as uncertain. Example (42) is coded as aspective in our database but is not greatly dissimilar:

(42) I am just now constructing a $_1$photometer about two feet in diameter, and two or three inches deep, with $_1$which I hope to appreciate the effect of heat in the feeble rays of the moon. (1825pond.s5b)

Both (41) and (42) serve to show that there is no clear boundary between restrictive and aspective.

One example, (43), is so bizarre as to resist coding for closeness of link:

(43) The author has assumed four successive thicknesses for the $_1$shell, viz. <proto-table>, and proceeding on the above principles has calculated the total annual contraction of the nucleus for each case. The partial mean coefficient of contraction adopted for that of the nucleus has been the mean between the two highest partial means shown in the curve and Table I above given, viz.

> 0.0000769 for 1° FAHR.
>
> The final results obtained are comprised in Table II., before referring to which, however, some explanation and reference to diagram fig. 2 are necessary.
> R being the radius of our globe = 3957.5 English miles, r = the radius assumed for the nucleus, $_1$whose thickness = R-r.
>
> Let the nucleus be assumed to contract by loss of its heat transmitted through the shell until its radius = r', the shell then, in following down after the contracted nucleus, must descend everywhere through a vertical height equal r-r'. (1874mall.s6b)

The piece details how heat loss from the nucleus ('core') of the Earth through its outer shell is calculated. From reading the entire contribution, it is clear that the word *thickness* in the relative clause can only apply to *shell*--not to *radius, globe,* or *nucleus*--but the nearest occurrence of *shell,* some 100 words earlier, is hardly available to act as a prototypical antecedent. The clause is thus not really grammatically integrated into the discourse and cannot therefore be treated as adnominal.[20]

Modeling the Distinction

Binary, Gradient, or Multi-dimensional?

When there are two terms, whether restrictive and non-restrictive, tense and lax, defining and non-defining, or any other contrastive pair, the simplest approach is to invoke the Law of the Excluded Third and treat the terms as mutually exclusive and exhaustive. That, whether implicit or explicit, is the line taken in many elementary handbooks and guides. Even with the introduction of a further, minor class or classes, as Olofsson (1981) and others contemplate, the several terms of the system can remain clearly delimited, now under the Law of the Excluded Middle (see here Aarts 2006:363, esp. fn.2).

Another theoretical model for a two-term system is one-dimensional gradience, with either continuous or step-wise variation from wholly restrictive to wholly non-restrictive (or *pari passu* for other labels). We explore this possibility below.

Finally we must consider the possibility of modeling our data in a multidimensional system where restrictive and non-restrictive (etc.) probably represent dominant clusters of properties, but where other combinations of properties are also represented, as hinted by Sigley (1997:129-130). Once again, each dimension of variation could in principle involve mutually exclusive classes or could vary in gradient fashion. Sigley is the most impressive proponent here, suggesting a space of variation which he presents graphically (1997:129), our Figure 1.

[FIGURE 1 HERE]

In this conception, intended to capture the PDE facts, there are essentially three dimensions of variation presented in a two-dimensional diagram. The y-axis represents a broadly semantic parameter. Three different parameters are packed into the x-axis. At the top of the diagram we have the traditional set-theoretic notion of ±restrictive, which is only roughly correlated with the two scales at the foot of the diagram, which are respectively to do with information packaging and phonology; the latter two are regarded as marching in lockstep and wholly correlated and can therefore share a single scale. The diagonal dividers allow for different groupings of data according to the two different horizontally plotted parameters, set-theoretic and informational/phonological.

While we are broadly sympathetic with Sigley's careful and innovative approach, we have not been able to adopt it for our study.[21] The main reason is that we found it very difficult to operationalize the concept of antecedent specificity and so could not make use of the vertical dimension. The horizontal dimensions make good sense, though for the reasons discussed above, for our historical, written data the one at the foot of the diagram is more

easily operationalized as ±essential (or equivalently, the converse ±parenthetic) than as tense-lax.

Revised Classification

We do not wish to multiply categories needlessly nor to create a more complex picture than our data analysis can support. Our tentative conclusion from this study of scientific texts from the seventeenth to the nineteenth centuries is that restrictive/non-restrictive is less helpful as a binary distinction for adnominal relatives than Huddleston, Pullum, and Peterson's (2002) integrated/supplementary. Their dichotomy makes a good starting point. But the binary distinction is not really workable. There are several subtypes which deserve recognition, attractors in the space of possible relative clause types, so that a more subtle picture might represent adnominal relative clauses as bands on a one-dimensional gradient. As illustrated in Figure 2, at one extreme are those clauses which are universally agreed to be restrictive. At the other extreme are continuative relative clauses, which are less tightly integrated into NP structure. In between we find aspective and ordinary non-restrictive/supplementary relatives. The large rounded box represents adnominal relatives. A fuller picture would bring in sentential relatives (which were outside the scope of this study); for now we have merely hinted at sentential relatives as an area largely to the right of the adnominal box and shown how two of our types straddle the adnominal-sentential boundary.

[FIGURE 2 HERE]

The other important feature of our classification is that the bands overlap. We have argued that certain examples are underdetermined and work equally well--and with much the same meaning--whether they are classified in one way or another: as restrictive or aspective, as aspective or ordinary non-restrictive, as ordinary non-restrictive or continuative. Alternative terms for underdetermined examples include "underspecified" and "vague." The phenomenon is quite widespread, and examples (24), (28), (41), and (42) above are only a sample.

Robert Sigley makes some useful observations on an earlier version of our Figure 2 (p.c. 21 Feb. 2012), four of which we quote here almost *verbatim*:

A. "Restrictive"/"aspective" overlap predominantly in cases where the antecedent could be conceptualized either as having an individual-level or class-level referent. (On this distinction please see note 16).

B. "Nonrestrictive"/"aspective" overlap mainly as a result of the subjective nature of "relevance."

C. The diagram implies there is no possible overlap between "restrictive"/"nonrestrictive" categories, which is probably true by definition, although surely ambiguity is possible (and is not shown).

D. The "integrated"/"supplementary" dichotomy--because it is a structural, not a functional, label--really should be represented as binary, rather than a continuum.

It seems to us that A is a helpful observation, and it is possible that following it through might reduce that particular overlap. We accept B too: recall that aspective clauses capture "an aspect of the antecedent that is extremely relevant to the content of the matrix clause" (Sigley 1997:127), a definition that involves a subjective judgment. We think B supports the case for genuine vagueness and explains the relatively large number of cases found in this group (see Table 6 below). As for C, it is precisely our point that uncertainty between restrictive and non-restrictive would be a matter of ambiguity (as discussed below), not vagueness: there is no overlap. On D, however, which certainly represents a widely-held view on the nature of structural analysis, we do not necessarily agree. A structural dichotomy in any constituent structure framework would indeed be binary, but constituent structure grammar is by no means the only kind of syntax out there, although it is dominant. We do not propose a formal model of syntax in this paper, but see for example Quirk (1965), who allows for the

possibility of one kind of gradience across a range of patterns, or those models (such as certain versions of Construction Grammar or Word Grammar) which allow for dual inheritance. It goes beyond the limitations of our data to carry through a rigorous distinction between functional and structural parameters.

What we are claiming is that a descriptively adequate picture of our data should allow for a minimum number of intermediate classes, and furthermore that to try to reassign members of such intermediate classes to one or other of the adjacent "established" classes would be an arbitrary decision for which there is no convincing evidence. Those examples are underdetermined and do not need to be resolved either by language users or linguists. We are less concerned to argue for a real continuum than for the possibility of specific areas of overlap between adjacent classes.

Ambiguity

Absence of a clear boundary is different from ambiguous cases caused by uncertainty about the potential intonation and the precise meaning intended. In such cases we cannot decide between two incompatible analyses, a problem due to missing information: unavailability of the original writer, historical distance, sentence constructions with elements intervening between antecedent and relative clause, and so on. An illustrative case is found in (44).

(44) And the $_1$Gentry, $_1$<u>that ride abroad</u>, do little mind these things.
(1675hook.s2b)

Without further context, example (44) is genuinely ambiguous on semantic/pragmatic as well as prosodic grounds; at such an early date, neither relativizer *that* nor punctuation can override this ambiguity. In the wider context an aspective reading looks more likely:

(44') The reason of it is (I think) that the Labouring part do seldom travel, or remove, so as to learn by other experience: And the $_1$Gentry, $_1$that ride abroad, do little mind these things. (1675hook.s2b)

Another ambiguous example is (45):

(45) But especially our Country $_1$men $_1$who are satisfied in the experience of it, should seriously bethink themselves, If there may not be easier and cheaper way of Conveyance, for a greater quantity thereof to be brought up into the middle of the Country. (1675hook.s2b)

The question here is whether the relative clause refers to all our countrymen or merely a subset. Both readings are possible.

In example (46) there is some uncertainty as to the antecedent: if it is *design*, then the relative clause must surely be non-restrictive; if, on the other hand, it is *part*, then the type is ambiguous between restrictive and non-restrictive/aspective (see Sigley 1997:120 on ambiguity arising from multiple possible antecedents). Even after checking the wider context and the illustration, we cannot resolve either point. (This example is therefore three-way ambiguous!)

(46) The Figures of the naked Snails are omitted in this Specimen, being not material to that $_1$?part of the $_1$?design, $_1$which is, (when other parts of these Tables are finish't,) to give the Reader an exact view of Animal-shells, as well as of Fossils figured like Shells, whereby he will be best able to Judge, what to think of their Original. (1674ano1.s2b)

Example (47) is an interesting one:

(47) Many other $_1$Instances I could name, $_1$which, if we had such a Man as Mr. Sellar, who could employ Workmen to perfect the Instruments, and to sell them off; it would (doubtless) procure us many Operatours, and many free Discoveries in some points of Philosophy, of which we have yet heard but little Tydings. (1675ray-.s2b)

The first relative clause minus various parenthetic elements appears to be "which...it would...procure us many Operatours...." If *which* is subject, then the singular pronoun *it* cannot be resumptive for a relative with plural reference. Alternatively, as Sigley suggests (p.c. 21 Feb. 2012), *which* is not subject but pseudo-locative (= *in which*). For our purposes, though, the more interesting problem is the type of relative. It seems to us that the relative clause under discussion is ambiguous between restrictive and continuative. But on our analysis, continuative relatives are even more detached from the antecedent than (other) non-restrictives. On a one-dimensional gradient scale, such an example could not be underspecified or vague but could only be truly ambiguous. It would be more convenient to dismiss (47) as an anacoluthon, and this may indeed be the case. Example (40) above is somewhat similar, although in that case we felt that the ambiguity was resolved in favor of a continuative reading.

Another apparent ambiguity between non-adjacent categories is illustrated by (48):

(48) The natural $_1$hollow $_1$which it occupies appears formerly to have been a $_2$lake, $_2$which in process of time became nearly filled by the continued growth and decay of marshy plants, and the consequent formation of peat. (1825weav.s5b)

The second relative clause in (48) appears to be essential information, which would argue for an aspective reading, but the PP “in process of time” makes explicit the narrative separateness of the relative clause: with looser juncture, the clause can be read as continuative. Our analysis requires the choice to be a question of ambiguity, but if so, the semantic difference between the readings is quite subtle.

Compare now example (49):

(49) SAUSSURE used a spirit-thermometer of REAUMUR S with a large $_1$ball, $_1$which he surrounded with a mixture of wax, resin, and oil 3 inches thick; and the whole was then placed in an iron-wire cage. (1874pres.s6b_tagged)

The main clause would be complete if the sentence ended with “ball.” The underlined relative clause is clearly non-restrictive, but it is arguable whether it is best regarded as “advancing the discourse”--in this case, moving on in time to the next action of the experimenter, like the “and…then” clause which follows--or whether it is part of the same situation as “used a…thermometer” and merely adds extra information to that, as non-restrictives generally do. We lean towards the former interpretation and therefore have coded (49) as continuative (in this context, see also the discussion of examples [29] and [33] above**)**.

Revised Coding of Corpus Data

We have suggested a revised classification of adnominal relatives with four main bands: restrictive, aspective, non-restrictive, and continuative. The bands overlap, which gives a further three intermediate classes for examples which are indeterminate (vague) between adjacent bands. There is also the possibility of ambiguity between two (or more) of these seven classes, and here we make a broad distinction between two sorts: whether the ambiguity is between adjacent classes or not. To see whether this classification is workable

and how the data are distributed, we have worked through our seventeenth-century corpus data in detail. The revised tabulation is given in Table 6.

[TABLE 6 HERE]

How does this compare with our initial analysis in Table 3?[22] It is easiest to visualize the two classifications in chart form.

[FIGURE 3 HERE]

[FIGURE 4 HERE]

The comparison shows minimal change in the proportion of restrictive and non-restrictive relative clauses (50.2 > 51.9 percent and 23.6 > 24.2 percent, respectively). We have separated out examples showing real ambiguity, distinguishing between adjacent and non-adjacent categories (4.6 percent in all). The real improvement is that the residue of questionable relative clauses (covering both ambiguous and otherwise "difficult" examples) has been reduced from 26.2 percent to a mere 1.1 percent (unclear) by the availability of a more fine-grained classification (plus reconsideration of some previously misunderstood examples). The adoption of the aspective type (altogether 15.5 percent at most if we include overlap on either side) has added an important new category to the traditional dichotomy. Furthermore, the recognition that classes overlap make the revised classification more representative of the amount of variation found in the data.

In our seventeenth-century sample we did not, as it happens, wind up with any examples in the overlap between non-restrictive and continuative relative clauses. However, the possibility should be allowed for. In addition to (49) above, consider such examples as (50) and (51), which at least admit of some doubt (even if in some cases resolved by us either as non-restrictive or as continuative):

(50) ...most of them being taken from $_1$Stars $_1$which are not in the British Catalogue, $_1$whose Places therefore are here determined, only by comparing them with some that were... (1724brad.s3b)

(51) The first Plant I shewed was the $_1$Laureola, grafted upon the Mezerion, and the Evergreen $_2$Oak of Virginia upon the common English Oak; both $_{1\&2}$which hold their Leaves all the Winter, and are in good State and flourishing, though grafted on Plants that drop their Leaves in Winter...(1724fair.s3b)

Both (50) and (51) were coded as non-restrictive.

Conclusion

We cannot be sure of all our data. There were quite a few examples which challenged our *Sprachgefühl* at first (though most could be resolved on closer inspection). The difficulties were partly due to sentence length and different structures, partly also because scientific texts can be hard for non-scientists to understand and therefore particularly difficult to analyze, with uncertainty as to what the antecedent is on top of uncertainties about the type of relative clause. As we have seen, there were many occasions for insecurity about the best classification for a particular example; the absence of intonational information is the most critical factor here. Historical distance is compounded by changes in genre conventions and publication processes, with earlier periods showing a greater tolerance of more loosely constructed sentences as well as greater sentence length and complexity in formal written language. Copy editing nowadays helps to reduce possible ambiguities, which is particularly pertinent to scientific writing.

Can we trust our intuition in analysis of historical data? Only with caution. Initial judgments were often biased, particularly by knowledge of present-day tendencies in

relativizer choice or conventions of punctuation. Quite a few examples provoked disagreement among the three annotators and sometimes prolonged discussion--individual examples have in some cases been reclassified several times in our database. With care, however, we can confidently assign the great majority of relative clauses appropriately, so long as we use the full range of available evidence. And assigning them appropriately in turn means to an appropriate class or--where necessary--to a pair of adjacent classes. Both points are crucial: taking the full range of evidence into account, and having a suitable classification.

Our initial procedure had involved a forced choice between restrictive and non-restrictive relative clauses, creating a substantial residue of clauses which were not clearly one or the other. Careful analysis of that residue confirmed that a simple restrictive vs. non-restrictive dichotomy was not sufficient, Huddleston, Pullum, and Peterson (2002) having already shown convincingly for PDE that the purely logico-semantic definition of restrictiveness is in conflict with a definition based on formal characteristics. The class of aspective relatives, as named by Sigley (1997), turns out to capture the problem area caught between the two kinds of definition. With this three-fold classification, the number of problematic analyses is greatly reduced. A fourth type, continuative, formed a coherent subset of non-restrictives and was therefore added to the classification. The remaining difficulties are in large part due to the fact that boundaries between adjacent classes are not Aristotelian. To insist that all examples must fall cleanly within just one of four classes is to impose an artificial neatness on the data, since certain examples are simply underspecified. This is why the overlaps need to be recognized. We are advocating neither an endless subdivision of the data nor a continuum, merely the minimum level of flexibility needed to accommodate the data that actually occur in the ARCHER science texts. Even then a residue inevitably remains, genuinely ambiguous or simply incomprehensible, but it is not large. And in principle it is no different from the residue of problem cases that one would encounter in a present-day dataset.

Science may seem to be an extreme genre as far as difficulty is concerned, but for PDE Sigley (1997) has shown more generally that the traditional dichotomy fails to capture the range of relative clauses in naturally-occurring data. We suggest that the classification called for by our data could be applicable to other kinds of text.

Acknowledgements

We wish to record our thanks to Gerold Schneider, whose parser automatically retrieved the relative clauses that formed the original basis of our database, to Anja Neukom-Hermann, for help with the removal of duplicates, for her meticulous coding of some seventeenth- and eighteenth-century data, and for her insightful questioning of categories, and to Gunnel Tottie and Robert Sigley, for their detailed comments on a draft; Sigley's in particular led to serious rewriting. Thanks too to Sigley for the use of his figure. We have taken account of the helpful comments made by two anonymous JEngL reviewers. None of the above can be held responsible for our final version.

Notes

1. Note that relative clauses as such are not a universal category, and that the distinction between different types of relative clause as a linguistic universal is also contested (Balthasar Bickel, p.c.); see also http://ressources-cla.univ-fcomte.fr/gerflint/Mondearabe7/grande.pdf for a view that argues in favor of a universal distinction, and http://wwwstaff.eva.mpg.de/~cschmidt/SWL1/handouts/Gensler.pdf, which contests it.

2. Suárez-Gómez (2006:38), in her study of Old English relative clauses, points out that the basic distinction between restrictive and non-restrictive relative clauses is there in the earlier periods, but also that it is essentially a semantic one (see also Traugott 1972:103; Mitchell 1985:II 167-177). Suárez-Gómez (2006:47) also concedes that the binary distinction only works for the prototypical instances but poses problems for peripheral ones.

3. We used a preliminary version of ARCHER 3.2. See http://www. manchester.ac.uk/archer/, http://www.helsinki.fi/varieng/CoRD/corpora/ARCHER/index.html for details of ARCHER= A Representative Corpus of Historical English Registers version 3.2. 1990–1993/2002/2007/2010/2013 . Originally compiled under the supervision of Douglas Biber and Edward Finegan at Northern Arizona University and University of Southern California; modified and expanded by subsequent members of a consortium of universities. Current member universities are Northern Arizona, Southern California, Freiburg, Heidelberg, Helsinki, Uppsala, Michigan, Manchester, Lancaster, Bamberg, Zurich, Trier, Salford, and Santiago de Compostela.

4. The parser uses a dependency-based grammatical model and was designed to cover the most frequent phenomena of standard PDE grammar.

5. In this paper we mark the relevant relative clause by underlining and use subscripts to indicate (possible) anaphoric reference between nominal antecedent and relativizer. Following the logic of the dependency-based parser, it is the head of the antecedent which is

subscripted. This has several advantages, not least consistency. For a determiner before the head it avoids prejudging whether the relative clause is non-restrictive, in which case in a constituency analysis the determiner would be part of the antecedent, or restrictive, in which case it would not. Similarly it does not require us to judge whether a post-modifying prepositional phrase is part of the antecedent of a relative clause (stacked post-modification) or not (two separate post-modifiers); see also note 19 below. Both points are illustrated in the following example:

(i) a $_1$Discourse upon this Subject, printed in the Philosophical Transactions, in $_1$which he observes, That… (1720perc.s3b)

6. A number of relative clause types were excluded by design. We did not include adverbial relativizers like *where, why, how,* etc., nominal relative clauses, non-finite relative clauses, or--for this paper--zero relatives. Like Sigley (1997:32) we did not consider conjoined relative clauses without an explicit relativizer:

(ii) ...but were succeeded by $_1$others $_1$which appear'd and $_1$Ø vanished again by turns... (1720cote.s3b)

We removed *it*-clefts, whose relative-like clauses do not lend themselves to discussion as (non-)restrictives.

7. Example (11) cannot be a relative because there is no gap (Sigley p.c. 21 Feb. 2012); likewise (12). On the dispute as to whether relative *that* is a pronoun or a complementizer, see for instance van der Auwera (1985) or Seppänen (1997); on the historical basis for the distinction, see Seppänen (2000).

8. A somewhat simplistic prescriptivist "definition" is: "A restrictive relative clause is essential to the grammatical and logical completeness of a sentence" (Garner 2003:782). The more descriptive (data-informed) usage guide by Peters (2004:469) concedes that the distinction between restrictive and non-restrictive relative clauses is far from clear at times, especially with indefinite antecedents.

9. For some formal proposals for modeling the syntactic distinction in generative grammars by means of structural differences see, e.g., Chomsky (1965:217), Haegeman (1994:202), or Fabb (2005).

10. Prescriptivists often maintain that the distinction is (relatively) unproblematic. Fowler (1965:626), for instance, claims that "[t]here is no great difficulty, though often more than in this chosen pair, about deciding whether a relative clause is defining [his term for 'restrictive'] or not."

11. In correspondence Sigley has suggested that aspectives should be regarded as a subtype of restrictive relatives where the antecedent NP has an individual-level rather than class-level referent (p.c. 6, 21 Feb. 2012).

12. Sigley is not convinced that the relative clause of (23) is essential and therefore aspective (p.c. 21 Feb. 2012).

12A. Actually Huddleston, Pullum, and Peterson's classification has two further members: "clefts," which we do not consider here, and "fused" relatives, which are not adnominal and hence are also irrelevant to our purposes.

13. Some additional properties that were also coded for initially turned out to be too subjective and prone to inter-coder variation and were therefore not considered further in the analysis.

14. ARCHER science only has one 50-year subperiod in the seventeenth century but two each in subsequent centuries, hence the discrepancy in total number of relatives in Table 3.

15. In example (28) the antecedent is preceded by the indefinite quantifier *some*, which actually makes an analysis of the relative clause as restrictive plausible, although on semantic grounds the other post-modifications weaken the case for a restrictive reading. At the same time, the intervening insertions make it harder to read this sentence with tight juncture of the relative clause. Punctuation in these examples is of no help either, quite apart from its general unreliability in older texts (next section), because the comma preceding the relative clause is required by the preceding parenthesis. See also note 11.

16. We argue that blood, urine, etc. are all clearly defined substances, and that the first relative clause in (29)--indicated with subscript 1--therefore provides essential information but is not set-delimiting, i.e., it is aspective. The antecedents of the second relative clause, however--subscript 2, the clause under discussion--are not clearly defined: some other parts, excrements and juices (which we have hesitantly taken as a list of three, rather than an antecedent *parts* followed by an apposition). In other words, the relative clause really does restrict the set to those that "afford store of acidity." Sigley generalizes the difference to whether the relative clause applies to a whole class already specified (class level, aspective) or each member of the class (individual level, restrictive) (p.c. 21 Feb. 2012).

17. Even in twentieth-century English, punctuation does not reliably distinguish between restrictive and non-restrictive relative clauses; Sigley (1997:111) even suspects that the difference in punctuation between the Brown corpus and the Wellington Corpus of Written New Zealand English might indicate ongoing change.

18. The term "junction" comes via Christophersen (1939:36-7) from Jespersen (1924:97).

19. Note that (35) is the only instance where the prepositional phrase adds up to five words separating the relativizer from its antecedent. Certain prepositional phrases which could be argued to be part of the antecedent are nevertheless counted as intervening words,

because we always count from the head of the antecedent phrase (see note 6). An example is the PP *of gold* in:

(iii) I took a $_1$Ball of Gold of an Inch in Diameter, $_1$<u>that had a little Stem of the same Metal</u>... (1721desa.s3b)

This increases the apparent separation in such cases.

20. In our original coding, the antecedent had been wrongly identified as *nucleus* and the relative clause accordingly classified as aspective.

21. Sigley's view is that his 1997 diagram oversimplifies a far more multidimensional situation and "confuses levels of 'how speakers operate' and 'what analysts can observe,' which can and should be more carefully separated" (p.c. 21 Feb. 2012). Nevertheless he was kind enough to send the image file reproduced here (p.c. 4 Apr. 2012).

22. Both individual figures and the total vary slightly because of extensive reclassification of examples.

References

Aarts, Bas. 2006. Conceptions of categorization in the history of linguistics. *Language Sciences* 28(4). 361-385.

van der Auwera, Johan. 1985. Relative *that* – a centennial dispute. *Journal of Linguistics* 21(1). 149-179.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Christophersen, Paul. 1939. *The articles: A study of their theory and use in English*. Copenhagen: Munksgaard.

Denison, David. 1998. Syntax. In Suzanne Romaine (ed.), *The Cambridge history of the English language*, vol. 4, *1776-1997*, 92-329. Cambridge: Cambridge University Press.

Fabb, Nigel. 2005. *Sentence structure*. London: Routledge.

Fowler, Henry W. & Ernest Gowers (ed.) 1965. *A dictionary of Modern English usage.* 2nd edn. London: Clarendon Press.

Garner, Bryan A. 2003. *Garner's modern American usage.* 2nd edn. Oxford: Oxford University Press.

Geisler, Christer & Christine Johansson. 2002. Relativization in formal spoken American English. In Marko Modiano (ed.), *Studies in Mid-Atlantic English*, 87-109. Gävle: Högskolan i Gävle.

Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory.* 2nd edn. Oxford: Blackwell.

Huddleston, Rodney, John Payne & Peter Peterson. 2002. Coordination and supplementation. In Rodney Huddleston & Geoffrey K. Pullum et al. (eds.), *The Cambridge grammar of the English language*, 1273-1362. Cambridge: Cambridge University Press.

Huddleston, Rodney & Geoffrey K. Pullum et al. (eds.) 2002. *The Cambridge grammar of the English language.* Cambridge: Cambridge University Press.

Huddleston, Rodney, Geoffrey K. Pullum & Peter Peterson. 2002. Relative constructions and unbounded dependencies. In Rodney Huddleston & Geoffrey K. Pullum et al. (eds.), *The Cambridge grammar of the English language*, 1031-1096. Cambridge: Cambridge University Press.

Hundt, Marianne, David Denison & Gerold Schneider. 2012a. Retrieving relatives from historical data. *Literary and Linguistic Computing* 27(1). 3-16.

Hundt, Marianne, David Denison & Gerold Schneider. 2012b. Relative complexity in scientific discourse. Special issue, *English Language and Linguistics* 16(2). 209-240.

Jacobsson, Bengt. 1994. Nonrestrictive relative *that*-clauses revisited. *Studia Neophilologica* 66(2). 181-195.

Jespersen, Otto. 1909-1949. *A Modern English grammar on historical principles.* 7 vols. Heidelberg: Carl Winters Universitätsbuchhandlung. Reprint, 1961, London: George Allen & Unwin.

Jespersen, Otto. 1924. *The philosophy of grammar*. London: George Allen & Unwin.

Lehmann, Christian. 1984. *Der Relativsatz*. Tübingen: Gunter Narr.

Mitchell, Bruce. 1985. *Old English syntax*, 2 vols. Oxford: Clarendon Press.

Montgomery, Michael. 1989. The standardization of English relative clauses. In Joseph B. Trahern, Jr. (ed.), *Standardizing English: Essays in the history of language change: In honor of John Hurt Fisher*, 113-138. Knoxville, TN: University of Tennessee Press.

Olofsson, Arne. 1981. *Relative junctions in written American English*. Gothenburg: Acta Universitatis Gothoburgensis.

Peters, Pam. 2004. *The Cambridge guide to English usage*. Cambridge: Cambridge University Press.

Quirk, Randolph. 1965. Descriptive statement and serial relationship. *Language* 41(2). 205-217.

Romaine, Suzanne. 1982. *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge University Press.

Rydén, Mats. 1974. On notional relations in the relative clause complex. *English Studies* 55(6). 542-545.

Rydén, Mats. 1984. När är en relativsats "nödvändig"? *Moderna Språk* 78. 19-22.

Schneider, Gerold. 2008. Hybrid long-distance functional dependency parsing. Zürich: University of Zürich Ph.D. dissertation.

Seppänen, Aimo. 1997. Relative *that* and prepositional complementation. *English Language and Linguistics* 1(1). 111-133.

Seppänen, Aimo. 2000. On the history of relative *that*. In Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg & C. B. McCully (eds.), *Generative theory and corpus studies: A dialogue from 10 ICEHL*, 27-52. Berlin: Mouton de Gruyter.

Sigley, Robert. 1997. *Choosing your relatives: Relative clauses in New Zealand English*. Victoria: Victoria University Ph.D. dissertation.

Suárez-Gómez, Cristina. 2006. *Relativisation in early English (950-1250): The position of relative clauses*. Bern: Peter Lang.

Sutherland, Kathryn. 2010. "Austen's points": Kathryn Sutherland responds. http://languagelog.ldc.upenn.edu/nll/?p=2811<accessed 29 Nov. 2010>

Traugott, Elizabeth Closs. 1972. *A history of English syntax: A transformational approach to the history of English sentence structure*. New York: Holt, Rinehart & Winston.

Wood, Frederick T. 1952. "That" and "which" as relative pronouns. *English Language Teaching Journal* 7. 8-14.

TABLE 1

Alleged Diagnostics for (Non-)Restrictives

	Restrictive	Non-restrictive
Relative clause restricts set of entities denoted by head	+	–
Relative clause is essential for full understanding of the matrix	+	–
Relative clause belongs to the same intonation contour as matrix in speech	+	–
Zero relativizer is a possible substitute (except in subject function)	+	–
Relativizer is *that*	+	–
Relative clause is semantically optional, an aside that provides additional information	–	+
Relative clause is preceded by a pause in speech, comma in writing	–	+
Relativizer is a *wh*-pronoun	–	+

TABLE 2

Initial Coding of Examples

Field	Possible values
Type of relative	restrictive, non-restrictive, continuative, aspective, questionable, ambiguous, unclear, other
Relativizer	*that, which, who, whom, whose*
Antecedent	[free]
Verb of relative clause	[free] – used mainly to identify particular relative clause in complex sentence
Punctuation before relative clause	comma, non-relative comma, none, other
Juncture	1 tight, 2, 3 uncertain, 4, 5 loose

TABLE 3

Restrictive vs. Non-Restrictive (Initial Figures)

Coded as	17th century	18th century	19th century
Restrictive	132	187	229
Neither clearly restrictive nor clearly non-restrictive	69 26.2%	100 21.3%	90 20.5%
Non-restrictive (including continuative)	62 (5)	182 (1)	120 (3)
Total	263	469	439

TABLE 4

Proportions of Clearly Restrictive and Non-Restrictive Relative Clauses without Punctuation

	Restrictive	Non-restrictive
17th century	92/132 69.7%	1/57 1.8%
18th century	155/187 82.9%	12/181 6.6%
19th century	218/229 95.2%	6/117 5.1%

TABLE 5

Distance (Number of Words) between Antecedent and Relativizer

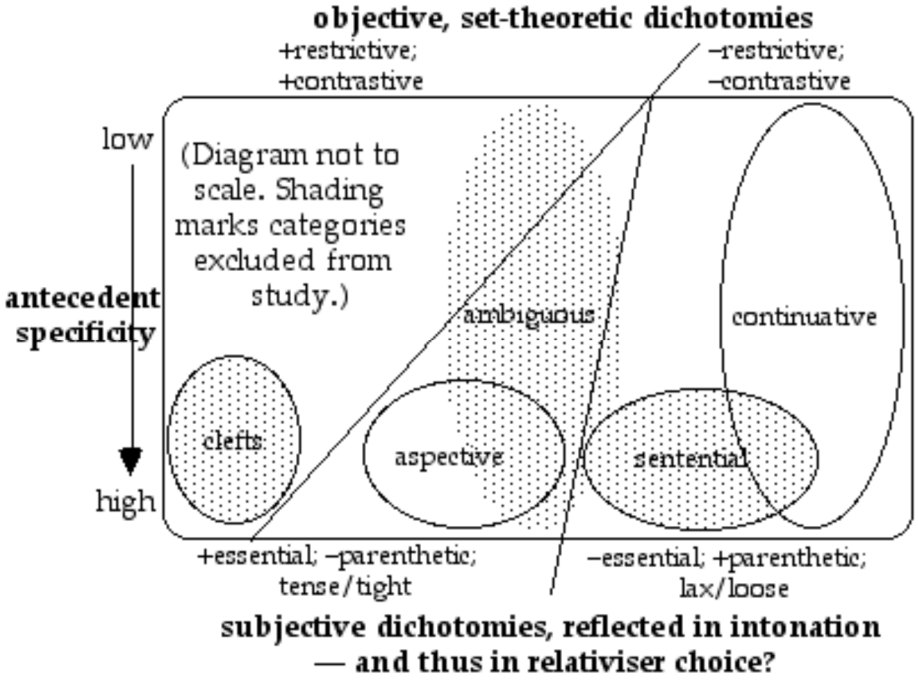
	Restrictive			Non-restrictive		
	≤ 3	4-9	≥ 10	≤ 3	4-9	≥ 10
17th century	129 (97.7%)	1	2	35 (61.4%)	15	7
18th century	182 (97.3%)	5	0	126 (69.6%)	38	17
19th century	217 (94.8%)	8	4	75 (64.1%)	24	18

TABLE 6

Classification of Seventeenth-Century Adnominal Relatives (Revised)

Coded as	N	percent
Restrictive	137	51.9%
Restrictive/aspective	3	1.1%
Aspective	27	10.2%
Aspective/non-restrictive	11	4.2%
Non-restrictive	64	24.2%
Non-restrictive/continuative	0	0.0%
Continuative	7	2.7%
Ambiguous between adjacent	5	1.9%
Ambiguous between non-adjacent	7	2.7%
Unclear	3	1.1%
Total	264	100.0%

Figure 1: Sigley's Diagram of Relative Types (1997: Fig. 5.2.1; reproduced by permission of the author)



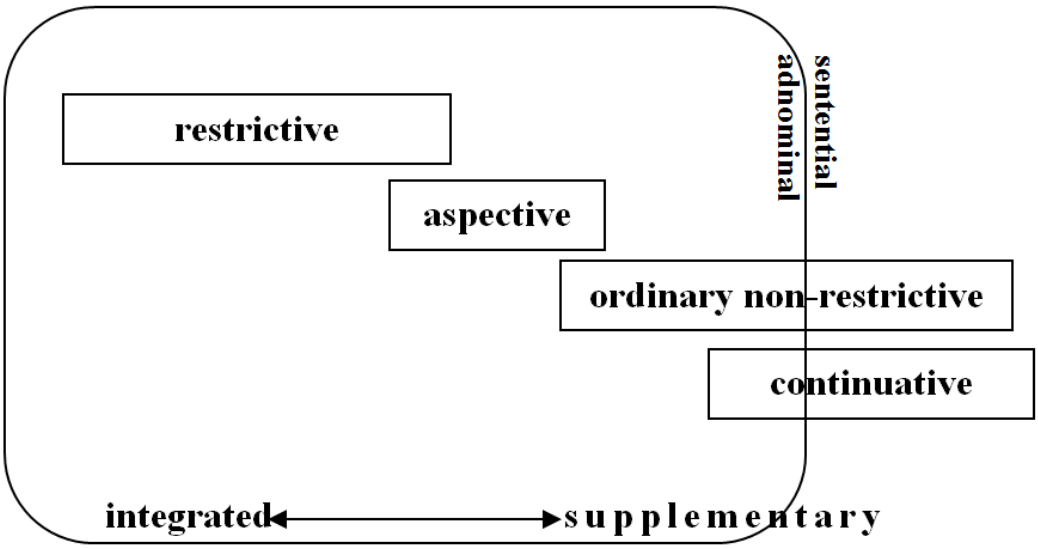


Figure 2: A Simple Gradient Model of Relatives

Figure 3: Original Classification of Seventeenth-Century Adnominal Relatives

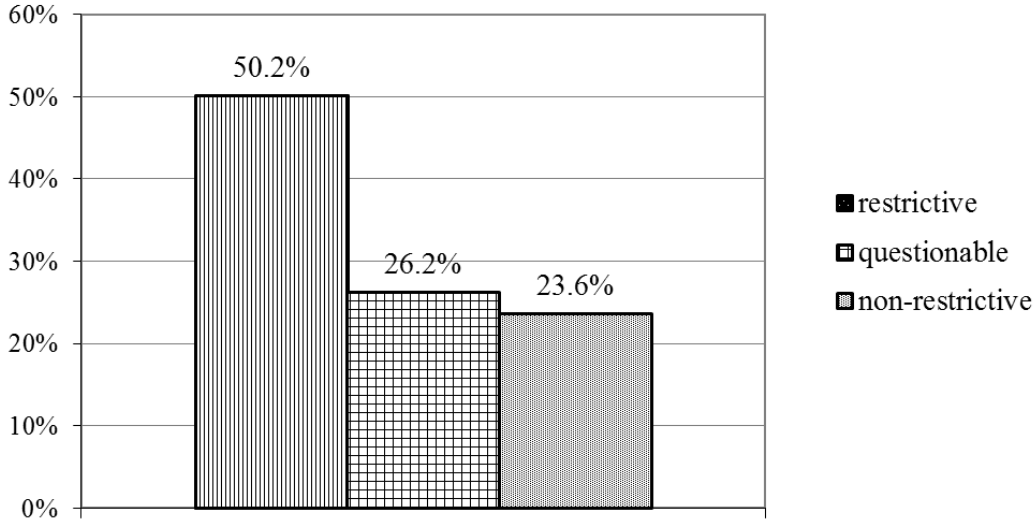


Figure 4: Revised Classification of Seventeenth-Century Adnominal Relatives