Analytical Tools for Toponymy: Their Application to Scottish Hydronymy

Jacob King MSc.

A Thesis Submitted for the Degree of Doctor of Philosophy

University of Edinburgh

2008

I declare that this thesis is entirely my own work.

Jacob King

Abstract

It has long been observed that there is a correlation between the physical qualities of a watercourse and the linguistic qualities of its name; for instance, of two river-names, one having the linguistic quality of *river* as its generic element, and one having *burn*, one would expect the *river* to be the longer of the two. Until now, a phenomenon such as this had never been formally quantified. The primary focus of this thesis is to create, within a Scottish context, a methodology for elucidating the relationship between various qualities of hydronyms and the qualities of the watercourses they represent. The area of study includes every catchment area which falls into the sea from the River Forth, round the east coast of Scotland, up to and including the Spey; also included is the east side of the River Leven / Loch Lomond catchment area. The linguistic strata investigated are: Early Celtic, P-Celtic, Gaelic and Scots.

In the first half of the introduction scholarly approaches to toponymy are discussed, in a Scottish and hydronymic context, from the inception of toponymy as a discipline up to the present day; the capabilities and limitations of these approaches are taken into consideration. In the second half the approaches taken in this thesis are outlined. The second chapter explains and justifies in more detail the methodology and calculus used in this thesis. The subsequent chapters examine the following linguistic components of a hydronym: generic elements, linguistic strata, semantics and phonological overlay. In each of these chapters the methodology is harnessed as an analytical tool to generate new findings for hydronymic research. The conclusion consists of a summary of the findings and a review of the performance of the calculus.

It emerges that these analytical tools are of use to the field of toponymy in two ways. Firstly, they formalise and challenge previously unquantified statements made in the field of toponymy. Secondly, they elucidate hitherto unnoticed phenomena. It is suggested that in the future this methodology be applied to other datasets (particularly hill-names) and to other regions in Scotland and the world at large.

Contents

List of Figures

List of Tables

Preamble

0.1 List of Abbreviations

G	Gaelic
OC	Old / Early Celtic
P	P-Celtic
Sc	Scots
SSE	Standard Scots English
W	Welsh
OI	Old Irish

Y	Generic element
X	Specific element
P	Any non-hydronymic element
H	Any hydronymic element
PN	place-name (non-hydronymic)
RN	river-name
AOS	Area of study
OS	Ordnance Survey

For example ‘YP of X’ means a non-hydronymic generic element of a specific element, such as ‘Mill of Redford’.

‘R’ as used in this thesis is not an abbreviation as such, but the name of a program (see page A.2 on page 242).

For abbreviations used for old forms in the database, see page 263.

0.2 List of Terms

The following terminology is used throughout this thesis:

watercourse:	any running body of water
water-basin:	any static basin of water
tributary:	any watercourse which flows into another watercourse
superordinate watercourse:	any watercourse which has a tributary
hydronymicon:	the group of lexical items which occur in RNs

0.3 Style Sheet

This thesis follows the MHRA style. Further to this the following conventions have been used:

- A lemma or root is in italics, e.g. 'Sc *rough* and G *garbh* are synonymous in meaning'.
- The language abbreviations are only used with lemma and not in general discussion, e.g. 'The occurrence of G *dubh*, 'black' as opposed to Sc *black* shows this area was until recently an area saturated with Gaelic speakers, with Scots the minority language until recently'. Occasionally the abbreviations are used in tables for economy of space.
- A root and meaning are always in the following format: G *dubh*, 'black', except in cases where the meaning is not the word itself, but a type, e.g. G *Bhaltair*, personal name.
- The form *na* will be used to represent 'of the' in Gaelic in cases where a particular name is not used, e.g. 'Names of the type *Allt na* X are common' could represent Allt nam Bothan etc.
- Where a specific RN is mentioned which occurs more than once in the database, its *id* number is mentioned after the name in brackets. An *id* number represents a unique identification number so it can always be identified within the database, e.g. Allt Dubh (2664) but simply: River Tay.
- Old forms are in brackets after the name in the format: data, source, reference. Where there is more than one, the forms are divided by a semi-colon, (which is rendered into a carriage return on the CD-ROM). The actual form is in italics, e.g. *Lochs of Achlee* (c. 1591 Pont map 10); *Loches of Auchlie* (c. 1750 Macf. V1 p70).
- Technical terms used only within this thesis are in italics, e.g. *geogscore*, *specelem*.

0.4 Acknowledgments

Special thanks to the following: My supervisors Ian Fraser, Prof. William Gillies and Dr Doreen Waugh; Simon Taylor for all his suggestions help and guidance over the years and Chris Yocum for all the technical support. Thanks to all the following other people for their suggestions, data and wisdom: Abigail Burnyeat, Graham Isaac, David King, Emily Lyle, Bill Nicolaisen, Oliver Padel, Peadar Morgan and Richard Cox. Thanks to the Norwich Town Close Estate Charity and Marjorie King for much valuable financial assistance, without which this thesis would never have been written. Finally I would like to express my thanks to my wife, Rachael King, for her patience and support.

Part I

Discussion

Chapter 1

Introduction

The purpose of this chapter is twofold. The first half contains a discussion of scholarly approaches to toponymy in a Scottish and hydronymic context, from the inception of toponymy as a discipline up to the present day. The shortcomings and advantages of these approaches will be discussed. Since the primary focus of this thesis is to create a methodology for studying the relationship between hydronyms and the watercourses they represent, the second half outlines the approaches adopted in this thesis.

1.1 A Brief History of Representational Approaches to British Toponymy

This section of the introduction gives special emphasis to the following phenomena pertinent to this thesis:

1. A representational approach to Scottish toponymy: This is defined for the purposes here as any which seeks to present data in a form that is not purely discursive, in other words, to portray the data in some sort of visual form, e.g. diagram, map, table or graph.

2. Classification systems: putting data into discrete groups, e.g. labelling 'specific element + generic element' forms with one arbitrary label and 'generic element + generic element' forms with another.

3. The observance of the phenomenon that the relationship between a place and its name obeys certain principles not detectable by only studying a single or limited number of place-names.

W. F. Skene may have been the first scholar to put into print points 1 and 3 above. Firstly, he compiled a three-page table containing a column of varying Scottish place-name elements, and a row of each Scottish county, with the amount of times the element occurs in each county.[1] Secondly, he

[1] W. F. Skene, *The Four Ancient Books of Wales* (Edinburgh, 1868), p. 162-164.

may have been the first person to have put into print the following concept: “When the names of places are applied to purely natural objects, such as rivers, mountains, etc., which remain unchanged by the hand of man, the names applied by the original inhabitants are usually adopted by their successors, though speaking a different language.”[2]

It would be appropriate here to mention a similar statement made by Johnston in 1934: “It will be found in Scotland, as in any other country, that the oldest place-names, the names which, like the hard granite, best resist weathering, are those of large rivers, mountains, promontories and islands.”[3] In the intervening half century little work had been done to investigate this phenomenon. Throughout the rest of this thesis, occasional ‘fuzzy’ statements such as these will be mentioned where the relevant phenomenon is being discussed.

Ekwall’s *English River-Names*[4] deserves mention at this point. Although not concerned with Scotland, the book does deal with English Celtic RNs, and in the introduction he discusses a number of pertinent topics. Under his section on pre-English names, he classifies Celtic RNs according to their semantic qualities, although he gives only a few examples for each group, and uses only uncompounded names.[5] Also of interest is his ‘Distribution of the various Types of Names’, which lists by county the main pre-English names, Early English names, Scandinavian names and back formations. In this case no real effort is made to analyse these data, beyond pointing out a few discrepancies as ‘noteworthy’.[6]

Enwau Afonydd a Nentydd Cymru,[7] meaning ‘The River and Stream Names of Wales’, is also a pertinent resource for hydronymy. The gazetteer is a useful resource in comparing Scottish and Welsh hydronyms.

In 1954 George R. Stewart published ‘A Classification of Place Names’.[8] This article is not a survey of names but is rather a discussion of a classification of different types of American names predominantly of English origin. The article is one of the first to recognise the virtue of classification of place-names by their meaning. The author offers nine classes, with sub-classes, offering representative examples for each class. Although the classes do not compare exactly with the semantic classifications used later in this thesis, his classes will be mentioned where they coincide. This work was expanded and refined in his 1975 work ‘Names on the Globe’.[9] The first part of this work discusses the philosophy behind naming systems, and expands on the sub-classes mentioned in the previous work. The second section gives examples from around the world and throughout history. Stewart was one of the first scholars to attempt to formalise the contexts within which names are coined and describe what would now be called the psycho-linguistic processes which underpin place-naming.

---

[2] Skene, *The Four Ancient Books of Wales*, p. 147.
[3] J. B. Johnston, *The Place Names of Scotland* (London, 1934), p. xiii.
[4] Eilert Ekwall, *English River-Names* (Oxford, 1928).
[5] Ibid., p. liv.
[6] Ibid., p. lvi.
[7] R. J. Thomas, *Enwau Afonydd a Nentydd Cymru* (Caerdydd, 1938).
[8] George R. Stewart, ‘A Classification of Place Names’, *Names* II no i (1954).
[9] George R. Stewart, *Names on the Globe* (New York, 1975).

The first attempts at representing Scottish toponymic data visually were by Kenneth Jackson in three maps, 'Distribution of the Place-Name Element Pit-',[10] 'Distribution of Certain P-Celtic Place-Name Elements Other than Pit-',[11] and 'Map of British River-Names'.[12] The first two maps were point-type distribution maps created on an analogy with archaeological distribution maps featured elsewhere in the book. These maps rendered the data in a way easily appreciable to the eye, combining maps and toponymic data. This approach has much more information than a sentence such as "aber... North of the Forth... is common on the east side as far as the Spey".[13] As for the 'Map of British River-Names', despite some possible criticisms about the extent and layout of this map, it was for its time an excellent visual representation of the different zones between English and the Celtic languages in England.

Worthy of mention here are two articles by W. F. H. Nicolaisen, which are also discussed in more detail below. The first is 'The Semantic Structure of Scottish Hydronymy'.[14] In this article, the author classifies a number of Scottish RNs according to their semantic content and in the final section compares this information against the linguistic strata and morphological structures to which these names belong. Whilst there are various toponymic classifications which had been previously created, this was the first time a classification had been applied to Scottish toponyms in any sort of systematic manner, unless one counts W. F. Skene's table mentioned above. Whilst this is a valuable and overlooked piece of research it necessarily has limitations, which are discussed below.

The last chapter in Nicolaisen's *Scottish Place Names*, 'Pre-Celtic Names',[15] discusses the phenomenon whereby larger watercourses have names from older linguistic strata compared to smaller ones. The author here discusses this in considerably more detail than the comments above by W. F. Skene and Johnston, though even in this article, due to space, no systematic attempt was made to investigate this topic in any detail. Nicolaisen was also perhaps one of the first people to realise the importance of gathering RN data according to how the rivers themselves interact. The hierarchical nature of rivers necessarily means that extra data should be recorded for each river, in terms of tributaries and parents, that has no equivalent for settlement names or natural features. Nicolaisen represents a small sample of this information as a schematic map;[16] this is discussed in more detail below.

In 1968 'A Preliminary Report on an Investigation into Pit Place-names'[17] was published; this was an important work and very relevant to this thesis in that it was written not by linguists but by geographers. It compared the precise location of various *Pit-* names with altitude and soil classification. The accompanying visual representations were not only distribution maps marked on maps showing

[10]Kenneth Jackson, *The Problem of the Picts* (1955), chap. The Pictish Language, p. 147.
[11]Ibid., p. 150.
[12]Kenneth Jackson, *Language And History In Early Britain* (Edinburgh, 1953), p. 220.
[13]William J. Watson, *The Celtic Place-Names of Scotland* (Edinburgh, 1993), p. 459.
[14]W. F. H. Nicolaisen, 'The Semantic Structure of Scottish Hydronymy', *Scottish Studies* 1 (1957), p. 211-240.
[15]W. F. H. Nicolaisen, *Scottish Place-Names*, 2nd edition (Edinburgh, 2001), p. 222-246.
[16]Ibid., p. 225.
[17]G. Whittington and J. A. Soulsby, 'Preliminary Report on Pit Place-Names', *Scottish Geographical Magazine* 84 (1968), p. 117-125.

altitude, but also graphs which displayed the data in a way instantly understandable to the eye. Although somewhat out of date now, in methodological terms this was an innovative and important work, but which unfortunately stands largely alone in its approach to place-name studies.

An often-overlooked article is ‘Studying the Place Names of Bernera’.[18] As the title suggests, this article is a survey of the place-names of Bernera. The names are divided into Gaelic and Norse names, and then for each of these languages a number of categories are imposed on the names by both meaning and syntactical structure. From the point of view of the present thesis, it seems that the author has combined the meaning and structure as being equivalent to one another, when this is not the case. This criticism aside, this is one of the few works which clearly derives from and builds upon Nicolaisen’s article on semantics discussed above.

Two years later an article appeared called ‘Place-name Analysis in the Geographical Study of the Rural Landscape of Wales’.[19] This created ‘a crude classification of field names’[20] with a basic semantic classification for field-names. This article is also noteworthy in that it contains a diagram which visually represents a number of different Welsh field-name generic elements by the altitude range in which the fields exist.[21]

The first systematic effort to gather place-names in the United Kingdom was of course part of the English Place Name Volumes.[22] This set about creating a classification system for usages of elements. Since Anglo-Saxon toponymy is somewhat structurally simpler than Celtic toponymy, however, it does not translate very well into Celtic toponymy. This point was noted by O. J. Padel,[23] who constructs a basic classification system for Cornish toponymic morphology.[24]

The Placenames of the Isle of Man[25] is a monumental seven-volume work equivalent to the English Place Name Surveys but covering the Isle of Man. Since Manx is a language closely related to Scots and Irish Gaelic, these volumes are useful for comparison of names. Some of the shortcomings in the classification system for usages of elements in the English Place Name Surveys are addressed here, but again, no attempt is made to analyse or categorise each of the names, the classification system is simply listed with a few examples.

Margaret Gelling’s and Anne Cole’s book, ‘The Landscape of Place-names’,[26] whilst concerned with English place-names, gives a detailed semantic analysis of various English place-name elements. Also of importance to this discussion is the methodology of displaying the places in pictorial format where they appeared in the landscape. This was a new approach and quantified for the first time the relationship between a place-name element and the topographical environment in which it could exist.

[18]Donald MacAuly, ‘Studying the Place Names of Bernera’, *Transactions of the Gaelic Society of Inverness* 47 (1970-72), p. 313-337.

[19]Colin Thomas, ‘Place-name Analysis in the Geographical Study of the Rural Landscape of Wales’, *Studia Celtica* 9 (1973-74), p. 299-318.

[20]Ibid., p. 303.

[21]Ibid., p. 317.

[22]Albert Smith, *English Place-Name Elements*, vol. 25-26 (Cambridge, 1956).

[23]O. J. Padel, *Cornish Place-Name Elements* (Cambridge, 1985), p. xiv.

[24]Ibid., p. xiv-xvii.

[25]George Broderick, *Placenames of the Isle of Man* (Tübingen, 2005).

[26]Gelling Margaret and Anne Cole, *The Landscape of Place-names* (Stamford, 2000).

Whilst the approach taken in this thesis is markedly different from that particular approach, they both have in common the fact that they compare geographical aspects of toponyms over a wide area in order to discern patterns not available by merely studying one particular area.

The Northern Ireland Place-Name Project, under the aegis of Dr Kay Muhr "was established with government funding in 1987 to research names of physical features and settlements on the Ordnance Survey 1:50,000 scale map. The project was to construct a computerised database of the names, giving their origin, meaning, and any historical or other relevant information."[27] In addition to this, seven volumes of place-name research have been generated covering a large area of Northern Ireland. The database, however, is not available online.

The Melville-Richards Archive or Archif Melville Richards[28] is now an online database of 328,778 records of Welsh place-names originally compiled by Melville-Richards himself on slips of paper last century. The information was derived from numerous sources over decades. In 2005 the database was put onto the web, under the supervision of Hywel Wyn Owen, and is a very useful resource for toponymists working in the field, not least because it documents many small settlements rather than focusing merely on ones with interesting names.

In the last decade or so many of these criticisms made above about Scottish toponymic research have been made invalid. Work carried out comprises on the one hand the Scottish Place Name Database and a number of published surveys on the other. The Scottish Place Name Database contains about 8000 names compiled predominantly from various studies made over the years. This is an important and relevant work, since it is stored in Microsoft Access, a simple relational database program, but due to various reasons funding for the continuation of the project on the scale it deserves has not been forthcoming and it is now continued now on a volunteer basis at time of print, primarily by Honorary Fellow Doreen Waugh, in conjunction with the Shetland Place Names Project.[29] As of 2007 the Royal Commission has agreed to host the database.

There have also been a group of excellent surveys;[30] all of which are comprehensive works cataloguing *every* name in a given area, with old forms gathered from manuscripts and names gathered from informants including precise coordinates and information about aspect and altitude and so on. An undertaking such as this thesis would have been much simpler if these surveys were extended to the whole of Scotland in the manner of the English Place Name Survey.

Whilst these works and work on the Scottish Place Name Database are obviously welcome, they can be problematic from the point of view of this study for a number of reasons:

1. Some of the areas of coverage stand outside the coverage of this thesis.

[27] `http://www.qub.ac.uk/schools/SchoolofLanguagesLiteraturesandPerformingArts/SubjectAreas/IrishandCelticStudies/Research/NorthernIrelandPlace-NameProject/2007`

[28] `http://www.e-gymraeg.co.uk/enwaulleoedd/amr/`

[29] `http://www.shetland-heritage.co.uk/amenitytrust/placenames/placenames.html`

[30] These are: Richard A. V. Cox, *The Gaelic Place-names of Carloway* (Isle of Lewis, 2002) (although the area covered means that no manuscript sources were used in this work); Simon Taylor and Gilbert Márkus, *The Place-Names of Fife*, vol. 1 (Donington, 2006); Adam Watson and Elizabeth Allan, *The Place Names of Upper Deeside* (Aberdeen, 1984); Roy Wentworth, *Gaelic place-names of Beinn Eighe National Nature Reserve* (Perth, 1999) and to a certain extent: A. MacDonald, *The Place-Names of West Lothian* (Edinburgh and London, 1941).

2. The database and surveys as they stand are not primarily concerned with hydronyms.

3. Including all data from these surveys can create an issue akin to the 'accident of survival' problem in archaeology. For instance, within Watson and Allan, *The Place Names of Upper Deeside*, valuable work has been done on gathering names of small features from local Gaelic-speaking informants, meaning that for almost every *burn* name there are a large number of variants, often with constructions such as *The Burn of* X. Whilst this work is of inestimable value, if these data were added 'as is' into my own database, it would create a slant towards a high degree of variation in that area and no other. One could theoretically end up with a distribution map, for instance, where *burn* names with the definite article were clustered round the Deeside area and nowhere else, which would only be because of the more detailed work done here than anywhere else. To remedy this, whilst these works have been used as secondary sources, the policy has been to generally disregard names sourced from informants which have not gained general currency in the area.

4. Whilst each work is internally consistent, there is no established classification system for toponymy. For example, there is no established method for giving grid coordinates or altitude for places and no established system for representing generic elements for place-names.

It should be stated that all the previous scholars worked within and extended the knowledge of their time as well as could be expected and are not responsible for limitations on technology or access of their time. On the whole, however, many toponymic works have a number of shortcomings which should be pointed out at this stage:

1. Before the time of the English Place-Name Survey, the concept of database storage and management was non-existent. The most consistent way for information to be stored was on slips of paper. Even in later times, with the English Place-Name Survey and various place-name dictionaries, the information was in a printed, alphabetical, discursive format. One manifestation of this was the lack of data accompanying distribution maps. There was no list of place-names and coordinates accompanying the maps, no way to find out which dot represented which place-name; the only way to check the data was to start from scratch oneself. Furthermore a place-name survey from which these maps and tables were derived was lacking.

2. Apart from the presentation of the database, the actual toponymic coverage was largely incomplete. Whilst total documentation of all place-names may not be practical, there has in Scotland not been the same degree of coverage as in England with the English Place-Name Survey. Previous Scottish surveys have tended to include mainly those names of particular antiquity or interest, or names of larger settlements or natural features. Manuscripts and old forms had not always been adequately documented.

3. Apart from some of those works mentioned above (which form a very small part of the whole corpus of work in this area) the majority of previous works have purely dealt with the linguis-

tic properties of a name. This can be expressed as: '...pre -twentieth-century, and much of the twentieth-century, onomastics has been predominantly diachronic, etymological in orientation'.[31] This is reflected in the concept of the place-names dictionary such as Johnston's *The Place Names of Scotland*, which strives to offer the meaning, etymology and linguistic stratum for each place-name entry.

Despite the relative paucity of work done in Scottish toponymy compared with its neighbours, the future looks to remedy some of this. Simon Taylor's work mentioned above is but one of five volumes covering the whole of Fife, and at time of writing work proceeds apace on the second volume at the University of Glasgow under the direction of Professor Thomas Clancy, with a PhD position, a research assistant and a large grant from the AHRC. This exciting opportunity promises to yield a further book: 'Gaelic in medieval Scotland: the evidence of names'. To quote the website: "This is a long-awaited boost for name scholarship in Scotland, and the Department of Celtic at Glasgow, along with colleagues in other departments, hope to be able to build on this in the future."[32]

Within the methodological context, there are signs that the toponymic community has begun to realise that it may now be time to move away from the approach in point 3 above:

> There continues... a strong tradition of work on name etymology and naming systems; but such work... shows increased awareness of interaction with wider historical concerns and of psychological and social motivations for name giving.[33]

1.2 Approaches of this Thesis

1.2.1 Principles

The primary focus of this thesis is to create, within a Scottish context, a more rigorous methodology for elucidating the relationship between various qualities of hydronyms and the watercourses they represent, that is, to explore in greater detail the phenomenon mentioned by W. F. Skene and Johnston and discussed by Nicolaisen above. This approach aims to resolve some of the shortcomings listed above in the following ways:

1. A database has been created which contains consistent physical data about rivers and linguistic data about the RNs.

2. In this survey an attempt has been made to include *every* RN evidenced in the coverage area (Within the caveats stated in point 3 on page 8 above.) This creates a somewhat different picture than previous surveys. For example, as far as semantic classifications are concerned there was a certain bias towards counting the larger watercourses, which slanted which semantic groups

[31] John M. Anderson, *The Grammar of Names* (Oxford, 2008), p. 88.

[32] `http://www.gla.ac.uk:443/newsdesk/newsletter/details.cfm?id=3949&issue=273&category=catresearch`

[33] Ibid.

were chosen. The cause of this was probably by influence from English place-name studies, where, because no Celtic languages have been spoken for a long time in most of England, the remaining Celtic RNs are all the older names, attached to larger rivers, whilst the smaller names are English in origin. Thus discussions of Celtic RNs in England are largely restricted to names such as the Dee, the Thames and the Stour and so on.[34] An example of this can be shown negatively: within this database, by far the commonest single specific element in RNs is G *coire*, 'circular hollow', yet this has not been recognised before. This is probably for two reasons, firstly because the average length is 2.7 km, and thus the watercourses are considered insignificant. Secondly, the names themselves are largely unproblematic in terms of etymologisation, thus they do not receive as much 'air time' in terms of discussion. This is a general issue with small, insignificant, linguistically transparent names.

3. This database seeks to gather more varied and specific physical information about each river; such factors include, altitude, soil fertility, location, length and so on.

1.2.2 Practicalities

1.2.2.1 Extent of the Database

Figure 1.1 on page 11 shows the coverage area comprising the Forth catchment area, and every catchment area which falls into the sea from that point, round the east coast of Scotland, and up to and including the Spey. Also included is the River Leven / Loch Lomond catchment area, except watercourses falling into the west side of Loch Lomond. Obviously, it would have been desirable to study the whole of Scotland, or even Britain, but time pressures prevented this. It was felt that a large representative area would still yield meaningful results. The area was chosen for a number of reasons:

1. Within the area there is a good cross-section of names from each linguistic stratum (except Norse). It has a large area of overlap with 'Pictland' as defined by the distribution of Pictish monuments[35] or distribution of *Pit*-names.[36]

2. Since it contains a large stretch of the 'Highland Line', there is also a large section of Gaelic and Scots names, with names both anglicised and non-anglicised, creating interesting opportunities for research.

3. It does not include areas of any particular Norse influence (due to the way the score system works, explained in the next chapter, it would be better to have none at all rather than a handful).

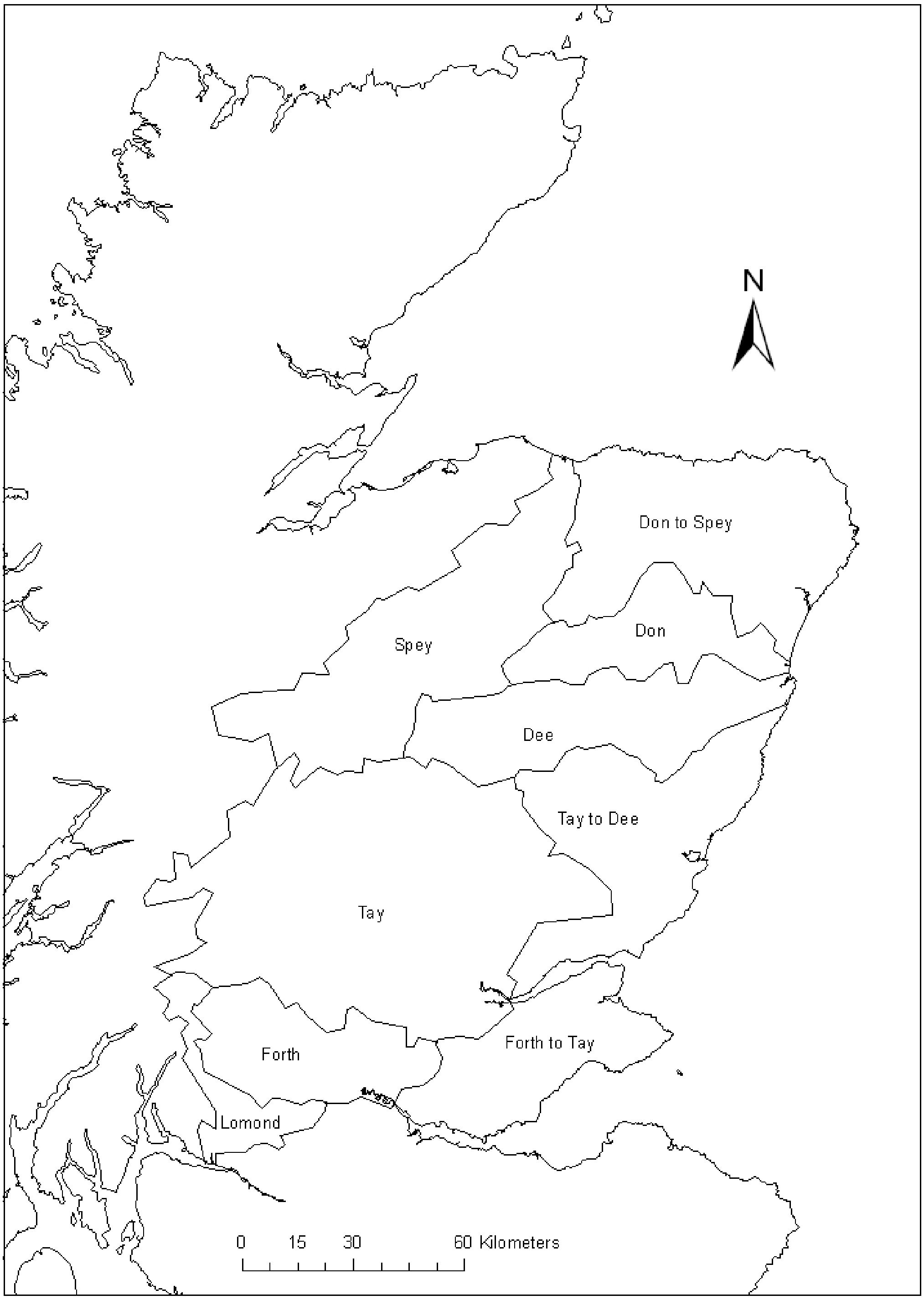
---

[34] For example, MacDonald, *The Place-Names of West Lothian*, p. 76-79.

[35] Peter B. G. MacNeill and Hector L. MacQueen, *Atlas of Scottish History to 1707* (Edinburgh, 1996), p. 53-56.

[36] Ibid., p. 51.

Figure 1.1: Area of Study Divided by Catchment Areas



1.2.2.2 Structure of the Database

The database has two main sections. As mentioned above this is stored in MySQL. The first section displays each RN hierarchically, that is, it gives a visual representation of which watercourses are tributaries of which other watercourse, and which are parents. This is called *hieralist*, a section is shown in table 1.1. See C.2 on page 262 for more details on this table.

Table 1.1: Example of Hierarchically Arranged RNs

1	2	3	4
Burn of Boyne			
	Bicker Burn		
	Park Burn		
	Burn of Corncairn		
	Beg Burn		
	Burn of Badenyouchers		
		Rae Stripe	
		Burn of Inverkindling	
Burn of Durn			
	Loch of Soy		
Burn of Fordyce			
Scattery Burn			
Burn of Cullen			
	Claypots Burn		
	Glen Burn		
		Burn of Darbreich	
			Badentarget Burn
			Burn Levenit
			Stripe of Holes
			Back Burn

In this table, the watercourses in column 1 flow into the sea, the watercourses in column 2 flow into the watercourses 'next up' in the corresponding cell to the left. Thus, for example: 'Sea' <Burn of Boyne <Burn of Badenyouchers <Burn of Inverkindling. This creates a simple diagrammatic representation of the watercourse hierarchy, that can be recognised by database programs, and can be easily manipulated and updated. Where the symbol '<' appears after a RN, it means the river is part of the same watercourse as its parent watercourse, but has a different name. In many cases, the exact point on a watercourse where one name stops and another begins is either not known or not exactly delineated. In these cases a 4 point grid reference is usually given, representing a larger area of one square km. Static basins of water (e.g. lochs) are treated in the same way in the hierarchy. For some of the larger lochs, the tributaries falling into the loch are divided by cardinal directions (e.g. East side, West side); obviously these references do not represent watercourses. It should also be stressed that this has been done for the whole of the database, not only a representative section.

The second part of this schema is a database containing every RN with various bits of information. This is called *list2* in the database (see C.1 on page 257). It was necessary to build this gazetteer almost from scratch for a number of reasons. Firstly the gazetteer is a compilation of all RNs attested, not just those on present OS maps (which is used in the hierarchy), such as names from old forms, reconstructed names, and names from gazetteers. Secondly the OS data are not internally consistent or available in a format to be used in toponymy. For instance, the OS data marks each name as it is written on OS maps, so any large watercourse, such as the River Tay is in the gazetteer twice, because the name is written on OS maps twice. The grid coordinates also denote where this name is written on the map, an arbitrary place on (and sometimes off) the watercourse. (In this gazetteer the coordinates are for the lowest point of the watercourse explained below.) Finally, the geographical data gathered for each watercourse is not available in a way that can be applied to the gazetteer; for instance the length of watercourses or extent of altitude of each river is not gathered as a discrete unit of data by OS.

The factors broadly divide into two sections: Linguistic information about the RN and geographical / geological information about the watercourse. The following sections do not discuss the specific column names in the database, but instead the general types of information. Consult appendix C.1 on page 257 for more detailed information as to how these relate to the database. The following geographical factors have been gathered:

1.2.3 Geographical Factors

Catchment area The river basin system or catchment area in which the watercourse is located. This is used rather than parish or county, because many watercourses do not sit wholly inside a region, but can flow from one into another, or can form a boundary.

Location Precise grid location of the watercourse. OS grid references where given to within 100m of the lowest part of a watercourse, or the outlet for a loch, where one watercourse flows into another, or the sea.

Soil class / altitude The Soil Survey of Scotland[37] has mapped the whole of Scotland by their system of soil classification. This measures not only information about the soil, but also the altitude. For each watercourse, a list was compiled of all the soil classes through which it ran; from this the information about the minimum and maximum altitude ranges was derived.

Length For watercourses, this represents the length in km; for water-basins, this represents the circumference. It is rounded to the nearest half km, with anything under 0.5 km rounded off to that length. The measurements express the extent, that is, not every meander on a small burn is measured precisely, nor is every small promontory in a loch. These data were extracted from digital and hard copy versions of the 1:250 000 scale Explorer OS maps.

Relation to other features A list of any pertinent non-hydronymic place-names.

Position The position in the hierarchical list, i.e. how far removed a watercourse is from the sea as corresponding to the column number in table 1.1.

Number of named tributaries The number of named tributaries which ultimately flow into the given watercourse, not simply those which flow into it.

[37] E. L. Birse, *Soil Survey of Scotland* (Aberdeen, 1970).

1.2.4 Linguistic Factors

Name This is the OS form if there is one (e.g. 'Caochan na Bruaich'). If the name is only known from old forms, the name as mentioned in the manuscript is given (e.g. 'B: of Aldchash', mentioned only from a Pont map). If the name is inferred indirectly and not extant, it is reconstructed in the least meddlesome way (e.g. 'hadden', inferred from Inverhadden).

Generic element Information about the generic element is stored here. This includes the identity of the element itself (e.g. whether *burn* or *stripe*), the stratum from which it derives (e.g. *burn* is a Scots generic element) and syntactic position (e.g. whether X *Burn* or *Burn of* X). This is done regardless of the specific spelling, so that for instance, Garrauld, Garvald and Garbh Allt all have the same information. This is discussed in more detail in chapter 3 on page 49.

Specific Element or Linguistic Stratum This denotes the linguistic stratum from which the specific element originally derives. Discussion of both the linguistic strata chosen and how a stratum is assigned to a specific element are explained on section 5.1.1 on page 185.

Semantics This takes semantic information about the specific element, and places it into one of a number of classes: e.g. colour, manner etc. This is discussed in much more detail in section 4.3 on page 133.

Phonology / Orthography The stratum by which the RN is most phonologically or orthographically influenced. Where this is different from the linguistic stratum, it generally means the name has undergone phonological change under influence of a language other than that from which it originated or the name is not written in the 'correct' orthography for its linguistic stratum.

Etymology The derivation of the specific element. The linguistic stratum of the name *may* differ from that mentioned in the specific element in cases where the watercourse relates to a settlement with an older name, although the RN is from a later stratum. Such an example is Burn of Ogilvie (4991) which is probably a Scots name, but relates to a settlement name which is probably P-Celtic or Early Gaelic in origin.[38]

Old forms A list of old forms from manuscripts, maps etc. The phrase 'old form' is used in any case where the RN is not from modern OS maps. Thus a name from 1880 is considered an old form.

Other names A cross reference to any other names by which the watercourse is known; these other names also have their own entry.

Other generic or pleonastic elements A note is made where the old forms have a variation in generic element but not specific element, or the name has a pleonastic element.

Since the main database (*list2*), contains names not present in OS maps, it necessarily contains more entries than *hieralist*. Where a watercourse has two names, both names are entered and the geographical

[38] Watson, *The Celtic Place-Names of Scotland*, p. 378.

information is the same for them. This includes any watercourses which can be reconstructed, so, for example, the watercourse commonly known as Abernyte Burn also has an entry of simply ‘nyte’; since all Aber- names derive from watercourses, it can be confidently thought that there was once a watercourse called ‘nyte’, with an unknown generic element. In these cases, the entries are linked by the columns *otherid* which refers to the id of the other RN, and *othernames* which refers to any other names in the database (this column is technically redundant but is added for ease of use).

Where a name derives from an old form and its exact location is unknown, the geographical information is filled out to the greatest accuracy possible, but in many cases has many blank fields. All naturally formed watercourses and basins have been included in the database; not included are: man-made watercourses and basins such as reservoirs, dams, lades and stanks. Bogs have on the whole been excluded, unless they are shown as ‘flowing’ on maps. Where a reservoir has altered the qualities of a loch (e.g. by making it larger), the original qualities have been entered in the database.

It was decided that the visual approach taken by Nicolaisen[39] would not be practical for this project for two reasons. Firstly it would become cumbersome to illustrate over 6000 rivers like this, and secondly, this information stored in this way cannot be interpreted or manipulated with a database program (altitude, coordinates etc.).

1.2.5 Strengths of this Approach

1. MySQL is a very powerful program which can easily and quickly build queries limited only by the structure of the database. Being able to do this obviously is a much more versatile method of accessing information than thumbing through the pages of a book or reading a printed table. In addition, statistical operations can be performed very easily, such as counting occurrences, or creating averages of data and other more complex procedures. For more detailed notes on MySQL and SQL in general see section A.1 on page 241 in the appendix.

2. MySQL can connect with a number of programs which can manipulate the data in a number of versatile ways. For the purposes relevant to this thesis, it connects to certain Geographical Information Systems (GIS) programs which can create maps and map-based diagrams very simply and quickly. For instance it is possible to create distribution maps ‘on the fly’, with any data the database contains. MySQL can also connect with the statistical environment called ‘R’ (see section A.2 on page 242 in the appendix for more details) which can create graphs and other visual representations of the data. In addition R can perform statistical analysis of the data, which is explained more fully in the next chapter.

3. It will be shown in following chapters that this ability to access and manipulate data will, on the one hand, solve a number of outstanding questions. For example, what was the original generic element for RNs only surviving as *Aber*- names, such as Abercairny? On the other hand it is

[39] Nicolaisen, *Scottish Place-Names*, p. 225.

hoped that by comparing the various factors mentioned above, previously unnoticed problems or phenomena will come to light.

1.2.6 Potential Drawbacks of this Approach

1. The storage of data like this does not lend itself well to indeterminacy, something common in toponymy. The main manifestation of this is a RN which has two equally likely but distinct etymologies. The field 'etymology' should only have one suggestion in it, since to have two etymologies would require two separate entries, which would distort the data. A number of things have been put into the database to offset these issues: In the instances of an uncertain etymology, the q column is an attempt to express the relative certainty or otherwise of an etymology. Likewise, the columns concerning phonology and orthography are attempts to clearly represent colouring of other strata onto a place-name, not representable in simply assigning a given stratum an integer. In the case of linguistic strata, intermediate categories have been proposed between P-Celtic and Gaelic, and Old Celtic and P-Celtic to prevent over-simplification (see page 187 for more discussion of this). All the problems are essentially that of turning a discussion of the possibilities of the derivation of a place-name into series of binary oppositions

2. One RN may have more units of data associated with it in one way than another RN. For example, some RNs have variant generic elements in old forms, whilst this has been allowed for in the database, presently this can only support one variant, but of course various old forms may have many different types. One solution would be to have extra columns for each variant, but this would create more and more columns which would be rarely used, and would make the whole database more cumbersome.

3. This approach is on a macro scale; that is, rather than investigating single derivations, or making very detailed surveys of particular parishes, a large area has been taken in order to perceive toponymic phenomena over a large area in order to detect patterns of nomenclature. Needless to say, a macro approach is based upon many micro pieces of information, and it has not been possible, for reasons of time, to gather thorough data for all areas, especially where little place-name work has been done, and this could be viewed as a drawback.

1.3 Concluding Remarks

This introduction has attempted to uncover some shortcomings in toponymic research to date, both in terms of coverage and approach. It has also outlined the basics of the methodology that will be employed throughout this thesis. In the next chapter I aim to show how this new methodology based on digital resources can contribute to knowledge in the field of hydronymy, on both a practical and theoretical platform.

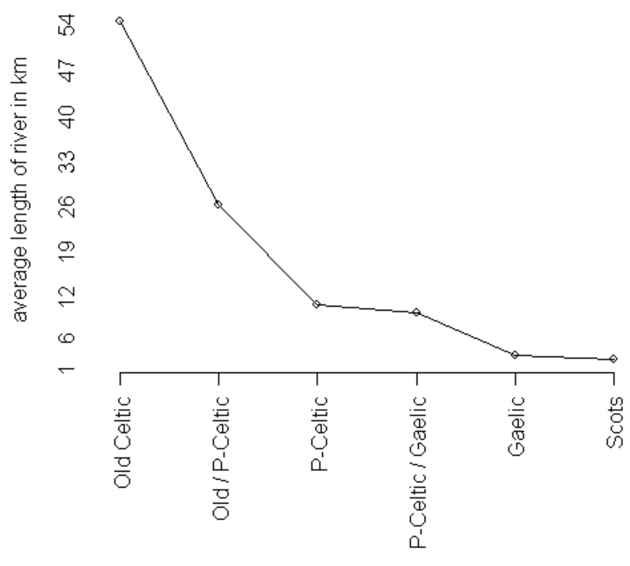
Chapter 2

Methodology

2.1 Introduction

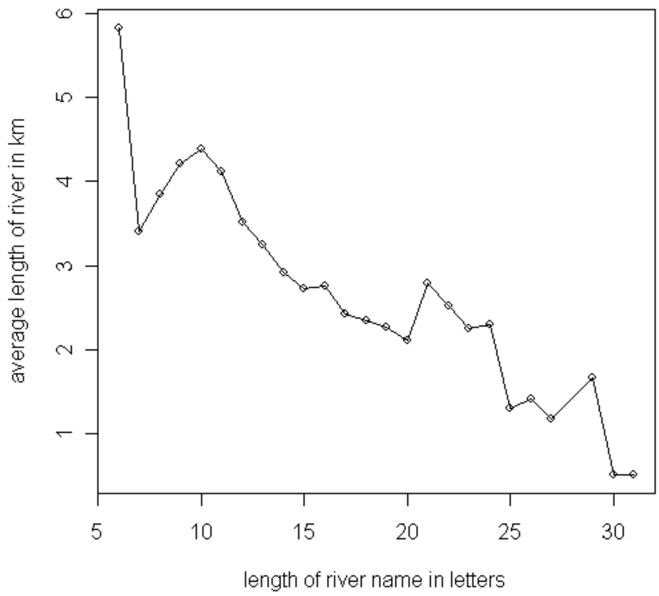
In the previous section, the concept that there is a relationship between a RN and its physical features was discussed. In the simplest form, there is a notion that the longer a river is, the more likely it is to have an older name, or even more basically that there is an inverse proportion between the length of a RN (in letters) and the length of its watercourse. Before the more complex testing, it would be desirable to see if these basic assumptions are true. Whilst 'age of name' is very difficult to quantify, it is possible to get the average length in km for watercourses in each linguistic stratum. This is done in figure 2.1. The linguistic strata used in this graph are discussed in more detail in section 5.1.1 on page 185.

Figure 2.1: Relationship between the Length of a Watercourse and its Linguistic Origins



The second assumption is much easier to test, as can be seen in the figure 2.2.

Figure 2.2: Relationship between the Length of a RN in Letters and the Length of its Watercourse



Whilst these extremely simplistic graphs do show a clear relationship, it is of course true that other factors besides 'length of name' and 'length of watercourse' contribute to this phenomenon. The aim of this section is to discuss what these factors might be, and then go on to create a methodology whereby these factors can be meaningfully interpreted.

The underlying philosophy within this thesis aims to more accurately quantify and test the assertions made discursively by the scholars discussed in the introduction. The steps involved in this are:

1. Justifying what data it is appropriate or possible to gather.

2. Gathering quantifiable data: This means gathering data which is comparable to other data in the set. An example of this is to measure the length of each watercourse as opposed to a discursive description of the watercourses, which might call the watercourses, 'considerable' or 'small', for example '...the majority of old names of rivers are pre-English... while names of small streams or brooks are to a great extent English or Scandinavian'.[1]

3. Converting the data into comparable formats with other data: For example, the watercourses are measured in km, yet due to the Soil Classification Survey, the altitude is measured in hundreds of feet. Beyond this, a system needs to be set up to compare completely unlike sets of data, i.e. geographical qualities and linguistic qualities, which effectively involves converting all data into a numerical range.

[1]Eilert Ekwall, *Introduction to the Survey of English Place-Names* (Cambridge, 1925), chap. The Celtic Element, p. 24.

4. Utilising the converted data in a meaningful way so that one can make predictions and make suggestions about missing sections of the data.

These steps form a basis for the rest of this chapter. For reasons which will be made clear below, the geographical data are discussed first, followed by an explanation of the linguistic data. Geographical data are those factors about the watercourse itself, regardless of name. The linguistic data are those factors inherent in the name, regardless of the physicality of the watercourse.

2.2 Geographical data

2.2.1 Justification of Factors Used

In this section I attempt to show a correlation between various factors, for example, between length of river and generic element. All the geographical factors mentioned below were mentioned in the introduction in the discussions and observations of scholars, which is used as a starting point for looking for factors. In this section, 'probable cause' that the factor is meaningful enough to be worthy of inclusion is sought for. This essentially means the factor must introduce new information into the system (i.e. the data are not predictable from other data, making it redundant) and it must be relevant. Also discussed in this section is the notion of 'direction of the data', this means that for the factors discussed, it must be ascertained at what end of either spectrum the score should run in. For instance, it is intuitively clear that a longer river is more 'important' than a short one, but it is not necessarily so that a river at a lower altitude is more or less 'important' than one at a higher altitude.

2.2.1.1 Km: Extent of the Watercourse

Figure 2.2 at the beginning of this chapter displays a correspondence between the length of a watercourse and the linguistic stratum from which the specific element of the name derives. Intuitively it is the factor most associated with importance. In figure 2.3 on page 20 the linguistic stratum is compared with the length of watercourse, showing a clear, direct correspondence.

2.2.1.2 Nont: Number of Named Tributaries

The number of named tributaries represents the number of all watercourses in the database (i.e. ones with names) in the catchment area of each watercourse. It does not simply measure the number of watercourses which fall into each watercourse. This was done because it more accurately reflects the size of a catchment area of a given watercourse. This obviously correlates with the length of a watercourse, as can be seen in figure 2.4 on page 20.

The peaks and troughs in this graph show that the correlation is not so strong that *nont* should not be considered redundant data. *Nont* also factors in watercourses which are in themselves short, but through which many watercourses flow (such as Abhainn Ghaig). This is an important factor, and one which 'length of watercourse' alone does not account for, neither does 'length of watercourse'

Figure 2.3: Relationship between the Length of a Watercourse in km and its Linguistic Stratum

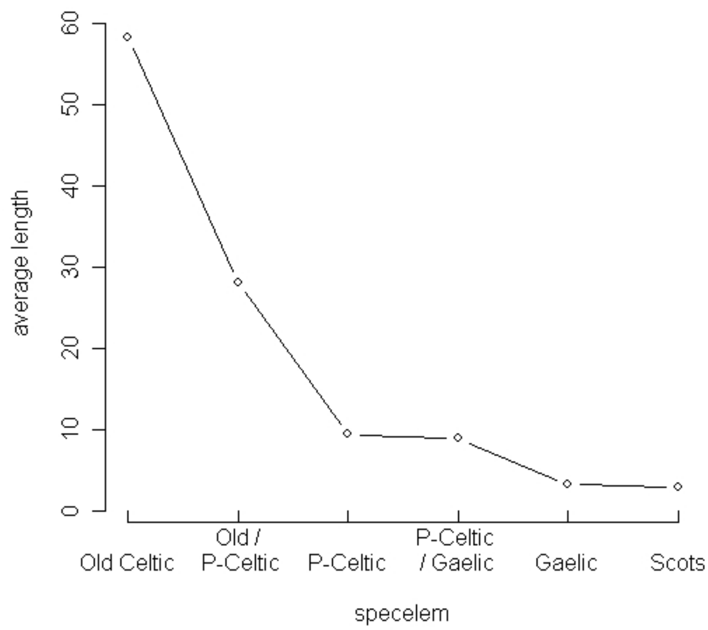
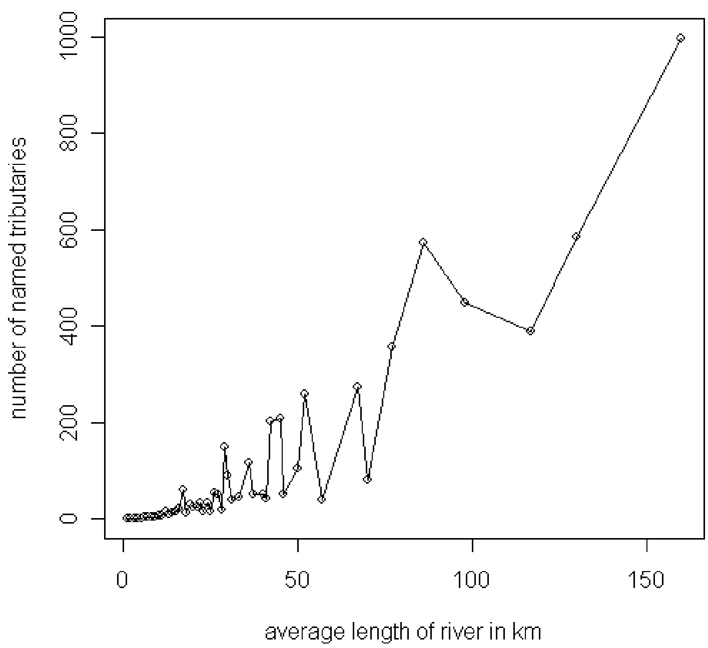


Figure 2.4: Relationship between the Number of Named Tributaries of a Watercourse and its Average Length in km

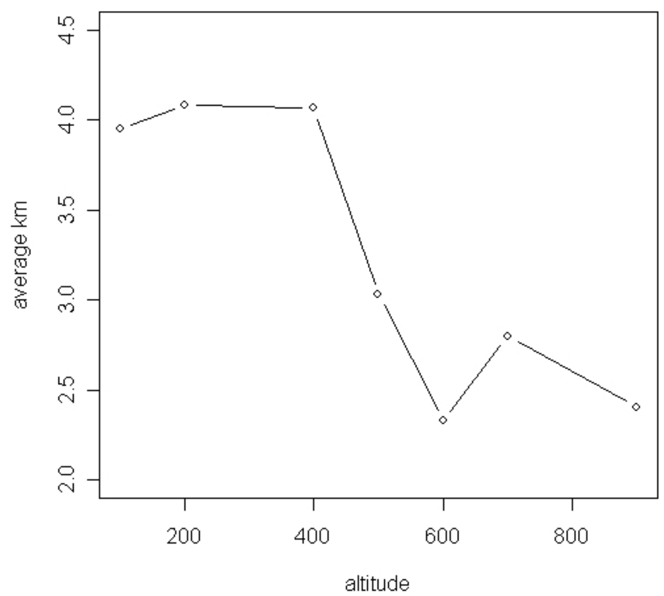


distinguish between two watercourses of the same length, one flowing through unused land, one with many tributaries named because the area is important to agriculture and habitation. Take for example River Lochay (4320) and Tarland Burn (5896). River Lochay is a well known river with a P-Celtic name, whilst Tarland Burn is a newer name of much less 'importance'. Both are 26.5 km long, and as such *km* would treat them as equivalent, but Lochay Burn has thirty-nine named tributaries and Tarland Burn has fourteen.

2.2.1.3 Alt: Altitude

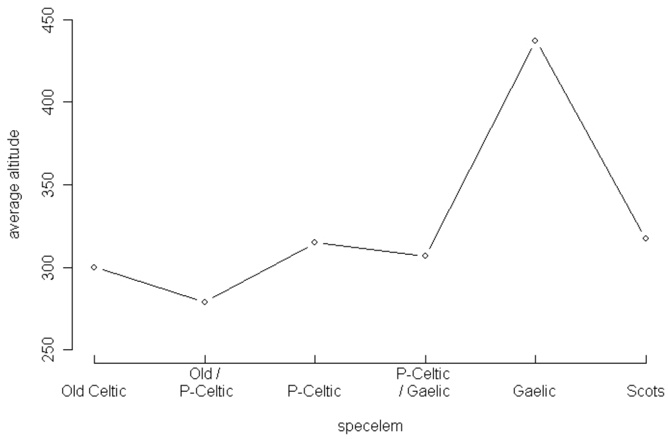
That altitude and / or the soil classification has a correlation with the size of a watercourse is shown in figure 2.5. The general trend also shows that the lower the altitude, the longer the watercourse.

Figure 2.5: Relationship between the Altitude in Metres and Length in Km of Watercourses



There is also a notion that the older the RN, the lower in altitude the watercourse. Figure 2.6 shows the correlation between linguistic stratum and altitude, showing a general upward trend, with some interesting phenomena with Gaelic and Scots. These figures show at least that altitude has a meaningful contribution to make on hydronymic nomenclature, and that lower altitude is equivalent to a longer watercourse, and to a certain extent, an early linguistic stratum is equivalent to a low altitude. Whether this reflects a situation of early coastal habitation or an accident of survival whereby RNs at higher altitudes were more readily renamed remains to be seen.

Figure 2.6: Relationship between the Altitude of a Watercourse and its Linguistic Stratum



2.2.1.4 Pos: Position

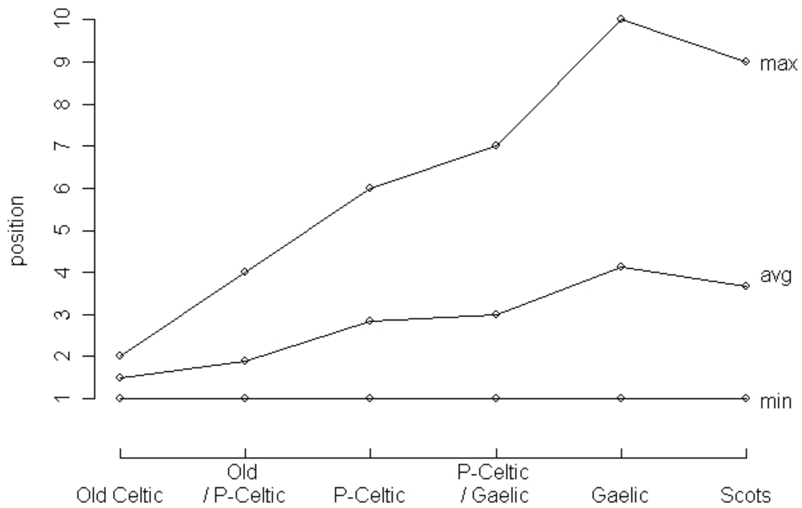
Position has a slightly different relationship to the other data. In the case of *km* and *nont* above, it is generally true that ‘bigger is better’, i.e. that the longer the watercourse the older the name etc. With the case of position, the range of length in watercourses at the lower end of the position hierarchy (i.e. closer to the sea) is much larger than those at the higher end (i.e. rivers far from the sea, and typical in mountainous areas), which are generally much smaller. So, for smaller watercourses, *pos* acts like *km* or *nont*, but for larger watercourses this does not necessarily hold as in figure 2.7. That there is a correlation, however, is not in doubt as can be seen in figure 2.8 on page 24, where a strong, but not exact correlation between *km* and *pos* is shown.

Figure 2.9 on page 24 shows the percentage occurrence of each position within each linguistic stratum. The implications and meaning of this graph is discussed in more detail in section 5.2.3 on page 190. Essentially it shows the direction of the data are that the lower the position (i.e. the closer to the sea) the more important the watercourse.

2.2.1.5 Concluding Remarks

These four factors fall into two groups. On the one hand, *alt* and *pos* relate to the location of the watercourse. In a certain sense these two factors act as a coordinate system, *alt* gives the height and *pos* gives the relative distance from the sea. On the other hand, *km* and *nont* reflect the size of the watercourse regardless of its location. This grouping is implicit in much of the analysis in the later chapters of this thesis.

Figure 2.7: Relationship between the Position of a Watercourse within a Catchment System and its Linguistic Stratum



2.2.2 Rendering of Factors

To create a comparable system involving varying sets of data, one must have the same range. I chose this to be a range between 1 and 10, since this is intuitively simple, but it should be stressed that it is arbitrary. The number 1 was chosen as the lowest possible number since 0 has a deleterious effect on data. (For instance 1 x 5 x 6 = 30, but 0 x 5 x 6 = 0) The policy is that 1 represents the smallest watercourses, and 10 the larger. This is simple for *km*, *nont* and *alt*, but *pos* is somewhat less so (see section 2.2.2.3 on page 26).

In all the geographical factors below, the outcome of the scores does not naturally come to a minimum of one and a maximum of ten. Equation 2.1 below takes a given range of numbers and converts it to another given range.

a = original minimum number	d = new maximum score
b = original maximum number	x = original number
c = new minimum score	y = new score

$$\left((d-c)\left(\frac{x-a}{b-a}\right)\right)+c=y \tag{2.1}$$

By way of example, *alt*, which is discussed below, naturally falls into a number between one and thirty-six. To turn this into a score between one and ten, the following numbers are input into the equation: $a = 1$, $b = 36$, $c = 1$, $d = 10$. To derive the maximum score, x, the original number, is 36, so that y, the new score, is 10:

Figure 2.8: Relationship between the Position of a Watercourse within a Catchment System and its Length

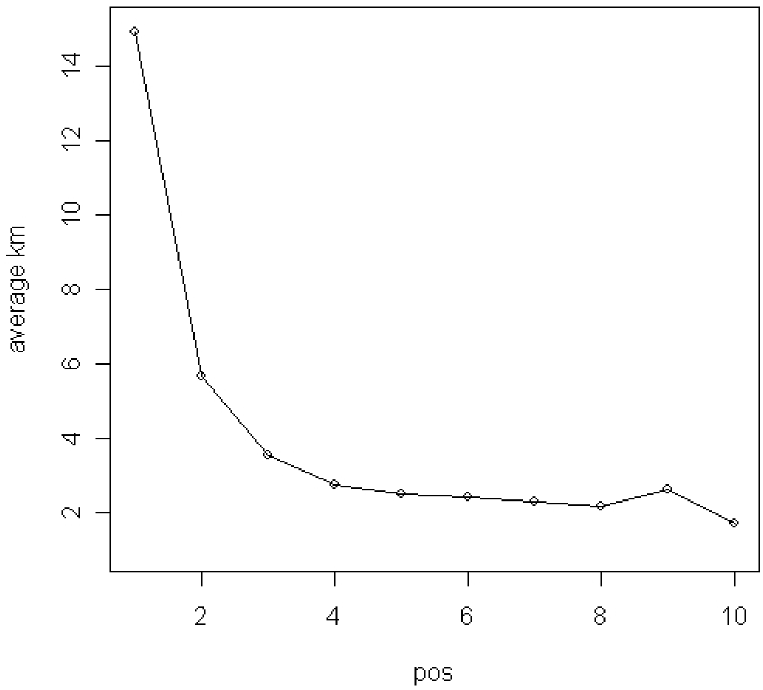
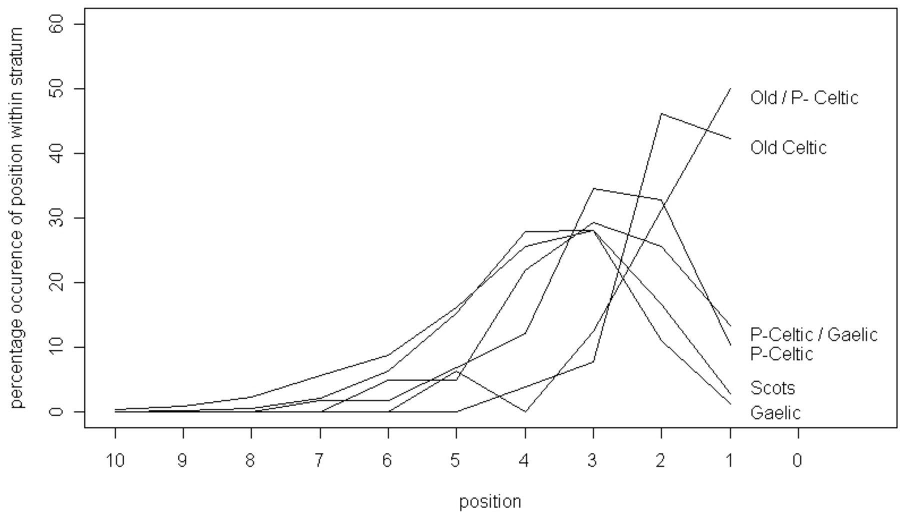


Figure 2.9: Relationship between the Position of a Watercourse and the Linguistic Stratum from which the Name Derives



$$\left((10-1)\left(\frac{36-1}{36-1}\right)\right)+1=10 \tag{2.2}$$

To derive the new minimum score, x, is 1 and y remains 1:

$$\left((10-1)\left(\frac{1-1}{36-1}\right)\right)+1=1 \tag{2.3}$$

To make each score comparable makes an implicit assumption that each factor is as important as any other. For instance, it may be that length of watercourse is a more important factor than, say, altitude within the mind of a speaker naming a given watercourse. Whilst this may be true, any other model would be undesirable, since as of now there is no apparent basis on which to alter the weighting. The data used to compile these factors are from the columns: *km* (length of the watercourse), *nont* (number of named tributaries), *pos* (position) and the *alt* columns (altitude).

2.2.2.1 Km

The *km* factor gave idiosyncratic results if taken 'as is', i.e. *km* was divided by a number to make it between 1 and 10, mainly because whilst almost a quarter of the RNs are up to 1 km long, the longest is 159.5 (River Spey). The lengths were weighted according to the algorithm in table 2.2 which accounts for more difference between shorter watercourses than long watercourses. This is because the difference in linguistic impact on a RN of a watercourse being 1 km long or 16 km long is much greater than that between a watercourse 120 km long and 135 km long, despite the fact they have the same difference in length (i.e. 15 km). This creates a score between 1 and 9. These scores are then processed through equation 2.1 on page 23, keeping 1 as the minimum score, but changing the old maximum number from 9 to the new maximum number, 10.

Table 2.1: Maximum and Minimum Lengths of Watercourses for Names from each Stratum

Stratum	minimum km	average km	max km
Old-Celtic	12.0	63.1	159.5
Old / P-Celtic	1.0	25.2	159.5
P-Celtic	0.5	9.4	77.0
P-Celtic / Gaelic	0.5	9.0	51.5
Gaelic	0.5	3.2	56.5
Scots	0.5	2.7	26.5

If *km* was simply scaled down to a score between 1 and 10 it would create 'bland' data, that is, for all except the longest eleven watercourses, the *kmscore* would be under 5, and 98.6% of the names would be less than 2. Whilst this is of itself significant, it homogenises the data, reducing the significance of the differences for all the medium and short length watercourses.

2.2.2.2 Nont

This factor ranges between 0 and 997 (for the River Tay), and even more so than *km*, the weighting is at the lower end of the scale, with 3832 watercourses with a *nont* of 0. For the same reasons as *km*, effectively the same calculus is used as in table 2.3. Since the maximum is also nine as above, the scores are processed through equation 2.1 on page 23 to make the maximum ten.

Table 2.2: Weighting of *km* Range

km range	*kmscore*
.5 - 1	1
1.5 - 2	2
2.5 - 4	3
4.5 - 8	4
8.5 - 16	5
16.5 - 32	6
32.5 - 64	7
64.5 - 128	8
128.5 -256	9

Table 2.3: Weighting of *nont* Range

nont range	*nontscore*
0 - 4	1
5 - 8	2
9 - 16	3
17 - 32	4
33 - 64	5
65 - 128	6
129 - 256	7
257 - 512	8
513 - 1024	9

The *nontscores* and *kmscores* were originally weighted using equation 2.1 instead of the system used, but this gave erratic results. This is because the above equation works best when each class is of equal importance, whilst the *nont* and *km* weightings are weighted towards the lower end of the scale.

2.2.2.3 Pos

In section 2.2.1.4 on page 22 the phenomenon was discussed that the more removed from the sea a watercourse is (using tributaries as a unit of measure), the more likely it is to come from a newer stratum, to be smaller, and so on. The number *pos* in *hieralist* represents this, with 1 representing those watercourses which flow into the sea, and 2 representing those which flow into watercourses which flow into the sea, and so on. Fortuitously, the highest *pos* is actually 10, meaning that the equation used above is not needed in this case. The 'direction' of *pos* is wrong however, in that a number is needed which is higher the closer to the sea a watercourse is and lower the further from the sea it is. In this case 11 was deducted from the score and the result was given an absolute value, making 1 into 10, 2 into 9 and so on. thus:

$$\text{if } x = pos \text{ and } y = posscore$$

$$|x - 11| = y \tag{2.4}$$

2.2.2.4 Alt

The next factor is altitude. The MySQL field 'class' (see table B.5 on page 247) contains a series of two-letter codes representing the type of land through which the watercourse runs. These codes describe two things: Firstly, they describe the soil classification (see table B.4 on page 246) such as

'Cold wet upland' or 'Warm dry lowland'. Secondly, they describe the altitudinal zones through which the watercourse runs (see table B.2 on page 245). In creating *altscore*, the soil classification code was used, because there are eighteen different soil classifications, but only four altitude ranges. There is a direct relationship between soil classification and altitude however, so using soil classification merely preserves more data than the broader altitude ranges.

The first letter of the two-letter soil classification codes denote what kind of altitude a given point on a watercourse can be in. This is shown in the *class* column of table B.4 on page 246; the *class2* column in the same table swaps the letter into a number between one and six. For each watercourse the lowest and highest points are measured (this is stored as *firstclass* and *lastclass*) in the database. With six different possibilities for the lowest point and six for the highest point, this gives a scale between one and thirty-six when multiplied. Within this range there are twenty-one possibilities, since each *firstclass* must necessarily be lower than or equal to its *lastclass*. Having a scale between one and thirty-six, this is then processed through equation 2.1 on page 23 above. Whilst this gives a score between one and ten, the direction is wrong, since at this point 10 represents watercourses in high mountainous areas, and 1 for those at sea level. Therefore the scores have equation 2.4 on page 26 above applied to it. This gives an *altscore* of 1 for watercourses at very high altitudes, and an *altscore* of 10 for those at sea level. It would be possible to average (as opposed to multiply) the *firstclass* and *lastclass*, but this would create over-homogenised data with only eleven, evenly spaced possible scores.

In other places in this thesis, the pure altitude ranges are used, as represented in the database as *altmin* (minimum altitude), *altave* (average altitude) and *altmax* (maximum altitude). It should be stressed these are not exact altitudes to within a few feet, but instead represent ranges of 200 feet.

2.2.3 Composition of Geographical Factors into Scores

There are several operations which one can perform with these scores. For the purposes of this section, it is desirable to combine them to create one overall score for 'importance' of the watercourse. The term 'importance' here is necessarily woolly; it is a representation of the underlying factors that prompt speakers to describe watercourses as 'large'; for example, the De Situ Albanie[2] from c. 1200 describes the Spey as *magnum et mirabile flumen quod vocatur Spe* ''the large and magnificent watercourse which is called the Spey'.

For reasons given below it is necessary to have a score between 1 and 10 as with the original scores. The logical course of action in this case is to average the scores to create *geogscore* (standing for *geog*raphical *score*), a number theoretically between 1 and 10,[3] counting altitude, length, position and number of named tributaries to quantify how 'important' a watercourse is. The results show the top five current RNs to be: River Spey, River Tay with Rivers Earn, Forth and Dee in equal third place.

It is of course possible to apply other operations on the scores. Multiplication was investigated

[2] W. F. Skene, *Chronicles of The Picts, Chronicles of the Scots* (Edinburgh, 1867), p. 136.

[3] In practice for a RN to have a *geogscore* of 1 it would need to have each component of *geogscore* as 1 as well, which does not happen. The same is true of RNs with a geogscore of 10.

at an earlier stage. This was actually a relatively valid method, and created results very similar to the averaging method. The main drawback, however, was that the data would need further altering, firstly because the data were not in the desired range (1 to 10,000) and secondly because the same phenomenon occurs as with the *km* data, whilst the possible range was 1 to 10,000 the vast majority was between 1 and 2,000 with only 3 entries over 6000. This is again the issue of 'bland' data. This is further illustrated below in figure 2.13 on page 42.

2.3 Linguistic Data

2.3.1 Justification of Factors Used

As in the section above,'probable cause' is proven for inclusion in the system. Since the study of names as opposed to places has been the focus of toponymic research as discussed in the introduction, this section is necessarily shorter, since in a sense the justifications have been made repeatedly in onomastic works.

2.3.1.1 Specelem: The Linguistic Stratum

The term *specelem* here is short for 'linguistic stratum of the *spec*ific *elem*ent'. This broadly represents what language is mentioned in the derivation, for instance Allt Dubh from G(aelic) *dubh*, 'black' represents Gaelic as the *specelem*. The specific strata used in this thesis, and the considerations about how each RN is assigned to a given stratum is discussed in much more detail in section 5.1.1 starting on page 185. The linguistic stratum from which a name derives is of course one of the central concepts when investigating place-names.

2.3.1.2 Genelem: The Generic Element

Genelem represents information about the *gen*eric *elem*ent of a RN. Generic elements often closely relate to the geographical qualities of a watercourse, as is discussed in much more detail in chapter 3. It could be claimed that this factor is redundant since it is similar to *specelem* above. This is not the case as can be seen by table 2.4.

Table 2.4: Occurrences of Combinations of Strata of Generic and Specific Element

	Specific Element					
Generic Element	Old Celtic	OC / P-Celtic	P-Celtic	P-Celtic / Gaelic	Gaelic	Scots
P-Celtic	0	0	**14**	3	1	0
Gaelic	2	3	7	11	**2468**	4
Scots	14	16	42	64	1809	**1819**

If there were a one-to-one correlation, it would be expected that only the cells with numbers in bold would have a number above 0. Of course, this is not the case, predominantly because generic elements from the more recent strata often attach themselves to RN from older strata, but other phenomena are at work here as well, which will be discussed below. Suffice to say, at this stage, there is a weak enough correspondence between *genelem* and *specelem* to justify the inclusion of generic elements.

2.3.1.3 Semtype: The Semantic Type

Semtype stands for *sem*antic *type* and gives information about the meaning or semantic content of the specific element. In other words, this score takes the meaning of the specific element and puts it into a group with other equivalent RNs. For instance, Muckle Burn and Allt Mòr both have the same *semtype*, because they both have a specific element meaning 'large', regardless of the actual terms used. That the meaning of a RN is related to the watercourse itself is one of the most basic assumptions in hydronymy, and in semantics itself.

2.3.2 Rendering of Factors

As with *geogscore* above, the policy is to create a score between 1 and 10. This new score is called *lingscore* which stands for *ling*uistic *score*. *Geogscore* was used as a basis for *lingscore* because of the nature of the data comprising each score, that is, the data for the physical attributes are largely numerical and incremental, for instance in terms of length, 2 is always longer than 1 and so on (unlike, for instance, stratum, where although it is received wisdom that a P-Celtic RN is older than a Scots RN, one would not wish to assume this), thus *lingscore* uses *geogscore* as the basis of its calculation. This is the reason why a *geogscore* between 1 and 10 was necessary as mentioned in the previous section. For each linguistic factor, the average *geogscore* was calculated for each specific score. This specific average was taken to be the score for each class. In doing this, one is not forced to make assumptions about these different groups, as to where each class should stand in relationship to another.

It might ideally be desired to derived *lingscore* purely from linguistic data, without having to resort to *geogscore*. There is, however, no basis for being able to do this. Take the name River Bervie; if someone did not know anything about this watercourse, one might guess it was a watercourse of considerable size because it has the generic element *river* and one might unconsciously think it was long because it has a short, lexically obscure specific element. How does one know a *river* is 'long', however? Because of our experience of the physical qualities of other watercourses with the same generic element. Without measuring other *rivers*, and purely treating *river* purely on linguistic grounds we are unable to show that *rivers* are 'long'. Looking at *river* purely as a generic element without looking at any actual examples of rivers gives one very little actual information. It would be possible, for example, to count the frequency of the term *river* in the AOS, but this would hardly tell us anything, since there are sixty-one examples of *river*, and sixty examples of *grain*. One could also look at the linguistic stratum from which the element derives or the types of syntactic structure which it occurs in, but without tying this to any given physical qualities the data tells us very little about the idea of *river*.

A separate MySQL view for each component of lingscore was created to handle the calculus. The average *geogscore* for each factor was calculated. Since these scores are derived from averages, they tend to 'bunch' themselves around the centre of the score spectrum, which is undesirable, since the *geogscores* are spread out, so these scores have been put through equation 2.1 on page 23 above, to make the minimum 1 and the maximum 10. These three scores were averaged together to generate *lingscore*. The following sections explain idiosyncrasies of the scores in a little more depth.

2.3.2.1 Specelem

The average *geogscores* for each linguistic stratum are referred to as *specelemscore* and are shown in table B.14 on page 254. The only factor involved in this is *specelem*, so for each *specelem*, all the *geogscores* are gathered and then averaged. Like the other components of *lingscore*, these seven scores (representing the seven strata) are then stretched so that the minimum score is one and the maximum is ten. The code for this is shown in section D.1 on page 271.

2.3.2.2 Genelem

The average *geogscores* for each generic element are referred to as *genelemscore* and are shown in section B.13 on page 254. As with *semtype*, there is more than one factor involved. In this case there are: *genelem1*, *genelem2*, *genelem3*. *Genelem1* denotes the linguistic stratum of the generic element (e.g. *burn* is Scots and *allt* is Gaelic). *Genelem2* denotes the identity of the element (i.e. *burn* or *allt*). *Genelem3* denotes the syntactic construction of the RN (e.g. to distinguish *Loch* X, X *Loch* and *Loch of* X). As expected, this creates an individual score for every element in every available position. For some calculations, this proved to be to restrictive; for instance, in some data mining, one may not care whether *loch* stands in first or last position, so *genelem2table* was also created, which excludes *genelem3*, i.e. syntactic position. For instance, the generic element construction with the highest *genelemscore* is that of the River X construction, as one might expect. This represents the string 745, the 7 stands for the fact that the term 'river' is Scots (as in table B.7 on page 248); 4 represents the specific term 'river' itself (as in table B.8 on page 249); finally 5 represents the syntactic position or the generic element in relationship to the specific element (as in table B.9 on page 249) . The average *geogscore* is then derived for all RNs of this 'River X' variety (which number sixty) to derive the *genelemscore* for this type of name. Like the other components of *lingscore*, the values of each of the *genelemscores* are then stretched so that the minimum score is one and the maximum is ten. The code for this is shown in section D.1 on page 269.

2.3.2.3 Semtype

The average *geogscores* for each semantic type are referred to as *semtypescore* and are shown in section B.15 on page 256. This is slightly more complicated to render than the previous factors because three factors are involved here, *semtype* (general semantic type), *pntype* (place-name type if semantic type relates to an external place) and *adjtype* (adjective type if the place-name relates to an adjective). The same procedure is carried out as with *genelemscore* in the previous section. So for example, the RNs with the lowest *semtypescore* are those with the string 3602; 36 represents RNs concerned with 'visbility' (as in table B.10 on page 250); 0 means the name does not derive from another place-name (as in table B.12 on page 251); 2 means the name means 'hidden' rather than 'exposed' (as in table B.11 on page 251). The average *geogscore* is then derived from for all RNs meaning 'Visibility: hidden' (which number nineteen) to derive the *semtypescore* for this type of name. Like the other components of *lingscore*, the values of each of these semantic classes are then stretched so that the minimum score is one and the maximum is ten. The code for this is shown in section D.1 on page 270.

2.3.3 Composition of Linguistic Factors into Scores

Once these three scores have been calculated, they can then be averaged together to create *lingscore*, or used in other functions. The specific information gleaned from these tables is discussed under the relevant sections. The top five RNs are: River Almond, River Ugie, South Ugie Water and North Ugie Water in equal third place, River Dee and River Don. Remember that if any value is not known, then the *lingscore* cannot be created. For instance, if the meaning (*semtype*) is unknown, the averaging of this with the other factors will create a null value, and the entry in question will not appear in the list. For instance in the case of the River Tay, the *specelemscore* is 10 and the *genelemscore* is 10 but the *semtypescore* is null, meaning that the score cannot be calculated, because an average of 10, 10 and NULL produces a NULL value in MySQL.

2.4 Factors Not Included

It is easy to think of several factors not included in this system particularly for the geographical aspects, for instance, type of vegetation or speed of flow or volume of water. The main reasons for any data being excluded are probably for one of the following reasons:

1. Many of these data are not stored in a way which makes it easy to put them in a database, and to extract much of these data was extremely time-consuming, so for reasons of time limitations not as much data could be gathered as was desired.

2. Geographical information is only relevant if it is known what the environmental conditions were like at the theoretical time of naming. This means that a reduction in biodiversity, deforestation and urbanisation have all altered the initial conditions under which a watercourse once existed, so while, for instance, it would be interesting to see if names with a derivation of 'woods' flow through more forested areas, the extant information may produce inaccurate results.

   This factor is already a problem in the data gathered, for instance the *km* of lochs has often been changed due to the creation of reservoirs and drainage etc. In all cases, an effort has been made to use data relating to the situation at the time of naming, i.e. in pre-industrial times.

3. Some information cannot be put into a score system, for instance names are changed because of the importance of the referent to man in a certain way, e.g. agriculture, religious use etc., which

can be made known to us by archaeology or historical sources. It would be extremely hard to quantify this into a score from 1 to 10.

Within the database the linguistic strata of the phonology and orthography of each name has been catalogued. It could be claimed that these factors could be taken into account in the same way as linguistic stratum or generic element. However, the orthographical or phonological overlay on a name is qualitatively different since it is not a factor that applies at the stage of naming, but rather is something that applies *after* the naming stage. The phenomenon of phonological overlay as applicable to this thesis is discussed in chapter 6.

Another factor which would be useful is that of distance from other similar names. As it stands the system does not factor in area whatsoever, and assumes naming styles to be uniform over the entire AOS. This is not hugely problematic, since the AOS was deliberately chosen because all areas contain Scots, Gaelic and P-Celtic names. Moreover it is easy to run queries across the whole data-set and compare these to queries run over a specific coordinate range or catchment area.

Another factor would be 'age of name'; this would involve gathering data on each name concerning the period in which the name was coined, which could then be used as a factor. Whilst this would be extremely useful, it has not been attempted since it is notoriously difficult to construct absolute chronologies in toponymy. There is only really a handful of names which have a documented inception and these are generally late.[4] The use of the date of the earliest documented mention of the name would not be acceptable either, since due to the accident of survival there are numerous RNs which are probably ancient but not documented until a much later date. The unevenness of survival / creation of manuscripts and maps etc. in the past along with the unevenness of scholarly research in different regions make the use of an absolute date methodologically unsound.

It should be said that the excellent Dictionary of the Older Scottish Tongue[5] does have dates for the first use of a term; this is unusable for the purposes here, however, not least because it only contains Scots terms, but also because the date of a term entering the lexicon is not necessarily the same as that of the date of entry into the onomasticon. This methodology is in an early stage and the future may hold further improvements.

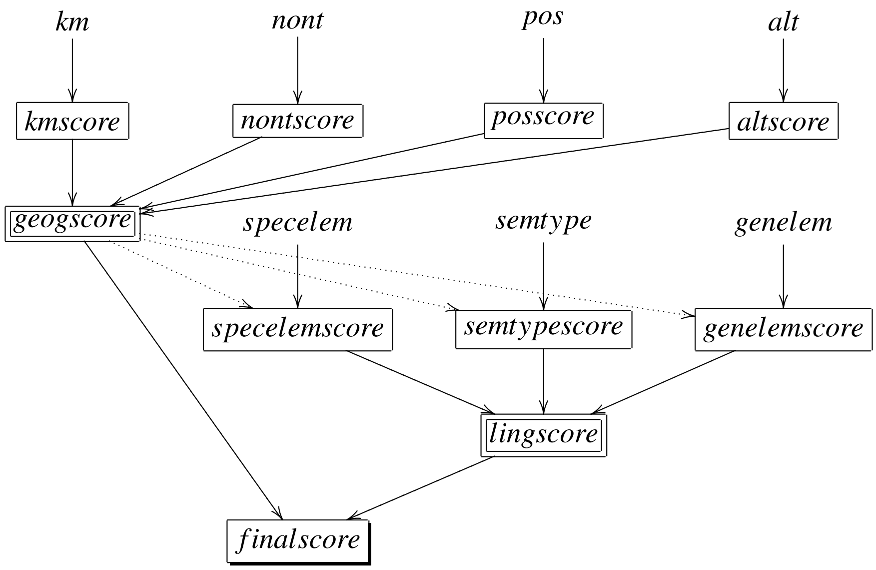
2.5 Combination of Lingscore and Geogscore

Once *lingscore* and *geogscore* have been derived, there are a number of operations possible. They can be averaged, resulting in a single *finalscore* which attributes a score to each database entry according to its physical and linguistic attributes. Since *lingscore* and *geogscore* are theoretically between 1 and 10, *finalscore* is also theoretically between 1 and 10. In practice the range is from 1.97 (For Lochan Uaine (6127)) and 8.78 (for River Almond). The top five RNs according to *finalscore* are: River Almond, River Dee, River Don, River Forth, River Ythan. The River Almond may not seem

[4] For example the creation of the name Friockheim: Nicolaisen, *Scottish Place-Names*, p. 86-87.
[5] 'Dictionary of The Scots Language' ⟨URL: `www.dsl.ac.uk`⟩.

like the most 'important' RN in the AOS, but it is the most important RN where all the linguistic qualities are known. The River Tay, for instance, has a larger *geogscore* but it is not included in the *finalscore* list because, since its meaning is not known, it has null *semtype*, giving it a null *lingscore*. Part of the purpose of this thesis is to fill in gaps in the database such as this (See table 4.36 on page 176 for discussion of the River Tay). The structure of the calculus is shown in figure 2.10.

Figure 2.10: Structure of the Score System



2.6 Practical Applications

The remainder of this thesis is essentially a comparison of the factors listed above. For the remainder of this chapter a method called *finalscore* is used. In essence, a score is created based purely on the physical aspects of the watercourse (*geogscore*), and then a score based on the linguistic attributes of the RN (*lingscore*), so each entry has two scores, these two scores are then compared to discover any correlations over the whole data-set. There are two predominant applications of this method; the first one is the *traditional approach*, where the method is used to make suggestions to fill in missing data. The other system works backwards from that; it looks through the data and finds anomalies which represent incorrect or erroneous data; this is commonly called *data mining*. The methods below are explained through several case studies.

2.6.1 Traditional Approach

The commonest methodology for the toponymist is to take a name, gather some old forms and then to posit an etymology. For a RN where the linguistic stratum and etymology is not known, in this thesis one would say the *semtype* and *specelem* was null, that is, empty. What is possible is to suggest sensible values to these empty data based on the data that already exists. This methodology is based on and often used in the scientific method in cases where data have been gathered, but some data are missing or obviously erroneous. The method is somewhat different here, either due to the nature of the missing data, and the fact that at all times one must use this method as a tool to aid 'old-fashioned' toponymic work, but it should not supersede it in importance. Following is an example of how this works in practice.

2.6.1.1 Frandy Burn

Frandy Burn[6] is a good example of an obscure name which could possibly be Gaelic, Scots or P-Celtic. Frandy Farm, called simply 'Frandy' by Stobie in 1783, is a nearby micro-settlement. The possibilities for the derivation of this name are:

1. G *(Allt a') Chrainn Duibh*, i.e. thus 'Black Tree Burn'. This interpretation involves a translation of *allt* into *burn*. The lenited version of *crann* is preserved however. Upon anglicisation, the [χ] has been changed into an [f], the changing of a voiceless fricative from a velar to bi-labial position.[7] The existence of an original G *duibh(e)* resulting in *-dy* or *-die* in modern forms is also seen in the forms for Eskandie on page 54.

2. A P-Celtic version of the above, with a word akin to W *prenn*, 'tree', a cognate of G *crann*.

3. Sc *frandie*, a diminutive of *frae hand*, 'free hand', meaning 'a pile of hay small enough to be picked up without a pitchfork'.[8]

There are a number of approaches open depending on the information available, in this case the *geogscore* (5.235) and the *genelemscore* (4.63) are known, whilst *semtype* and *specelem* are missing. If one looked at all the other RNs with the same *geogscore*, table 2.5 is the outcome.

Table 2.5: The Linguistic Strata of RNs with the same *geogscore* as Frandy Burn

stratum	frequency	percentage
P-Celtic	1	1.42%
P-Celtic / Gaelic	1	1.42%
Gaelic	45	64.29%
Scots	23	32.86%

It might be felt that only taking into account RNs with exactly the same *geogscore* as Frandy Burn is too restrictive and retrieves too little data. Table 2.6 shows the same information but for RNs with a *geogscore* $\pm$ 0.3 of that of Frandy Burn. As is evident, although there are more RNs posited, the

[6] This is briefly discussed in Angus Watson, *The Ochils: Placenames History Tradition* (Perth, 1995), p. 70.

[7] Other examples of this sound change are: Fummery from G (Allt a') Chomair (discussed on page 208); Afforsk, from G Achadh a' Chroisg (see Watson, *The Celtic Place-Names of Scotland*, p. 486) and Affleck, from G *Achadh + Leac (see George F. Black, *The Surnames of Scotland* (Chippenham, 1946 reprinted 2004), p. 8-9.

[8] M. Robinson, *The Concise Scots Dictionary* (Aberdeen, 1987).

percentages are similar. Both these tables lead to the suggestion that Frandy Burn was in origin a Gaelic name in the sense that other watercourses with similar physical qualities have names which are Gaelic in origin. If this is accepted, it is then possible to access all the RNs with the same *geogscore* and same *specelem* (6, i.e. Gaelic) and return each *semtype*, as in table 2.7 overleaf which suggests the derivation is ‘specific natural feature’.

Table 2.6: The Linguistic Strata of RNs with the same *geogscore* as Frandy Burn within a tolerance of $\pm$ 0.3

stratum	frequency	percentage
Obscure / None	13	0.77%
OC / P-Celtic	4	0.24%
P-Celtic	23	1.37%
P-Celtic / Gaelic	22	1.31%
Gaelic	998	59.37%
Scots	621	36.94%

It is of course true that most watercourses relate to a specific natural feature, so rather than each percentage of each semantic class being judged in the same way, the data are judged when compared to all Gaelic names, if there is a notable difference in the percentage occurrence in a particular semantic class when comparing the Frandy data to that of all the Scots data, there may be a candidate. This is done in table 2.8 on page 37 where the percentages in the Frandy data have been deducted from the data across the board.

This still suggests that the likeliest derivation for Frandy is ‘specific natural feature’ which fits in with the etymon posited earlier. Under the first semantic class listed, Frandy Burn could have derived from Allt a’ Chrainn Duibh, a burn relating to a nearby natural feature. Under the second class, Frandy Burn could have been a secondary development from the settlement (or man-made area) of Frandie, which itself derived from Y a’ Chrainn Duibh.

It was mentioned above on page 32 that a potential drawback in this system is that it ignores diachronic variation; whilst this is true, it is possible to mitigate this somewhat by rerunning the same method above, but only using entries from a particular geographical area, either county, catchment area or coordinate range. If one runs the same table, but using only watercourses situated within a certain distance of Frandy Burn, a much simplified, but essentially identical situation is shown as in table 2.9.

These are a few examples of how a judicious use of statistics can aid one in the discipline of toponymy. Other paths could have been taken here, the method above may not always be appropriate. Of course it will never be possible to simply press a button and get answers automatically, but armed with toponymic good sense, this method can guide someone seeking a derivation by suggesting likelihoods.

Table 2.7: The Percentage Occurrence of *semtypes* of RNs with the same *specelem* and *geogscore* as Frandy Burn

meaning	amount	percentage
Specific natural feature	133	29.95%
Specific man-made area	91	20.50%
Land around	33	7.43%
Specific person / occupation	21	4.73%
Flora	19	4.28%
Colour	17	3.83%
Concavity e.g. Glen	13	2.93%
Convexity e.g. Ben	13	2.93%
Relation to other features	12	2.70%
Use to man, agriculture	12	2.70%
Fauna	11	2.48%
Water feature	9	2.03%
Manner	8	1.80%
Effect / Character	8	1.80%
Material / Object	5	1.13%
Course	4	0.90%
Dimensions	3	0.68%
Sound	3	0.68%
Non specific settlement / building	2	0.45%
Bed	2	0.45%
Other	2	0.45%
Other	1	0.23%
Crossing	1	0.23%
Boundary	1	0.23%
Smell / Taste / Feel	1	0.23%
Weather / Air	1	0.23%
Temperature	1	0.23%
Age	1	0.23%
Supernatural entity	1	0.23%

Table 2.8: The Adjusted Percentage Occurrence of *semtypes* of Gaelic RNs with the same *specelem* and *geogscore* as Frandy Burn

meaning	difference	all Gaelic names	Frandy
Specific man-made area	9.92%	10.58%	20.50%
Land around	2.08%	5.35%	7.43%
Specific person / occupation	1.91%	2.82%	4.73%
Relation to other features	1.01%	1.69%	2.70%
Effect / Character	0.97%	0.83%	1.80%
Use to man, agriculture	0.87%	1.83%	2.70%
Water feature	0.71%	1.32%	2.03%
Concavity e.g. Glen	0.57%	2.36%	2.93%
Material / Object	0.55%	0.58%	1.13%
Bed	0.22%	0.23%	0.45%
Other	0.20%	0.25%	0.45%
Temperature	0.18%	0.05%	0.23%
Manner	0.09%	1.71%	1.80%
Age	0.07%	0.16%	0.23%
Non specific settlement / building	0.01%	0.44%	0.45%
Smell / Taste / Feel	0.00%	0.23%	0.23%
Supernatural entity	0.00%	0.23%	0.23%
Weather / Air	-0.07%	0.30%	0.23%
Other	-0.19%	0.42%	0.23%
Flora	-0.26%	4.54%	4.28%
Sound	-0.29%	0.97%	0.68%
Course	-0.35%	1.25%	0.90%
Crossing	-0.42%	0.65%	0.23%
Convexity e.g. Ben	-0.61%	3.54%	2.93%
Boundary	-0.72%	0.95%	0.23%
Dimensions	-0.89%	1.57%	0.68%
Colour	-2.42%	6.25%	3.83%
Fauna	-3.79%	6.27%	2.48%
Specific natural feature	-8.98%	38.93%	29.95%
Event		0.09%	
Dryness / Moistness		0.19%	
Elevation		0.23%	
Visibility		0.19%	
Number		0.12%	

Table 2.9: The Percentage Occurrence of *semtypes* of RNs with the same *specelem* and *geogscore* as Frandy Burn within a certain area

meaning	place-name meaning	frequency	percentage
Specific natural feature	Convexity	56	13.6%
Specific man-made area	Other	41	9.9%
Land around	None	31	7.5%
Specific man-made area	Land used for agriculture	29	7.0%
Specific natural feature	Concavity	29	7.0%
Specific person / occupation	None	20	4.8%
Specific natural feature	Body of water	20	4.8%
Flora	None	18	4.4%
Colour	None	17	4.1%
Fauna	None	14	3.4%
Concavity e.g. Glen	None	13	3.1%
Use to man, agriculture	None	12	2.9%
Convexity e.g. Ben	None	11	2.7%
Specific natural feature	Land not used for agriculture	10	2.4%
Relation to other features	None	10	2.4%
Water feature	None	9	2.2%
Effect / Character	None	8	1.9%
Specific man-made area	Land not used for agriculture	8	1.9%
Specific natural feature	Land used for agriculture	7	1.7%
Manner	None	7	1.7%
Course	None	5	1.2%
Material / Object	None	5	1.2%
Specific natural feature	Riparian area	4	1.0%
Sound	None	3	0.7%
Specific man-made area	None	3	0.7%
Dimensions	None	3	0.7%
Specific man-made area	Riparian area	3	0.7%
Non specific settlement / building	None	2	0.5%
Specific natural feature	None	2	0.5%
Bed	None	2	0.5%
Specific man-made area	Body of water	2	0.5%
Temperature	None	1	0.2%
Other	None	1	0.2%
Smell / Taste / Feel	None	1	0.2%
Crossing	None	1	0.2%
Specific natural feature	Other	1	0.2%
Boundary	None	1	0.2%
Supernatural entity	None	1	0.2%
Weather / Air	None	1	0.2%

2.6.2 Data Mining: The Non-Traditional Approach

In this method, rather than taking a RN as a starting point and attempting to deduce information about it, all the data are studied to see if any are erroneous; any anomaly is then investigated, either with a view to altering it if it has been misinterpreted, or justifying it if the interpretation is correct.

A useful tool at this point is *standard deviation*[9] which can discern how grouped or disparate a given set of data are, i.e. to detect whether any given comparison of two types of data shows any erroneous data, otherwise known as an outlier.[10] This has the effect of drawing discrepancies from the data which can then be analysed and if needed, corrected. The predominant application of this method is checking the standard deviation for each *lingscore* grouped by *geogscore*. In plain terms, one looks at all the rivers with a similar *geogscore* (i.e. rivers of comparable physical attributes) and looks at all the corresponding *lingscores* to ascertain whether they are similar or not. The more rounded the score is, the larger the data-set retrieved.

Of course a certain degree of deviation is to be expected from an organic network such as hydronymy. For instance, the names Toúaisis and Spey both represent the same river, but have a fairly high amount of deviation (of course it is possible to challenge the identification of the Ptolemaic Toúaisis with the Spey). Nevertheless, those names with a larger amount of standard deviation are worthy of investigation with a view to whether they have an inaccurate *lingscore* or *geogscore*. Such entries in nearly all cases were found to be inaccurate or dubious in some way. This method can be shown most easily visually; *geogscore* and *lingscore* were plotted in an x,y axis to render the relationships between the two factors and to detect any 'outliers' or erroneous data as in figure 2.11 on page 40. When this was done, the author expected an approximate line or zone showing increasing *lingscore* and *geogscore*. This is shown, but with decreasing amount of data and accuracy as the respective scores become larger.

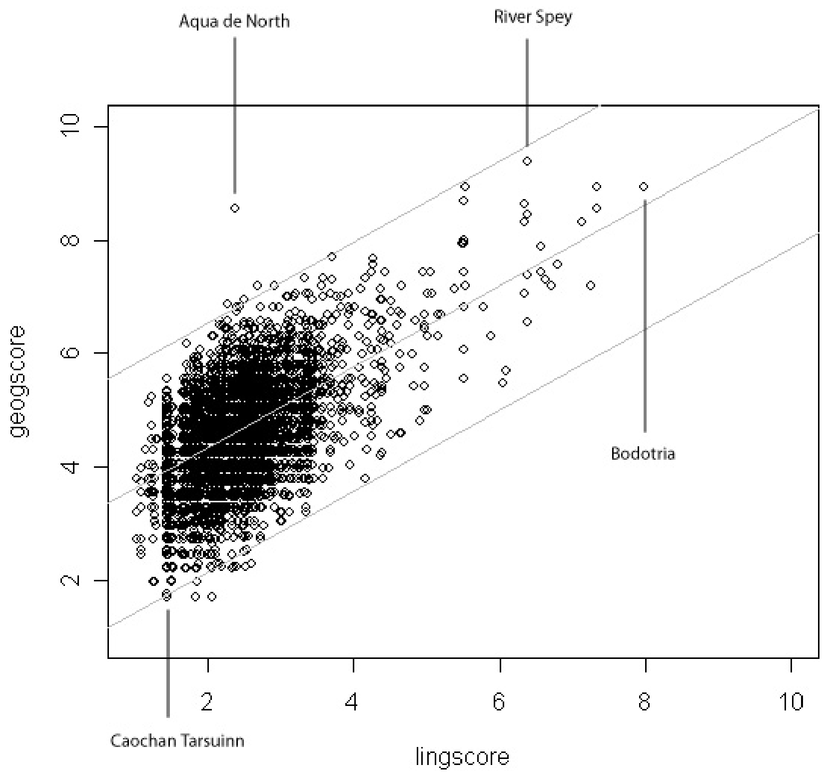
Since in this model *lingscore* is in part derived from *geogscore*, a criticism could be made that the ordering of the data into the zone of best fit actually only reflects the way the data are organised, rather than the value of the data themselves. To test this, a database was made of exactly the same structure, but with random data within the same range as that in the real database. These random *geogscore* and *lingscores* were then plotted. One would expect a random distribution of dots around a central point due to averaging. This is exactly the result produced as in figure 2.12 on page 41 with *lingscore* plotted with random values. The idea of multiplying rather than averaging the individual scores was discussed above. To further illustrate the issue of creating bland data, the same graph is shown, but with multiplication instead of averaging to show the outcome in figure 2.13 on page 42.

The same information can be generated in list form, with the largest divergence from the norm listed first as in table 2.10 on page 43. This approach is better for individual entries than the visual approach, because it displays idiosyncrasies that do not only lie outside the main body of the data, but also those that lie within the normal ranges although they are still inaccurate.

[9] Deborah Rumsey, *Statistics Workbook for Dummies* (Indiana, 2005), p. 57.

[10] The term 'outlier' here is used in its statistical sense, rather than its toponymic one.

Figure 2.11: *Finalscore* Graph with some Pointers



The top few entries from this list will now be examined to show how this method can be applied. This method is used throughout many subsequent sections to demonstrate certain points, or as a tool to calculate a range into which missing or uncertain data can be inserted which would fit the rest of the pattern (such as *semtype*).

2.6.2.1 Case study: Abhainn Dubh vs River Forth

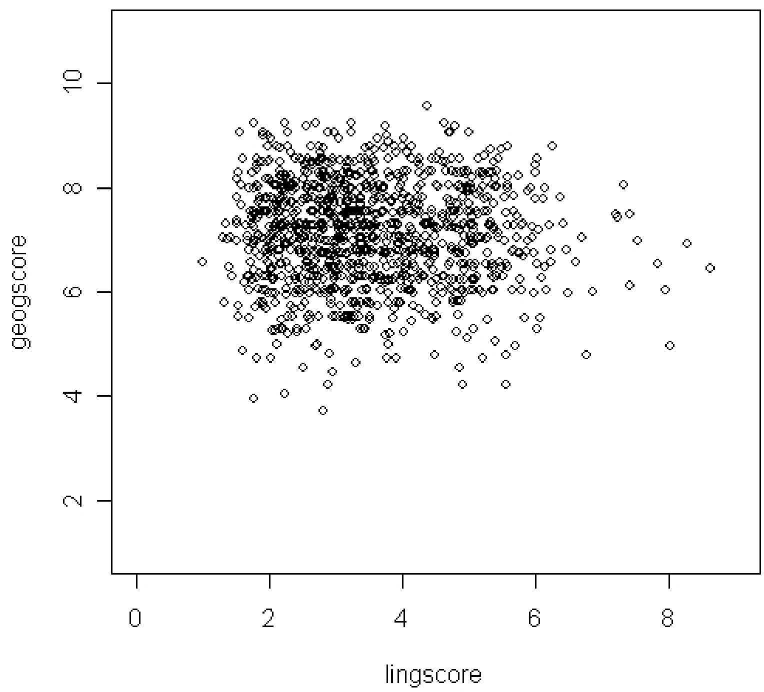
One of several names for the Forth is Abhainn Dubh, meaning 'Black River'. This is mentioned in Dwelly's list, and Gaelic informants have reported this in the Twentieth Century, although Dwelly seems to be the earliest known reference to this name. Table 2.10 on page 43 shows Abhainn Dubh to be an outlier, and the standard deviation for Abhainn Dubh is 4.338, the highest amount. This would make it worthy of investigation. There are a number of possible reasons for this idiosyncrasy:

1. The name 'Forth' should be reinterpreted to fit its *geogscore*, specifically, the *specelem* should be altered.

2. The idiosyncrasy is illusory since not enough data have been gathered.

3. The standard deviation shown is within an acceptable level

Figure 2.12: *Geogscore* and *Lingscore* Plotted with Random Values



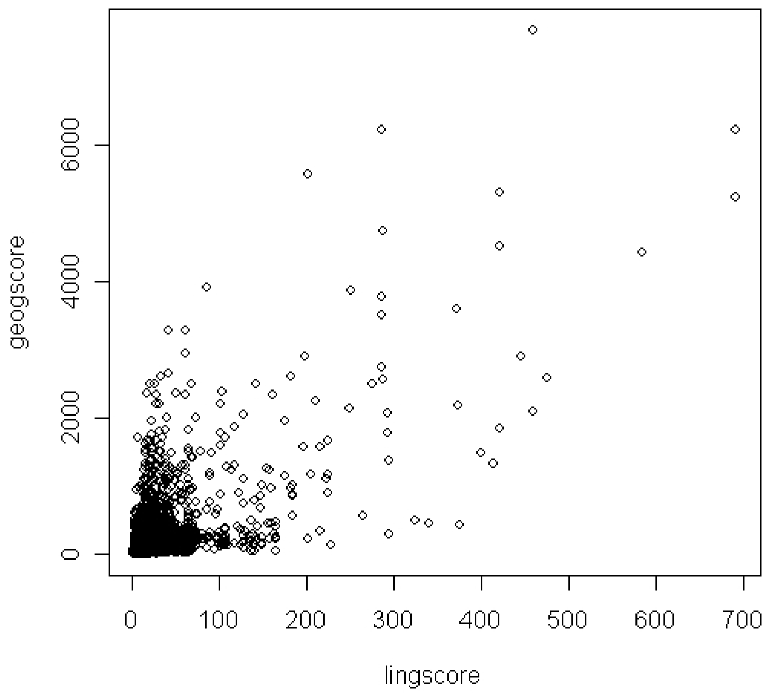
4. The name 'Abhainn Dubh' and 'River Forth' do not relate to exactly the same physical entity, i.e. Abhainn Dubh originally represented only a section of the Forth.

Each of these points will be dealt with in turn.

1. As it stands the name Forth has a *specelem* of 5 (P-Celtic / Gaelic), as it is thought to have come from a P-Celtic name such as **uo-rit-* and to have been subsequently gaelicised, or to have been influenced by early Gaelic. The preceding name Bodería is interpreted as 2 (Early / Old Celtic), because whilst it is Celtic it cannot be ascribed to any particular substratum. The Forth has a *lingscore* of 5.97 whilst Abhainn Dubh is 4.82. If one were to posit the Forth as a purely Gaelic name, the Forth's *lingscore* would become 5.53. Whilst this approaches the score of Abhainn Dubh, it is hardly satisfactory from a linguistic standpoint, since old forms give *weryd*, almost certainly representing the P-Celtic version of the name 'Forth'. If one made Forth a purely Pictish name, with an overlay of Gaelic phonology, the *lingscore* becomes 6.00, even further away from the standard. It would seem that arbitrarily tweaking the linguistic qualities of the name 'Forth' is not the solution.

2. This may have some validity since the generic *abhainn* is not especially common in the AOS, with only 8 names containing *abhainn* on OS maps, with a further twelve inferred from old

Figure 2.13: Score System using Multiplication



records and pleonastic names and so on. This situation changes of course in Argyll and the Western Isles where *abhainn* is the *de facto* term for a larger watercourse. This point, however, can be used to prove the opposite; since *abhainn* rarely appears in this area, one might wish to say the use of this term so far south and east is spurious. Figure 2.14 on page 44 is a distribution map of *abhainn* as a generic in Scotland. As said before, the symbol representing *abhainn* according to OS maps is not completely accurate, nor is it complete, but does show the general area of incidence for the element.

The westernmost point represents the lowest point of the Forth as Abhainn Dubh, an outlier. Of course distribution maps such as this do not show diachronic variation, but in general these maps show that if the whole of the Forth represented Abhainn Dubh, then it would be the southernmost name to do so.

3. Whilst the amount of standard deviation for the names Abhainn Dubh and Forth are not extraordinarily large, if other data are examined, one can see this identification causes large discrepancies, predominantly in the treatment of the generic *abhainn*, which has a high standard deviation, entirely because Abhainn Dubh is by far the largest watercourse to have the generic *abhainn*.

4. Having discounted the other possible solutions, it is the author's aim to prove that the conundrum can be solved by positing that Abhainn Dubh does not relate to the whole of the watercourse

Table 2.10: Top Ten Entries with Greatest Standard Deviation

RN	id	*lingscore*	*geogscore*	standard deviation
Abhainn Dubh	2667	2.703	8.935	4.338
Aqua de North	4982	2.377	8.55	3.948
River Spey	5785	4.743	9.373	3.066
Lochan a' Chroin	2148	2.072	1.708	2.675
Cart Burn	1270	5.652	4.428	2.533
Lochan na Beinne	578	1.845	1.708	2.512
Loch an Easain	2876	2.615	2.285	2.489
Goat Burn	3653	2.527	2.25	2.46
Burn of Carn an t-Sagairt	1241	2.397	2.25	2.367
Allt Fèith nan Sac	3106	2.335	2.208	2.364

known as the Forth.

Since *abhainn* as a generic element is predominant beyond the Highland Line, in areas of higher altitude, one could posit that Abhainn Dubh relates to the upper part of the Forth from around the area where the Duchray joins it. If one looks at the related names around the Forth, the names relating to the Forth stop at NS670961 with the Fords of Frew, whilst west of the confluence with the Duchray there is the Duchray itself, as well as Lochs Dubh and Dhu, and Gleann Dubh, all containing the G element *dubh*, 'black'. In addition, in Macfarlane's Geographical Collections, the lower part of Loch Ard is referred to as 'Burndow or Blackwater'.

If one updates the database to reflect these changes, one sees a change in *geogscore* from 8.935 to 6.8075. This standard deviance for *lingscores* with this *geogscore* is 0.756, much lower than before. When one looks at the standard deviance for generic elements, a section investigated below, it can be seen that the standard deviation for *abhainn* before was 1.9 and now it is 0.8. In plain English, Abhainn Dubh, if it is considered the same length as the Forth, is by far the longest *abhainn*, but when reanalysed in this way, whilst it is still the longest *abhainn*, it is only the longest by 2 km rather than 62. This would also bring the distribution maps into order, by removing the southernmost *abhainn* name in the area.

As a further note, a very similar case study could easily be done on a comparison between the Forth and *Scottewatre*[11] an old form for the Forth mentioned only once. The comparison would of course be between other hydronyms with the generic *water* and the River Forth.

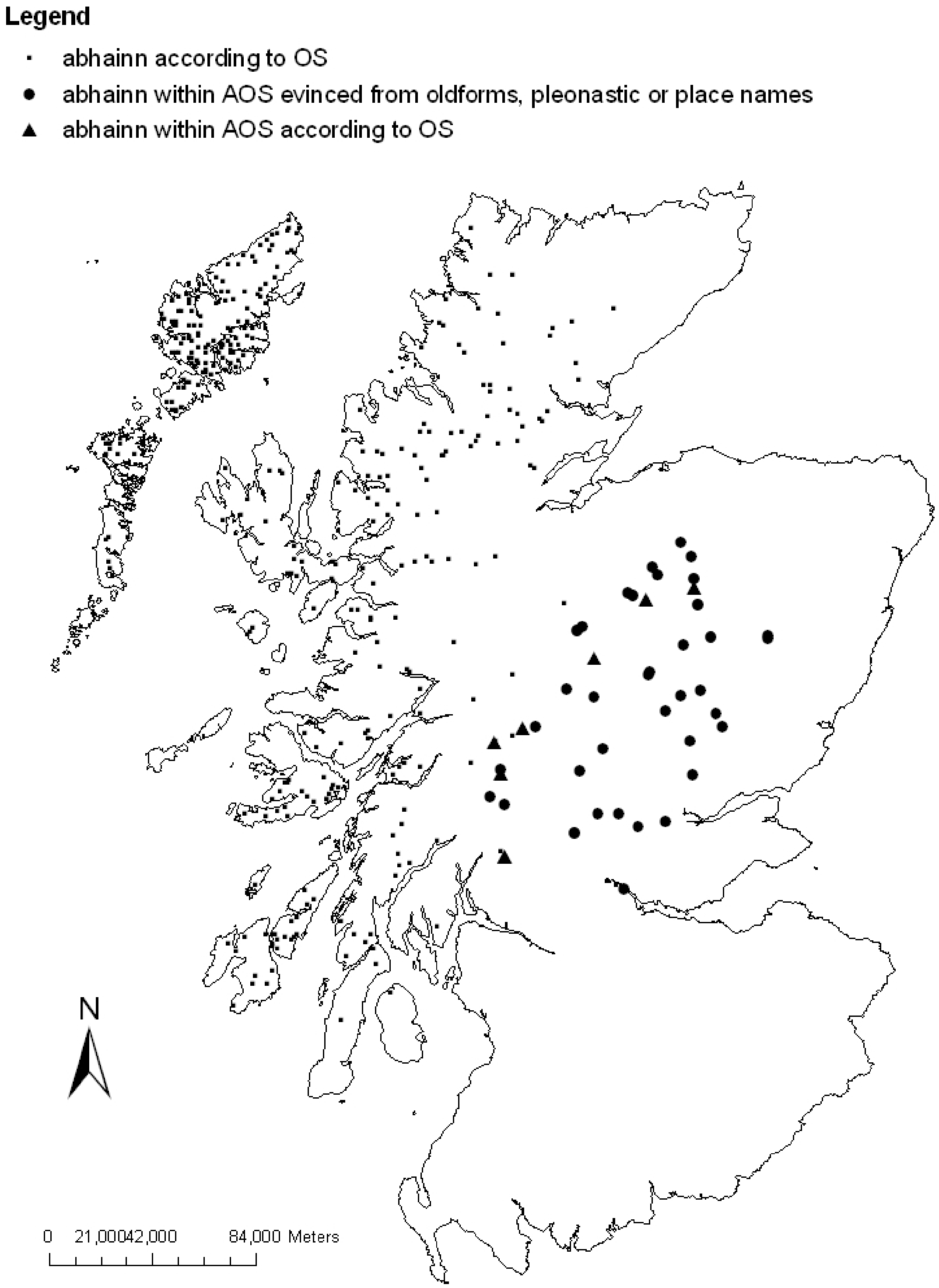
2.6.2.2 Case study: Aqua de North

The highest outlier is 'aqua de north'. This is another name for the River Don, known from a single record:

aqua de *north* que *Done* dicitur (1155 REA i p.112)

[11] Marjorie O. Anderson, ed., *Kings and Kingship in Scotland* (1973), p. 241.

Figure 2.14: Distribution of G *abhainn* in Scotland



It is an outlier because as a Scots name, meaning 'north', it is a very large watercourse; moreover, other names with the specific element *north* tend to be small burns (as discussed in section 4.34 on page 168) but the records show it is not well evidenced, suggesting the name was not used to any great degree. This cannot be removed from the database, but it does show that of the names with a high amount of standard deviation, there is a higher incidence of names which are from old forms which are not well in evidence.

2.6.2.3 Case study: River Spey

The River Spey has up till now commonly been said to mean 'hawthorn'.[12] If this is taken as the derivation, it creates a high amount of standard deviation, predominantly because it is such a large watercourse, and has a *semtype* representing 'flora', which in nearly all other cases relates to watercourses with a smaller *geogscore*. The name Spey is in Gaelic Uisge (or Abhainn) Spè. The earliest documented form of this name is (flumina) *Spe* (c. 1200 De Situ Albanie)[13] and I know of no other forms which suggest that the modern Gaelic name does not accurately reflect the underlying Pictish name.

A different derivation is posited here, based on the OC root **sk^{w}ei-*, 'vomit'. This root gives OI *sceïd*, *scé*, 'vomit'[14] and Welsh *chwydu*, 'vomit'.[15] Within a specifically Pictish context it is not known how initial OC *sk^w-* developed. In Welsh it became *ysb-*, *ysp-* or *chw-* as it did in this instance.[16] It is not unreasonable, however, to posit *sp-* as an acceptable outcome for this cluster. The vowel *ei* becomes *ẹ̄* in Brittonic, and then *wy* in Welsh.[17] There is no reason, however, to believe the latter development occurred in Pictish. The name may have had a feminine ending common in RNs, such as *-iā*, producing something like OC **sk^{w}ei-iā* >Pictish **spē(a)* >Gaelic *Spè*. To propose this derivation does not imply that other watercourses are not related to the Gaelic or P-Celtic root meaning 'hawthorn'. The River Spean outside the AOS but also rising near the source of the Spey may only superficially resemble the Spey and may represent the 'hawthorn' element. Alternatively it may also be related to the 'vomit' element, with either a Gaelic diminutive *-an* suffix or the common *-onā* / *-anā* suffix. If the latter were true, this pairing, of a root simply in *-ā* and another with *-onā* / *-anā* would mirror the relationship between the Dee and the Don. This derivation for the Spey is more intuitively satisfactory, and also brings the name further into line in the score system. This is discussed in section 4.3.15 on page 151.

2.6.2.4 Case study: Cart Burn

A. Watson[18] makes this comment about Cart Burn: 'If this name is not as simple as it seems, then W. F. H. Nicolaisen... derives the River Cart, an affluent (sic) of the Clyde from an Indo-European root

[12] Watson, *The Celtic Place-Names of Scotland*, p. 474.
[13] Skene, *Chronicles of The Picts, Chronicles of the Scots*, p. 136.
[14] 'Dictionary of The Irish Language' ⟨URL: www.dil.ie⟩.
[15] Various, *Geiriadur Prifysgol Cymru* (Aberystwyth, 1950-2002) the form *chwydu* is a form derived from the verbal noun.
[16] Jackson, *Language And History In Early Britain*, p. 527-528.
[17] Ibid., p. 330-334.
[18] Watson, *The Ochils: Placenames History Tradition*, p. 40.

meaning ‘hard, stone, stony’. Within the database this would be formalised as in table 2.11 overleaf.

Table 2.11: Alternate Scenarios for Cart Burn

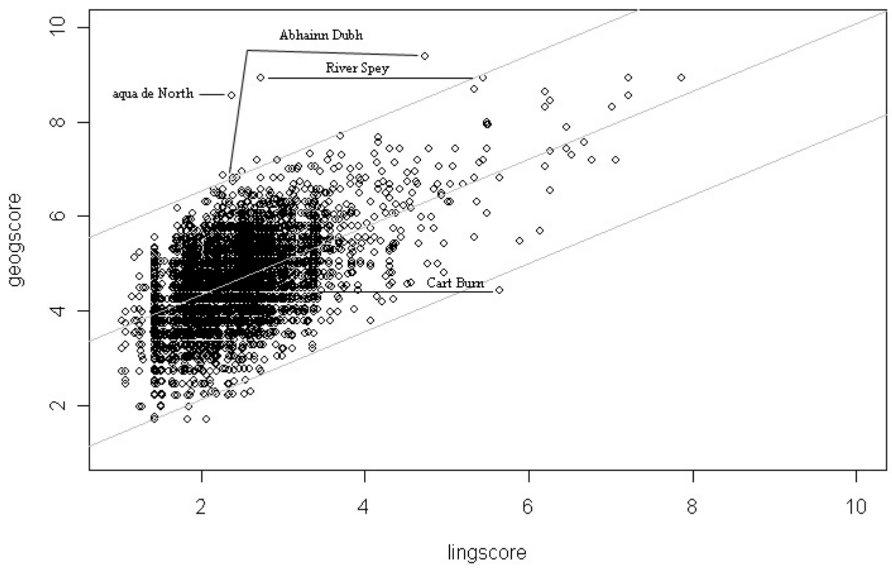
etymology	stratum	*semtype*	standard deviation
Sc *cart*, ‘cart’	Scots	Use to man, agriculture	0.4993
OC **kart*-, ‘hard, stony’	Old Celtic	Bed	2.553

This strongly suggests that Cart Burn is indeed a Scots name rather than an Old Celtic one.

2.6.2.5 Concluding Remarks on Case Studies

The following graph shows those RNs previously discussed, with a line connecting the two dots, from their original position of high deviation to the reinterpreted one, where they sit inside the zone of best fit (except in the case of Aqua de North, which should probably be dismissed). After these changes to the database have been made, table 2.10 on page 43 looks like table 2.12 on page 47.

Figure 2.15: *Finalscore* Graph Showing Updated Points



When one investigates these outliers, it can be seen that in most cases these are either names from poorly documented old forms, or names whose specific qualities are unclear. In fact it seems that for some reason the maximum amount of deviation seems to be 2.2, beyond which few names exist, and under which the names are valid (i.e. one cannot put forward reasons why the deviation should be what it is).

This method works well for smaller watercourses, whose names are definitely Gaelic or Scots, but perhaps whose orthography is obscure, since there is a much larger data set to work from. The system

Table 2.12: Updated Top Ten Entries with Greatest Standard Deviation

RN	id	*lingscore*	*geogscore*	standard deviation
Lochan a' Chroin	2148	2.072	1.708	2.675
Lochan na Beinne	578	1.845	1.708	2.512
Loch an Easain	2876	2.615	2.285	2.489
Goat Burn	3653	2.527	2.25	2.46
Burn of Carn an t-Sagairt	1241	2.397	2.25	2.367
Allt Fèith nan Sac	3106	2.335	2.208	2.364
West Water	6259	2.692	7.178	2.348
Shochie Burn	5637	2.267	6.873	2.348
Little Loch Etchachan	3003	2.348	2.25	2.332
Eas Buidhe	999	1.857	1.958	2.27

tends to become less predictable when one wishes to investigate larger watercourses with older names such as the Spey and the Tay. This is for two reasons:

1. Smaller watercourses seem to have more predictable scores, and seem to be bunched together more closely.

2. There is a much larger data-set of smaller watercourses than larger watercourses, so any inference is likely to be more accurate.

The methods used in data mining can also be applied to more specific factors. As will be displayed later, it is possible, for instance, to apply the same method, but using *specelem* and *genelem*, or by generating scores for particular specific elements, rather than just their semantic group. The data in table 2.12 are actually no longer extant. That table was made at an earlier stage and has been updated in line with the findings from the data mining.

2.7 The Hierarchical Network

The hierarchical network when applied to hydronymy is the way in which the linguistic qualities interact with each other concerning the specific position of the watercourses they represent. The column *posid* in the database represents the *id* of the watercourse into which the given watercourse flows. With this for instance, it is possible to count the number of combinations of a watercourse with a specific linguistic stratum flowing into another watercourse of a different stratum (see table 5.8 on page 201). This is somewhat different to *posscore* in that it will show a watercourse's direct neighbours, which *posscore* will not do.

In each chapter in the rest of this thesis a section will be included discussing the structure of combinations of the given linguistic quality. Whilst this does not contribute directly to the score methodology, it is extremely useful in uncovering how hydronyms interact within the hierarchical network of tributaries.

2.7.1 Other Applications

The compilation of *lingscore* and *geogscore* is not the only application of these data. It is of course possible to analyse comparisons between specific scores, such as, for instance, *altscore* and *semtype*, that is, altitude and semantic type. Of course, not all comparisons are appropriate, so listed in table 2.13 are the ones marked that make sense. The columns and rows in the table below do not have an exact correspondence to the fields in the database, but instead represent general concepts for study. Each following chapter will tackle one linguistic quality, and will be studied in respect of its comparison with other factors.

Table 2.13: Valid Comparisons for Study

	Land classification	Location	Length	Position	Tributaries
Generic element	Yes		Yes		Yes
Linguistic stratum	Yes	Yes	Yes	Yes	Yes
Phonology / orthography		Yes		Yes	
Semantic	Yes	Yes	Yes		

2.8 Conclusion

These few case studies have shown that despite the possibility of showing issues with the overall model, it is certain that at its core the method creates meaningful results, or at least shows meaningful patterns. The remaining chapters of this thesis take each component of *lingscore* in turn and uses the methodology discussed above to compare aspects of the given component with other factors, and in doing so, make a contribution to established knowledge in the field of hydronymy.

Chapter 3

Generic Elements

3.1 Introduction

In this chapter, an investigation will be undertaken into the principles which influence the relationships between generic elements and the components of *lingscore* on the one hand and the other components of *geogscore* on the other. In other words, for a given generic element (such as *burn*), all occurrences of RNs with this element will be studied, by looking at the RNs' other linguistic qualities (such as semantic content of the specific) and by looking at the physical qualities of the watercourses they represent (such as their altitude). The methodology used in the previous section will be used as a basis. The overall aim here is to use visualisation techniques to show what conceptual space the various generic elements occupy within the hydronymicon.

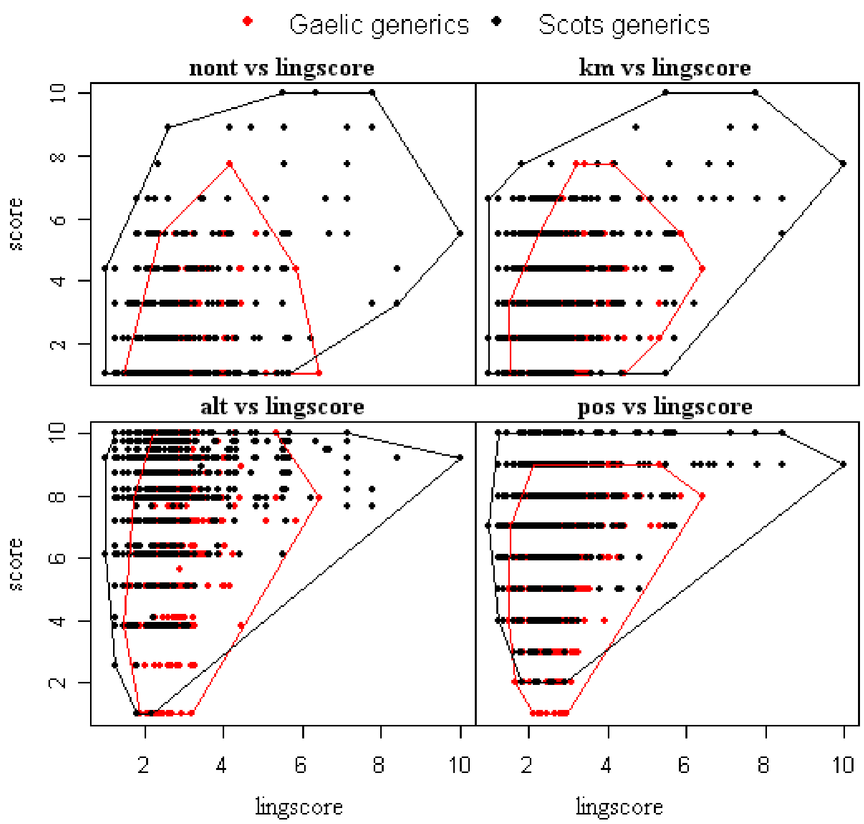
The sections in the first part of this chapter comprise the various linguistic strata. Within each of these sections is a survey of the various generic elements belonging to that stratum. Between each of these sections is a discussion of the relationship between the linguistic strata: for instance, a comparison between certain Gaelic and Scots generic elements. Whilst traditional etymologies and linguistic equivalents for each generic element are discussed, the focus is on the necessary and sufficient conditions for each element's attachment to any given watercourse. This chapter will also take into account the syntactic relationship between generic and specific elements (as expressed in the *genelem3* column in the database). For example, within the section for the element *burn*, the differences between *Burn of* X and X *Burn* constructions are discussed.

Following this is an investigation of all generic elements, comparing them to components of *geogscore*. The comparison of generic elements with the *specelem* and *semtype*, the other components of *lingscore* are not dealt with in this chapter. The comparison of *specelem* and *genelem* and the comparison with *semtype* are dealt with in following chapters.

3.2 Some Conceptual and Methodological Considerations

In order to compare generic elements and to discern the conceptual space they occupy, I wrote an R script called *lingscorecomparer*. This script queries the database to retrieve a certain set of RNs with qualities in common and then generates four graphs. Each of these plots the *lingscore* for each RN which has the given parameters against a different component of *geogscore*, that is: *nontscore*, *posscore*, *altscore* and *kmscore*. As such each graph is a version of graph 2.11 on page 40 but instead of the y-axis representing *geogscore* in general, four graphs each represent the four components of *geogscore*. A line is drawn round the outer points creating a convex polygon. This polygon represents the conceptual space occupied by the generic element. In most cases two elements are plotted against each other for comparison. An example in graph 3.1 simply plots Gaelic generic elements against Scots generic elements.

Figure 3.1: Comparison of Gaelic and Scot Generic Elements



The x- axis represents the *lingscore* of that name as explained in chapter 1. The y- axis, depending on the graph, represents one of the components of *geogscore* as explained in chapter 1. To interpret this graph, a number of factors should be taken into account:

1. Since *nont* and *km* are necessarily linked (i.e. there is usually a close correlation between the

length of a watercourse and the number of tributaries it has), the two top graphs are usually similar. Of course, if they are not similar, this is in itself worthy of study.

2. The same can be said of *alt* and *pos* (in this case, the factors are linked because altitude generally decreases as one approaches the coast).

3. The x-axis, *lingscore* is not of such importance as the y-axis. Its main use is to separate the dots out.

4. The y-axis, the component of *geogscore* is the main area of study here.

If our attention is turned to the example in figure 3.1, it can be seen that the extent for the *nontscore* and *kmscore* graphs, the coverage of Gaelic generic elements is entirely within that of Scots. Moreover, the Gaelic range is situated at the smaller extent of the Scots range. This means that ignoring its location, smaller watercourses can have either a Scots or Gaelic generic element, but larger watercourses have only Scots generic elements. This intuitively concurs with what is known: larger watercourses have Sc *river* and Sc *water* for generic elements, yet smaller watercourses can have any number of Gaelic or Scots generic elements, G *allt* and Sc *burn* being the commonest.

Looking at the lower graphs, it can be seen that almost the same situation exists, but that there is a larger number of names with Gaelic generic elements that stand outside the range of Scots generic elements. The dots representing names falling in this zone are essentially those further from the coast and in more mountainous areas, precisely where one finds a greater occurrence of purely Gaelic names. The names falling within the Scots zone and outside the Gaelic zone are of course those names closer to the coast, on lower lying ground, in traditionally Scots / SSE speaking areas, outside the Highland line.

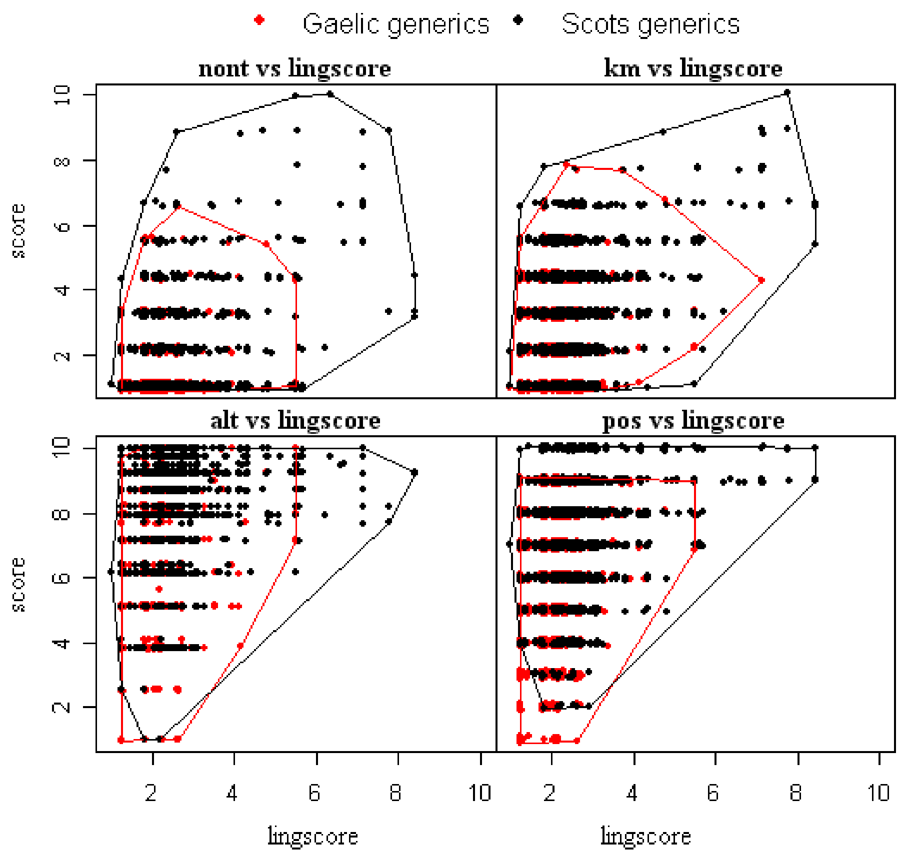
This set of graphs can of course be used for more than generic elements. One can input any of the components of *lingscore* or *geogscore*, or any combination thereof. Concerning the interpretation of these and other graphs in this thesis, it would be possible to use a more complex set of statistical tools to measure the similarity between elements. In these cases I have largely left the interpretation to the eye as opposed to creating another tool to quantify any relationship. This course was taken chiefly because it is unclear how far down this path this nascent methodology should be taken. As stated above, one needs to employ good toponymic sense alongside these tools in order to avoid pitfalls. In the future, once it can be ascertained how and where these pitfalls might appear, extra statistical tools may be employed. For now it is left to the reader's eye to decide how similar the elements are in these graphs.

There are two drawbacks to the graph above, which the R script can remedy. Firstly, if more than one dot inhabits the same space, one cannot see how many there are. There could potentially be a large number of dots in one area. This can be seen above, where there are lines of black dots which obscure the red dots behind it. To remedy this, a jitter has been added. This randomly moves the dots slightly within a very small area, which then makes it obvious how many dots there are within a given area. The red dots which sit behind the black dots have a slightly larger jitter range to compensate.

Secondly, as mentioned above, the x-axis represents the *lingscore*, which is itself an average of the three components, *specelemscore*, *semtypescore* and *genelemscore*. In the example above however, *genelemscore* itself is being tested. In some cases it would be desirable to remove the *genelemscore* from the *lingscore* equation. The *adjust* function will do this. In the above case this is not as important, but often is when comparing specific etymologies from different linguistic strata, where one does not want to factor in the *specelemscore*, but look purely at the meaning (see figure 3.2 below).

A final function which is not a drawback *per se*, is the ability to remove outliers. If there is a particular dot which is further out of range of the rest of the set, this can be removed. The algorithm for this is as follows: If the maximum number in a given data range (such as *kmscore* or *lingscore*) only occurs once in that data range (i.e. if the maximum is 9.76 and only occurs once), then this is removed. The same is done for the minimum in the data range. This is done for *lingscore* and all components of *geogscore*. This function should be used with care, since the existence of an outlier should be explained rather than removed. Figure 3.2 is the same as figure 3.1 but with the functions above applied. In reality, one would probably not want to remove the outliers, but it is shown here for explanatory purposes.

Figure 3.2: Comparison of Gaelic and Scot Generics with Jitter



As can be seen, the jitter makes it clearer just how many dots occupy a particular point, and whether red dots are behind them. The dot with a large *lingscore* (River Almond) has been removed, giving a smaller polygon (but at the cost of accuracy). The data have also been adjusted by ignoring *genelemscore*; this is actuated by all the Gaelic data points shifting 'left'.

In the rest of the graphs of this type, the default will be to adjust the graph. Outliers will not be removed unless otherwise justified and stated. It will be obvious where jitter has been used, but generally will only be present where the data-set is large.

3.3 P-Celtic Generic Elements

This is a problematic section with a number of pitfalls. Firstly, the amount of data for P-Celtic names is far less than for more recent names, and on the whole the methodology used here depends on large amounts of data for accuracy. Secondly, since generic elements are less resistant to linguistic change, and Pictish and other Scottish P-Celtic languages are extinct, no extant name has a P-Celtic generic. Thus one can only know about them through old records and pleonasms. This necessarily alters our approach in that the majority of the P-Celtic names of smaller watercourses must have been utterly forgotten, and where they do survive, do so only in their specific elements. That there were such names can be confirmed from, for instance, a mention of *Dobur Artbranani* from the Life of St Columba,[1] which seems to refer to a small (albeit unidentified) watercourse which is named after a specific person (No hydronym coined from a specific person survives from this stratum). It should be understood then, that for P-Celtic generic elements, this method is not as useful as it is for later linguistic strata.

This section also seeks to unravel the relationship between P-Celtic generic elements and certain ancient hydronymic specific elements, for example the difference between G *abhainn*, P *afon* and the various watercourses called Avon or variants.

3.3.1 P Esk

Much has been written about the hydronym *esk* and its cognates within the British Isles.[2] This discussion does not seek to address these issues in detail, but to investigate specifically the Scottish instances within the AOS where this element appears only twice, as River South Esk and River North Esk. These RNs have generally been identified as being part of a larger set of names, existing throughout Britain and Mainland Europe.[3] Discussion of the original meaning and role of this element goes beyond the remit of this section. Nicolaisen,[4] however, separates the Scottish 'River Esks', and some Irish hydronyms as being cognate with OI *esc*, 'water' and G *easg*, 'fen' (although this word is poorly attested in Old Irish and Gaelic). For present purposes, it remains to be seen if the hydronyms

[1] Adomnan of Iona, *Adomnan's Life of St Columba* (Edinburgh, 1961), pp. bk 1, ch.33.

[2] For example Williams Caerwyn, 'WYSG (river-name), WYSG, HWYSGYNT, RHWYSG', *Studia Celtica* 21 (1990), p. 670-678.

[3] A. L. F. Rivet and C. Smith, *The Place-Names of Roman Britain* (London, 1979), p. 376.

[4] W. F. H. Nicolaisen, 'Die alteuropäischen Gewässernamen der britischen Hauptinsel', *Beiträge Zur Namenforschung* 8 (1957), p. 241-42.

in question relate to these terms or belong to the larger, older set of hydronyms. It may be that the element *esk* is present in other instances. RNs such as Burn of Deskford, Deskie Burn and settlement names such as Desky may well represent a P-Celtic **dubh esc*, but could also simply represent G *dubh uisge*.

These names cannot unfortunately be compared to the Esk names in Britain, since they are obviously outside the AOS. One can see however, that they have one thing in common: all the known incidences of ‘Esk’ flow through relatively low ground, and are near the sea.[5] Whilst this evidence may suggest a relation between Scottish Esk and the other Esks, on linguistic grounds the names are somewhat different. The non-Scottish names are all clearly specific elements. The early record of names such as Isca Dumnoniorum for Exeter,[6] (also containing this element in some form) suggest that Isca was a specific element rather than a re-analysed generic. If the names such as Desk mentioned above indeed contain the element in question, it is clearly acting as a generic in these cases. Much the same can be said for North and South Esk, which now have specific elements of North and South, but once had different generic elements as old forms for South Esk show:

aquas de Tay et *Suthesk* (1370 RMS i no. 308)
aquas de Northesk et *Suthesk* (1370 RMS i no. 313)
South Esck flu (1595 Mercator Scotiae Regnum South)
Southesk (c. 1591 Pont text 139r)
South Esk R (c. 1591 Pont map 29)
South-Eske River or Esken-Duy, Eskenduy R: (c. 1591 Pont map 30 front)
Inveriskandie = *Inneskandie* (1638 Retours (Forfar) no. 242)
Esca Australis (1654 Blaeu text p.84)
aquas de Northesk et *Eskandie*, Innereskandie (1699 Retours (Forfar) no. 553)
water of *Esque* (1751 Macfarlane Geog. Coll. V1 p275)

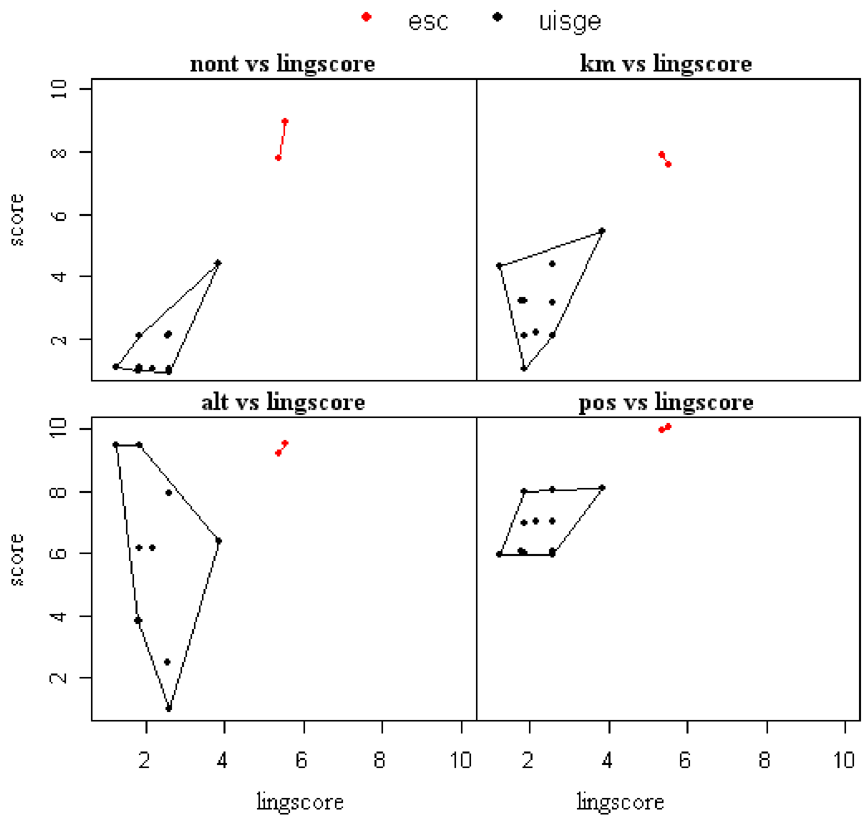
These forms suggest the old name of River South Esk was once called Eskandie. This could reflect two underlying forms: The first would be a genitival construction such as **esk an die* or **esk-an duibh(e)*. The second element could represent a slenderised form of G *dubh* or a P-Celtic cognate, as possibly seen in the RN Divie (possibly from **dub-iā*[7]). The second situation could be that the generic *esk* here has a suffix of some sort, in the same development as *aβ* <*aβonā* such as *esk-on* or *esk-an*. Regardless, this name suggests that the South Esk once had a specific element, thus suggesting Esk was once a generic.[8]

Once *Esk* has been posited as originally being a generic element, it would be prudent to compare these names with G *uisge*, ‘water’, a term whose relationship to P *esk* is unclear. Figure 3.3 on page 55 shows the differences between G *uisge* and the three names in question. This suggests that whilst the derivation for these words may be etymologically linked to G *uisge*, they do not take up the same

[5]Within the terminology of this thesis, one would say they have high *altscores* and *posscores*, although, being outside the AOS, they have not been measured.
[6]Rivet and Smith, *The Place-Names of Roman Britain*, p. 378.
[7]See Watson, *The Celtic Place-Names of Scotland*, p. 438 and Eric Hamp, ‘DOUBS’, *Études Celtiques* XXV (1988) with Eric Hamp, ‘Varia’, *Scottish Gaelic Studies* XVI (1990), p. 193 for more discussion of this element.
[8]The same situation exists with North and South Ugie.

conceptual space within the hydronymicon.

Figure 3.3: Comparison of G *uisge* and P *esk*



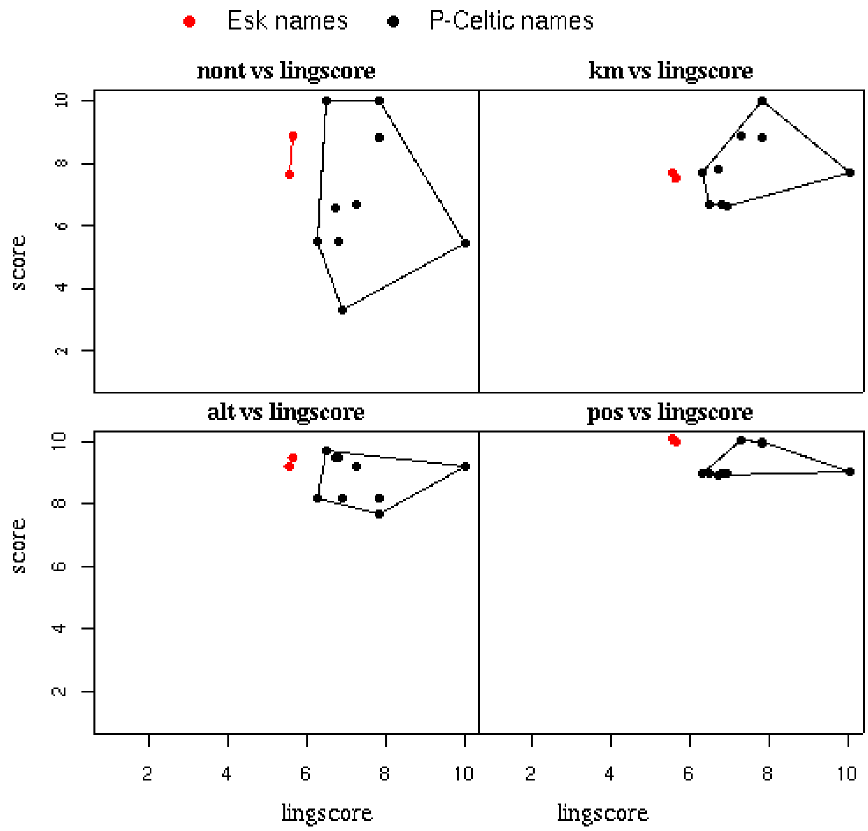
If these names are not Early Celtic and not versions of G *uisge*, then it is sensible to suggest a P-Celtic derivation for these names. Within the P-Celtic hydronymicon, it seems that this generic element relates to relatively large watercourses (with perhaps the exception of the relatively low *nontscore* of *Desk). Figure 3.4 shows the relationship between the Esk names and other names of P-Celtic origin. Although the two polygons do not overlap, they are on the same horizontal plane, which is the geographical y-axis, showing that *esk* names have similar physical qualities to P-Celtic names, perhaps suggesting *esk* names are of P-Celtic origin. The fact that the polygons differ on the x-axis is because, as it stands, the database reads *esk* names as being of Old Celtic origin, and thus differentiates the names from P-Celtic names.

3.3.2 P Dobhar

Dobhar seems to have been the default generic for a watercourse in pre-Gaelic times. As mentioned above, in the Life of St Columba,[9] a small burn is mentioned as *Dobur Artbranani*, perhaps Scotland's first native hydronymic generic element.

[9] Adomnan of Iona, *Adomnan's Life of St Columba*, pp. bk 1, ch.33.

Figure 3.4: Comparison of P *esk* and all P-Celtic RNs



Semantics: This element of course extends beyond the bounds of the AOS, and thus outwith the bounds of this enquiry. Table 3.1 on page 57 shows the specific elements used with *dobhar* from the AOS, and whilst there are other names with *dobhar* as a generic not mentioned above, this cross-section is a typical selection. The specific elements are all adjectival and are fairly typical of P-Celtic terms in Scotland in that they pertain primarily to adjectives, such as colour. The term *ail*, rock as seen in Alder is problematic. It could denote the qualities of the water or river bank, rather than denoting the particular 'stone' or nearby area of land. Alternatively the derivation is uncertain for this name, and moreover the term *ail* in a P-Celtic context is not well evidenced.

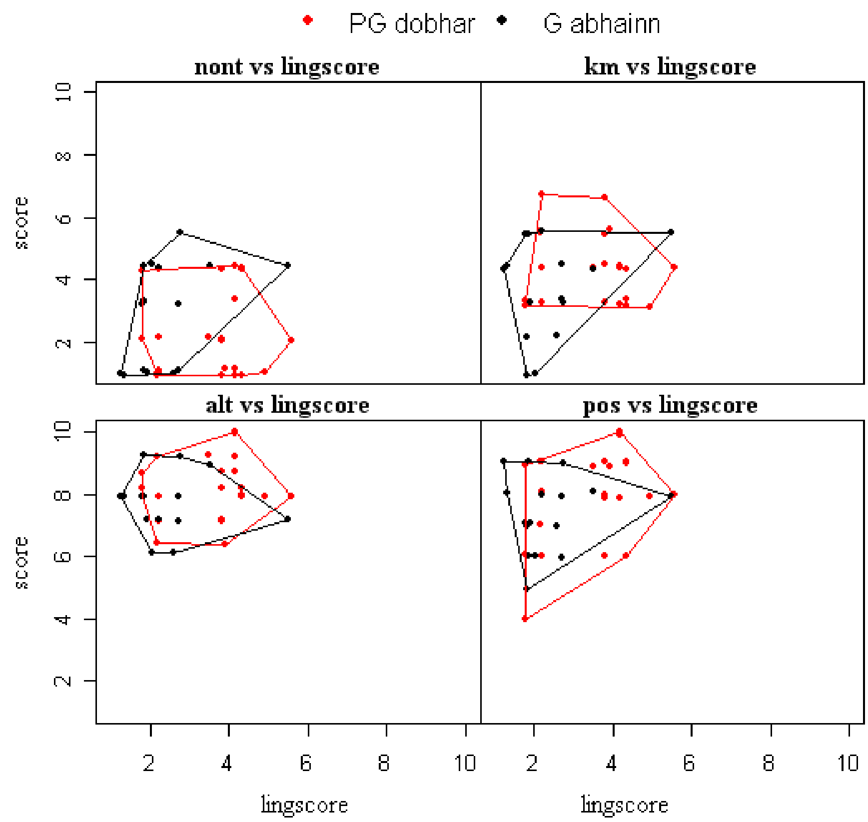
One term which is almost conspicuous by its absence is a cognate of G *dubh*, black. The closest synonym is the term cognate with G *ciar*, 'dark', which is relatively rare. This absence of a Pictish term for 'black water' may seem odd given the high incidence of the occurrence of this type of name in other languages in Scotland. This could perhaps be explained by the fact that *dobhar* and *dubh* are in fact etymologically linked. *Dobhar* derives from **dubro-* itself from OC **dubu-*, 'black'. *Dubh* also derives directly from this root. It is possible that *dobhar* originally meant itself 'deep, dark water', and that a P-Celtic and Early Gaelic equivalent of the combination *dubh* + *dobhar* would have been tautological, in the same sense 'Wet Water' would be in English.

Table 3.1: Elements Associated with P *dobhar*

etymology	frequency
P *cal*, 'hard'	2
PG *geal*, 'white'	2
P *fionn*, 'white'	2
G *àrd*, 'height'	2
PG *ciar*, 'dark'	1
PG *glan*, 'clean'	1
PG *ail*, 'rock'	1

3.3.3 P Dobhar vs G Abhainn

Figure 3.5: Comparison of P *dobhar* and G *abhainn*



A comparison of these two elements shows a close similarity in terms of physical features, as shown in figure 3.5 on page 57. This does not necessarily mean that G *abhainn* is a translation of P *dobhar* however, since *abhainn* is of course used in Ireland and in areas in Scotland where such a survival would be very unlikely, such as the Western Isles. It is more likely that they simply fill the same conceptual gap, having similar meanings.

3.3.4 P Glas

Table 3.2: Elements Associated with P *glas*

etymology	frequency
PG *fo*, ‘sub’	4
P **cun*, ‘lovely’	2
P *finn*, ‘white’	1
G *dubh*, ‘black’	1

Semantics: The situation is the same as that of *dobhar*; the names predominantly represent adjectives. The main exceptions to this are the names such as ‘Fowlis’, said to represent *fo* + *glas*, ‘sub stream’.[10] At first glance it seems that this name is in a class of names categorised by their relation to other features. It would seem strange that, whilst all other P-Celtic names relate to the qualities of the watercourse, these names relate to their relative position, sharing their conceptual space with names such as Allt Cùl and Back Burn. Whilst there is not enough information to ascertain the specific implication of *fo* in these names, it could represent the concept of a branch or tributary rather than a watercourse situated on low ground, that is, ‘sub’ in the sense of ‘subsidiary’ as opposed to ‘low’. Whilst there is not enough data at present to make such an assertion more strongly, it does fit with the little data there is, that in no cases do Fowlis type watercourses flow into the sea (i.e. they have a *nont* of 1).

3.3.5 P Glas vs P Dobhar

When comparing two generic elements from the same linguistic stratum, it is common (as will be seen below) that the conceptual space will overlap somewhat, but that both elements will have their ‘own’ areas. This essentially means that for two generic elements, there will be certain types of watercourses to which either element can be attached, but that there will also be some watercourses to which only one can be attached, usually dependent on size, i.e. some watercourses will be too small for a particular generic element, and some too large. This can be seen in figure 3.6, where in fact the *geogscore* and *lingscore* for elements *dobhar* and *glas* are remarkably similar.

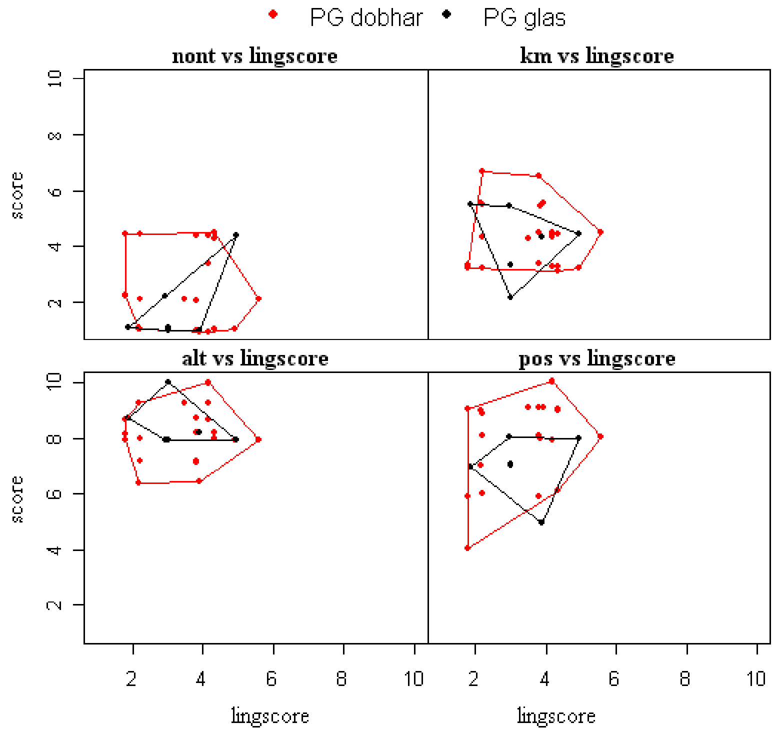
3.3.6 P *Pol, G Poll, Sc Pool and Sc Pow

Derivation These terms are all etymologically related, but nonetheless distinct, terms. According to Robinson, *The Concise Scots Dictionary*, Sc *pow* is a Scots borrowing from G *poll*, whilst Sc *pool* (or *puil*) is a native term from OE *pōl*. The Celtic origins of G *poll* and P **pol* are not clear. They exist as terms in OI and Welsh, but the fact that the OI term contains a [p] means it cannot be native.

Semantics: P-Celtic *pol* and its Scots version *pow* have been discussed as being described as a ‘sluggish stream’, ‘often referring to a Burn which moves slowly through peaty ground of tidal salt

[10] Watson, *The Celtic Place-Names of Scotland*, p. 458.

Figure 3.6: Comparison of P *glas* and P *dobhar*



marshes'.[11] Dwelly[12] glosses G *poll*, 'pond, pool, bog, deep stagnant water, dark and deep part of any stream'. *Pow* is described as 'a slow-moving, ditch-like stream, flowing through carse-land'.[13] *Pool* is not listed as a Scots term in the Dictionary of the Scots Language.

Whilst the interaction between these terms is complicated and largely beyond the scope of this thesis, figure 3.7 shows that Sc *pow* and P *pol* are more akin to each other than G *poll*, which represents a *pool*. The Sc term *pool* then should be compared to G *poll* check for similarity as in figure 3.8 on page 61, which shows a close correspondence. This concurs with previous work[14] which suggests that Sc *pow* is a continuation of P *pol* rather than the Gaelic cognate. Within the AOS it can be seen that in a few circumstances names with *pow* either as a generic or specific element have a certain P-Celtic colouring. The two most robust examples are The Pow (5158) which is the RN associated with Aberfoyle, and Pow Water (6395) which is on record as Polpefferie. The specific element *pefferie*, almost certainly reflects a P-Celtic element cognate with W *pefr*, 'shining'.[15]

Table 3.3 starting on page 63 shows all the names in the database with the Scots generic *pow* or

[11] G. W. S. Barrow, *Uses of Place-Names* (St Andrews, 1998), chap. The Uses of Place-names and Scottish History - Pointers and Pitfalls, p. 59.

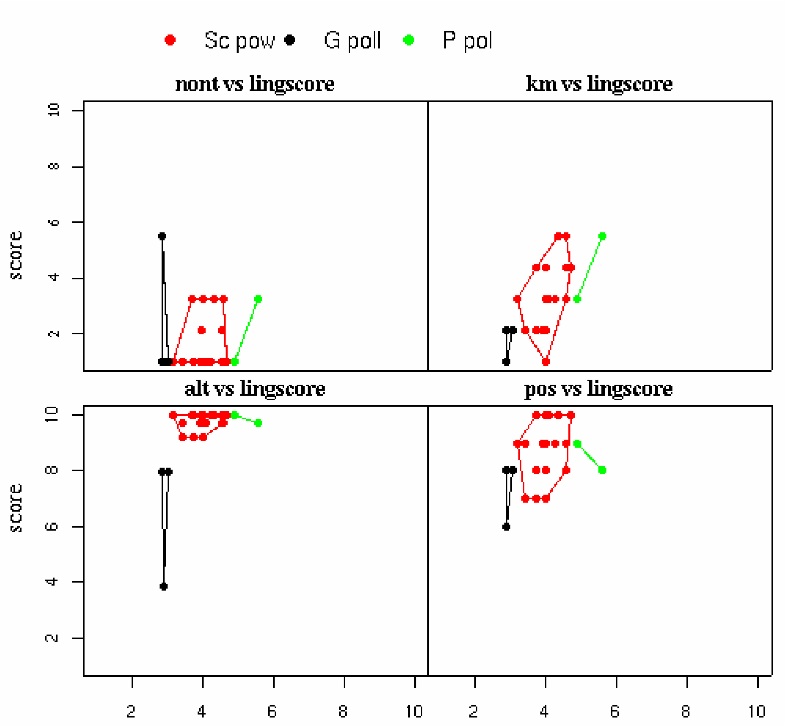
[12] E. Dwelly, *The Illustrated Gaelic-English Dictionary* (Edinburgh, 1901-11).

[13] 'Dictionary of The Scots Language'.

[14] See Barrow, 'Uses of Place-Names', p. 59-61 for further discussion outside the AOS.

[15] Watson, *The Celtic Place-Names of Scotland*, p. 452 discusses this term, although he translates it as 'radiant, beautiful'; 'shining' is a better translation.

Figure 3.7: Comparison of G *poll*, P *pol* and Sc *pow*



P-Celtic generic *pol*. For each name it shows the soil conditions for the lowest and highest points of the watercourse according to the Soil Classification System as in shown table B.4 on page 246. It also shows *altave*, a column in the database denoting the derived average altitude of a watercourse as explained in section C.1 on page 258. This table shows that they flow through similar types of land as suggested by figure 3.7.

Figure 3.8: Comparison G *poll* and Sc *pool*

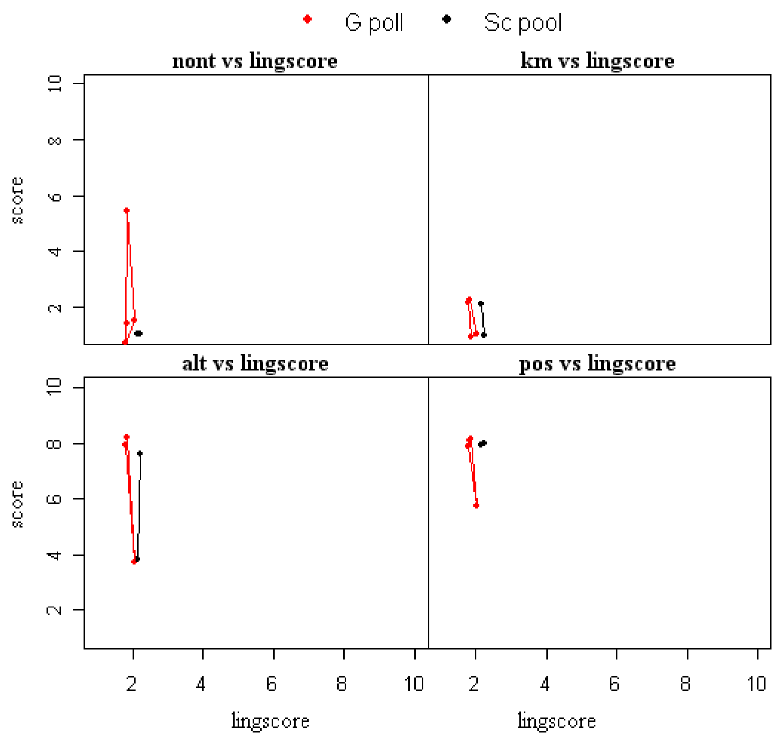


Table 3.3: Occurrences of Sc *pow* and the Soil Classification of the Land Through Which they Flow

RN	to	from	*altave*
Bogmill Pow <	Warm rather dry lowland	Warm rather dry lowland	100
Cairnie Pow	Warm rather dry lowland	Warm rather dry lowland	100
Pow of Clackmanan	Warm moist lowland	Warm rather dry lowland	100
Deepston Pow <	Fairly warm wet lowland and foothill	Warm wet lowland	200
Dipple	Warm rather dry lowland	Warm dry lowland	100
East Pow	Warm rather dry lowland	Warm rather dry lowland	100
Pow of Errol	Warm rather dry lowland	Warm rather dry lowland	100
Erskine Pow	Warm rather dry lowland	Warm rather dry lowland	100
Pow of Glencarse <	Fairly warm moist lowland and foothill	Warm rather dry lowland	200
Grange Pow	Warm rather dry lowland	Warm rather dry lowland	100
High Moss Pow	Warm rather wet lowland	Warm moist lowland	100
Pow of Inchaffray	Fairly warm rather wet lowland and foothill	Warm moist lowland	200
Inchmartine Pow	Warm rather dry lowland	Warm rather dry lowland	100
Pow of Lindores	Fairly warm moist lowland and foothill	Warm moist lowland	200
Little Pow	Fairly warm dry lowland	Fairly warm dry lowland	100
North Pow	Fairly warm rather dry lowland	Fairly warm dry lowland	100
Polpefferie	Fairly warm rather wet lowland and foothill	Warm moist lowland	200
Pole Burn	Fairly warm rather wet lowland and foothill	Warm rather wet lowland	200
Poles Burn	Fairly warm moist lowland and foothill	Fairly warm moist lowland and foothill	200
Pow Burn <	Fairly warm moist lowland and foothill	Fairly warm moist lowland and foothill	200
Pow Burn	Fairly warm moist lowland and foothill	Fairly warm moist lowland and foothill	200
Pow Burn	Fairly warm moist lowland and foothill	Fairly warm moist lowland and foothill	200

Table 3.3: Occurrences of Sc *pow* and the Soil Classification of the Land Through Which they Flow

RN	to	from	*altave*
Pow Burn	Fairly warm rather wet lowland and foothill	Fairly warm moist lowland and foothill	200
Powburn	Fairly warm moist lowland and foothill	Fairly warm moist lowland and foothill	200
The Pow	Fairly warm wet lowland and foothill	Warm wet lowland	200
Pow Burn	Warm moist lowland	Warm moist lowland	100
Pow Burn	Fairly warm rather wet lowland and foothill	Fairly warm rather wet lowland and foothill	200
Pow Burn	Fairly warm dry lowland	Warm dry lowland	100
Powgavie Burn	Warm rather dry lowland	Warm rather dry lowland	100
Powie Burn	Warm dry lowland	Warm dry lowland	100

3.3.7 P Aβon

A possible candidate for a P-Celtic generic is the term cognate with Welsh *afon* and G *abhainn*, both meaning, 'river'. A derivative or cognate of this element also appears throughout Britain as the specific 'Avon'. The only definite example of this within the AOS is River Avon (282) (a tributary of the Spey) so that there is not enough data to study this element using the usual methodology. That said, a discussion of this River Avon shows a number of interesting points.

Inveravon = *Inuerouen*, Strathavon = *Strathouen* (1187-1203 REM no. 16)
Inveravon = *Inuerhouen* Strathavon = *Strathouen* (1206-23 REM no. 50)
Strathavon = *Strathouen* (1208-15 REM no. 46)
Strathavon = *Strathouen* (1215 REM no. 47)
Inveravon = *Inuerhouen* (c. 1215 REM no. 49)
Inveravon = *Inuerouen* (1224-42 REM no. 62)
Strathavon = *Strathouen* (1224-42 REM no. 81)
Ain R: (c. 1591 Pont map 26)
Avin river (c. 1591 Pont text 138r)
Inveravon = *Innerawine* (1621 Retours (Forfar) no. 130)
super aquas de Die, Dy, Feuch et *Aven* (1637 Retours (Kincardine) no. 70)
Strathaven, *Glenaven* (1638 Retours (Banff) no. 78)
water of Awen (1656 Retours (Banff) no. 105)
aquam de Aven (1662 Retours (Banff) no. 112) *Awen* (1662 Blaeu text p.106)
Strathavon = *Strath-aviniam* (1662 Blaeu text p.110)

Nearly every other occurrence of the name seems to be situated along a route that could have reasonably been considered once to have been a boundary between speakers of Old English and of insular P-Celtic. The implication here is that 'it may be that some of these rivers had qualifying names[16] which were not taken into English'[17] The name 'Avon' in Scotland applying to this watercourse must have been first used at a date long before any speakers of any Germanic language were present and thus cannot be said to be a pleonastic name of some sort.

It could be claimed that this Avon simply reflects an underlying G *abhainn*, either originally as a simplex or as a generic with a now lost specific. The existence of Inveraven, suggests that that neither of these explanations are likely, since Inver- names in the AOS in all other cases only reflect the specific if it is Gaelic. Moreover, the existence of Loch Avon would be difficult to explain in this case, since it would reflect an underlying *Loch Abhainn, 'Loch River' which would be nonsensical. It is also exceptional for a Gaelic RN to reflect only a generic element. Simplex names effectively do not have a generic element, but the opposite is rarely, if ever, found.

These points suggest that the River Avon in our AOS is substantially different from the other names, and that, rather than being a pleonastic RN, deriving from linguistic contact between separate linguistic strata, the element in Avon represents a *bona fide* P-Celtic hydronymic specific element. The following graphs corroborate this conclusion. Figure 3.9 on page 66 compares the Avon with the

[16] I.e. the specific element.
[17] Rivet and Smith, *The Place-Names of Roman Britain*, p. 239.

P-Celtic specific elements and shows that in all cases the Avon falls in or very near the conceptual space occupied by other P-Celtic *specific* elements. In the *nont* graph, the point representing River Avon could be seen to be outside the area of the polygon, if River Avon were not counted to make the polygon. This can be addressed in two ways. Firstly, it can be seen that the actual *nontscore* of the River Avon is the maximum, but it is equal to an existing maximum *nontscore* for P-Celtic names (the River Forth). Thus it does not as such stand outside the exisitng range of *nontscore*. Secondly, if the River Avon did actually stand outside this range, one would then have to make a decision as to whether this undermined the whole theory about the River Avon or not. When new numeric data are added into a system, the minimums and maximums will necessarily be altered as more data are entered, this does not necessarily mean they are invalid.

Figure 3.10 on page 67 compares the Avon to P-Celtic *generic* elements and shows the Avon to occupy a very different conceptual space to P-Celtic generic elements. The main difference is in graphs 3.9 and 3.10, showing that the Avon is much longer and has many more tributaries than other watercourses with P-Celtic generic elements. It could be claimed, of course, that the historical record and linguistic change has removed those P-Celtic generic elements, and no doubt this has occurred, but in cases where hydronyms have vanished, the available evidence and common sense suggests that it is the smaller watercourses that become renamed rather than the larger ones. As far as these graphs are concerned, under this phenomenon, if more information about P-Celtic hydronyms was discovered (for instance if the identity of *Dobur Artbranani* were found) these data would create dots to the left and below the polygons created by the P-Celtic generic elements and specific elements in each graph. This would not undermine the point here, since the dot representing the Avon is generally situated to the right and higher than the polygons.

The horizontal location of the dots representing the River Avon in these two graphs are different. This is simply due to the *adjust* function mentioned above on page 52. As can be seen, the values of the components of *geogscore* are identical in both graphs.

It is unclear what *aβon* as a specific element would mean, but this is hardly problematic, since many of these types of names must have had nuances in meaning lost to us now. Compare the equivalent *Abha*, found in such names as Loch Awe,[18] which is certainly a *bone fide* specific element.

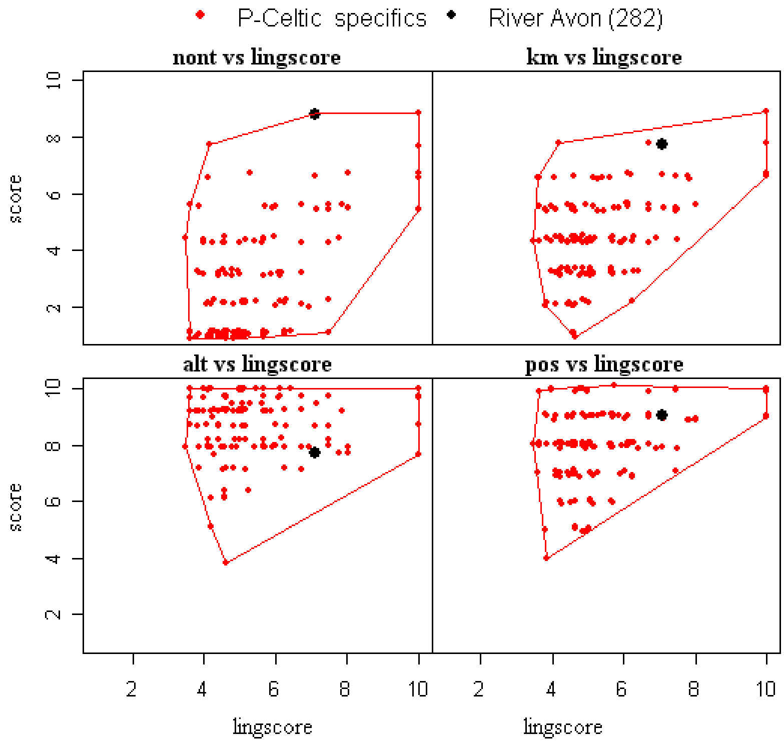
Another argument is negative: To the author's knowledge there is no old form from pre-Gaelic times which clearly shows *aβon* being used as a generic element. The usage is also not in evidence indirectly, through reconstructions of RNs and so on. For instance, P *dobhar* is inferred from a number of Aber- names, such as Aberarder and Aberchalder, but *aβon* is never inferred in this way.

This name has a number of similarities with the River Avon in Gloucestershire:

- Every other occurrence of this name in Britain represents a watercourse that flows into the sea . This Avon is a tributary of the Spey whilst the Gloucestershire Avon is also a tributary of a large river, in this case the Severn (in the terminology of this thesis it has a *pos* of 2).
- The pronunciation of the Speyside Avon is *aːn* sometimes written as A'an. Watson gives a

[18] Watson, *The Celtic Place-Names of Scotland*, p. 477 which also gives an example of *Abhainn Abha*.

Figure 3.9: Comparison of River Avon and P-Celtic specific elements

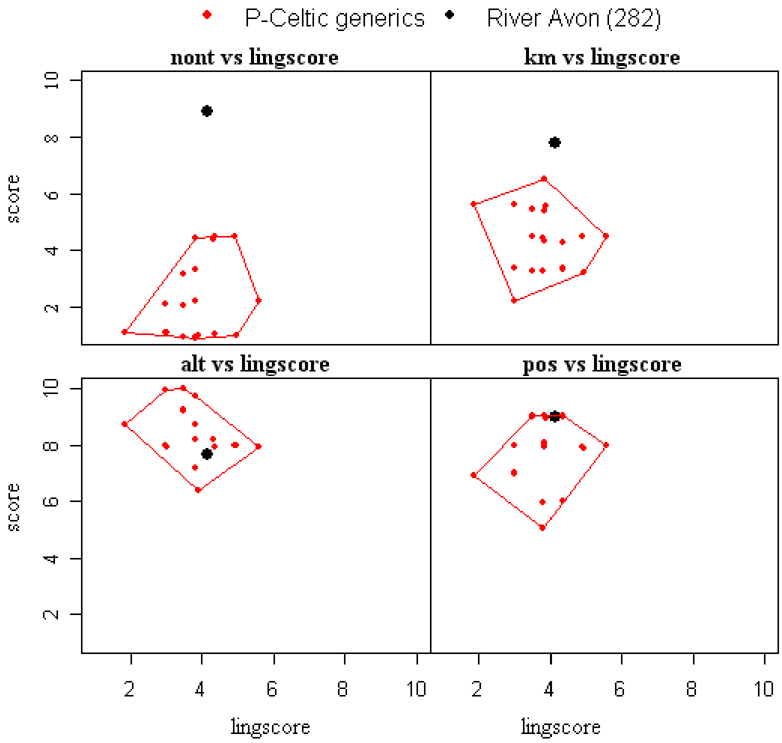


Gaelic spelling of Ath'inn.[19] These pronunciations possibly point to a different suffix than the one occurring in *aβonā*. There is evidence that the Gloucestershire Avon also derives from a non *aβonā* form. A Ravenna form identifies this river as ELAVIANA, which is usually reanalysed as Fl(uvia) Aviana, i.e. River Avon.[20] Whilst the reanalysis of EL to Fl(uvia) is orthographically justified, the alteration of the final vowel from *-iana* to *-ona* is not. Since the modern pronunciation of the Scottish Avon and the reconstructed forms of the Gloucestershire Avon belong to some of the earliest and latest strata of evidence, there is not enough evidence to form a definite opinion. It is possible, nonetheless, that these two names are specific variants of the *aβonā* name, unique to the British Isles.

This evidence suggests that the P-Celtic languages of the Old North and Northern Scotland did not use *aβon* as a generic at all, but only that form, or one similar to it, as a specific. This is in contrast to the rest of Britain, which also had this term (as suggested by the Gloucestershire Avon) but also had names in Avon created along linguistic boundaries.

[19]Watson, *The Celtic Place-Names of Scotland*, p. 451.
[20]Rivet and Smith, *The Place-Names of Roman Britain*, p. 239.

Figure 3.10: Comparison of River Avon and P-Celtic Generics



3.3.8 Preposed P-Celtic Generic Elements

In the cases where names have surviving P-Celtic generic elements, all bar one name (Polpefferie)[21] have a preposed construction, i.e. the generic element comes last. Such names as Douglas (a P-Celtic cognate of G *dubh glas*), or *Arder from Aberarder (a P-Celtic cognate of G *àrd dobhar*) are by far the commonest type of construction. This is the reverse situation from most Gaelic names (where, for example, Allt Dubh is more common that Allt Dubh); the reason for this is not clear. Nor is the relationship clear between these P-Celtic names and preposed Gaelic constructions (e.g. Dubh Allt) which are generally thought to be older than the later sort (e.g. Allt Dubh).

3.4 Gaelic Generic Elements

In order to construct a Gaelic generic element system, it is appropriate to start with *allt*, the commonest, default term for a watercourse. Other elements are then discussed in terms of their relationship with this term.

[21] Outside the AOS there is of course also *Dobur Artbranani*, mentioned above.

3.4.1 G Allt

Semantics and derivation: This term is usually derived from OC **alto-* meaning 'height, cliff or hill'. Eric Hamp, however, convincingly suggests the correct underlying form is **allto-*, with the OI form *allt* as opposed to *alt*, which is consistent with the modern Scottish Gaelic form.[22] This term chiefly means 'cliff' in OI and Irish, which is a secondary meaning in Scottish Gaelic. In the P-Celtic languages it seems to predominantly mean 'hill', as in Welsh *allt*, Breton *pen-an-aout*, 'top of the hill' and Middle Cornish *als* 'cliff, slope', with a secondary hydronymic meaning of 'sea-cliff, beach'.[23] Despite the secondary developments within a P-Celtic context, it is only within a Scottish Gaelic context that it comes to mean 'watercourse'. Whilst there is evidence for this word in non-Celtic languages, it is interesting that as a Celtic word, there is only evidence for it in an insular context, except for the weak exception of the Breton example above. Whether this is an accident of survival or not is a matter of debate.

Whilst the conditions under which *allt* underwent this semantic development is unclear, the following points can be made: The topography of Argyll and the area around which the Irish first brought Gaelic to Scotland is of course very mountainous and has many sea lochs and watercourses. It is a geological fact that nearly all watercourses in this region run through steep glens with cliffs on each side, and that every steep glen with cliffs has a watercourse running thorough it. In this geological environment a word for a valley with cliffs and a word for a watercourse could have become synonymous.

Secondly, it should be pointed out that the productive equivalent lexeme in Welsh has followed the same semantic development: W *nant* developed from the meaning 'valley' into 'watercourse'.

Thirdly, *allt* meaning 'watercourse' should be considered a purely Scottish semantic development; that is, there is no evidence for it occurring in Ireland, and its distribution shows there is no evidence that the development was due to a P-Celtic substrate.

The original meaning of 'height' was also preserved as a secondary meaning in Gaelic in both a lexical and a toponymic context.[24] As Gaelic spread East, out of the highlands and into the lower lying lands, the usage of *allt* as a mountain stream seems to have started to lose its mountainous aspect, since there is evidence for its use in relatively flat lands, such as Fife when Gaelic was spoken there. In the present day *allt* is extant as a name only where it existed at the time of the compilation of OS maps, yet evidence exists for it in Fife and Angus and other areas where Gaelic has not been spoken for a long time, or was not spoken for very long. This process is called *metonymy*,[25] which employs 'a figurative use based on association'.[26]

Table 3.4 shows the ratio between extant and extinct names with *allt* for each catchment. The areas with the highest ratio or no existing names in *allt* are those largely outside the Highlands, but where

[22] Eric Hamp, 'Varia IV: Goidelic *alt* and *allt*', *Ériu* 43 (1992).
[23] Padel, *Cornish Place-Name Elements*, p. 4.
[24] Watson, *The Celtic Place-Names of Scotland*, p. 242-243.
[25] Alan Cruse, *Meaning in Language* (Oxford, 2000), p. 112.
[26] Another example of this would be 'wheels' to represent a car.

Table 3.4: The Ratio of Extant and Extinct *allt* Names for each Catchment Area

ratio	current *allt* names	extinct *allt* names	catchment
NA	0	1	Forth to Tay
NA	0	6	Tay to Dee
1.06	16	17	Don to Spey
.23	157	37	River Dee
.23	43	10	River Don
.16	62	10	River Forth
.11	718	79	River Tay
.07	21	4	Loch Lomond
.05	496	25	River Spey

Gaelic was nevertheless once spoken, i.e. Forth to Tay (largely Fife), Tay to Dee and Don to Spey (the East coast).

3.4.1.1 G Allt vs Altitude

Because of the original meaning of *allt* and its definition in some dictionaries as 'mountain stream', it would be useful to look at the altitude and soil classes of those RNs which have *allt* as a generic. Figure 3.11 shows the percentage occurrence of each altitude for Gaelic RNs with *allt*, and for all RNs as a comparison. Essentially it confirms that a greater proportion of *allt* names are at a high altitude when compared to the entire data-set.

Figure 3.11 is useful in that it proves that *allt* relates more to higher altitudes than lower ones, but it does not take area into account. The AOS straddles a latitude from wholly Gaelic speaking areas to mainly non-Gaelic areas. Since Gaelic spread basically from West to East, one might wish to make the same comparison across different latitudes as in figure 3.12 on page 71. Each point on the x-axis relates to a 10 km range of latitude, ranging west to east from left to right. It shows that west of about latitude 340000, *allt* has a higher altitude than other Gaelic names, but east of this the difference is reduced; this is probably because of geographical conditions: there is simply less altitude in Eastern Scotland than further West.

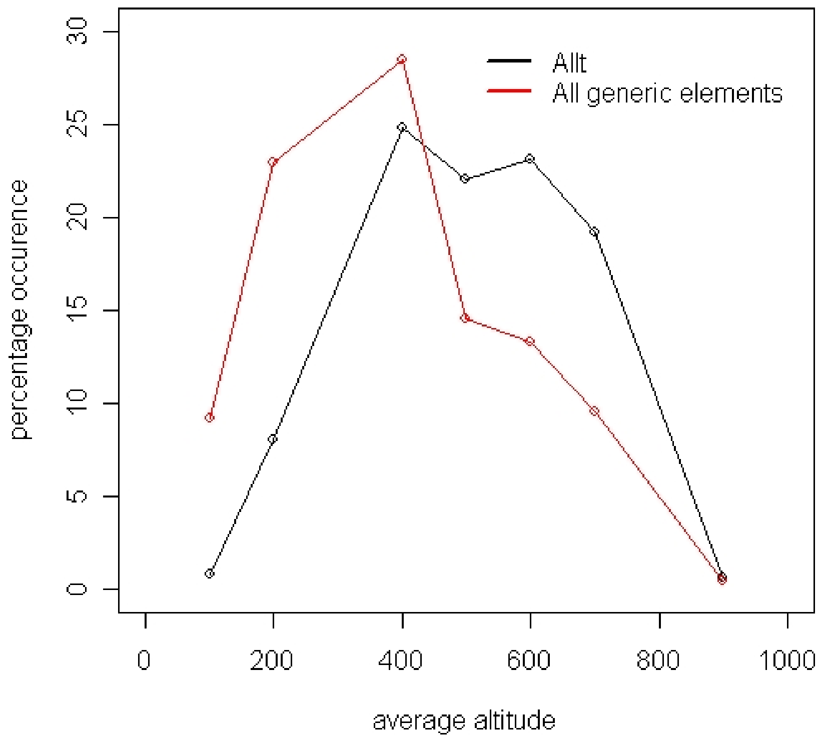
3.4.1.2 G Allt vs G Abhainn

In order to discern where in the hydronymicon *allt* stands, one should compare it to another generic element, in this case *abhainn*. In figure 3.13 on page 72 the components that comprise *geogscore* are compared for each RN where the generic element is *abhainn* and *allt*.

Geographical attributes: A number of points can be seen from figure 3.13:

- In nearly all cases the extent of *abhainn* is less than the extent of *allt*; that is, any RN with the generic element *abhainn*, could also have the generic element *allt*, but not *vice versa*.
- For *nontscore* and *kmscore*, the ranges are essentially the same between the two names, but tend to apply to larger watercourses in *abhainn* than *allt*.

Figure 3.11: Comparison in Altitude of G *allt* and other Generic Elements



- For *altscore* and *posscore*, *abhainn* only applies itself to watercourses with relatively higher scores. In plain English, watercourses with the generic element *abhainn* only flow through land at a lower altitude, and are generally closer to the sea in the tributary network, whilst names with the generic element *allt* are not restricted in this way, but occur at all altitudes and positions.

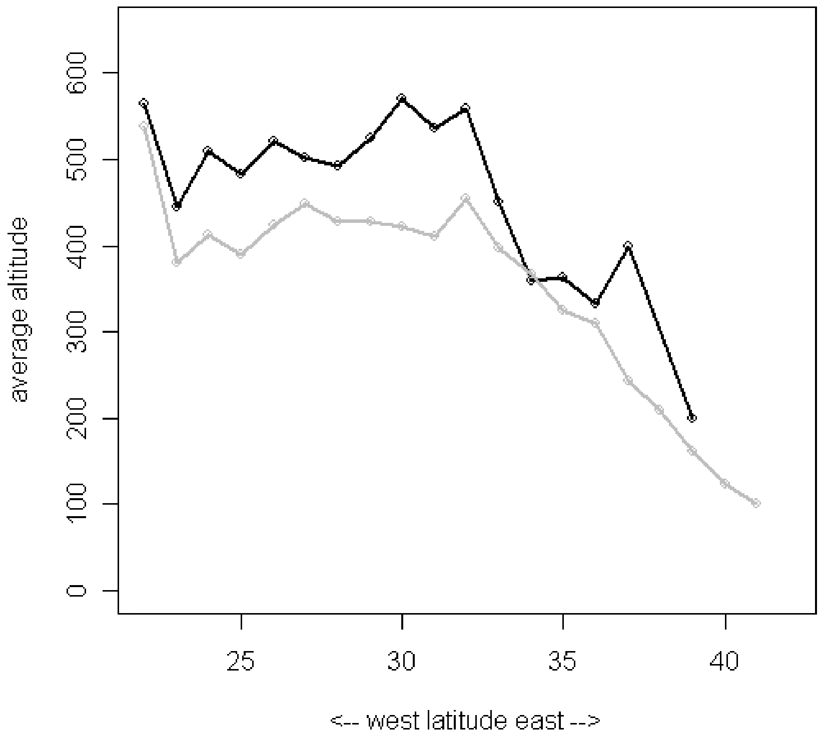
The only consistent *abhainn* name to lie outside the *allt* space here represents Abhainn Dubh as discussed in the previous chapter. It can be seen here again as a clear outlier. Figure 3.14 on page 73 shows the same data without Abhainn Dubh, showing a much more consistent situation.

3.4.2 G Uisge

Semantics: G *uisge*, simply meaning 'water', is a common term in G hydronymy, with the same meaning in the Gaelic lexicon as it is in the Scots / SSE translation. Within the AOS there are only eight occurrences of the name. Of these, four have the specific *garbh*, 'rough' and the remainder have colour terms as specific elements. This generic element has the highest altitude of all, but since this data are only based on eight terms it is not desirable to read too much into this.

The element *uisge* does not fit neatly into the generic element structure. Its low *nontscore* suggests

Figure 3.12: Comparison of Average Altitude by Latitude



it is more akin with names such as *caochan* and *lochan*, yet it has a narrow *pos* range (3-5). This can be seen in figure 3.15 on page 74, with *abhainn* as a comparison.

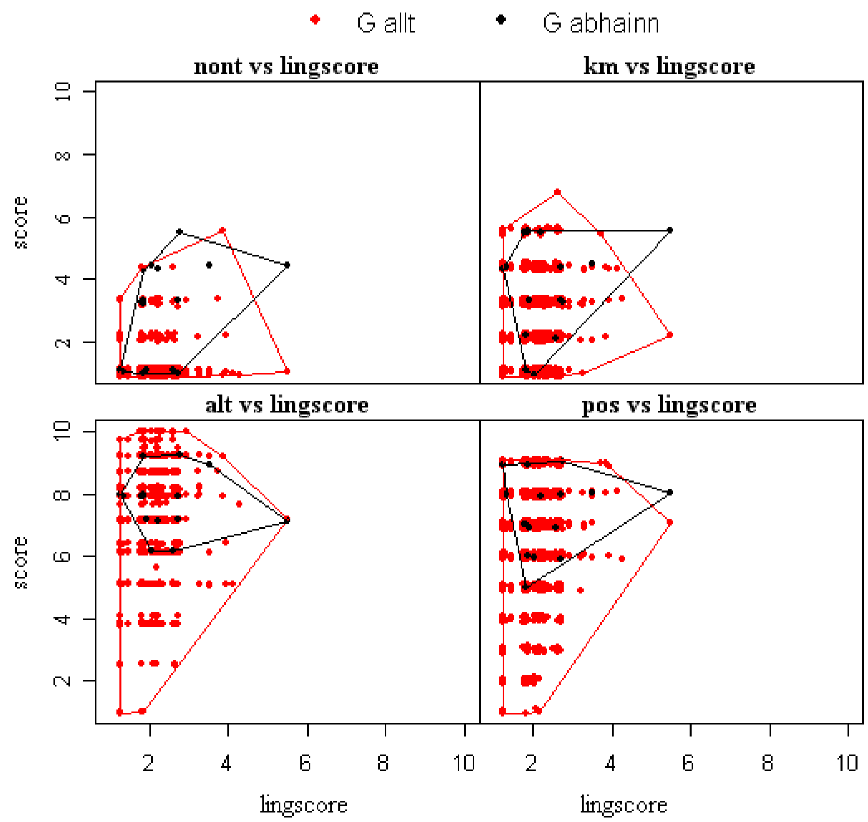
3.4.3 G Abhainn

Cognate with Latin **amnis*, the nominative in OI is actually *ab* or variants, with the accusative singular as *abinn*, later *abuinn*, *abainn* from which the Scots Gaelic form derives.[27] *Abhainn* is actually quite rare in the AOS, as its zone of influence sits further to the West, in areas where Gaelic was more commonly spoken at the time the first OS maps were made. As has been noted, generic elements are easily removed by an incoming linguistic stratum (in this case Scots). The smaller watercourses (such as those traditionally called *allt*) are more resistant to change, whilst generic elements for larger watercourses (such as *abhainn*) are more susceptible to change from the older stratum to the later.

Semantics: Amongst the bilingual Gaelic-speaking population, *abhainn* has become a translation, in some instances, of *river*, such as Abhainn Tuirc for River Turk. Also known is the name Abhainn Dubh for the River Forth, which is evidenced in Dwelly and sometimes heard by modern Gaelic speakers. See the discussion on the Forth in section 2.6.2.1 on page 40 for this, but it is questionable

[27]R. Thurneysen, *A Grammar of Old Irish* (Dublin, 1946), p. 213.

Figure 3.13: Comparison of G *allt* and G *abhainn*



whether this was a term used by Gaelic speakers in the time of monolingualism, and also whether it was meant to represent the entire length of the Forth as it is now known. If it were to represent only the upper portion, this would make more sense in terms of distribution as suggested above.

It would seem that *abhainn* has an epexegetic role similar to 'river' in English, i.e. it is a common lexical item that is often not very thoroughly 'stuck' to its specific element, for instance in English we may call most rivers *River* X or *The* X, as in 'The Tay' or 'River Tay'; the same might be said of *abhainn* in Gaelic as in the name Abhainn Tatha, its Gaelic name.

It is sometimes said that RNs ending with a vowel + n, have *abhainn* as their final element, albeit phonologically weakened.[28] Whilst it is possible that this is the case in some instances in Scotland, within the AOS there is only one very weak example of this, in Allt Tarabhan, which is Tarbh Abhainn in 1867 (OS 1st edition). This underlying element here is probably G *tarbhan*, 'little bull'.[29] Moreover, all the names with *abhainn* have that element in the initial position (i.e. *Abhainn* X not X *Abhainn*), unlike the names proposed here. It is probable that the majority of Pre-Gaelic RNs ending in vowel + n represent a suffix of some kind, rather than a reduced form of *abhainn*.

[28]For instance: D. Ross, *Scottish Place-names* (Edinburgh, 2002), p. 120, where the derivation for the Kelvin (given under the entry for Kelvinside), is given as *caol abhainn*, 'narrow river'; this is very unlikely.

[29]Watson, *The Celtic Place-Names of Scotland*, p. 453.

Figure 3.14: Comparison of G *allt* and G *abhainn* without Abhainn Dubh

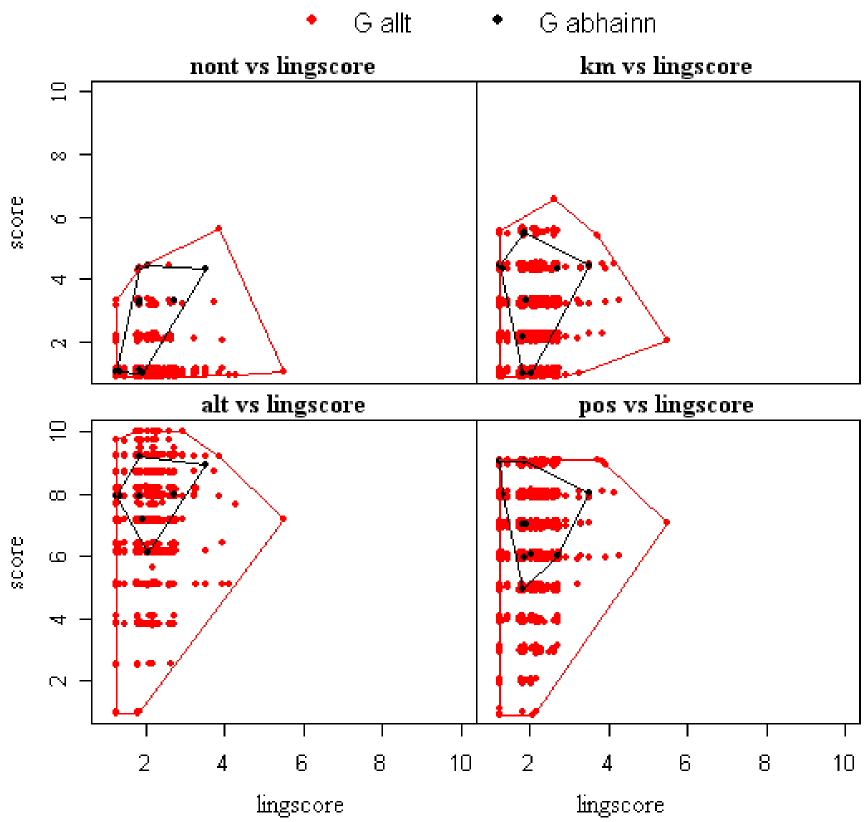


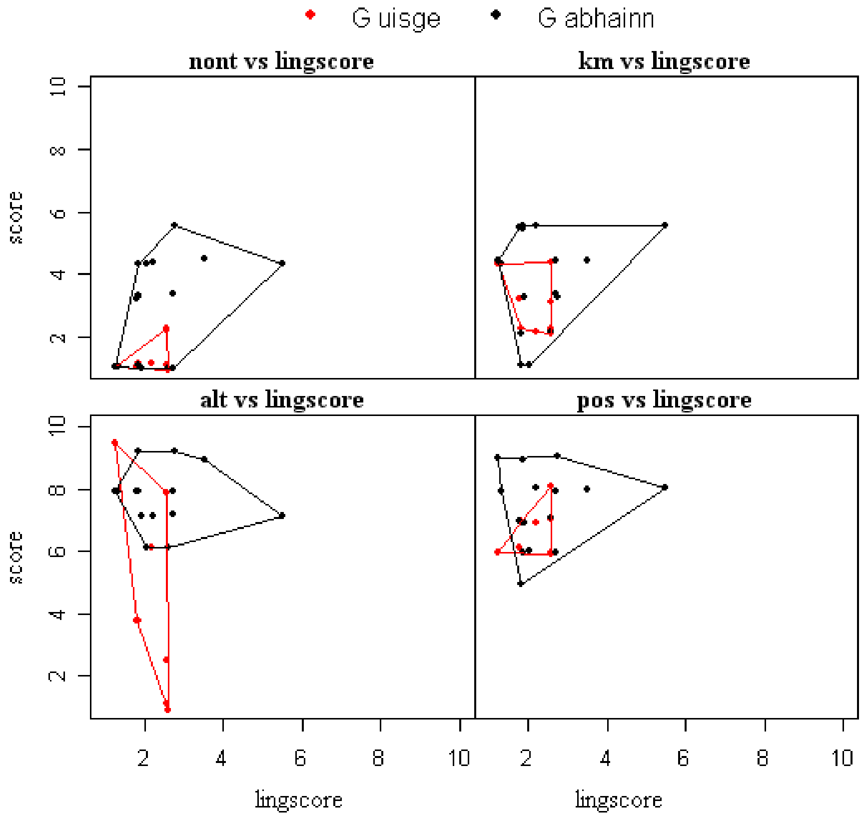
Table 3.5: The Incidence of *abhainn* by Current Counties According to OS Landranger

county	frequency
Argyll and Bute	69
Highland	121
Na h-Eileanan an Iar	142
North Ayrshire	1
Perth and Kinross	2
Stirling	2

3.4.4 G Gleann and Sc Glen

Semantics: *Gleann* seems to have become a hydronym in a handful of cases in the AOS. Since the course of a glen and its river are generally identical, it is hard to determine whether these names genuinely represent a secondary development of *gleann*, or whether this is a matter of (mis)interpretation on the part of the Ordnance Survey. Dwelly certainly does not offer any aquatic secondary meaning. That said, it should be noted that many, if not most of the most popular generic elements in Gaelic hydronymy originally derived from non-hydronymic elements, such as G *allt*, 'cliff'; G *meur*, 'finger, branch'; G *caochan*, 'little blind one' and G *fèith*, 'bog'. Therefore these names could possibly repre-

Figure 3.15: Comparison of G *uisge* and G *abhainn*



sent the inception of a new hydronym, but likely they simply present a situation where an unnamed watercourse flows through a named glen, or a glen name has ousted the hydronym as the *de facto* term for the area.

Replacement of *glen* or *gleann* names with their watercourse or *vice versa* is not as common as is generally supposed. Semantically, most *glen* or *gleann* names are closer to names of equivalent natural features such as hills, rather than to watercourses.

Ignoring name pairs such as ‘Aultmore’ (i.e. Allt Mòr) and ‘Glen Mor’ (i.e. Gleann Mòr) where both the names derive from the same element, names with the structure Glen + X + Y or Glen + HX, where the specific is a hydronym or hydronymic element, are relatively rare. The only two certain cases where a hydronym derived from a *glen* name which itself clearly derives directly from a hydronym are Glenfender Burn (from **finn dobhar* and Glen Lochsie Burn (from a PG *lòch*, ‘black / shining’ with an -*s*- suffix[30]). It is possible these two names relate to a Pictish cognate of the W term *glann*, ‘strand, river-bank’. Another less well attested name is Allt Arder, the specific here is perhaps from G *àrd* + *dobhar*, ‘high water’ and is on record as *Ald Glenardure* (c. 1591 Pont map 26).

Watercourses with this element also tend to have a high number of tributaries with Gaelic names,

[30]Jacob King, “Lochy’ Names and Adomnan’s Nigra Dea’, *Nomina* 28 (2005), p. 75.

suggesting that many of these names were originally G *gleann* and became Sc *glen* at the hydronymic stage, rather than being transferred into the hydronymicon directly as Scots *glen*. See table 5.6 on page 200.

3.4.5 G Fèith

Semantics: Dwelly gives two meanings for this term: in the first instance it is a 'sinew' or 'vein', with a secondary meaning of 'Rents in moor- or bog- land made by water'; another meaning is 'bog, quagmire or bog-channel'.[31] This is a similar semantic development to that of *beck* in England, where a name for a watery patch of ground has come to mean a watercourse. Semantically, the names are fairly typical of small watercourses in Gaelic, as in table 3.6. In the instances in this database, only names in *fèith* are gathered where they specifically relate to a flowing watercourse.
Distribution: Whilst it is true that all watercourses with Gaelic generic elements are in the Highlands and thus at higher altitudes, it should be pointed out that this is more pronounced with names in this element than others.

Table 3.6: Semantemes Associated with G *fèith*

name	frequency
Colour	17
Fauna	10
Specific natural feature	7
Flora	5
Land around	4
Dimensions	3
Convexity e.g. Ben	2
Use to man, agriculture	2
Specific person / occupation	2
Water feature	1
Sound	1
Smell / Taste / Feel	1
Course	1
Material / Object	1
Concavity e.g. Glen	1

3.4.6 G Eas

G *eas*, 'waterfall' is used both to mean a specific point on a watercourse where a waterfall exists, but it can also extend to represent the entire watercourse.
Distribution: *Eas* as a generic element is a south-western Gaelic phenomenon, with a concentration in and around Argyll.
Semantics: Fairly typical for smaller names, with a relatively high amount named after particular people.

[31] Dwelly, *The Illustrated Gaelic-English Dictionary*.

Geographical attributes: These watercourses are all very short, the longest being 3.5 km. The *nont* is also very low, the highest being 5 for Dubh Eas (2719). The old forms for this name however, are: *Dow-visk / Dow-viisk* (c. 1590 Pont text 150v) suggesting the modern *eas* is a misinterpretation of *uisge*.

3.4.7 G Lag

The three instances of G *lag*, 'hollow' as a hydronym probably represents OS errors in attributing feature types. This is probably like *gleann*, where the name of a concave area has been mistaken for the watercourse running through it.

3.4.8 G Meur

Semantics: This is described in Dwelly[32] as 'branch of a river', which is an apt description. These names are typical of small watercourses with Gaelic names: There are three occurrences each of G *lòn*, 'meadow, lawn' and G *cùl*, 'back' as specific elements where *meur* is the generic element.
Distribution: All but two of these elements are in the Spey Catchment Area, the other two existing in the Don Basin. The defining characteristic of *meur* however, is not location but altitude. All but one name, Meur an Loin (4372), exist in upland or mountain, and average an altitude of 631 m. This means they are confined to the Cairngorm area and other high regions of the Grampians. Being this high, it follows that they are all short, the longest being just 4.5, the average being 1.97. It also follows that the *posscore* is low (i.e. that they are far from the sea), the average being 4.56. *Meur* also occurs to a high degree in pairs of names where the variation is purely in the generic element in table 3.7.

Table 3.7: Instances of G *meur* and their Superordinate Watercourses

Name	Tributary of
Meoir Veannaich <	Allt Veannaich
Meur an Loin	Coul Allt
Meur Crionach	Allt Blairnamarrow
Meur a' Chuil	Muckle Fergie Burn
Meur a' Chraisg	Muckle Fergie Burn
Meur an Loin	Muckle Fergie Burn
Meur a' Chois	Muckle Fergie Burn
Meur Cul na h-Eige	Burn of Little Fergie
Meur an Loin	Burn of Little Fergie
Meur Luachaireach	Burn of Little Fergie
Meur Gorm	Caol Ghleann
Meur Shuas	Allt na Cuilce
Meur na Cuile	Allt na Cuilce
Meur Meadhonach	Allt na Cuilce
Meur Aillig <	Allt Aillig

[32] Dwelly, *The Illustrated Gaelic-English Dictionary*.

3.4.9 G Ailnig

Dwelly claims this term is a diminutive of G *ailn*, 'loch', but the source for this is only from Alexander Carmichael.[33] The term **ailn* does not seem to otherwise exist; it does not appear in the AOS. Moreover, the combination *-ln-* is unknown in Gaelic, this cluster in OI became *-ll-* in modern Gaelic. Watson[34] derives the Water of Ailnack from G *Ailneág*, 'little stony one', cognate with G *ail*, 'stone'. The examples in the AOS are Ailnack Beg, which is a tributary of Water of Ailnack; and Ailnac Bhuilg which is a tributary of Builg Burn. Watson's derivation seems more likely than being a diminutive of **ailn*, not least because all these examples are watercourses rather than waterbasins. This term is mentioned on page 121 below.

3.4.10 Diminutive Elements

This section uses the methodology described above to research the relationship between a generic element and its diminutive version. This primarily relates to *loch* vs *lochan* and *allt* vs *alltan*.

3.4.10.1 Loch vs Lochan

In the first instance, it would be desirable simply to ascertain the *geogscore* ranges of each generic element as in table 3.8. This shows what is expected: that the average *geogscore* is smaller for a *lochan* than a *loch*. The next step is to discover whether any specific component of *geogscore* is responsible for this difference more than any other. The two elements are compared in figure 3.16.

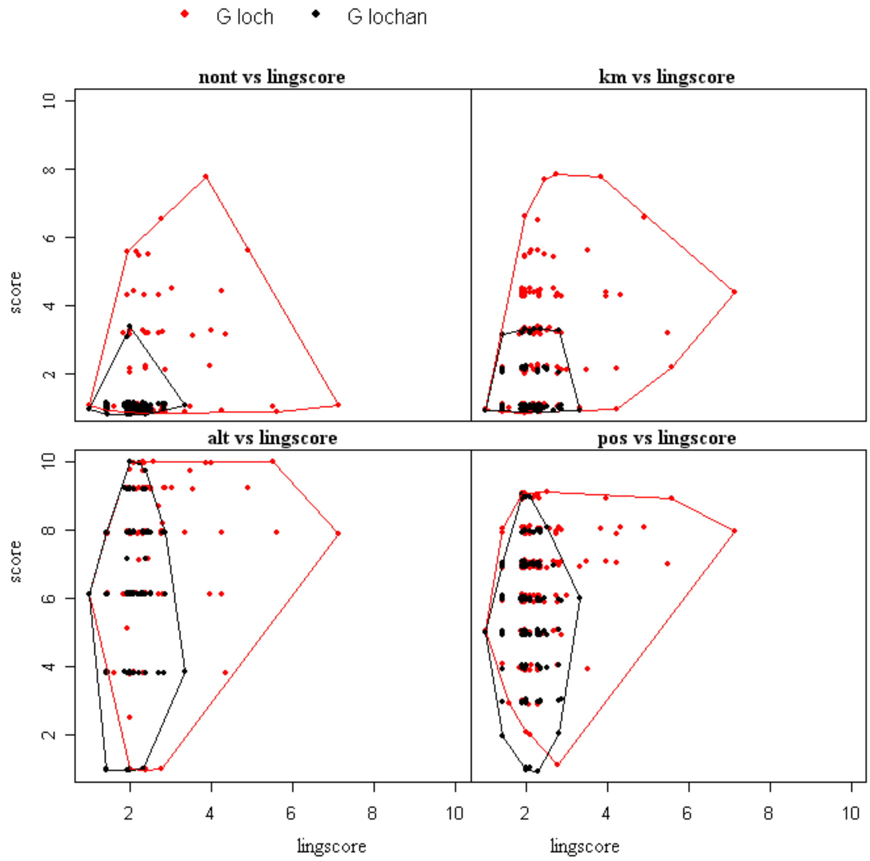
Table 3.8: *Geogscores* of G *loch* and G *lochan*

element	minimum	average	maximum
loch	2.2	4.4	8.3
lochan	1.7	3.6	5.2

Geographical attributes: The graphs for *altscore* and *posscore* show no variation in their respective scores when judged against *lingscores*. This essentially means that altitude and position have no effect on whether a body of water is called a *loch* or a *lochan*. The scores for *nontscore* and *kmscore* however, show that whilst a body of water with any *nontscore* or *kmscore* can be called a *loch*, only those bodies of water with a low *nontscore* or *kmscore* can be called *lochan*. This essentially means that a necessary condition of a body of water being designated a *lochan* is that it *must* be small (i.e. have a low *kmscore*) and must have very few tributaries (i.e. have a low *nontscore*). In fact, one can go further than this, and say that except for one case, within the AOS a *lochan* has only one tributary, the watercourse which it feeds. The exception is Dubh Lochan, which is a small water-basin on a watercourse, which is fed by other watercourses.

[33] Thanks to Prof. Ó Maolalaigh pers.comm. for bringing this to my attention.
[34] Watson, *The Celtic Place-Names of Scotland*, p. 449.

Figure 3.16: Comparison of G *loch* and G *lochan*



3.4.10.2 Allt vs Alltan

Alltan is the diminutive of *allt*. The same methodology as that taken for *lochan* is followed here.

Table 3.9: *Geogscores* of G *allt* and G *alltan*

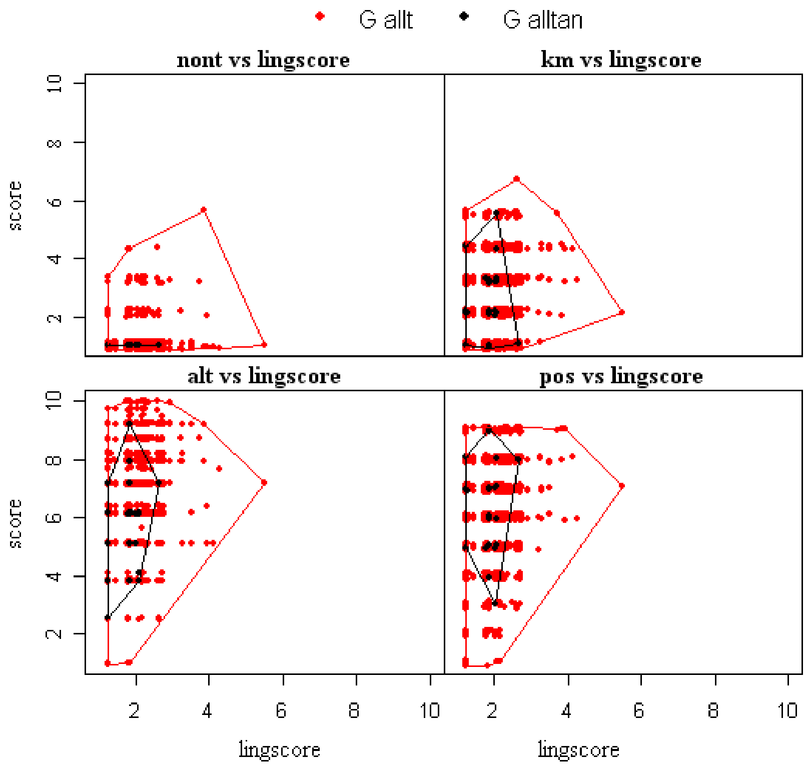
element	minimum	average	maximum
alltan	2.7	3.9	5.0
allt	1.7	4.2	6.5

Geographical attributes: These show a similar situation to *loch* and *lochan* above, but with one major difference, the minimum *geogscore* of *alltan* is greater than the minimum *geogscore* of *allt*. This is difficult to explain, but can probably be ascribed to the fact that whilst there are 1716 names containing *allt* in the database, there are only 26 containing *alltan*, suggesting that more examples of this name might alter the situation.

Figure 3.17 compares the two elements.

- The graph confirms our supposition that in general names with *alltan* are smaller than those with *allt*.

Figure 3.17: Comparison of G *allt* and G *alltan*



- The range of *alltan* is entirely bounded by that of *allt*. In other words, it shows that whilst any *alltan* could be an *allt*, the reverse is not true.
- The most striking feature in the data are the uniformity of *nontscore*, which is 1 across the board. This is supported by table 3.10 which counts the actual number of named tributaries for *alltan*.

Table 3.10: Occurrences of *alltan* for each *nont*

nont	frequency
0	18
1	4
2	2

Distribution: Names with *alltan* cover a surprisingly small area, with no outliers, nearly entirely in the Grampian Mountains. The average *altave* is 591. Part of this distribution is probably due to the term not having a clear equivalent in Scots (there is no single word for ‘little burn’ which derives from burn, e.g. *burnie*[35]), which causes it to be anglicised into *burn*. Unfortunately a search for any candidates does not yield anything likely in the same distribution area.

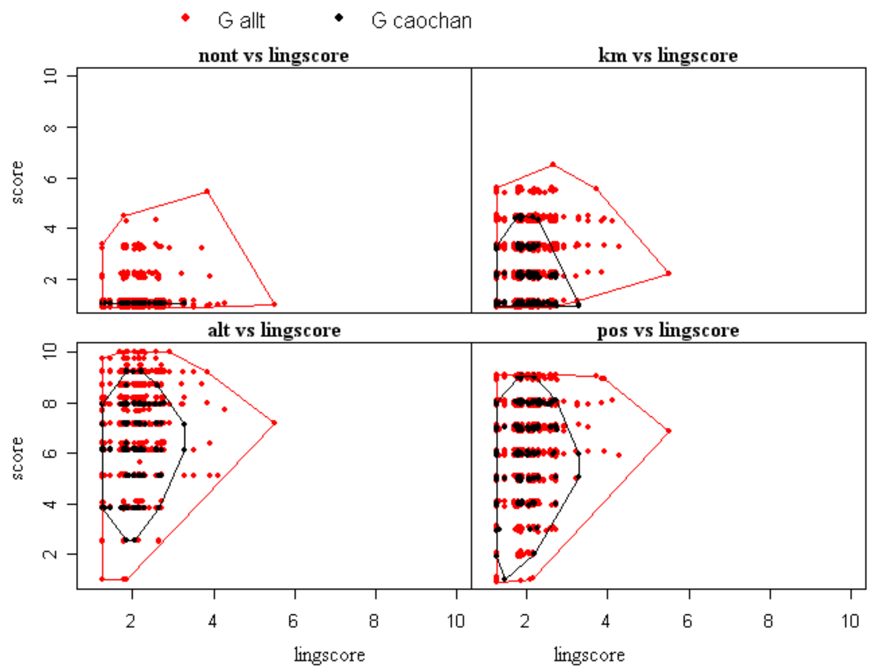
[35]This is how Watson and Allan, *The Place Names of Upper Deeside* translates *alltan*.

3.4.11 G Caochan

Semantics: Like *allt*, *caochan* exists in Irish but has no hydronymic meaning. It derives from OI *caech*, 'one eyed, blind', cognate with Welsh *coeg* of the same meaning. Watson explains it as 'a rivulet so overgrown with herbage it cannot see out of its bed'.[36] It could perhaps be better described not as itself being blind but as people being blind to it. In terms of distribution, the name is situated almost exclusively in the Gàidhealtachd, with the majority (69%) within the Spey catchment area.

The term *caochan* is linguistically a diminutive; the base-form *caoch* is attested in East Perthshire with the meaning 'small burn'[37] but this does not seem to be present as a generic element within the hydronymicon. The diminutive form *caochan* can be plotted against *allt* to see if it behaves in a similar way to the other diminutives. It should be stressed that the comparison with *allt* is merely as a control, since *caochan* is in no sense a diminutive of *allt*.

Figure 3.18: Comparison of G *allt* and G *caochan*



Geographical Attributes: Figure 3.18 shows a very similar pattern to *loch* vs *lochan* or *allt* vs *alltan*. The extent of *caochan* is predominantly less than that of *allt*, and the strongest correlation is that of a low *nontscore*. This is broken down as in table 3.11 overleaf.

[36]Watson, *The Celtic Place-Names of Scotland*, p. 442.
[37]Máirtín Ó Murchú, *East Perthshire Gaelic* (Dublin, 1989), p. 303.

Table 3.11: Occurrences of *caochan* for each *nont*

nont	frequency
0	188
1	21
2	10
3	2

3.4.12 Concluding Remarks on Diminutive Generic Elements

Looking at these relationships with *allt* and *alltan* on the one hand, and *loch* and *lochan* on the other, certain tendencies can be seen in common.

- The critical factors are number of named tributaries and length.
- The range of the diminutive is generally less than that of the base-form. This may explain why the minimum *geogscore* of *alltan* is greater than that of *allt*.

3.5 Genitival and Adjectival Constructions of Gaelic Generic Elements

The relationship between a specific element and a generic element is approximately divided into two categories: genitival and adjectival. Genitival constructions are the normal method for combining nouns in modern Gaelic, for example Allt an t-Sionnaich combines G *allt* + G *sionnach*, 'fox' hence 'fox burn'. The adjectival category combines adjectives and nouns together, for example Allt Dearg, with G *allt* + G *dearg*, 'red', hence 'Red Burn'. The adjectival type of combination has a number of variants. Firstly there is variation in word order, as in the so-called preposed adjectives, or inverse compounds, discussed at greater length below, e.g. Dearg Allt, with the same meaning as Allt Dearg (although it is possible there was a subtle difference in meaning as discussed below). The specific element does not have to be an adjective in the adjectival type of names, however. A number of formally incorrect names seem to represent a type of midway point between these names of the construction generic + specific but the specific is in the genitive case, without an article, such as Allt Mhairc from G *marc*, gen. *mairc* 'horse'. Often in these cases old forms show variations such as in the case above: *Alt Marck* (c. 1591 Pont map 20) which could show the genitive plural *marc*. The modern term could be in error for Allt a' Mhairc.

Throughout this thesis I have used the notation for specifics and generics as explained on page xxi. This was adapted from the Nicolaisen's usage of X for a specific.[38] The terms X and Y denote the syntactic role the element plays in the RN, regardless of the part of speech actually used. Thus an X (a specific) can denote an adjective or a noun or a verbal noun.

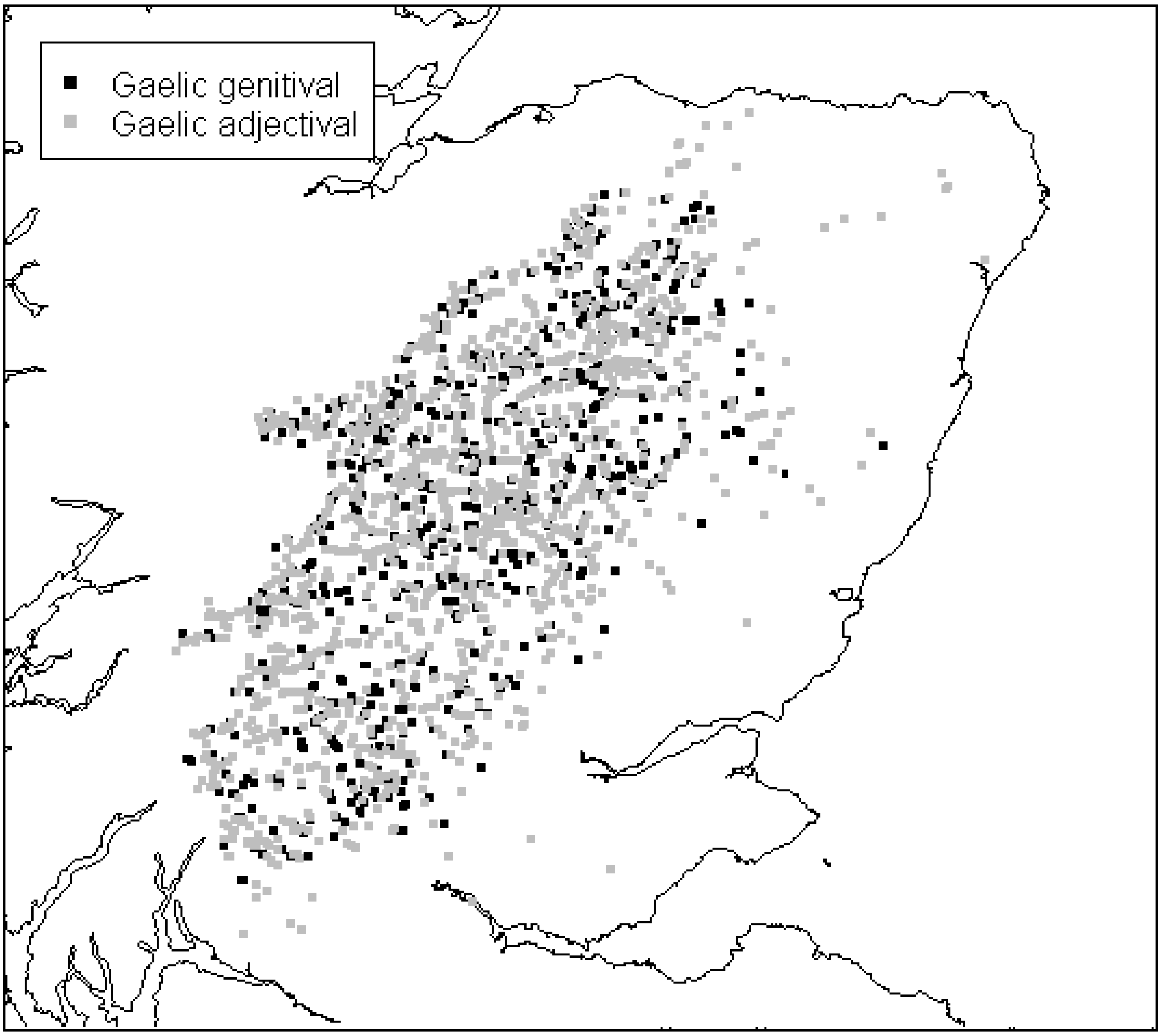
The next section discusses the various ways in which a specific element and a generic element can combine and the factors on which they depend.

[38]Nicolaisen, *Scottish Place-Names*, p. 74.

3.5.1 Location

Figure 3.19 shows the distributions of Gaelic adjectival and genitival constructions according to the AOS database, whilst figure 3.20 shows the same (for watercourses) according to OS 1:150000 in Scotland. It is clear that the distributions are essentially the same, with *Allt* X somewhat more common. It is clear then, that the phenomenon is not particularly related to region.

Figure 3.19: Gaelic Genitival and Adjectival Constructions



3.5.2 Geogscore

Table 3.12 shows *geogscores* for the various types of Gaelic constructions, showing vary little variation. This suggests that the physical qualities of the watercourse do not effect the construction of Gaelic RNs.

Figure 3.20: Gaelic Hydronymic Genitival Constructions

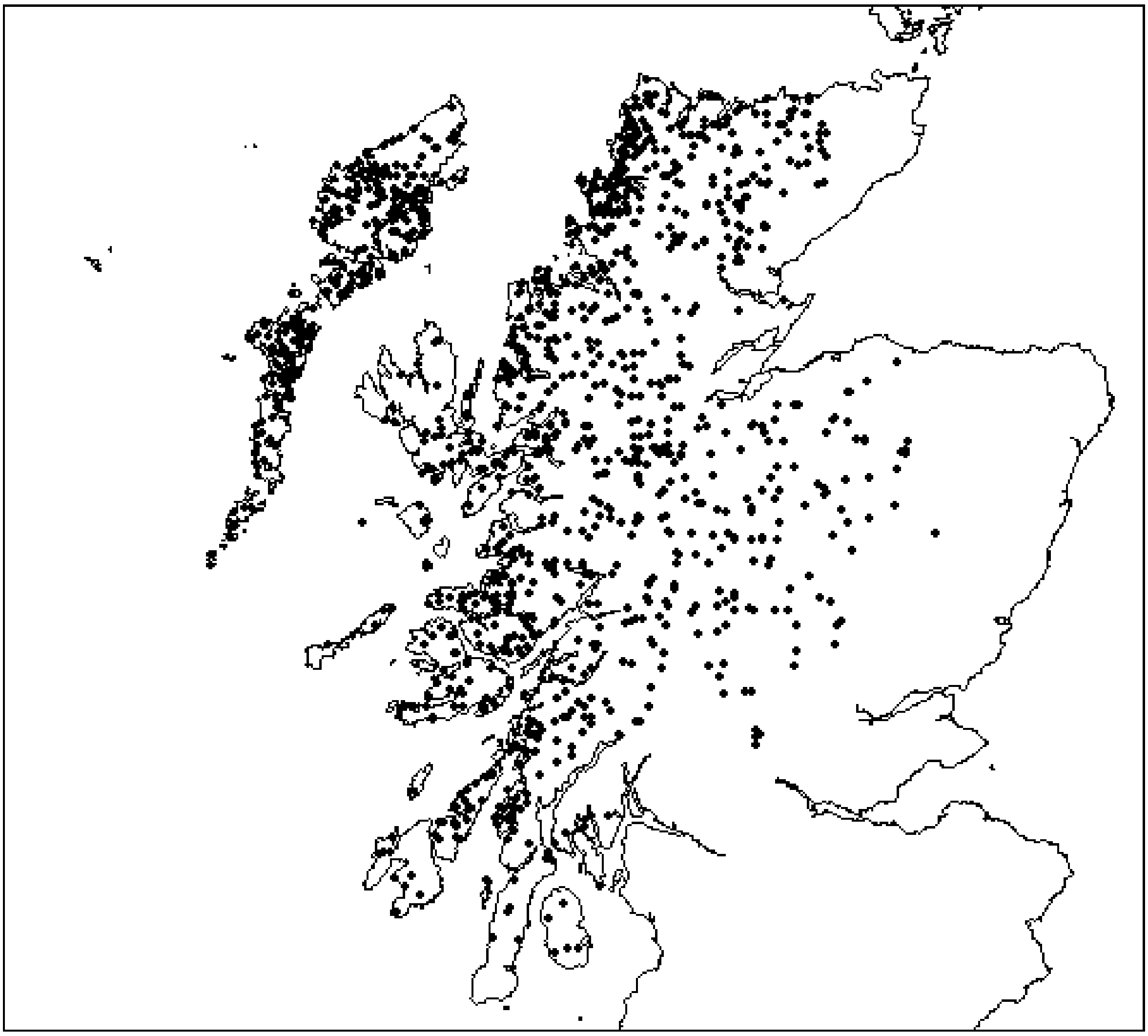


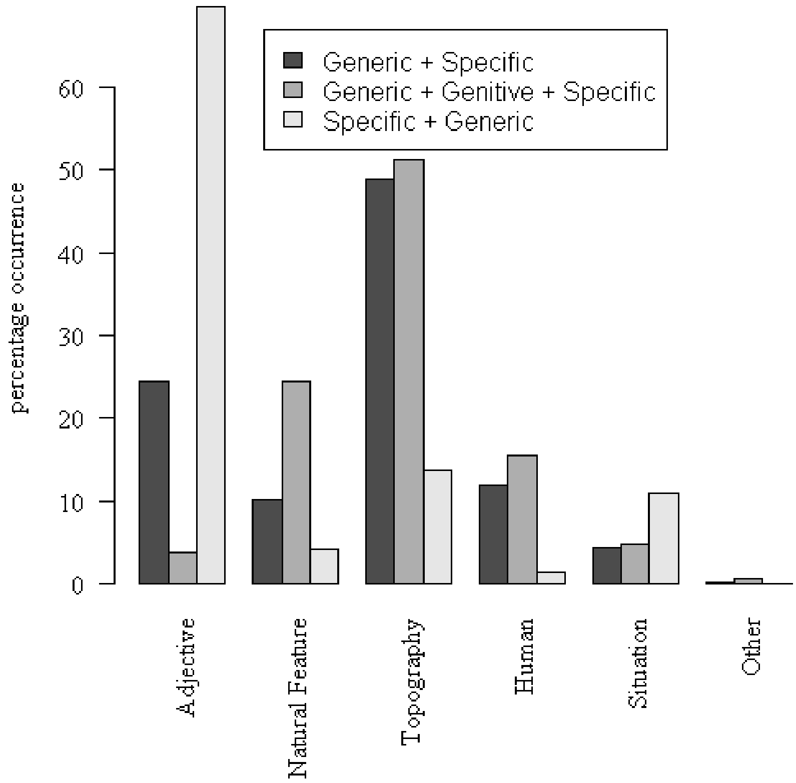
Table 3.12: Average *geogscores* for Gaelic Constructions

average score	RN example	*genelem3*
4.12	Dubh Allt	1
4.18	Allt Dubh	5
4.12	Allt an Achaidh	6

3.5.3 Semtype

Figure 3.21 on page 84 plots the three main types of constructions against the predominant semantic types as described in section 4.3 starting on page 133. One can see that there is a great degree of variation in some of the types, most strikingly between ‘adjectives’ and ‘topography’. RNs with an Y *na* X construction may generally be said to have as a generic element a ‘noun’, be that a geographical entity, an item of flora or fauna, or a man-made thing, or a person. This is hardly surprising since the purpose of a genitival construction is to bind nouns together. Conversely, and as one would expect the X Y and Y X constructions are predominantly adjectival.

Figure 3.21: Gaelic Generic Elements



3.5.3.1 X Y: Preposed Adjective Constructions

Although semantics are largely not discussed in this section, they must be taken into account for X Y constructions. This type is much rarer and on the whole is restricted to a few specific terms as shown in table 3.13. A few anomalies in this table should be explained:

Tarabhan This more likely reflects *Allt Tarbhan 'little bull burn'. See page 72 for more discussion of this name.

Eiginn Allt This name seems to denote a noun rather than an adjective and is anomalous. It possibly reflects *Eag an Uillt, 'notch of the burn'.

Cruaidh Allt These two instances have a noun rather than an adjective as a specific and as such are anomalous.

Feur Lochan Given that G *fiar*, 'crooked' is a term that can be preposed, it could be that *fiar* is the original term from these two near homonyms. That said, Dwelly gives *feur-lochan* as a phrase meaning 'grassy pool'.

The situation is essentially the same throughout the rest of Scotland, the only term absent here being Dearg Allt and perhaps Fiar Allt. The semantic content here is almost opposite of Y *na* X constructions: Primarily adjectives describing the water, or if not, the land immediately around it. Three names represent colour, namely, *dubh*, *fionn* and *glas*, with *dubh* being by far the most popular. Also of note are the three terms, *fiar*, *cam* and *crom*, all meaning, 'crooked' or similar.

Table 3.13: Preposed Adjectives in Hydronyms

typical name	etymology	frequency
Dualt	G *dubh*, 'black'	25
An Garbh-allt	G *garbh*, 'rough'	9
Crom Allt	G *crom*, 'crooked'	6
Coul Allt	G *cùl*, 'back'	5
Leault	G *leth*, 'half'	3
Cam Allt	G *cam*, 'crooked'	3
Cruaidh Alltan	G *cruaidh*, 'hill side'	2
Feur Lochan	G *feur*, 'grass'	2
Glas Allt	G *glas*, 'grey-green'	2
Schenaven	G *sean*, 'old'	2
Caol Ghleann	G *caol*, 'narrow'	1
Crion-alltan	G *crìon*, 'withered'	1
Eag Uillt	G *eag*, 'notch'	1
Eiginn Allt	G *èiginn*, 'steep hillside'	1
Finalty	G *fionn*, 'white'	1
Geal Loch	G *geal*, 'white'	1
Mas-chaochan	G *màs*, 'bottom'	1

These preposed adjectives exist in the lexicon in all Celtic languages. In Gaelic, *sean*, 'old' and *leth*, 'half' are used only in an attributive sense and in the hydronymicon precede the noun in both the lexicon and hydronymicon (e.g. *Allt Leth and *Allt Sean seem to be inadmissible both as a RN and in general speech). Likewise, the other terms seem to be possible in both preposed and postposed position in RNs (e.g. Allt Dubh and Dubh Allt); in general speech these terms can be used both attributively and predicatively.

One linguistic feature of note here is that the meaning of an adjective can change depending on its position, in OI for instance, *math*, 'good' can mean 'gentle' when placed before the noun. This structure is comparable to poetic usage, where word order can be altered to fit metre. It may be that this is occurring here, with nuances now lost.

It would be tempting to posit the phenomenon of preposed adjectives as having some sort of P-Celtic origin, since there exist a number of Welsh terms with a similar role as these Gaelic terms[39]

[39] Such as *hen*, 'old' (cognate with Irish, G *sean* of same meaning); *hoff*, 'favourite' and *prif*, 'main' (cognate with Irish *príomh*, G *prìomh*, 'first').

but the Scottish distribution does not concur with this, being equally popular in the Highland and Argyll area. In fact the distribution is similar to that of *allt* in general. Also, in Ireland the same phenomenon appears to have existed, with the same elements in Scottish Gaelic as in Irish, for example *cam* in Camowen[40] (i.e. Cam Abhainn) and with *sean* in Shanowen[41](i.e. Sean Abhainn), although these names may be in this order due to the qualities of the adjective in a lexical sense as opposed to its hydronymic status. The precise origins of these preposed adjectives are not clear; since the phenomenon exists in Ireland it must be at least partially Q-Celtic in origin. Until an equivalent study has been done for Irish toponymy the precise relationship cannot be fully known. It is also possible that they have been influenced by a P-Celtic substrate, (as discussed in section 3.3.8 on page 67) since the majority of surviving P-Celtic generic element constructions are preposed. In any case, it appears that this construction dates back to an early stage in Scottish Gaelic toponymy.

There is perhaps a correlation between these names and RN pairing. A relatively high proportion of the names appear as part of a pair, i.e. Easter and Wester Shenalt, Glas Allt Mor and Beag, and Dubh Allt Beag and Mòr. As well as this, some of the other terms are those which traditionally can appear as part of a pair, even if they do not in these instances, those are: Dubh Allt and Fionn Allt, Sean Allt with Allt Nuadh, and possibly Glas Allt with Dearg Allt / Allt Dearg (see below for section on colour). Lastly, G *leth*, 'half' really means 'one of a pair'. It may not be that these names are directly linked with pairing as such, but it is more likely that the phenomena of inverse compounds and pairing both belong to the earliest period of Gaelic hydronymy in Scotland and as such often occur together.

3.5.3.2 Y na X Compounds

Comparing Y *na* X compounds to X Y compounds yields what one would expect: these names are much more likely to contain nouns as specific elements, by definition. Limiting ourselves to the twenty most popular elements gives table 3.14.

3.6 Interface between Gaelic and Scots Generics

3.6.1 The Commonest Types of Alternation

Table 3.15 on page 88 lists in order the commonest types of alternation between generic elements (ignoring single occurrences). As it stands this only represents the change where the specific element remains the same. The commonest type of interchange involves *allt* and *burn*. This occurs either as a replacement of one of these elements by another (e.g. *Allt* X and X *Burn*), or it occurs where the same element is preserved, but the position changes (e.g. X *Burn* and *Burn of* X). This type of change is the phenomenon of 'variation' as defined by Simon Taylor whereby the generic element is changed between one language and another, when 'people still understood the meanings of the

[40] P. W. Joyce, *The Origin and History of Irish Names of Places*, vol. 2 (London, 1973), p. 420-421.
[41] Ibid., p. 481.

Table 3.14: Commonest Elements in G Y *na* X compounds

typical name	etymology	frequency
Allt a' Bhreac-choire	G *coire*, 'circular hollow'	81
Allt-na-Creich	G *crìoch*, 'boundary'	19
Allt a' Chaoirnich	G *caorunn*, 'mountain ash or rowan tree'	12
Loch nan Eun	G *eun*, 'bird'	10
Lochan na h-Aon Chraoibh	G *craobh*, 'tree, bush'	9
Allt a' Bhealach Chumhainn	G *bealach*, 'pass'	9
Allt na Caillich	G *cailleach*, 'woman'	9
Aldachuie	G *cùl*, 'back'	9
Allt an Dubh Shluic	G *sloc*, 'pit, hollow'	9
Allt na Beinne	G *beinn*, 'hill'	8
Lochan a' Chait	G *cat*, 'cat'	8
Aldnecrage	G *creag*, 'crag'	8
Allt an Lochain Duibh	G *lochan*, dimin *loch*, 'small lake'	8
Allt nan Gabhar	G *gabhar*, *gobhar*, 'goat'	8
Fèith na Mad	G *madadh*, 'wolf, dog'	8
Allt a' Mhuilinn	G *muileann*, 'mill'	8
Allt nan Seileach	G *seileach*, 'willow'	8
Allt na Bà	G *bò*, 'cow'	7
Allt a' Bhreac-ruighe	G *ruighe*, 'slope / summer shieling'	7

elements involved'.[42] One major difference between the hydronymic data and the settlement name data gathered from Fife by Simon Taylor is that he interprets the variation in generic elements between a single specific element as the fact that they originally 'referred to different places, or at least to different parts or aspects of the same place. They ended up becoming interchangeable only after the lexical meanings had become lost or unimportant'.[43] The hydronymic equivalent of this would be a watercourse with its upper part having a different generic to that of the lower part. This type of change is shown in table 3.27 on page 120. This section rather investigates diachronic change (i.e. change over time) between generic elements.

[42]Simon Taylor, 'Generic Element Variation, with Special Reference to Eastern Scotland', *Nomina* 20 (1997), p. 8.
[43]Ibid., p. 9.

Table 3.15: Commonest Types of Diachronic Change between Generic Elements

modern example	old forms example	frequency	modern	old
Ardchyle Burn	*Auld Ardchyle* (1783 Stobie)	29	Specific + Burn	Allt + Specific
Acharn Burn	*Burn of Acharn* (1783 Stobie)	19	Specific + Burn	Burn + Genitive + Specific
Burn of Angels	*Angels Burn* (1869 OS 6 inch 1st edition)	17	Burn + Genitive + Specific	Specific + Burn
Loch Dochart <	*lacu Lochdochyre* (1377 RMS i no 605)	12	Loch + Specific	lacus + Specific
River Almond	*Almont fl.*(1654 Gordon: Braid-Allaban)	12	River + Specific	Specific + fluvium
River Ardle	*Auon Arle* (c. 1591 Pont map 27)	8	River + Specific	Abhainn + Specific
Alyth Burn <	*Water of Elycht* (c. 1591 Pont map 27)	7	Specific + Burn	Water + Genitive + Specific
Black Water	*Darth flu.* (1659 Jansson's map: Scotia Provinciae)	6	Specific + Water	Specific + fluvium
Allt na Caillich	*Auld Calloch* (1783 Stobie); Alt Chaillach (c. 1591 Pont map 20)	5	Allt + Genitive + Specific	Allt + Specific
Allt Charmaig	*Allt a' Charmaig* (1867 OS 6 inch 1st edition)	5	Allt + Specific	Allt + Genitive + Specific
Allt Anndeir	*Auon Indaur* (c. 1591 Pont map 19)	5	Allt + Specific	Abhainn + Specific
Ardtalnaig Burn	*Auon Tolnaig* (c. 1591 Pont map 18)	5	Specific + Burn	Abhainn + Specific
Allt Bad na h-Earba	*Burn of Bad na-Earba* (c. early 1800s Invercauld Papers)	5	Allt + Specific	Burn + Genitive + Specific
Loch Glow	*lacu Lochglo* (1692 Retours (Fife) no. 1340)	5	Loch + Specific	lacus + Specific
Artney Water	*Auon Artnay* (c. 1591 Pont map 21)	4	Specific + Water	Abhainn + Specific
Loch Cairny	*Loch of Craiglush* (1783 Stobie)	4	Loch + Specific	Loch + Genitive + Specific
Water of Ailnack	*avon Ailnaig* (c. 1591 Pont map 7)	3	Water + Genitive + Specific	Abhainn + Specific
Loch Achray	*Achray Loch* (1751 Macfarlane Geog. Coll. Vol 1 p337)	3	Loch + Specific	Specific + Loch
Dreel Burn	*Driel W* (1681 Adair map 7)	3	Specific + Burn	Specific + Water
Ey Burn	*Ey fl.* (1654 Gordon: Braid-Allaban)	3	Specific + Burn	Specific + fluvium
Lochan Achlarich	*Loch Monachachclarach* (1783 Stobie)	2	Lochan + Specific	Specific + Loch
Loch Poll Dubh-ghlas	*Loch of Puldowglash* (1783 Stobie)	2	Loch + Specific	Loch + Genitive + Specific
Caochan Dubh	*Allt Dubh* (1874 OS 6 inch 1st edition)	2	Caochan + Specific	Allt + Specific
Fèith Thalain	*Allt Fèith Chalan* (1912 Bartholomew: Survey Atlas of Scotland)	2	Fèith + Specific	Allt + Specific
Garchory Burn	*Allt a Garbh Choire* (1952 Alexander)	2	Specific + Burn	Allt + Genitive + Specific
Burn of Dalvey	*Auon Dauey* (c. 1591 Pont map 26)	2	Burn + Genitive + Specific	Abhainn + Specific
Burn of Care	*Cair W* (1681 Adair map 6); Turnpike Burn (Awoch)	2	Burn + Genitive + Specific	Specific + Water

3.6.2 Lochs: Gaelic / Scots Generic Elements

The problem with the element *loch* is that it is spelt and largely pronounced identically in SSE and Gaelic. The problem is adroitly summed up in the Gaelic Names Liaison Committee's Orthographic Principles:[44]

> In order to apply consistent Orthographic Principles, it is important to understand whether the generic element within a name is in Gaelic or Scottish Standard English orthography. This is difficult with the generic term 'loch' which is spelt identically in Gaelic or Scottish Standard English. Word order is indicative of which language is involved, but is not in itself conclusive, since Scottish Standard English has taken on the Gaelic word order of putting the generic first in many loch-names.
>
> In the case of the generic element loch, which could be either Gaelic or Scottish Standard English, if it is likely from the context that loch in a particular name is Gaelic rather than Scottish Standard English, then the specific element is only written in Gaelic orthography if it is an existing name that is recorded elsewhere by Ordnance Survey in Gaelic orthography.
>
> For example, Loch Insh in Badenoch, where there are many names of natural features still in Gaelic orthography on the modern maps, is believed to represent Gaelic loch, but because the settlement (and parish) name appears only in Scottish Standard English orthography, then the specific of the loch name is also in Scottish Standard English orthography, remaining Loch Insh, and not converted to the Gaelic orthography of Loch Innis.

Within the context of this thesis the attribution of the element 'loch' to either Gaelic or Scots has depended on the principles mentioned in the quote above. For instance, given a name such as Dow Loch (2572), there is seemingly no reason not to assign the generic here to the Gaelic stratum, as has been done with the other examples of *Dubh Loch*, with which it is identical in origin. To do so, however, would make this the only example of a current hydronymic Gaelic generic element in Fife,[45] thus, it has been assigned as a Scots / SSE generic element, despite the fact it was once obviously Gaelic.

A comparison of the two versions of this element as in figure 3.22 on page 90 shows what one would expect: a broadly similar set of *geogscore* components, with the highest lochs being exclusively Gaelic.

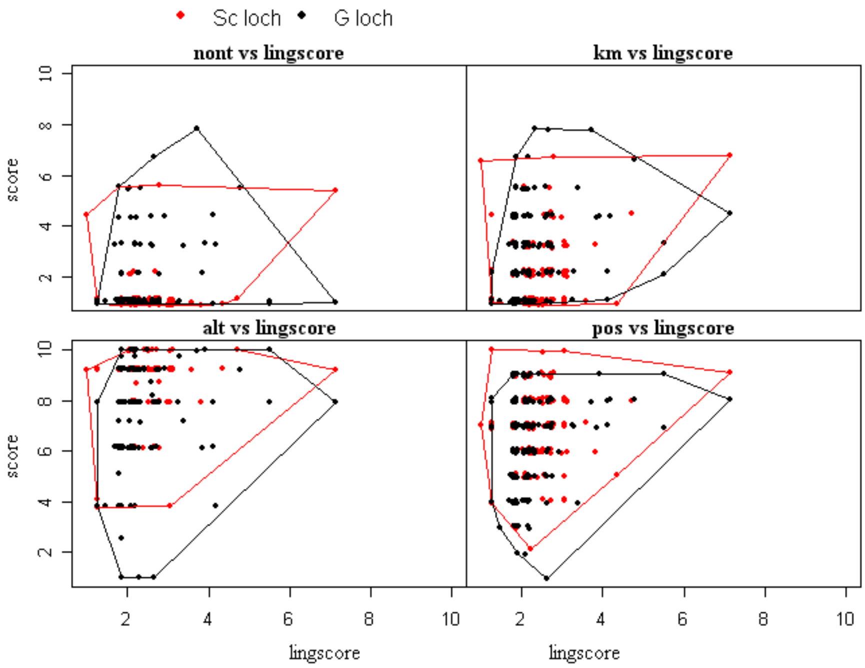
3.6.3 G Uisge vs Sc Water

Since G *uisge* has the same meaning as Sc / SSE *water* it might be prudent to investigate whether there is any evidence that the latter is a translation of the former.

[44] `http://www.ordnancesurvey.co.uk/oswebsite/freefun/didyouknow/placenames/docs/GNLCprinciples.pdf`

[45] This is not to say that plenty of other *loch* names in Fife were not once Gaelic.

Figure 3.22: Comparison of G *loch* vs Sc *loch*



Distribution: Within the AOS there is not really enough evidence to decide, but outwith this area, the name occurs throughout the British Isles, rarer in Wales, but commonest in the Scottish Borders as far North as the Clyde. This is clearly a Scots / SSE phenomenon then. This is backed up by the fact that there is a slight paucity in the area around Dumfriesshire, once a Gaelic speaking area; one would expect this for a Scots element, but not for a Scots element which is a translation of a Gaelic element. North of this area, the situation is essentially the same, a Scots distribution. In fact, *water* is similar in distribution to the *Burn of* X distribution group mentioned elsewhere, and in Shetland and Caithness at least the names have been influenced by Norse.

If the components for *geogscore* are compared between *uisge* and *water*, it is plain they barely inhabit any of the same conceptual space as in figure 3.23. Moreover, there is no evidence for any current RNs with *water* having *uisge* in old forms. The evidence does not support a direct relationship between *uisge* and *water*.

3.6.4 G Abhainn vs Sc Water

If *abhainn* and *water* are compared it is clear they are often applied to the same types of watercourses as in figure 3.24. This is also supported by old forms which show a number of names, now with the generic *water*, which were once *abhainn* in figure 3.16 on page 92.

Although this table is predominantly compiled from a single source, Pont, it still has validity. As mentioned above, if one compares *water* with *uisge* in a like manner, there are no results.

Figure 3.23: Comparison of Sc *water* vs G *uisge*

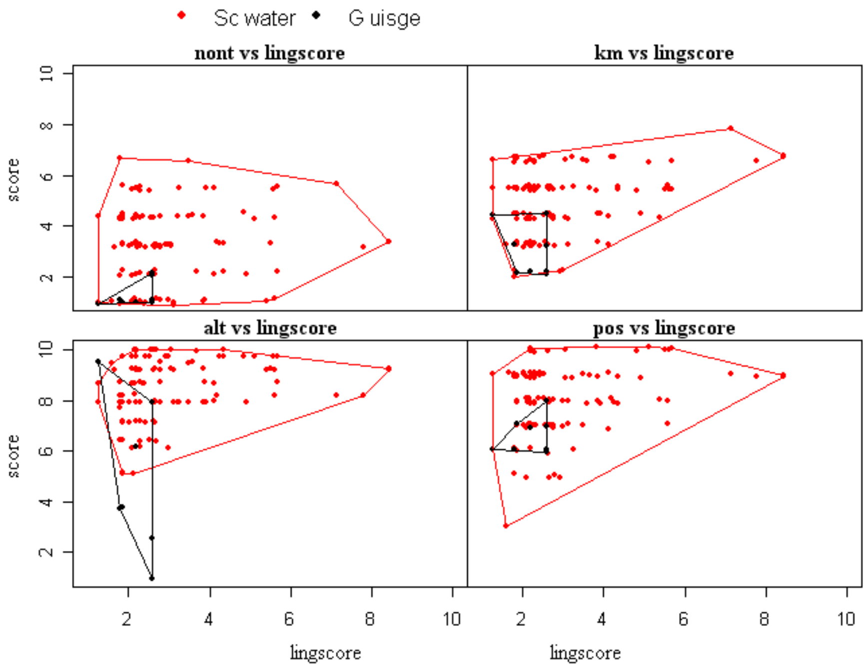


Figure 3.24: Comparison of Sc *water* vs G *abhainn*

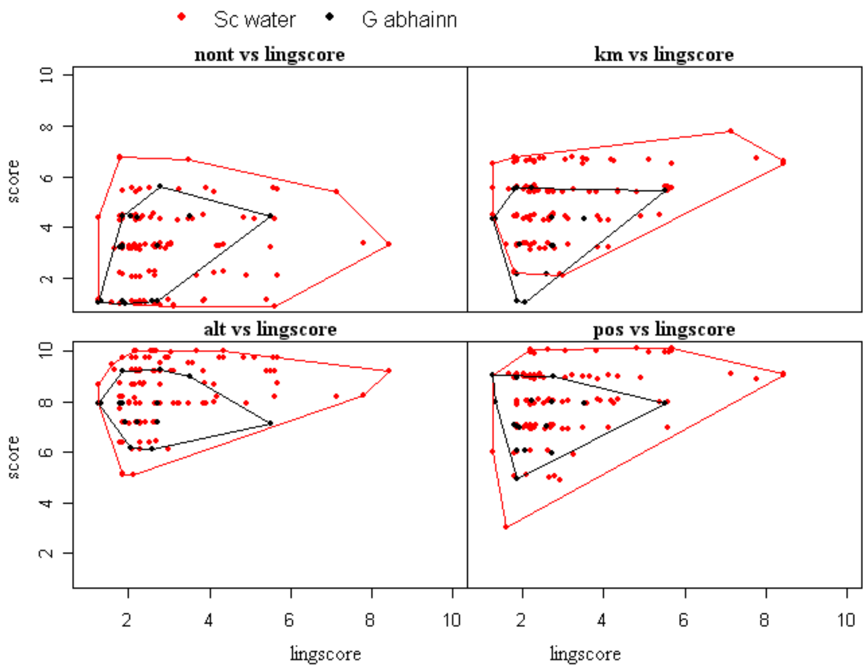


Table 3.16: Diachronic Interchange between *abhainn* and *water*

id	RN	oldforms
46	Water of Ailnack	*avon Ailnaig* (c. 1591 Pont map 7)
90	Water of Allachy	*Auon Ellachy* (c. 1591 Pont map 7)
198	Artney Water	*Auon Artnay* (c. 1591 Pont map 21)
1328	Chabet Water	*Avon Cheabak* (c. 1591 Pont map 7)
3289	Water of Gairney	*Auon Gairny* (c. 1591 Pont map 7)
4476	Machany Water	*Auon Machay* (c. 1591 Pont map 21)
4572	Water of May	*Auon May, Inner-May* (c. 1591 Pont map 21)
5205	Quoich Water	*Avon Coich* (1654 Gordon: Braid-Allaban)
5895	Tarf Water	*Arf A.* (1654 Gordon: Braid-Allaban); *Avon Tarf* (c. 1591 Pont map 20)

3.6.5 G Meur and Sc Grain

The primary meaning of *grain* is a branch of a tree, but it has many secondary meanings, like the word 'branch' in English and *meur* in Gaelic, the latter of which is also a hydronym. Its sense here seems, then, to be 'tributary', as attested in the forms such as "Not to slay salmon fish in the Die and Don or granes thereof".[46] Figure 3.25 on page 93 shows that *grain* and *stripe* have similar distributions, both akin to that of *Burn of* X constructions (this is discussed in the conclusion). Half of *grain* names are on River North Esk. If *grain* and *meur* are compared, it is seen that they have very similar roles within the landscape as in figure 3.25. Where the altitude ranges of all the generic elements are compared below in figure 3.43 on page 115 it can be seen that *grain* is unique amongst the Scots generic elements in that it cannot occur at the lower altitudes. It perhaps should not be assumed that *grain* is a translation of *meur* however. It is more probable that they simply fulfil the same role in the human landscape. This is supported by the lack of evidence of any interchange between *grain* and *meur* as generic elements within the hydronymicon.

3.7 Scots / SSE Generic Elements

3.7.1 Sc Burn X

There are fifteen names in the AOS which have *burn* in initial position but are not *Burn of* X constructions. These are in table 3.17 on page 93. Syntactically, they have the same structure as the default *Allt* X names. The distribution of the names are as in figure 3.27 on page 95.

Distribution: It can be seen that nearly all the names exist in the flat lands of the North-East, in fact if one studies the names by altitude and soil class, they are all at a low altitude and tend to flow through lowland and foothill areas. This distribution is similar to *Burn of* X names.

The specific elements are all Gaelic but with a Scots phonological overlay, aside from this they are all typical of Gaelic specific elements. Two seem to be in the genitive: Burn Gauly, Burn Hervie. These

[46] 'Dictionary of The Scots Language'.

Figure 3.25: Comparison of Sc *grain* vs G *meur*

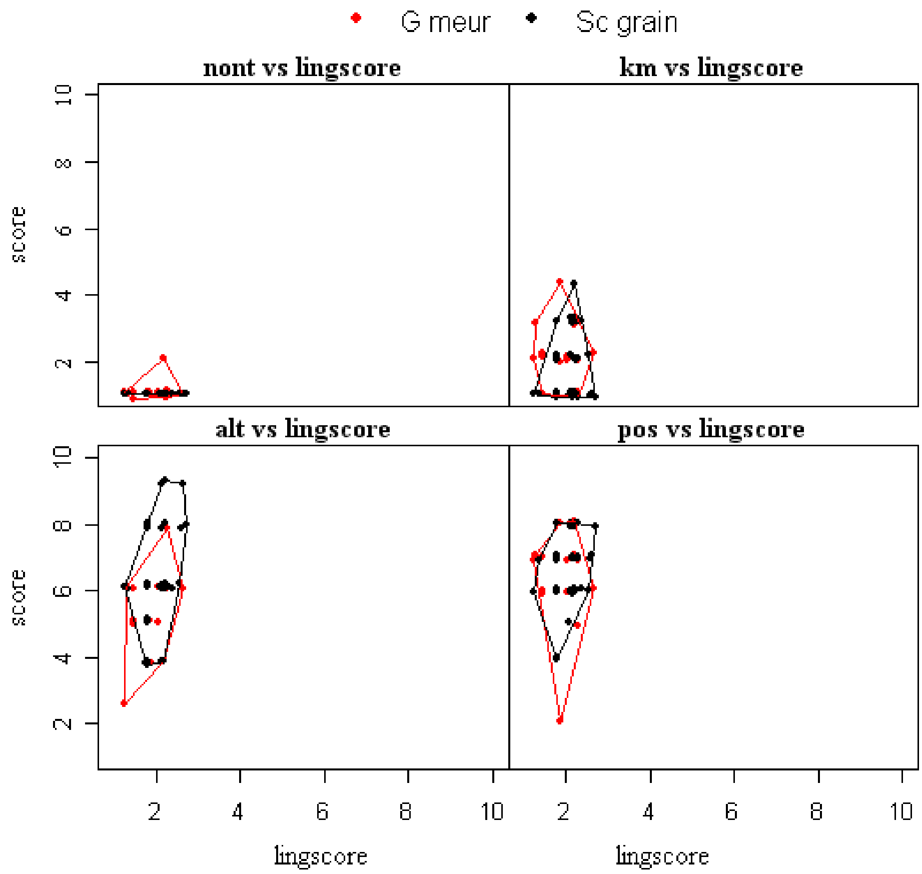
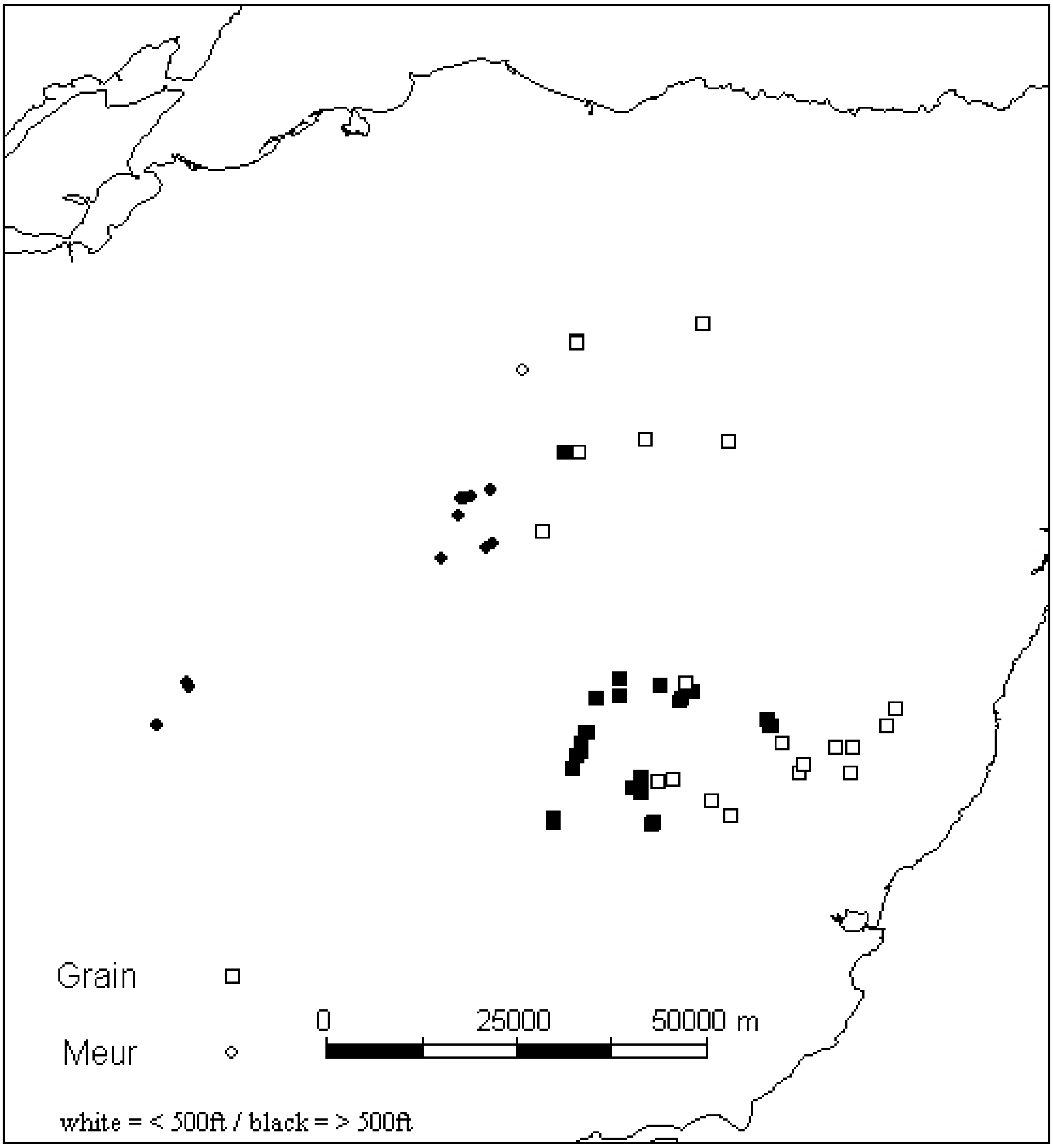


Table 3.17: *Burn* X Constructions

RN	etymology
Burn Beg	G *beag*, 'small'
Burn Crooks <	G *cruach*, 'hill'
Burncruinach	G *cruinneachd*, 'convexity'
Burnervie	G *eirbhe*, gen. *h-eirbhe*, 'wall, boundary'
Burngarnie	G *goirneag*, 'little crier'
Burn Gauly	G *gobhal*, gen. *goibhle*, 'fork'
Burn Hervie	G *eirbhe*, gen. *h-eirbhe*, 'wall, boundary'
Burn Levenit	G *leamhnach*, 'abounding in elms'
Burn Loishkean	G *loisgean*, 'pimpernel' or G *loisgeann*, (variant of *losgann*) 'toad'
Burn Mackarty	G Mackarty, personal name
Burnorrachie	G *uar*, 'waterspout'
Burn Roy	G *ruadh*, 'red'
Burnshangie	G *seang*, 'slender'
Burn Taick	G *taic*, 'prop, support'
Burn Treble	G obscure

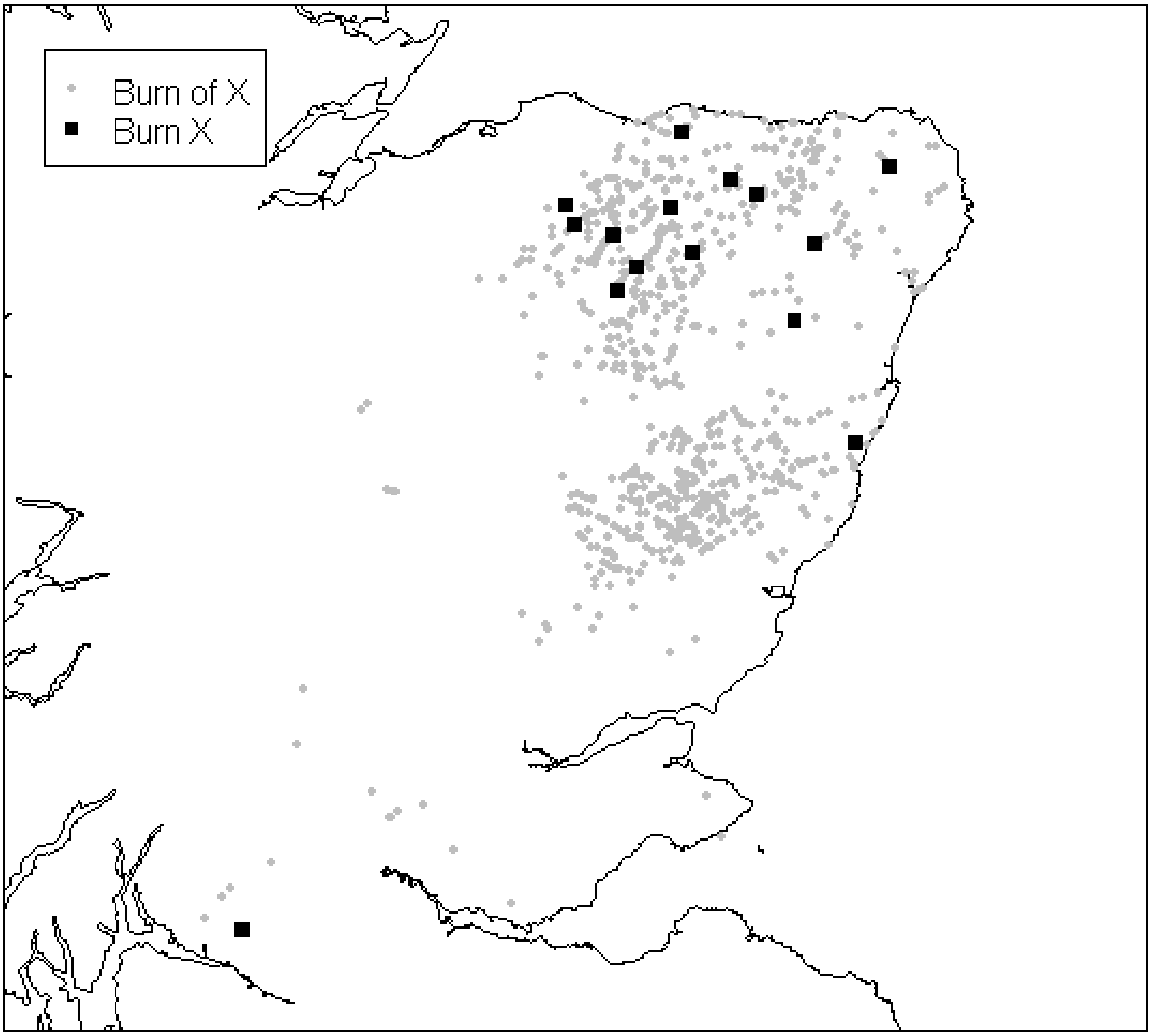
Figure 3.26: Comparison of *grain* and *meur* Distribution



names broadly sit within the distribution zone of the *Allt na* X construction. With the exception of Burn Mackarty, the remaining names, such as Burn Beg have adjectives as specific elements. Figure 3.28 on page 96 plots the semantic types of *Burn* X and *Allt na* X. It shows that apart from 'adjective', the names are broadly similar.[47] The graph also shows that *Burn* X names all sit within the conceptual space of *Allt na* X.

[47] Since there are only fifteeen names, small variations in the count will make large changes on the graph.

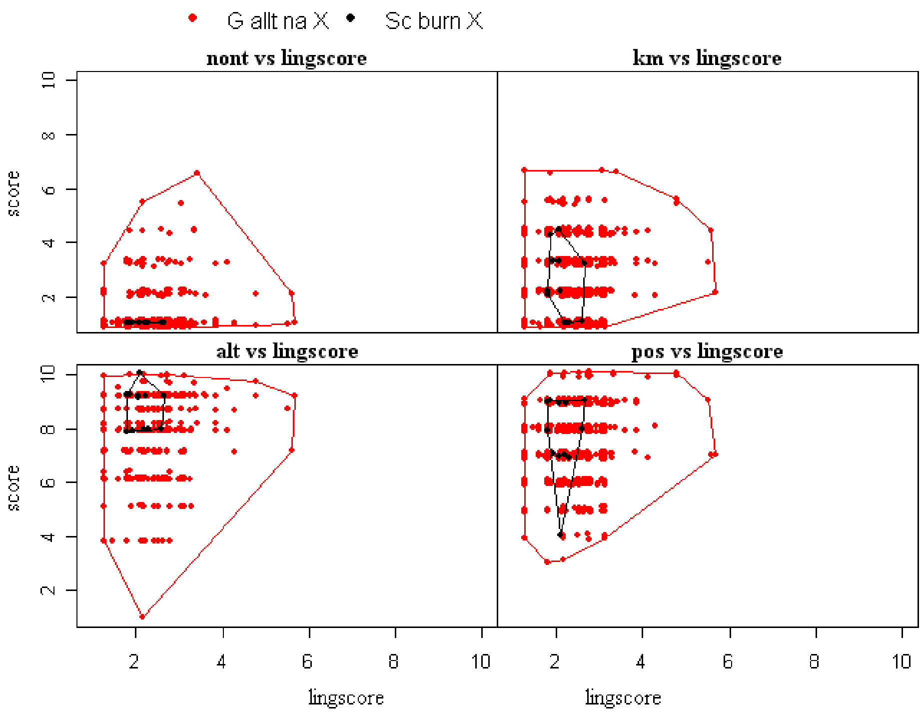
Figure 3.27: Comparison of *Burn of* X and X *Burn* Names



The placing of *burn* as a first element is reminiscent of the Scots term *loch*, borrowed from Gaelic. In this case, since the borrowed term was pronounced identically as the Gaelic term, there was a large degree of overlap, which helped *Loch* X constructions to become almost standard even for Scots names. A similar phenomenon has occurred here. It would seem they are a midway point of transliteration or translation between *Allt na* X / *Allt* X and *Burn of* X constructions. This is corroborated by two comments made by Alexander, the first states: 'names like Burnshangie, consisting of Sc *burn* followed by a Gaelic word, date from the bilingual period'[48] and under the entry for Burngarnie: 'now valley is called *alltgarney*'.[49]

[48]William Alexander, *The Place-names of Aberdeenshire* (Aberdeen, 1952), p. 28.
[49]Ibid.

Figure 3.28: Comparison of *Allt na* X and *Burn* X Names



3.7.2 Sc Y of X Constructions

One cannot study *Burn of* X constructions in isolation; these constructions only make sense within the context of other hydronymic Y *of* X constructions, and, further afield, non hydronymic Y *of* X constructions. This section starts with these general Y *of* X constructions and then deals more specifically with hydronyms. This construction is far commoner than in England or Wales, where it predominantly only exists in very modern names (such as Isle of Wight Farm, or University of Essex etc) or represent features on OS maps which are not really names as such (Mouth of the Severn, Source of Usk etc). According to OS Landranger there are 3969 names using the 'of' constructions in Scotland. The twenty commonest features are as in table 3.18.

There are three relatively distinct areas for this type of name, the tip of Caithness and the Northern Isles, the Lowlands of Scotland, and Dumfriesshire. In Dumfriesshire, the commonest elements in this construction are *water* (18), *rig* (18) and *mains* (14). For the Northern distribution, the commonest elements are: *loch* (183), *hill* (168) and *burn* (114) with many if not most of the less common elements being Norse in derivation. This distribution is generally accounted for as being from a Norse substrate.

The Lowland area, the area with the highest density of names, has as the commonest four elements: *mains* (344), *hill* (253) *burn* (176) and *mill* (146). It is this distribution, within the AOS that will be the main focus of this study.

HY *of* X follows this distribution rather closely, but in the AOS the names exist predominantly in the higher altitudes, whereas other Y *of* X names are at lower altitudes. Therefore a number of

Table 3.18: Twenty Commonest Elements in Y *of* X Constructions

element	frequency	element	frequency
Hill	425	*Water*	49
Mains	349	*Bridge*	49
Burn	291	*Hillhead*	43
Loch	207	*Moss*	41
Mill	147	*Sound*	36
Point	95	*Ward*	36
Bay	92	*Geo*	32
Newton	77	*Falls*	32
Milton	65	*Head*	32
Ness	65	*Holm*	30

areas which have PY *of* X names, do not have, or have very few, HY *of* X names. Such areas are approximately: Fife, South Angus, Lowland Perthshire, East Gordon, and much of Royal Deeside. As Nicolaisen has stated,[50] this distribution is that of Scots encroaching into Gaelic speaking areas, where Scots was the productive language as opposed to SSE, but this does not account for the discrepancies mentioned above.

The Southern Y *of* X distribution (the one predominantly in the NO tile range) has a relatively higher average altitude (414) but this is less true of the Northern Distribution (308). Why there should be this discrepancy, and why there should be an absence North of the Dee until the Ythan is a mystery.

When comparing this distribution with the distribution of traditionally Gaelic speaking areas, one is struck by the contrast between the area North and South of the Forth-Clyde Line. North of this line, the Y *of* X distribution largely avoids the area of Gaelic. This is of course not surprising, since Y *of* X is a Scots / SSE phenomenon. South of this line, however, the opposite is true. The Borders area, previously an area of Anglian settlement, now an entirely Scots / SSE speaking area, is largely devoid of Y *of* X names. Conversely, Dumffriesshire and Galloway, which was once an area of Norse settlement, then going on to briefly become a Gaelic speaking area, now a Scots / SSE speaking area, is full of these types of name.

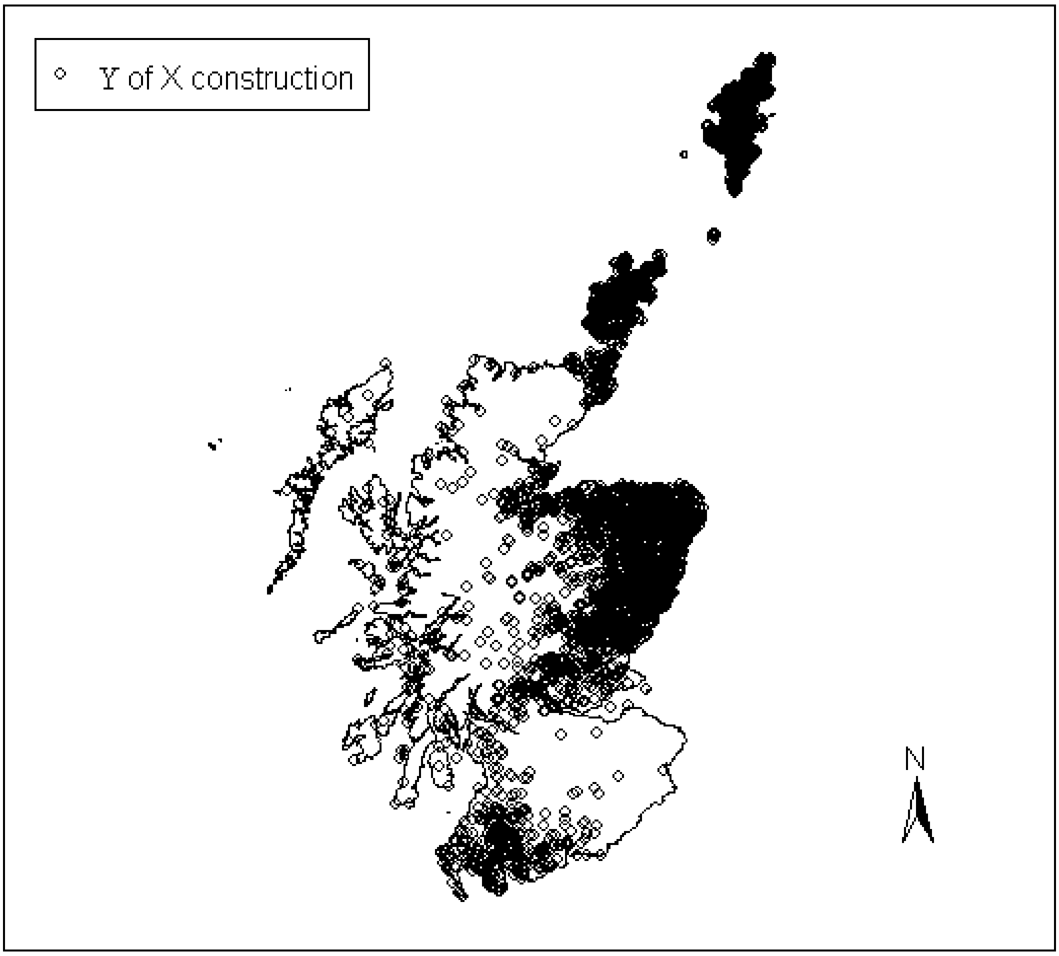
The reason for this distribution goes outside the area of this thesis somewhat, but must lie in the relationships between Gaelic and Scots / SSE in both the areas. In the South-West, where Gaelic has not been spoken for as long, or as recently, it would be tempting to speculate that the current Y *of* X names are anglicised Gaelic, but, with the possible exception of *water of* names, the existing names are all Scots / SSE specific elements with generic elements such as *rig*, *mains* and *fell*.

Burn of X is of course not the only hydronymic genitival structure, there are a number of other types, which will be called here HY *of* X as in table 3.19. This table shows that HY *of* X is broadly even across the spectrum of Scots generic elements, that is, the percentage occurrence of each generic is roughly the same across the whole data-set as compared to HY *of* X constructions.

There also exists a phenomenon which is that HY *of* X constructions may also be at least in part

[50] Nicolaisen, *Scottish Place-Names*, p. 84.

Figure 3.29: Comparison of Y *of* X Constructions for All Features



brought about by syntactical constraints. It seems that there is a higher ratio of HY *of* X constructions as opposed to simple X Y constructions where the generic element is plural as in table 3.20. Of the nineteen names with plural generic plural elements, ten have the Xs *of* Y construction, i.e. a little over 50%, compared to 34% of names with a Scots phonology having an X *of* Y construction.

It is also of note that the watercourses these names represent are all small (average *geogscore* is 3.86), which makes sense, since it is easy to talk about smaller features in the plural than larger ones (e.g. no one would say *Rivers of* X since the watercourses would be so large as to demand an individual toponymic identity).[51]

Another linked phenomenon is that where names with directional secondary specific elements encourage genitival constructions, for example East Burn of X is more common than East X Burn. A possible reason for this is that the East X Burn construction can create confusion, for example from the name West Seaton Den it is unclear whether the *den* is named after West Seaton or if there are

[51] It is interesting to note that this phenomenon does not seem to occur with other features. For instance, with protuberances it is common to have collective names representing large areas such as the Cheviot Hills or the Grampian Mountains.

Table 3.19: HY *of* X Constructions

element	frequency
Burn	824
Loch	35
Water	33
Stripe	21
Pow	5
Grain	4
Latch	3
Den	2
Doups	1
Pools	1
Lake	1
Slack	1
Pond	1
Glen	1
Bog	1

Table 3.20: Plural HY *of* X Constructions

	Burns	Lochs	Grains
Xs *of* X	Burns of Allalees Burns of Allalees	Lochs of Achlee Lochs of Allt na Craoibhe-caorainn	East Grains of Allachy West Grains of Allachy Grains of Auchterwhaile Grains of Tanar Grains of Coralea Grains of Slochd Chaimbeil
X Xs	Burnt Burns Burnt Burns Three Burns	Lazywell Lochs Pitcastle Lochs	Benty Grains Cot Grains The Grains The Grains

two *dens* named after Seaton, differentiated by East and West. This is formally shown in the trees in figure 3.30.

Figure 3.30: Underlying Structures of Scots Place-Names with Cardinal Directions

Equivalent Gaelic evidence is thinner but still present: The situation is somewhat different due to the differences in Gaelic and Scots word order.

3.7.2.1 Allt na X vs Burn of X

Nicolaisen states that the North Eastern *Burn of* X constructions are the result of translated *Allt na* X names within the AOS.[52] If this hypothesis is correct, one might expect to find evidence of variation between RNs with *Burn of* X as a modern form and *Allt na* X in the old forms, or *vice versa*. In the entire database there is only one clear example of this: Allt a' Mhòirneas (4777) which has as an old form, *Burn of Moreinch* (1783 Stobie). The careful reader will notice that this name has *Burn of* X in the old forms, with *Allt na* X as the current form. This is the opposite situation which one would expect from Nicolaisen's hypothesis.

By far the commonest way for a modern day *Burn of* X name to have derived from an *allt* name is in a pleonastic context: for example Burn of Allnaharvy from *Ald na heruy* (c. 1591 Pont map 7), but this is a separate process from that of generic element transliteration. This is discussed in section 4.16 on page 160.

Distribution: The general areas overlap slightly mainly in the area between the Upper Don and the Spey (this is also where the two examples listed above are broadly situated). This overlap, however, is not really meaningful, since it is the same as that of all Gaelic and Scots generic elements.

3.7.2.2 X Burn vs Burn of X

The largest amount of variation between old forms and modern forms, is simply of that between *Burn of* X and X *Burn*. The variation goes both ways: there are sixteen names with *Burn of* X as a modern form but X *Burn* in an old form with the same specific element, and 19 the other way round. This information could be unreliable however, since fourteen of the sixteen names in the first group are solely recorded in Ainslie, and to a lesser extent the same is true of the other group, but with Stobie. This strongly suggests that individual map makers may have influenced the data as it is seen here. However, evidence from non-manuscript sources, such as local informants[53] shows a fair degree of variation occurring in the same toponym and even sometimes with the same speaker, suggesting that the variation is perhaps analogous to variation in names with *river* as the generic element, where River Tay is the correct Ordnance Survey term, but in common parlance is generally referred to as 'the Tay'. This may explain the forms with Stobie.

3.7.3 Sc Burn, Sc River and Sc Water

3.7.3.1 Burn

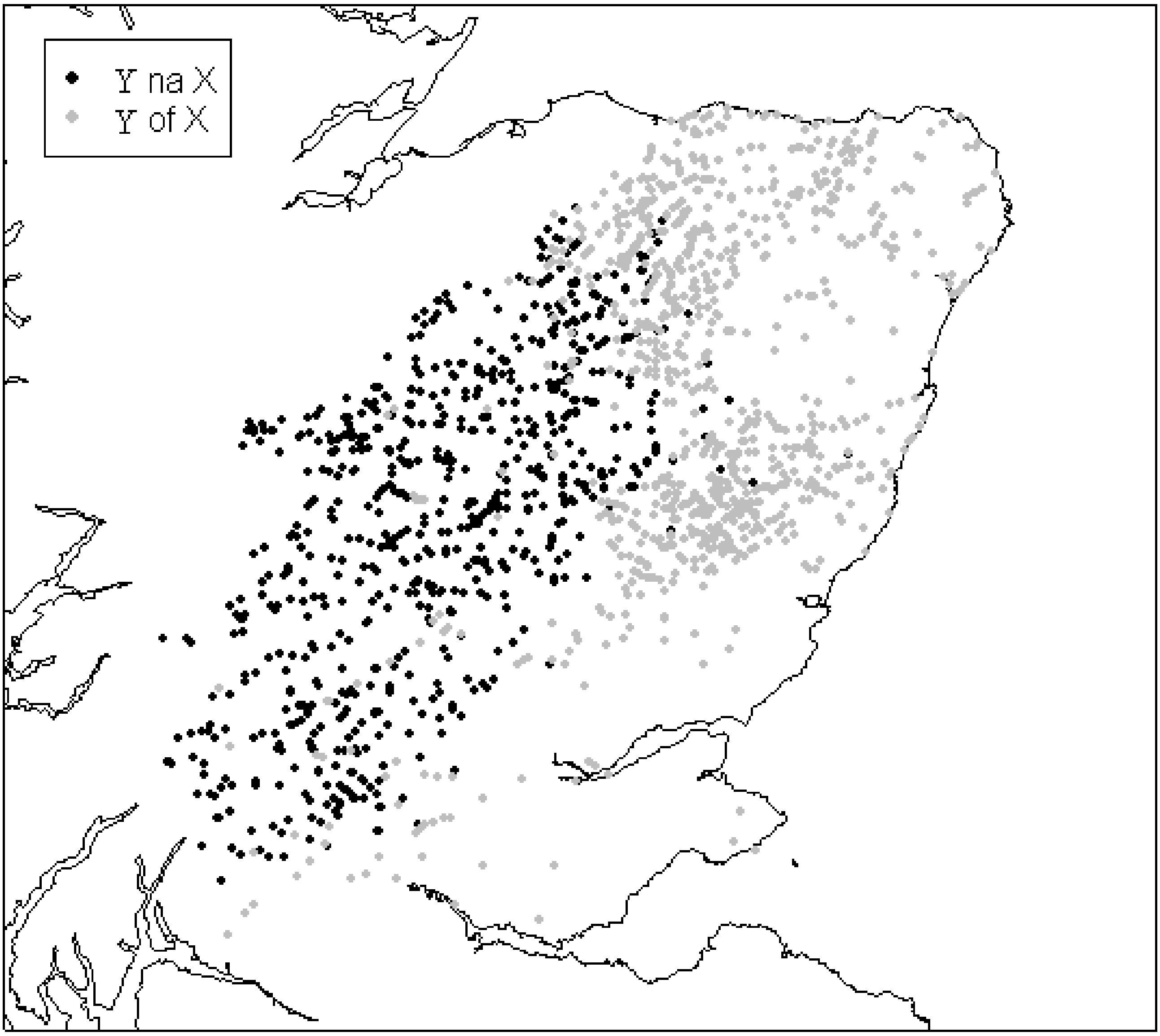
Semantics: The element *burn* is the *de facto* term for smaller watercourses in Scotland, in contrast to *river* and *water* for larger watercourses, as shown in figure 3.32 on page 102.[54] The Dictionary of the

[52]Nicolaisen, *Scottish Place-Names*, p. 77.

[53]For example: Watson and Allan, *The Place Names of Upper Deeside*.

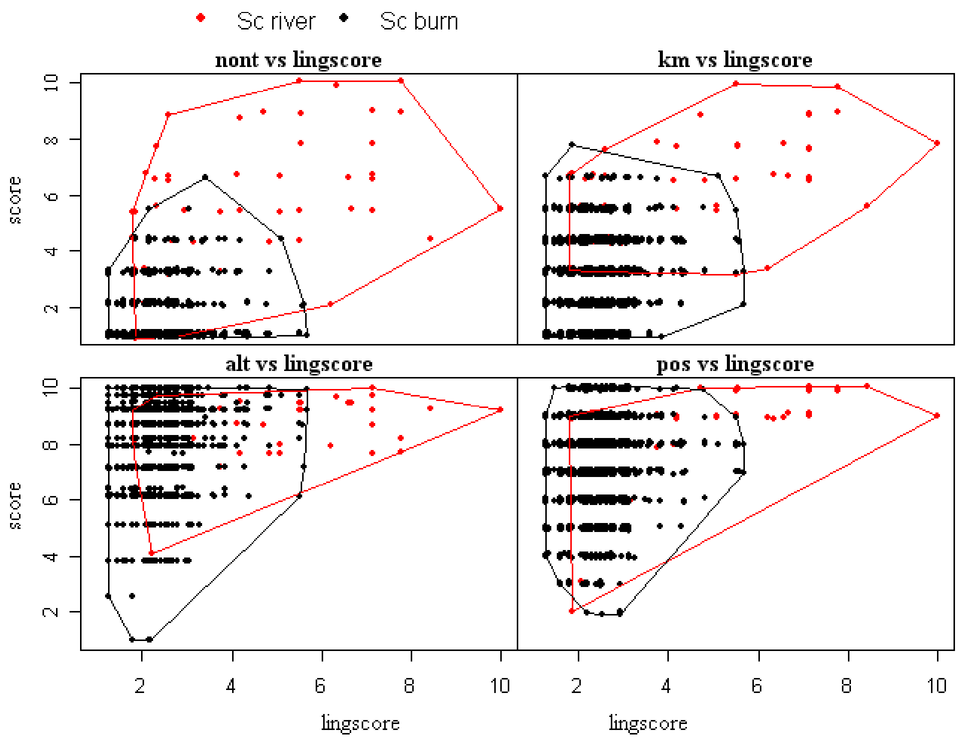
[54]This concurs with a comment made in J. G. Johnston, *Come Fish With Me* (London, 1948), p. 207: 'A "water" is something bigger than a burn, but not large enough to be called a river.' and a comment from Ane Description of Scotland in Walter MacFarlane, *Geographical collections relating to Scotland made by Walter Macfarlane* (1906), p. 146 'a glen where

Figure 3.31: Distribution of G Y *na* X and Sc Y *of* X



Scots Tongue simply describes it by its equivalent English terms: 'A brook or stream'.[55]

Figure 3.32: Comparison of Sc *burn* vs Sc *river*



3.7.3.2 Water

Semantics: 'A large stream, usu. thought of as intermediate in size between a Burn and a river, freq. a tributary of a main river or occas. applied to the upper reaches of what becomes a larger river'.[56]
Distribution: This generic exists throughout the whole of Britain, although it is commonest in Scotland. Within Scotland it is commonest in the borders, with some in the AOS but very rare outside, except for Sutherland and the Northern Isles, especially Shetland, probably under Norse influence. The average *geogscore* for this element is the second highest of all the generic elements (6.02 before adjustment).

3.7.3.3 River

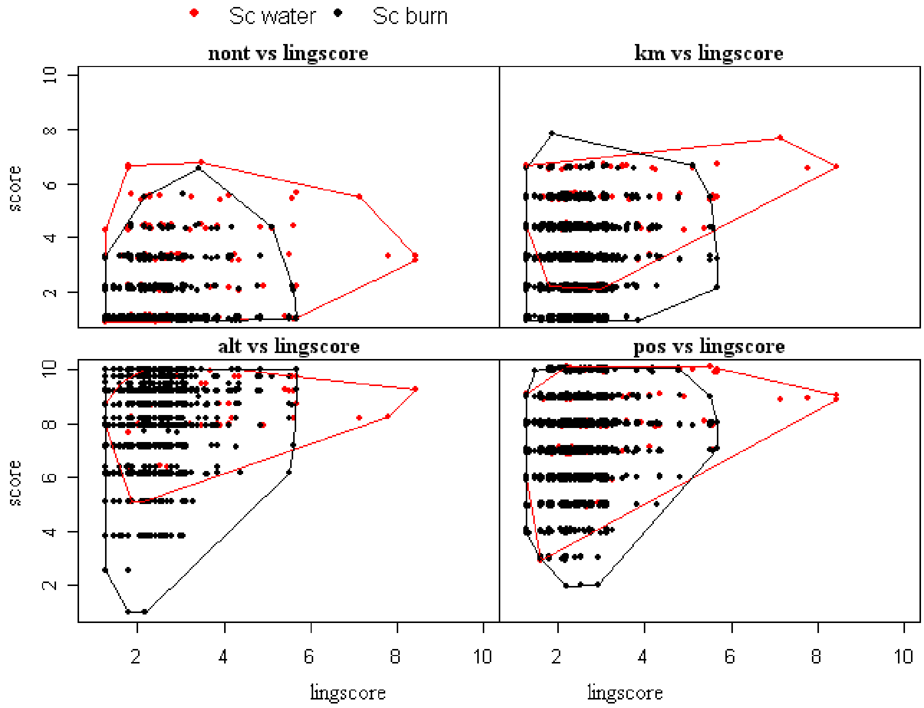
River is the *de facto* term for larger watercourses, and as such occurs throughout Scotland, even in some Gaelic speaking areas. River is an epexegetic term, which means the generic is less tightly bound to the specific. This is reflected in the variation in Scots and SSE in the names of *rivers*, one can

throw this *water* or *litle river* doeth flow'.

[55] 'Dictionary of The Scots Language'.

[56] Ibid.

Figure 3.33: Comparison of Sc *burn* vs Sc *water*



equally say 'The Tay' or 'The River Tay'. Obviously watercourses called *river* are the largest, with the highest *geogscore* as in table 3.29 on page 124.

Rivers tend to be the watercourses flowing through glens with tributaries of small mountain streams. Where watercourses called *river* are short, they link other watercourses, and as such have a large amount of water flowing through them, such as River Druie; as such, the RNs all have high *nont*. Of the thirty-four *rivers* six names have a *nont* of under ten, with the rest over twenty.[57]

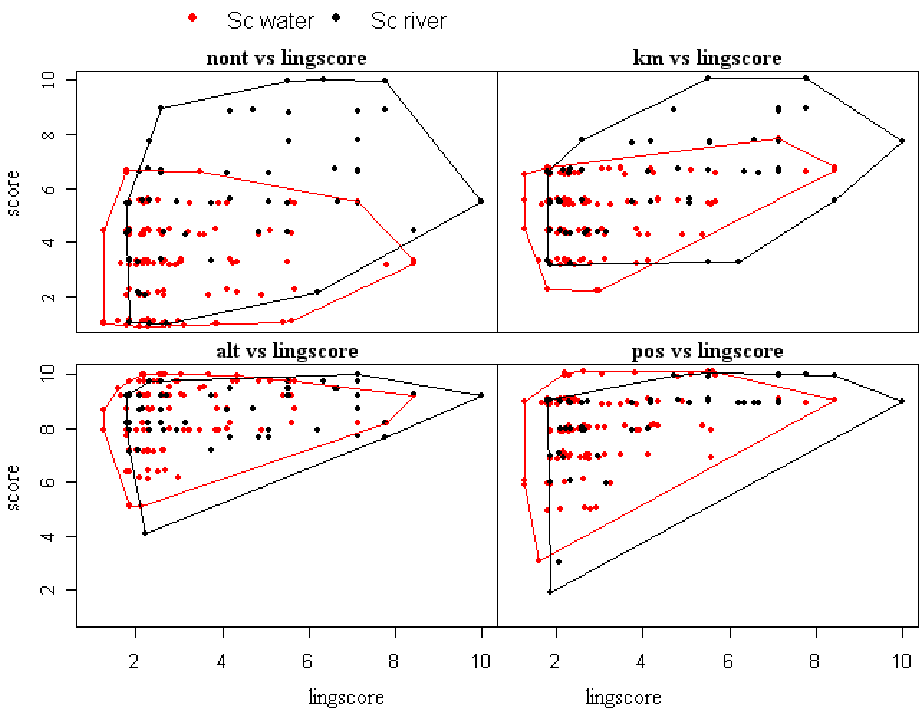
A comparison of Sc *water* and Sc *river* as in figure 3.34 shows what one might expect. Whilst there is an overlap in size of *waters* and *rivers*, the largest watercourses are called *rivers* and cannot be called *water*, whilst the smaller watercourses in the set can be called *water* and not *river*.

3.7.4 Sc Stank

Semantics: The primary definition is given as 'pond', with a secondary definition of '[a] stretch of slow-moving water, a ditch; a sluggish stream or river.'[58] *Stank* only exists in two entries: Black Stank and Teuchar Stank. *Stank* also appears as a specific element in Scots speaking areas of Scotland. Presumably stanks are common throughout rural Scotland, but are so insignificant as features that only a random few gain toponymic status, resulting in the fairly random distribution.

[57] This is a good example of how *nont* and *km* differ substantially, as discussed on page 21.
[58] 'Dictionary of The Scots Language'.

Figure 3.34: Comparison of Sc *water* vs Sc *river*



3.7.5 Sc Slack, Sc Den and Sc Stripe

These three elements all have comparable distributions as can be seen in figure 3.35 on page 105. They all describe small watercourses.

3.7.5.1 Sc Slack

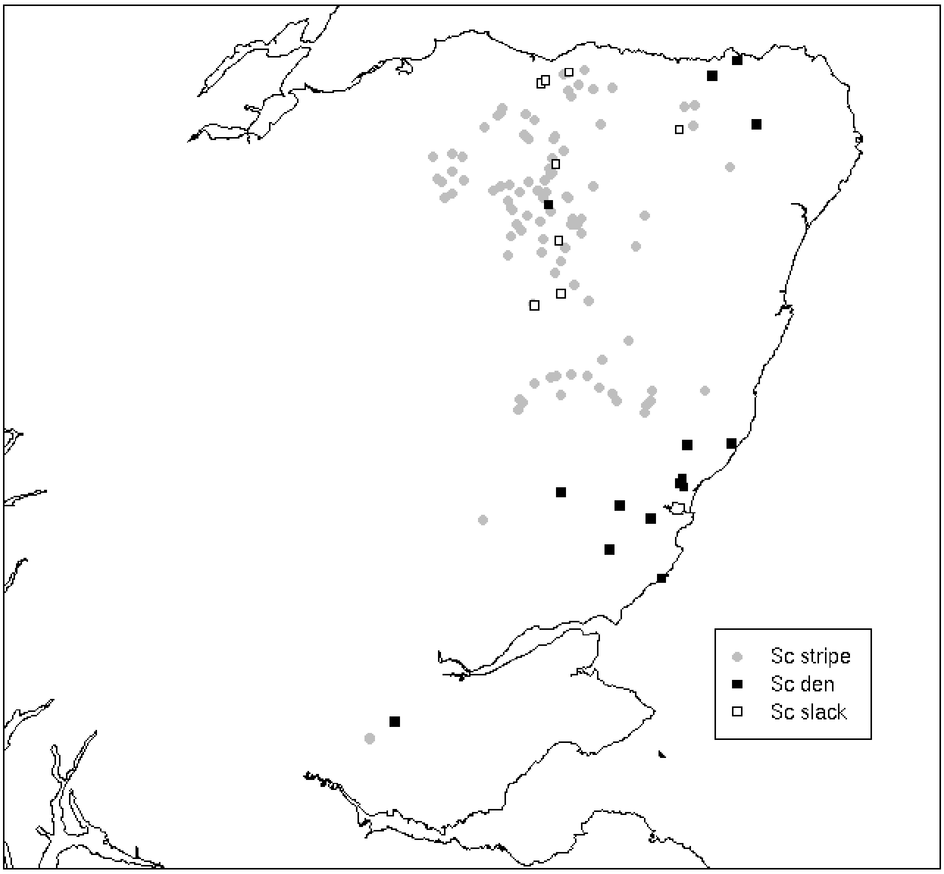
Semantics: The definition of this term is: 'A hollow or declivity, esp. between hills, a saddle in a hill-ridge, a defile, dell, pass'.[59] which makes it clear that, like *glen* and *den* this term is primarily a name for a concavity in the landscape which has gained a secondary meaning of a watercourse. **Distribution:** This element is entirely restricted to Banff, Moray and Gordon. This is precisely the same area as the northern *Burn of* X cluster.

3.7.5.2 Sc Den

Semantics: 'A hollow where the ground slopes on both sides; generally, such a one as has a rivulet running through it; a small valley'.[60] The primary meaning of *den* is not as a hydronym. In this survey, only ones 'in blue' on the OS maps have been counted, i.e. only ones considered a watercourse by the OS are counted. The specific elements are all Scots in origin. The possible exception is Dowelly

[59]'Dictionary of The Scots Language'.
[60]Ibid.

Figure 3.35: Comparison of Sc *slack*, Sc *den* and Sc *stripe*



Den, perhaps reflecting an element G **dubh*, 'black' with some other element or suffix, or from G **do-bhaile*, 'bad farm'[61] (although there is no evidence of a settlement here). Alternatively the name could also be a Scots surname, or reflect Sc *dowly*, 'sad, doleful'. The semantic make-up of these names is typical for small watercourses: One third represent people or professions, and another third specific places, with the remaing third miscellaneous semantemes.

Distribution: The majority of *den* names as hydronyms are between the Tay and the Dee, but the main defining environment for a *den* is low altitude. They all have their lowest points around sea level. In addition, the *dens* are all small features, the longest being 7.5 km.[62] The distribution is similar to that of *Burn of* X and others above, however the Tay to Dee names are generally further South. Although it seems that *den* has a somewhat different distribution, from *stripe* and *slack*, it will be shown below that their distributions are in fact closely related (see figure 3.47 on page 123).

[61] See Taylor and Márkus, *The Place-Names of Fife*, p. 306 for other examples of names with this derivation.

[62] *Geogscore* is not particularly relevant here, since names at low altitude are marked up in *geogscore*, yet short watercourses are marked down, cancelling each other out.

3.7.5.3 Sc Stripe

Semantics: 'A small stream, a rivulet, rill... a small channel crossing a sandy beach'.[63] The specific elements are fairly typical of a Scots generic element. Many of the names at higher altitudes are coined from hill names, whilst 'Blind Stripe' is also popular. One anomaly is that there appears to be more White Stripes (4) than Black Stripes (2).
Distribution: *Stripe* has a similar distribution to *Burn of* X constructions, there being two main clusters, one in Banff, Gordon and Moray area, and another in Kincardine and Angus. There is even a smaller cluster in the Ochils, where Burn of Sorrow and Burn of Care are. There is also Hunt Stripe and Darn Stripe.

The *geogscore* range is between 3.28 and 5.06, but the main condition for a watercourse to be called a *stripe* is its length, the *km* range is between .5 and 6, with the average at just 1.27 *km*.

3.7.6 Sc Pond and Sc Lake

Pond is not productive in Scotland, there being only two names in the AOS, Bennybeg Pond and Pond of Drummond on the Earn. OS Landranger shows another 5 ponds throughout the rest of Scotland.

It is sometimes said that the Lake of Menteith is the only *lake* in Scotland, but there are in fact five others in the AOS and a few more on OS Landranger, all in the Borders.[64] Three of the names are in the Ythan catchment system. They are all on the east coast in Scots / SSE speaking areas and represent relatively small bodies of water (all the *geogscores* are between 4.55 and 5.69).

There are two possible answers here. Firstly, the term *lake* represents a Scots word which has become a generic element. *Lake* is said to be a Scots word for a 'pool or pond' or a 'small stagnant pool, esp. one formed at ebb-tide on the shore'.[65] Secondly, the term *lake* is an anglicisation (in this case meaning English not SSE) of the word *loch*, and as such should be considered an orthographic variant of that term, in the same way Sc *auld, old* can be a variant of G *allt*. The two terms are compared in figure 3.37 on page 108 which shows that *lake* sits within the boundaries for *loch* in each graph, meaning that any *lake* could be a *loch*.

Since the definition of *lake* suggests the term is a synonym for *pond*, a comparison of these two terms would show if the comparison extends to the hydronymicon from the lexicon. Figure 3.36 shows this is the case.

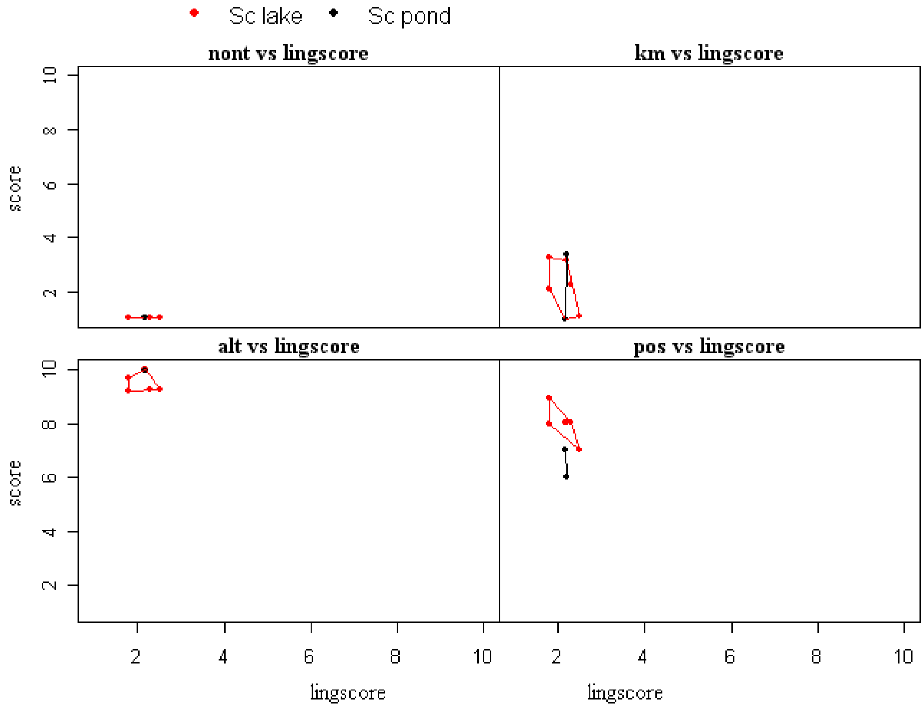
3.7.7 Sc Loch

Having ascertained that *lake* and *pond* are synonyms representing similar water-basins, it would now be prudent to compare these with Sc *loch* as in figure 3.37 on page 108, which shows that as far as *nont* and *km* are concerned, *pond* and *lake* act as diminutives of Sc *loch*.

---

[63] 'Dictionary of The Scots Language'.
[64] Moreover, the Lake of Menteith is probably a misinterpretation of Sc *laich*, 'lowland'.
[65] Ibid.

Figure 3.36: Comparison of Sc *lake* vs Sc *pond*



3.7.8 Sc Pow

Pow appears both as a generic and specific element and is noteworthy because it seems to retain the same sense in both uses. *Pow* is typically a slow moving stream, over a flat area and as such only exists in the Lowland area. The altitude is always low; this is quantified in section 3.43 on page 115.

3.8 The Relationship between Generic Elements and Specific Elements

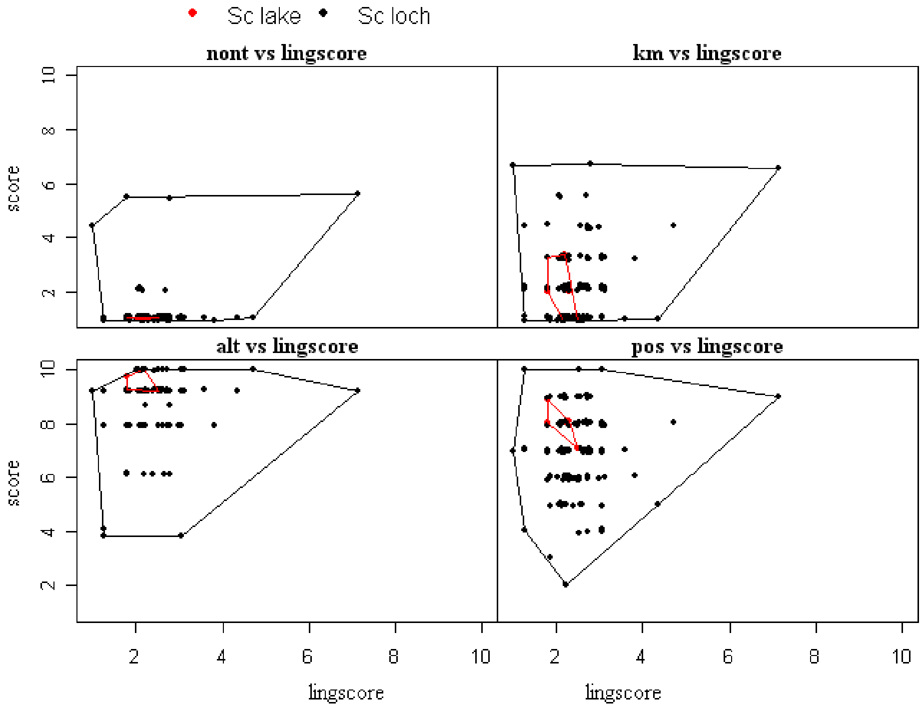
3.8.1 Simplex Names with the Definite Article.

This section discusses the phenomenon of RNs in Scots which occur as *The* X or in Gaelic as *An* X or variants.

3.8.1.1 Scots Names: The X

Within a Scots context, names simply consisting of the definite article followed by a specific element fall into two categories. The first group are those where the specific element can also act as a generic element as in table 3.21. It is interesting to note that the Dour, Aberdour's burn, despite being P-Celtic in origin, acts in the same way as the Scots names. The same could possibly be said of the Pow, which

Figure 3.37: Comparison of Sc *loch* vs Sc *lake*



has at its confluence Aberfoyle, suggesting the Pow is ultimately P-Celtic in origin.

Table 3.21: *The* X Constructions with Generic Elements as Specific Elements

name	frequency
The Lead	3
The Grains	2
The Strip	2
The Den	1
The Dour	1
The Latch	1
The Pow	1
The Slack	1

The second group contains specific elements which are *bona fide* specific elements and are Gaelic in origin, as in table 3.22. This shows two terms: G *slob*, and G *slug* are lexical items describing water features but not terms in common parlance in the Gaelic hydronymicon. The first term is idiosyncratic.

Table 3.22: *The* X Constructions with Non-Generic Elements as Specific Elements

name	etymology
The Shevock	G *sèamhag*, ‘little quiet one’
The Slobach	G *slob*, ‘pond, dam’ + *ach*, ‘place of’
The Slogs	G *slug*, ‘puddle’

3.8.1.2 Gaelic Names: An X

There are many Gaelic names with this structure, but only three within the AOS are actually hydronyms. Names with this structure are more common in areas where Gaelic has until recently been spoken or is still spoken.

Table 3.23: G *An* X Constructions

id	RN	etymology
553	Am Beanaidh <	G *beannach*, 'fork'
1078	A' Chaim <	G *caim*, 'loop or curve'
2927	An t-Eileach <	G *eileach*, 'mill-dam, bank'

It could reasonably be said that these three names relate to a feature of or near the watercourse and are not true hydronyms. This type of semantic process is called *automeronymy*, whereby a part and a whole are named the same thing, but with distinct senses.[66] Within Gaelic hydronymy as a whole, the definite article is not generally applied directly to hydronyms without generic elements. Thus, whilst in Highland areas settlement and non-hydronymic natural feature names without generic elements take the definite article, this is not true of RNs. For example, Am Ploc, the Gaelic name for Plockton means in Gaelic 'the pimple'. The reason for this situation is not clear, but one factor may be because within Gaelic toponymy transparent names are generally given the article (as in the example above, and for names such as Am Baile Meadhanach, 'the middle town', often anglicised to Balmeanach or Balmeany etc.), with Gaelic hydronyms, a large amount of these are not transparent, and it may be that to not use the article was seen as the norm due to this high number; this was then extended to transparent names such as Loch Dubh etc.

It is interesting to note that P-Celtic (e.g. The Pow and The Dour) and Scots names (e.g. The Den) of this type fall into one category whilst Gaelic hydronyms with the article are rarer or non-existent.

3.8.2 Generic Elements as Specific Elements

In several cases, generic elements are also specific elements in a way that does not denote a pleonastic name or a specific natural feature, instead the element is used in a simplex way inside the specific element, e.g. Slack Burn, where *slack* is used as a specific feature, but elsewhere can be used as a generic e.g. Dry Slack. As might be expected, the commonest elements to be used like this are those names which have once meant some feature related to a watercourse but have since become hydronyms themselves, These are *fèith* (an example of this term as a generic is Fèith Mhòr, and as a specific: Fee Burn) and *eas* (an example of this term as a generic is Eas Domhain, and as a specific: Loch Eas Domhain). In terms of distribution, only these terms *fèith* and *eas* really have comparable distributions between the two types of occurrences, but in general a similarity is that the term as a specific is more widely distributed than that of the generic. This is not surprising, since it is a general rule, on linguistic contact, generic elements are translated and specific elements transliterated.

[66] An example of this is 'arm', which can represent the whole arm including the hand, or the limb from shoulder to wrist only.

Figure 3.38: Comparison of G *eas* as a Generic and G *eas* as a specific

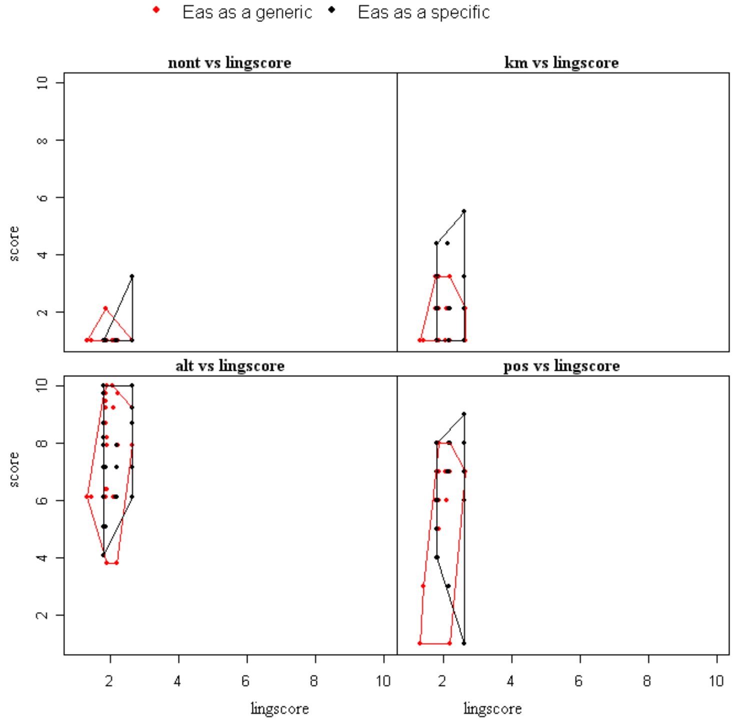


Figure 3.39: Comparison of G *fèith* as a Generic and as a Specific

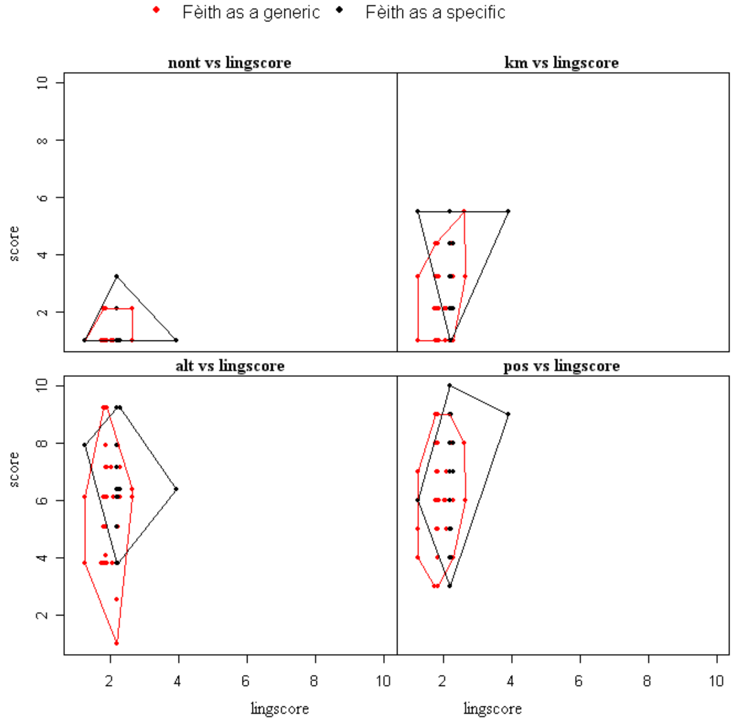
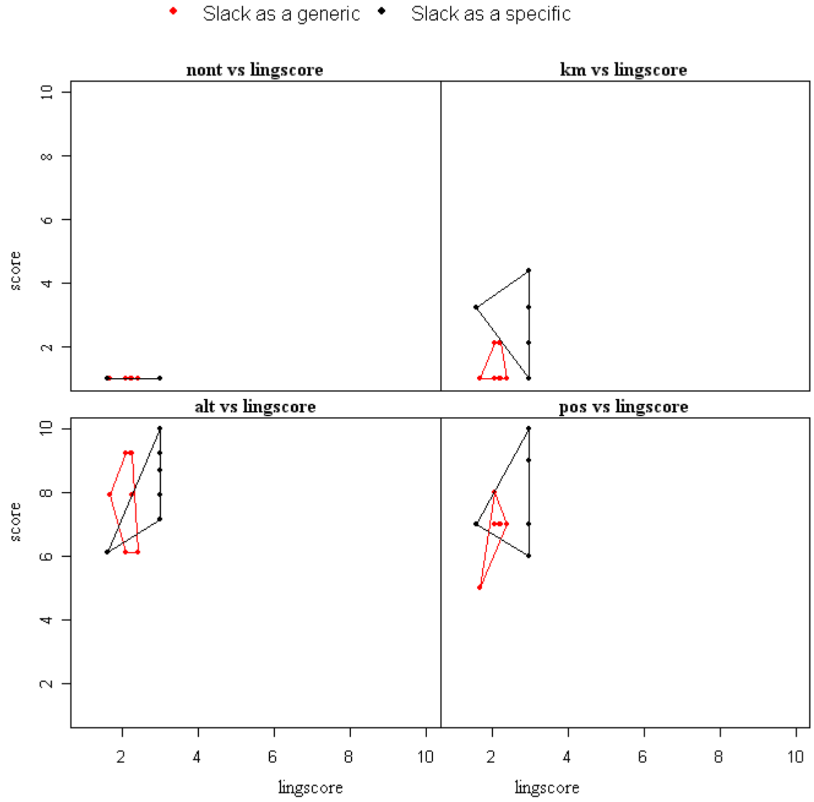


Figure 3.40: Comparison of Sc *slack* as a Generic and as a Specific



A comparison of these elements as a generic and as a specific show similarities for the element *eas* as shown in figure 3.38 on page 110, for *slack*, figure 3.39 on page 110 and in figure 3.40 on page 111 all show that each of the elements is interchangeable between its occurrences as a generic and a specific. Figure 3.41 on page 112 is an example showing that the two treatments of the element *eas* also broadly have the same distribution, especially in the south-west of the AOS. This similarity in distribution tells us something about the semantic development of *eas* from generic to specific. It would seem that in the west, *eas* began to describe watercourses by a process of *automeronymy* discussed above. Thus Dubh Eas came to represent the waterfall but also the whole watercourse. As the term spread further east however, the distributions become less equivalent. It would seem in these cases that *eas* was coming to be understood as a generic element in its own right. This process is a type of *metonymy* whereby the whole is known from the part.

3.9 Generic Elements and the Statistical Method

In order to seek the generic element for a particular specific element, it is necessary to calculate what possible generic elements could be attached to that specific which would not alter the existing ranges for the generic elements. A hypothetical example of this would be a specific of a watercourse which was 1 km long; it would not be appropriate to posit this watercourse as having a generic element of

Figure 3.41: Comparison of G *eas* as a Generic and as a Specific

river, since the existing shortest watercourse with the element *river* is 3 km long. It would be wise to posit a generic element for which there already exists a watercourse of the length of 1 km, e.g. *burn*.

This line of reasoning can be extended beyond just length to the other components of *geogscore*. To calculate this range for any particular watercourse, the R function *genelemfinder* has been written. The main application for this function is to uncover what possible generic elements could exist for RNs only known via Aber- or Inver-names.

3.9.1 Abercairny: A Case Study

As a case study, the name Abercairny shall be studied. This is situated near the foot of Muckle Burn (4870), a Scots name. Since it has been shown that Aber- nearly always associates itself with a RN, one can confidently state that *-cairny* is a P-Celtic specific element.[67] The name exists in at least one other place, as Abercairnie[68] and perhaps as *Abyrcardon* (1224-1233 Moray Register p.83).[69] The old

[67] Watson, *The Celtic Place-Names of Scotland*, p. 463.
[68] See Ibid. and Erskine Beveridge, *The 'Abers' And 'Invers' of Scotland* (Edinburgh, 1923), p. 6.
[69] Ibid.

form: *Abircairdney* (1617 RMS vii no. 575) suggests the specific is related to the Pictish root related to W *cardden*, often glossed as 'copse'.

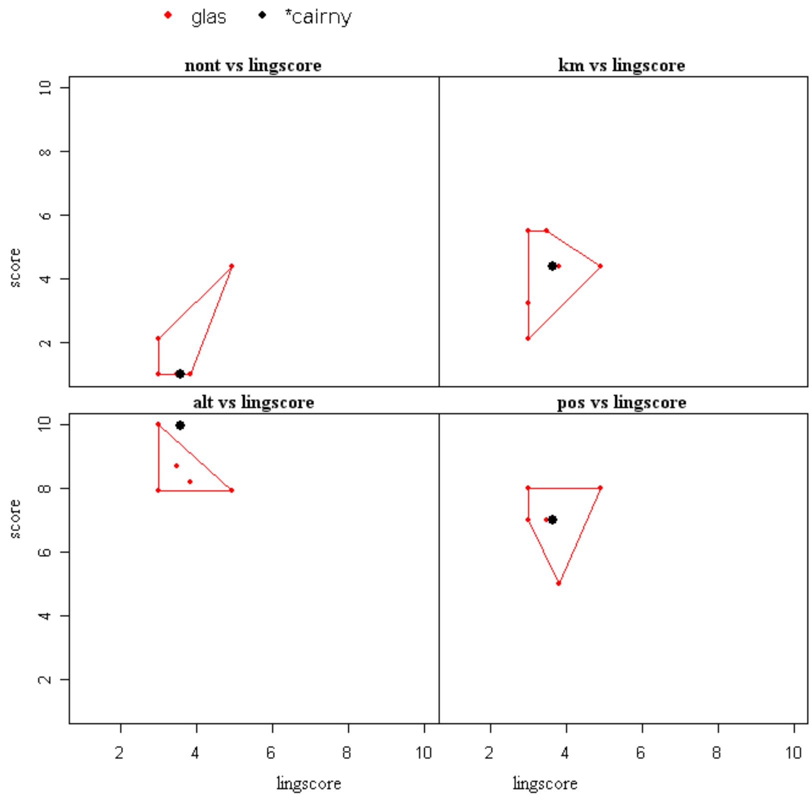
Table 3.24: Generic Element Possibilities for **cairny*

element	minkm	maxkm	minnont	maxnont	minalt	maxalt	minpos	maxpos
glas	2.00	5.00	1.00	4.00	7.94	10.00	5.00	8.00
Allt	1.00	6.00	1.00	5.00	1.00	10.00	1.00	9.00
Loch	1.00	8.00	1.00	7.00	1.00	10.00	1.00	9.00
Other	1.00	4.00	1.00	3.00	2.54	9.23	2.00	8.00
Burn	1.00	7.00	1.00	6.00	1.00	10.00	2.00	10.00
Den	1.00	4.00	1.00	1.00	7.94	10.00	5.00	10.00
Loch	1.00	6.00	1.00	5.00	3.83	10.00	2.00	10.00
Water	2.00	7.00	1.00	6.00	5.11	10.00	3.00	10.00
The X	1.00	6.00	1.00	2.00	3.83	10.00	5.00	10.00
Glen	1.00	3.00	1.00	1.00	6.14	10.00	6.00	9.00

Table 3.24 on page 113 is produced when *genelemfinder* is run on this name. For the record, the scores for **cairny* are: *kmscore* = 4.38, *nontscore* = 1, *altscore* = 10, *posscore* = 7. From here one can employ a certain amount of common sense. The generic will not be *loch* or any other body of water. The generic element type 'Other' reflects miscellaneous names, and as such do not form a group. The original generic will also not be from a later stratum than the specific element, so the Scots names can be ruled out. This leaves: *glas* and *allt*. Since the name Abercairny is P-Celtic, perhaps the only P-Celtic generic in the list is viable, *glas* leading one to posit **glas carden* or **carden-glas* as the original RN. It is possible that the original name was a simplex, since the database has no information on Pictish simplex RNs, as it does with Scots and Gaelic names. It is not known whether this type of name was even permissible in Pictish.

Figure 3.42 on page 114 compares the current name of this watercourse, Muckle Burn, with the conceptual space of *glas*. Ignoring the x-axis, which represents the linguistic qualities of Muckle Burn (irrelevant here), here labelled as **cairny*, it can be seen that **cairny* fits into the conceptual space of *glas*. There is a possibility that the name Muckle Burn and **cairny* are not of the same extent, that is, that **cairny* did not relate to the fullest extent of the existing watercourse. Within the terminology, this would mean that **cairny* had a lower *kmscore* than otherwise anticipated. The existing *kmscore* as mentioned above is 4.38. Table 3.24 shows that the minimum and maximum ranges for exiting *glas* names are two and five respectively. This means that **cairny* is at the higher end of the range, and if its length were shortened, it would still comfortably fit between the minimum and maximum values. In fact, the *km* length of Muckle Burn/**cairny* is 5.0; it could be reduced to a length of 1.5 km and it would still be a viable candidate for having the generic of *glas*.

Figure 3.42: Comparison of P *glas* and **cairny* (= Muckle Burn)



3.9.2 Next Section

In the next four sections, a somewhat different approach is taken. Rather than investigating the elements on a case by case basis, the whole range of generic elements will be analysed by comparison to a particular component of *geogscore*.

3.10 Generic Element Compared to Components of Geogscore

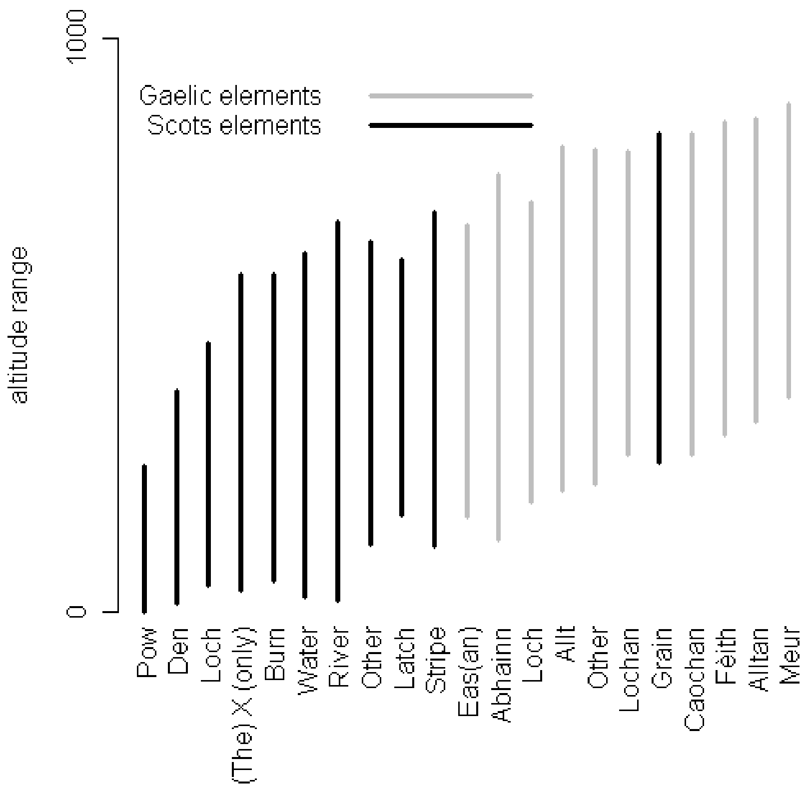
3.10.1 Generic Element vs Altitude

Figure 3.43 on page 115 shows the average altitude range (not *altscore*) for every generic element except those where the element appears less than 10 times, since there is not enough data for these elements to accurately draw any conclusions. A number of points can be discerned:

- Watercourses with Scots generic elements are in general all at a lower altitude and those with Gaelic at a higher altitude.
- The obvious outlier here is the element *grain*. This has been discussed on page 92.
- *Pow* and *den* are restricted to the lower altitudes as discussed in section 3.3.6 on page 58.

- *Meur* and *alltan* are restricted to higher altitudes. These points are discussed on pages 76 and 79 respectively.

Figure 3.43: Generic Elements by Altitude Range



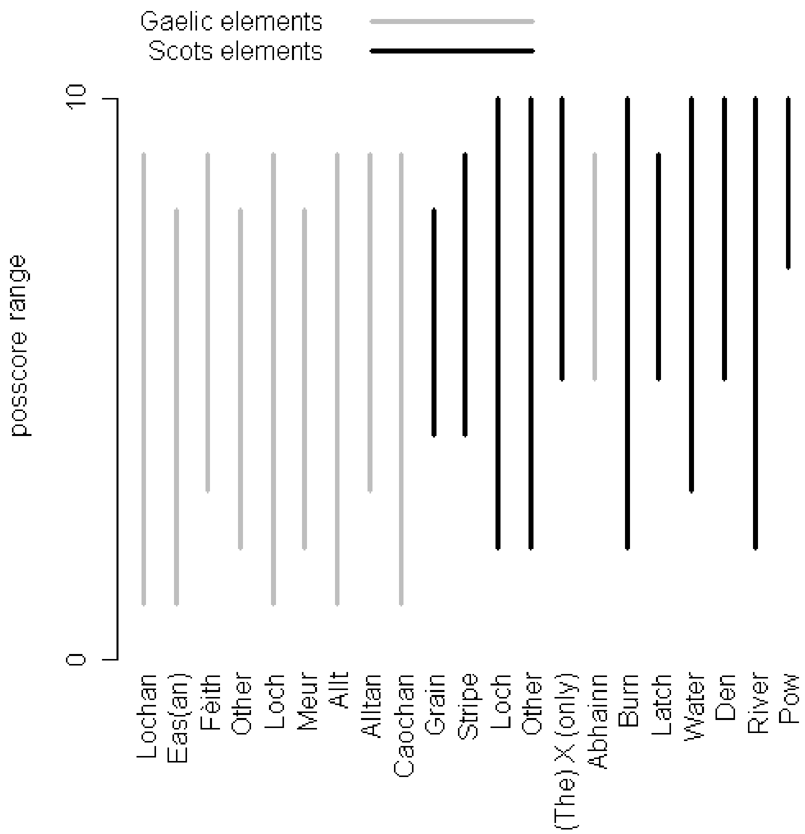
3.10.2 Generic Element vs Position

Figure 3.44 on page 116 shows the following points:

- The same phenomenon as explained in point 1 in the previous section can be observed with position (the elements seem back to front because in the altitude graph above *altscore* is not being measured but the actual altitude).

- The outlier in the graph is *abhainn* which has an uncharacteristically high *posscore* range for a Gaelic generic element. This is perhaps a somewhat illusory phenomenon. Firstly, there are other Gaelic elements such as *uisge* that also have a high average *posscore* but are excluded from the graph due to there being too few examples. Secondly, as stated, the AOS does not contain the core distribution for *abhainn* which is further to the west.

Figure 3.44: Generic Elements by *posscore* range



3.10.3 Generic Element vs Length

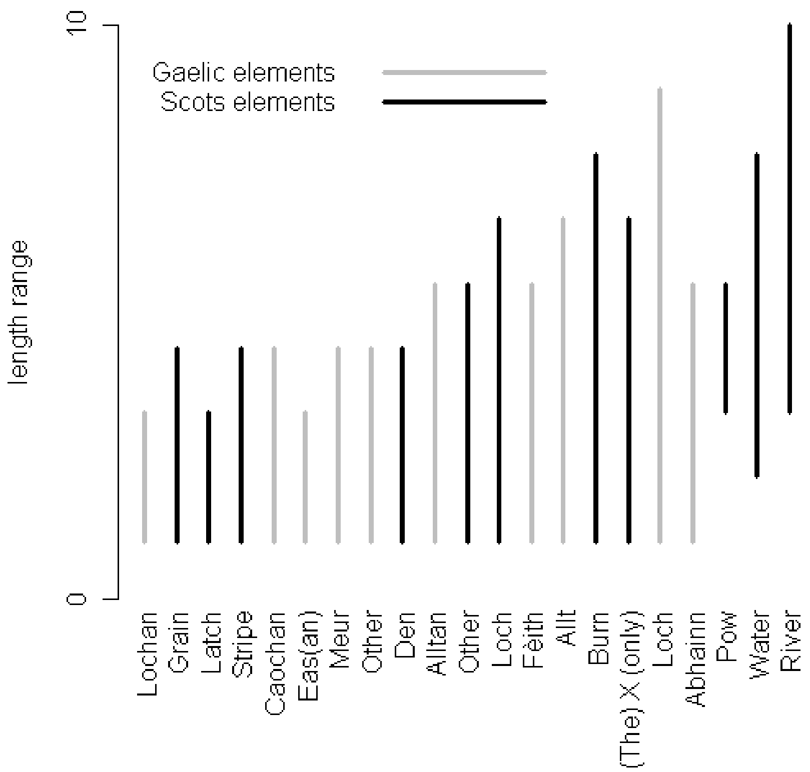
Figure 3.45 on page 117 shows a number of points:

- The vast majority of generic elements can represent short watercourses. The three that do not are *pow*, *water* and *river*. This is to be expected for the final two terms, but perhaps not in the case of *pow*. This is doubtless due to the geographical correlation between watercourses which flow through low altitudes being longer, whilst mountain streams are shorter.
- As expected, the diminutive versions of elements are all shorter than their non-diminutive counterparts, e.g. *alltan* vs *allt*; *lochan* vs *loch*.
- The default generic elements, (e.g. *river*, *burn* and *allt* for a watercourse and *loch* for a water-basin) are those with the greatest difference between their minimum and maximum.

3.10.4 Generic Element vs Number of Named Tributaries

Figure 3.46 on page 118 suggests these points:

Figure 3.45: Generic Elements by *kmscore* Range



- It can be clearly seen that the necessary conditions for many generic elements is that they have very few or no named tributaries. These elements are the same as those at high altitude for their respective strata (i.e. Sc *stripe* and *latch* are at a high altitude for Scots names, but not overall). This is to be expected to a certain extent: watercourses at high altitude must be short as mentioned in the previous section.
- The minimum *nontscore* for all generic elements is 1. This means that for all generic elements, including those such as *river*, it is possible for the watercourse they represent to have none or very few tributaries, which is in contrast to *kmscore*.

3.11 Genelem and the Hierarchical Network

Table 3.25 on page 119 shows the forty commonest combinations of watercourse confluences where the watercourses in question have a particular generic element. The commonest structure is for a watercourse with the element *burn* to flow into a watercourse also with the element *burn*, the second most is for a watercourse with the element *burn* to flow into a watercourse with the element *river*, and so on.

Figure 3.46: Generic Elements by *nontscore* range

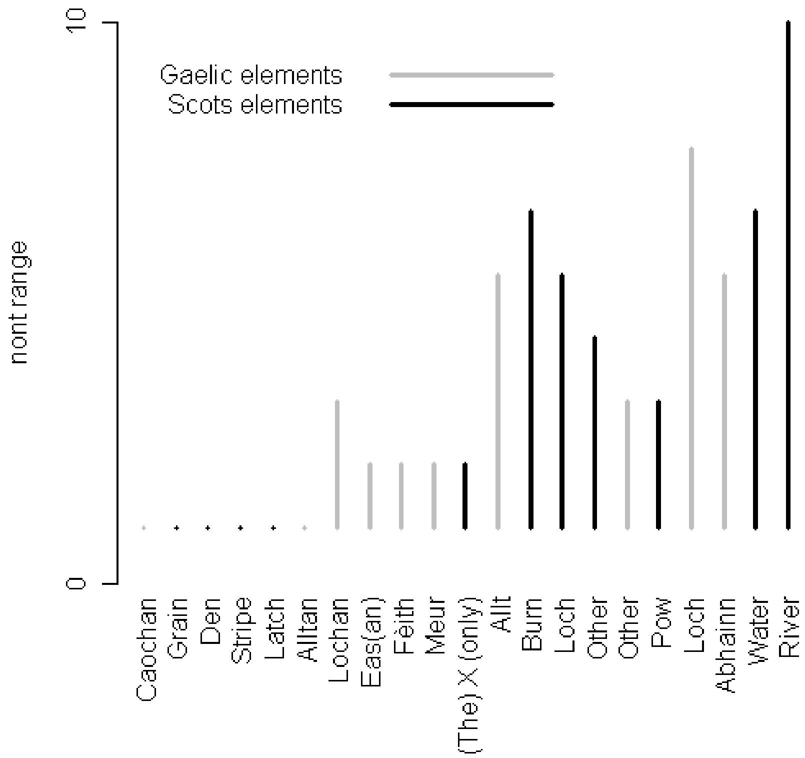


Table 3.25: Forty Commonest Combinations of Generic Elements between Main Watercourse and Tributary

main	tributary	frequency
Burn	*Burn*	1491
River	*Burn*	691
Water	*Burn*	522
Allt	*Allt*	494
River	*Allt*	399
Burn	*Allt*	271
Water	*Allt*	132
G *Loch*	*Allt*	105
Allt	*Caochan*	67
Burn	Sc *Loch*	65
Burn	*Stripe*	60
Allt	*Lochan*	57
Burn	*Caochan*	57
River	*Water*	55
Allt	G *Loch*	50
Allt	*Burn*	50
G *Loch*	*Burn*	46
Burn	G *Loch*	42
River	*Caochan*	39
Other	*Burn*	38
Burn	*Grain*	36
River	*River*	35
Sc *Loch*	*Burn*	35
G *Loch*	*Allt*	33
Allt	*Fèith*	28
Burn	Other	28
River	G *Loch*	28
River	*Lochan*	21
Burn	*Lochan*	19
Water	*Caochan*	17
River	Sc *Loch*	16
Water	*Water*	16
Lochan	*Allt*	14
Water	*Grain*	14
Water	*Stripe*	14
Allt	*Stripe*	13
Caochan	*Caochan*	13
River	Other	13
Water	G *Loch*	12
Water	*Fèith*	12

Table 3.26: Generic Element Hierarchy

Burn
River
Water
Allt
G *Loch*
Sc *Loch*
Other
Fèith
Other
Abhainn
Caochan
Eas(an)
Lochan
An X (only)
Pow
(The) X (only)
Uisge
Gleann
Alltan
Den
Stripe
Meur
Grain
Slack
Lake
Poll
Pool
Glen
Pond

The table shows a fairly standard network of generic elements. Some elements appear predominantly in the left 'main' column, whilst others occur predominantly in the right 'tributary' column. *Burn*, *river*, *water* and *loch* occur first and several times in the 'main' columns, whilst names such as *allt*, *caochan* and the other generic elements representing smaller watercourses all occur generally on the right hand side. If the generic elements are ordered by the first time they appear in the 'main' column, the hierarchy as in table 3.26 is shown. A number of points are notable here:

- G *loch* and Sc *loch* are both next to each other, suggesting a similar, if not identical, role in the network.
- G *meur* and Sc *grain* are next to each other, corresponding with what is mentioned above.
- G *uisge* and Sc *water* are not near each other, which concurs with the discussion and conclusion above.

This methodology can also tell us the number of occurrences where two neighbouring names have the same specific element but a different generic element. An example of this is Lemno Water and Lemno Burn. Table 3.27 shows the number of combinations of this, only in instances where one watercourse flows into the other watercourse.

Table 3.27: Hierarchy of Generic Elements for Watercourses Where One Flows into the Other, with Different Generics and the Same Specific

Upper element	Lower element	frequency
Loch	*Burn*	12
Loch	*River*	8
Burn	*Burn*	8
Loch	*Allt*	7
Loch	*Burn*	6
Grain	*Burn*	5
Lochan	*Allt*	4
Caochan	*Allt*	3
Loch	*Water*	3
Water	*River*	3
Lochan	*Caochan*	3
Stripe	*Water*	2
Lake	*Burn*	2
Grain	*Water*	2
Burn	*Water*	2
Loch	*River*	2
Stripe	*Burn*	2
Water	*Loch*	1

Upper element	Lower element	frequency
Lochan	*Fèith*	1
Lochan	*An* X (only)	1
Grain	*Allt*	1
Eas(an)	*Loch*	1
Other	Other	1
Allt	*Lochan*	1
Allt	*Allt*	1
Allt	*River*	1
Burn	*Loch*	1
Loch	*Abhainn*	1
Fèith	*Allt*	1
Allt	*Burn*	1
Water	*Burn*	1
Allt	*Loch*	1
Ailnaig	*Burn*	1
Other	*Burn*	1
Meur	*Allt*	1
Bog	*Burn*	1
Loch	*Pow*	1

This shows that the commonest combinations occur with a watercourse and a water-basin. Examples of these are numerous in Scotland such as Loch Dubh and Allt Dubh and are discussed above. The second commonest combination, of *burn* and *burn* seems contradictory but in fact represents pairing names where one or more of the two RNs have two specific elements, one of which is in common with the other RN, such as the pairing of West Burn of Builg and East Burn of Builg. Of the remaining names, a number of patterns can be seen:

- Combinations tend to contain only one stratum. One could say that combinations such as River Tay and Loch Tay are equivalent, since *loch* was originally a Gaelic term. As discussed above, this is largely a matter of interpretation. Apart from this, the only occurrences of a combination of a Gaelic and a Scots generic element in this situation are:

  Grains of Slochd Chaimbeil with Allt Slochd Chaimbeil As discussed the element *grain* operates somewhat differently from other names and may be equivalent in some way to *meur*. Indeed, if *grain* were interpreted as *meur* here, the relationship would seem more typical as in section 3.7 on page 76.

  Builg Burn with Ailnac Bhuilg Ailnac represents G *ailnig* which is a water-basin and as such can be interpreted as the *loch* and *burn* type.

  River Deveron with Allt Deveron Allt Deveron is the upper part of River Deveron. This is a genuinely odd construction. The naming of an upper part of a large watercourse in this way is idiosyncratic. It cannot be a modern name as the old form: *Auld-overane* (1662 Retours (Aberdeen) no. 363) shows.

- The group of elements which have a low *geogscore* are generally also present in the left hand column, such as *stripe*, *grain* and *caochan*.

To be added to this list are also occurrences of combinations where the two watercourses have a lowest point within 1 km of each other; that is, they both meet their parent river at the same point. Table 3.28 shows watercourses with this quality, where they have the same specific element and different generic element. Although there is not much information here the situation is essentially the same as in table 3.27 on page 120 once the water-basin names and names of the same type have been removed. The most common types are those with default terms *burn* and *allt* and other terms for small watercourses: *stripe*, *caochan*, *grain* and *lochan*.

3.12 Conclusion

3.12.1 Various Generics with a Similar Distribution

Figure 3.47 on page 123 shows a number of elements mentioned above that have a similar distribution. This graph is designed to show the similarities between the various distributions. The individual distributions can be seen in figure 3.35 on page 105 for *slack*, *den* and *stripe*; figure 3.25 on page 93 for *grain* and *meur* and figure 3.29 on page 98 for Y *of* X features. The altitude of these watercourses tends to be somewhat higher than other equivalent names. The average altitude is 369.3 whereas

Table 3.28: Hierarchy of Generic Elements for Watercourses which have the Same Lowest Point, with Different Generics and the Same Specific

element	element	frequency
Stripe	*Burn*	3
Allt	*Caochan*	3
Burn	*Grain*	2
Caochan	*Lochan*	2
Allt	*Loch*	1
Water	*Stripe*	1
Allt	*Lochan*	1
Bog	*Grains*	1
Den	*The* X	1
Other	*Allt*	1
Burn	*Water*	1
Burn	*Lake*	1

for Scots names in general it is 324.4. Since this distribution relates to research appearing in other chapters, the distribution is discussed in a wider context in the final conclusion.

3.12.2 Some General Points

- Watercourses attached to a generic have a relatively specific range of physical features.
- These 'ranges' of an element are not exclusive. With the exception of the very largest watercourses, (which can only be *rivers*) all other watercourses would be able to have a number of generic elements attached to them.
- The conditions for a generic to exist roughly fall into two categories: size (represented here as *km* and *nont*) and location (represented here as *alt* and *pos*). Of course, this is largely a matter of how the data have been presented here.

3.12.3 Generic Element Hierarchy

Graphs 3.26 on page 119 and figure 3.27 on page 120 show a generic element hierarchy. It is useful to see that the ordering of the generic elements by their *genelemscores* for each stratum shows a similar image to the one that has emerged above as in tables 3.29 on page 124, 3.29 on page 124 and 3.12.3 on page 124, These tables are implicit in much of the discussion above.

Figure 3.47: Distribution of Various Generic Elements

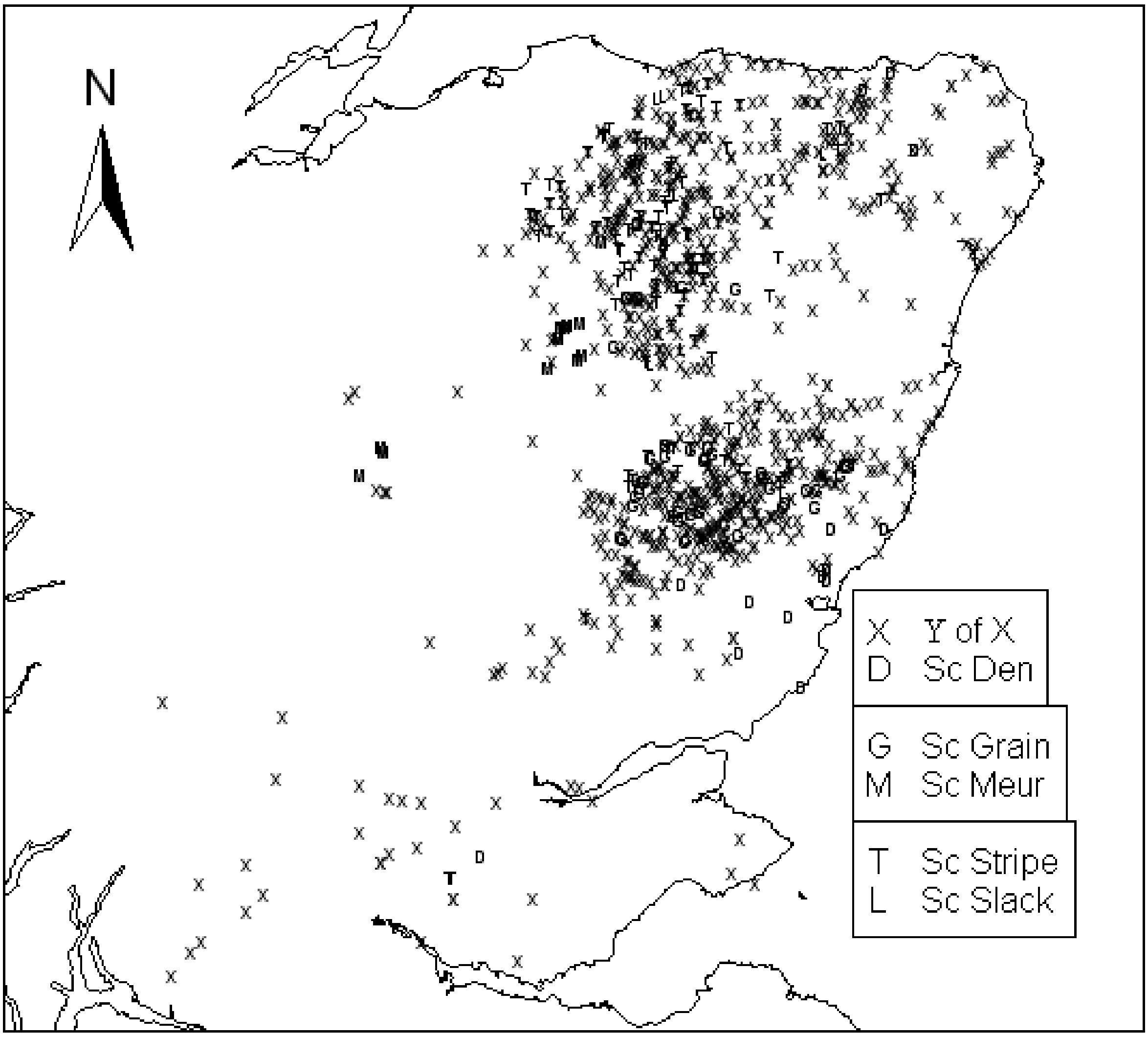


Table 3.29: *Genelemscore* for Scots Generic Elements

element	*genelemscore*
River + Specific	7.69
Water + Genitive + Specific	6.29
Specific + *Water*	6.10
Specific + *Pow*	5.68
Loch + Genitive + Specific	4.75
Specific + Den	4.67
Burn + Specific	4.52
Specific + *Burn*	4.52
Loch + Specific	4.48
Specific + Specific + *Burn*	4.33
The X	4.31
Burn + Genitive + Specific	4.28
Specific + *Burn* + Genitive + Specific	3.91
Specific + *Loch*	3.91
Specific + *Latch*	3.89
Specific + Other	3.82
Specific + *Slack*	3.46
Stripe + Genitive + Specific	3.44
Specific + *Stripe*	3.41
Specific + *Grain*	2.76

Table 3.30: *Genelemscore* for P-Celtic Generic Elements

P-Celtic Generic Elements	
element	*genelemscore*
Specific + *esk*	6.97
pol + Specific	6.79
Specific + *pol*	5.75
Specific + *dobhar*	5.51
Specific + *glas*	4.97

Table 3.31: *Genelemscore* for Gaelic Generics

element	*genelemscore*
Abhainn + Specific	5.12
Loch + Specific	3.97
Specific + *Allt*	3.70
Eas(an) + Specific	3.48
Specific + *Allt* + Specific	3.46
Allt + Specific	3.43
Allt + Genitive + Specific	3.43
Other	3.27
Other + Specific	3.10
Caochan + Genitive + Specific	309
Meur + Genitive + Specific	3.02
Fèith + Specific	2.99
Allt + Specific + Specific	2.97
Loch + Genitive + Specific	2.92
Alltan + Specific	2.89
Caochan + Specific	2.87
Specific + *Alltan*	2.74
Fèith + Genitive + Specific	2.70
Specific + *Loch*	2.69
Lochan + Specific	2.55
Specific + *Lochan*	2.49
Lochan + Genitive + Specific	2.35
Meur + Specific	2.32

Chapter 4

Semantics

4.1 Introduction

This chapter has a threefold purpose:

1. An attempt will be made to quantify the notion of 'semantic distance'. This is more fully explained below.
2. A brief survey of the various semantic classes will be made, modelled on the previous chapter, using the same analytical tools.
3. Other analytical tools, equivalent to those described in the previous chapter will be employed.

4.2 Some Conceptual and Methodological Considerations

This section tackles some conceptual issues which permeate much of this chapter. Following this is a brief explanation of the relationship between Nicolaisen's approach and the approach taken here.

4.2.1 The Taxonymic Approach

A taxonymic approach to semantic groups has been taken here, whereby a given term belongs to a number of different groups within larger sets. For instance, in general semantics, a 'poodle' belongs to a referent set of 'dog', which is part of a set called (or is a hyponym of) 'mammal', part of the set 'animal', part of the set 'living things', and so on. (Conversely 'living thing' is a hyperonym of 'animal'.) An example from our database might be Little Socach Burn, named after the place-name 'The Socach' from G *socach, soc*, 'point of land jutting between two rivers'. This RN is part of the semantic set called 'topography', which contains the set (or is a hyperonym of) 'specific natural feature', within this group is the group of names coined from a 'riparian area'.

Within the database, some shortcuts have been used, the value *semtype* represents the second level mentioned, and no specific field exists for the 'topography' type, however it can be easily inferred. The values *pntype* and *adjtype* further clarify names relating to other place-names, and adjectives respectively. The terms ‘class’, ‘ type’ and ‘secondary type’ are used here simply for clarity. The situation is set out in table 4.1, with another example, this time of the semanteme ‘black’:

Table 4.1: Semantic Hierarchy Examples

Level	Name	Example
Top level	Class	Adjective
Middle level	Type	Colour
Bottom level	Secondary Type	Dark

4.2.1.1 Semantemes Used

This section contains a brief description of the semantic taxonymy used, and how the semantemes can be further subdivided. See below for actual discussion of the semantemes within Scottish toponymy. Table 4.2 on page 127 shows how the various semantic classes practically combine to make the existing semantic categories. The ‘class’ column represents MySQL table B.10 on page 250. The ‘adjective’ column represents the *adjtype* MySQL table B.11 and the ‘PN type’ column represents the *pntype* MYSQL table B.12 on page 251. Both *adjtype* and *pntype* denote hyponyms of *semtype*.

Every RN in the database for which the meaning is known falls into one of these classes. The classes are largely self-explanatory but the following points should be made:

- The classes, ‘Concavity’,‘Convexity’, ‘Land Around’ and ‘Water feature’ relate to non-specific semantemes. For example, Glen Burn is an example of ‘Concavity’, but Glenfender Burn denotes a specific place, so is ‘Human: Specific man-made area: Concavity’.
- The term ‘adjective’ represents not just adjectives in the linguistic sense, but any specific element that describes the watercourse. This means for instance that a name such as Burn of Dararach, from G *dararach*, ‘rattling sound’ is categorised as ‘Sound: Loud’ despite the fact that *dararach* is a noun.
- Specific man-made area represents settlements regardless of the original derivation of the settlement name. Thus table 4.26 on page 162 contains elements such as *ruighe* and *inbhir* in this list may seem to represent natural features, but in fact they represent settlements such as Rintarsin Burn (from G **Ruighe an Tarsainn*) or Invergeldie Burn (from G **Inbhir G(h)eallaidh*) .

4.2.1.2 Relationship between Nicolaisen’s Table and the Table Used in this Thesis

The semantic classification used here is based on Nicolaisen’s mentioned in the introduction.[1] As can be seen from table 4.3 on page 129, in the main the new table has followed Nicolaisen’s, but with several divergences in details. The new classification has expanded on Nicolaisen’s categories or has interpreted the available data somewhat differently.

[1] Nicolaisen, ‘The Semantic Structure of Scottish Hydronymy’, p. 211-240.

Table 4.2: *Combinations of Semantic Classes*

Class	Type	Adjective type	PN type
Adjective	Colour	Light	
Adjective	Colour	Dark	
Adjective	Smell / Taste / Feel		
Adjective	Manner	Gentle	
Adjective	Manner	Rough	
Adjective	Sound	Quiet	
Adjective	Sound	Loud	
Adjective	Temperature	Cold	
Adjective	Temperature	Hot	
Adjective	Other		
Adjective	Course	Crooked	
Adjective	Course	Straight	
Adjective	Effect / Character		
Adjective	Dimensions	Big	
Adjective	Dimensions	Small	
Adjective	Bed		
Adjective	Number		
Adjective	Moistness	Dry	
Adjective	Moistness	Wet	
Adjective	Age	Young	
Adjective	Age	Old	
Adjective	Visibility	Exposed	
Adjective	Visibility	Hidden	
Adjective	Elevation		
Ecosystem	Flora		
Ecosystem	Fauna		
Topography	Concavity e.g. Glen		
Topography	Convexity e.g. Ben		
Topography	Land around		
Topography	Water feature		
Topography	Specific natural feature		Concavity
Topography	Specific natural feature		Convexity
Topography	Specific natural feature		Body of Water
Topography	Specific natural feature		Land not used for Agriculture
Topography	Specific natural feature		Land used for Agriculture
Topography	Specific natural feature		Riparian area
Topography	Specific natural feature		Other
Topography	Material / Object		
Topography	Weather / Air		
Human	Non specific agricultural area		
Human	Non specific settlement / building		
Human	Specific man-made area		Concavity
Human	Specific man-made area		Convexity
Human	Specific man-made area		Body of Water
Human	Specific man-made area		Land not used for Agriculture
Human	Specific man-made area		Land used for Agriculture
Human	Specific man-made area		Riparian area
Human	Specific man-made area		Other
Human	Agricultural object / structure		
Human	Specific person / occupation		
Human	Supernatural entity		
Human	Event		
Situation	Relation to other features		
Situation	Boundary		
Situation	Crossing		
Other	Other		

The adjectival semantic classes, i.e. those describing the quality of the water itself, are basically the same with the following exceptions:

1. As explained above, in section 4.2.1 the adjectives have been divided roughly into opposites. Where Nicolaisen has ‘colour’, in this system there are also markers for ‘dark’ and ‘light’.

2. Nicolaisen’s ‘form of the bed of the stream’ (B b.), has been reanalysed. Some of his names, such as Broad Burn, are now classified under ‘dimensions’, and others such as Cam Alltan (derived from G *cam*, ‘crooked’), which relate to the course of the stream are classified under ‘course’, while yet others which relate to valleys or glens are classified under ‘concavity’.

3. ‘Number’ has been added, this is a very rare class, containing names such as Three Burns.

4. ‘The geological nature of the bed of the stream’ (B. d): Names in this section such as Allt Garbh have been moved to ‘Manner’. The remaining names have been divided between ‘bed’ if the geological feature is underwater, and ‘land around’ if above water.

5. Trees and bushes have been put into one category: ‘flora’.

6. ‘Water courses connected with names of hills, valleys and human settlements’ are split between concavity, convexity, specific natural feature and human settlements. This distinction, between a specific natural feature and a generic one, is discussed below.

7. ‘The situation of the watercourse’ has been divided into three categories: ‘relation to other features’, ‘boundary,’ and ‘crossing’.

4.2.2 The Notion of Semantic Distance

The concept of ‘distance’ represents an order of how conceptually far an element is from the original meaning of ‘water’. So while, ‘water-word’ means simply any word just meaning ‘water’, ‘adjective’ denotes a hydronym which relates to the qualities of that water somehow. ‘Ecosystem’ is further removed by denoting flora or fauna which dwell in, on or around the river, and ‘topography’ denotes hydronyms related to features entirely within the area of influence of the watercourse the names represent. Still further removed from this are ‘human’ and ‘situation’ which say nothing about the watercourse beyond their use or relevance to man.

An assumption is often made, though rarely stated, that the larger the watercourse the stronger the tendency for the meaning of the name to be conceptually closer to the water. This means that, for instance, whilst the original meaning of the name of the River Tay is unknown, one would expect it to relate to an adjective of some sort, describing perhaps the manner in which it flows. Conversely, one would not expect the Tay to be named after a settlement. (It has been claimed that the second element in Dundee derives from the same root as Tay, but never *vice versa*.) In the terminology of this thesis, if the above phenomenon is correct, one would expect to see some correlation between the average

Table 4.3: Comparison of Nicolaisen's Semantic Categories and Those Used Here

Class used here	Nicolaisen's ref	Nicolaisen's Class
Water word	F	'Water-words'
Colour	A a.	The colour of the water
Smell / Taste	A b.	The taste or smell of the water
Manner	B c.	The speed and movement of the flowing water
See item 4	B d.	The geological nature of the bed of the stream
Sound	A d.	The noise of the water
Temperature	A c.	The temperature of the water
Other		no equivalent
Course	B b.	The form of the bed of the stream
Effect / Character	A e.	The effect of the water
Dimensions	B a.	The size and length of the stream
Bed	B b	The form of the bed of the stream
See item 4	B d.	The geological nature of the bed of the stream
Number		no equivalent
Flora	C b.	Tree vegetation associated with the watercourse
Flora	C c.	Plants, other than trees, associated with the water-course
Fauna	C d.	Animals, birds, fishes etc. associated with the stream
Concavity e.g. Glen	B b.	The form of the bed of the stream
Specific natural feature: Concavity	E c.	Water-courses named from the names of valleys
Convexity e.g. Ben	E b.	Water-courses named from the names of hills
Land around	B d.	The geological nature of the bed of the stream
Land around	C a.	The terrain through which the stream flows
Water feature	E d.	Water-courses named from the names of lakes
Specific natural feature: Water feature	E d.	Water-courses named from the names of lakes
Other		no equivalent
Use to man, agriculture		no equivalent
Specific person / occupation	D b.	Water-courses associated with human beings
Specific man-made area	E a.	Water-courses named from the names of human settlements
No equivalent	E e.	Water-courses connected with primary RNs
Agricultural object / structure	D a.	Water-courses associated with human institutions / human beings
Non specific settlement / building	D a.	Water-courses associated with human institutions / human beings
Supernatural entity	D b.	Water-courses associated with human beings
Event		no equivalent
Relation to other features	C e.	The situation of the watercourse
Boundary	C e.	The situation of the watercourse
Crossing	C e.	The situation of the watercourse
Other		no equivalent

geogscore (or components of *geogscore*) and the class of *semtype* (i.e. adjective, topography, etc.). This section will explore whether this supposition is correct, and whether it can be quantified.

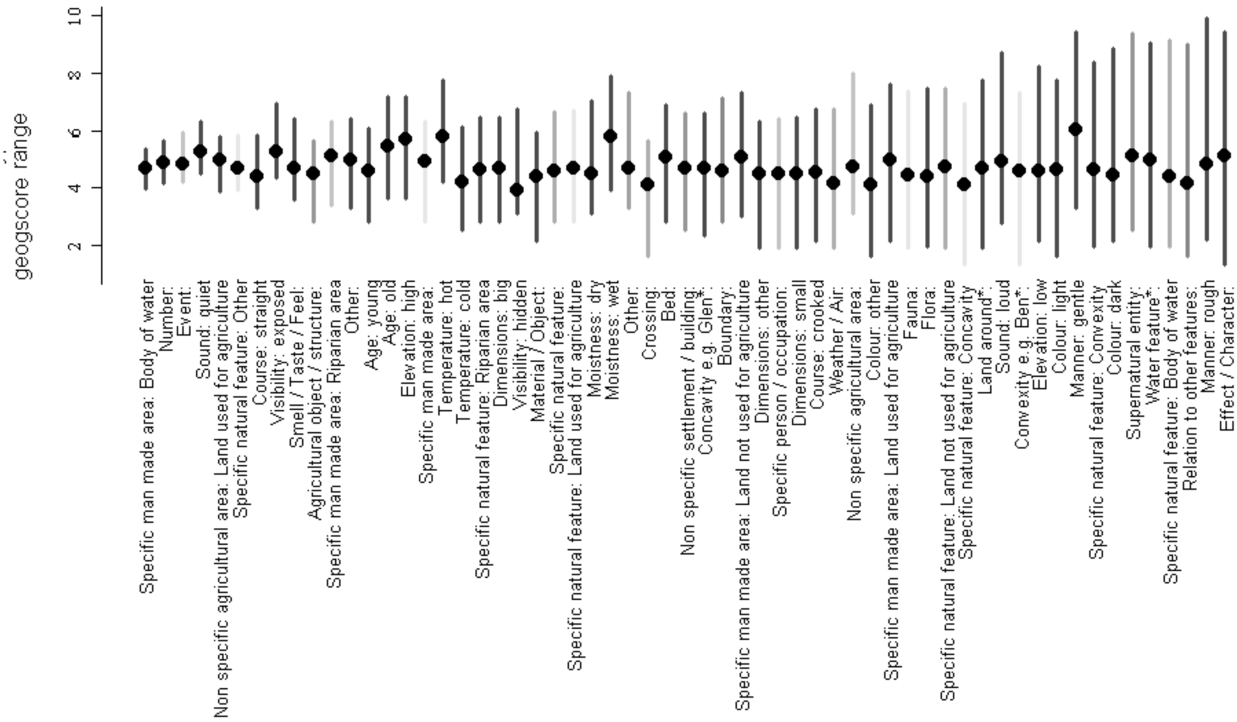
Using the taxonymic approach above, and taking a specific G *sean*, 'old' as an example, the question arises where on the following scale it is most meaningful to regard the element:

1. Element which is an adjective.
2. Element which denotes age in general.
3. Element which specifically denotes old age.

The answer is probably that for each semantic group, the answer varies. It is clear that some groups of meanings are fairly equivalent across linguistic strata and specific elements within a stratum (e.g. see terms for 'boundary' below) whilst some nominally have the same meaning but seem to be used in very different ways (e.g. colour terms). With these caveats in mind, some tentative attempts are now made to discern a hierarchy of conceptual distance.

The distance is calculated by retrieving the average *geogscore* across all RNs which possess a given *semtype* as in B.15 on page 256 in the tables section. This table however only shows the average scores, whereas the maximum and minimum score are also of interest, shown in figure 4.1. This is ordered by smallest range to largest range, whilst 4.2 is the same data ordered by average score.

Figure 4.1: *Geogscore* Range Ordered by Difference between Maximum and Minimum



As a general pattern, it can be seen in figure 4.1 that names with a larger range, to the right of the graph, represent semantemes that occur with all sizes of watercourse (e.g. 'effect / character', 'rough manner', 'dark colour'). These names occur across all the linguistic strata as discussed below. Conversely, names at the left end represent more restricted semantemes such as 'event', 'number' etc. Many of the semantemes in this area are relatively rare. Comparing this to figure 4.2, ordered by average, it can be seen that the terms with the highest *geogscore* are all adjectives, in descending order: gentle, wet, hot, old, exposed and quiet. This group of semantemes will be discussed in the conclusion. The differences in the averages are of course very small, but this is a common phenomenon when comparing averages.

Figure 4.3 is a different visualisation of the same data. Each dot represents a RN; for each *semtype*, the *geogscore* has been plotted. A jitter has been added to the graph to show where a dot represents one entry, or several identical ones. Where the previous graphs showed only the maximum and minimum ranges, these graphs will show if the maximum or minimum is an outlier. Two outliers in the 'Supernatural entity' column can be seen. These are of course the Dee and Don.

If ordered by the maximum amount in each *semtype* one sees a more expected range. This is because the smaller the watercourse, the more variation there is with the *semtype*, i.e. a watercourse of any size can be named after its colour or manner, but larger watercourses have a more restricted range

Figure 4.2: *Geogscore* Range Ordered by Average

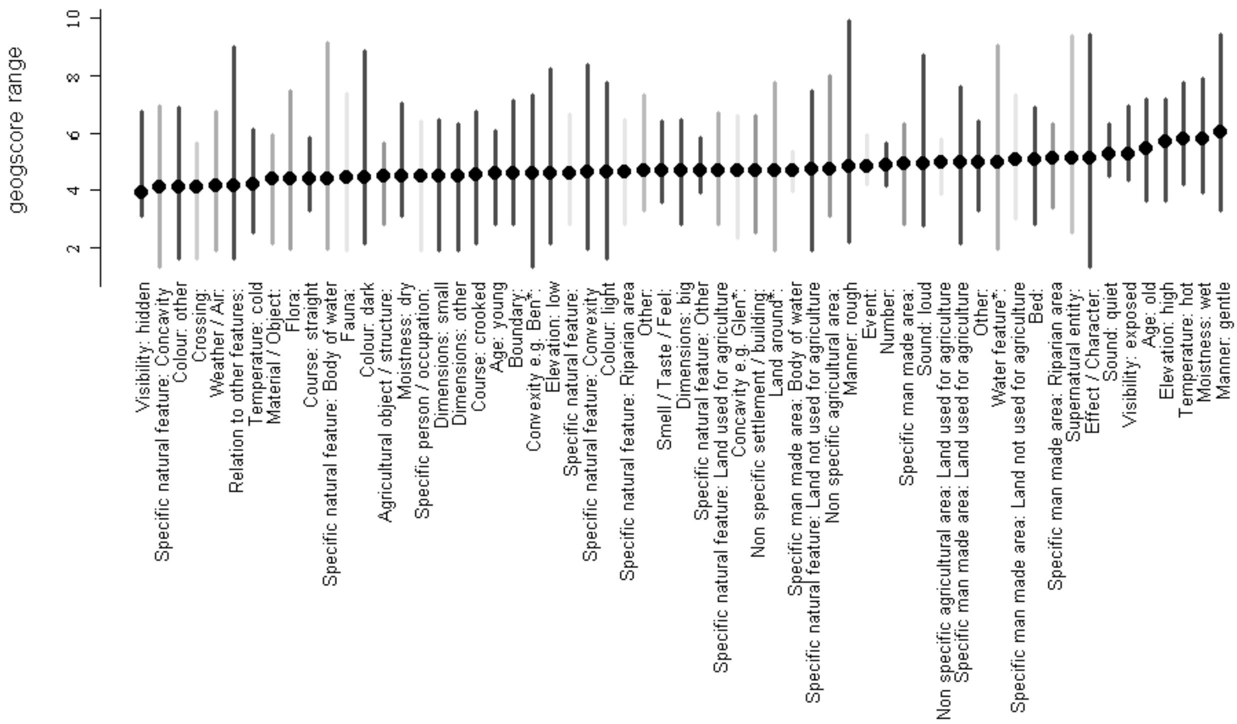
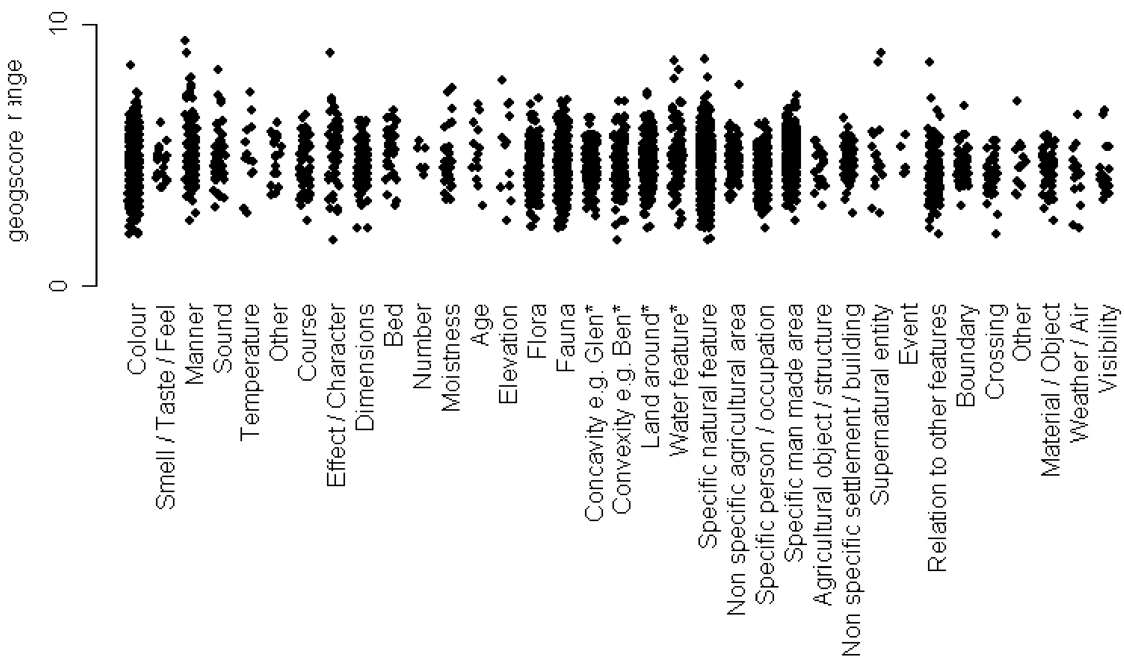


Figure 4.3: *Semtypescore* with Dots



of types of lexical items from which they draw their names.

Concerning the notion of semantic distance and linguistic stratum, table 4.4 broadly corroborates what was already known: that hydronyms from older linguistic strata tend to describe the qualities of the watercourse itself, and later names tend to describe less ‘basic’ things.

Table 4.4: Number of Hydronyms Belonging to Each Semantic Group by Stratum

	Scots	Gaelic	P-Celtic / Gaelic	P-Celtic	OC / P	OC	total
Adjective	273	622	50	60	11	8	1024
Human	617	699	2	0	0	2	1320
Ecosystem	145	467	6	4	0	2	624
Other	2	11	1	0	0	0	14
Situation	147	142	6	0	0	0	295
Topography	600	2263	22	10	7	0	2902

Of the Old / Early Celtic names, nearly all relate to water-words or adjectives, the two entries under ‘human’ denote the Dee and the Don under names relating to supernatural.

4.2.3 Semantic Defaults

Many semantemes which exist in binary pairs have a default value. That is, watercourses are *generally* assumed to have a specific value unless otherwise described. An obvious example of this would be the fact that there are thirty-one instances of the semanteme ‘dry’ and only five examples of ‘wet’ or ‘damp’. This is obviously because all watercourses are considered to be wet. The ones that dry up are a small subset of all the watercourses, and as such are named for this special quality, giving names such as Dry Burn. This is related to the semantic notion of *markedness*. An example of this is the word *lion*, an unmarked term in English. This can be used for a lion of any gender, but *lioness* specifically denotes a female lion and as such is marked. For the following adjectival semantemes, the default and non-default value tend to differ in some or all of the following ways:

1. The non-default names tend to be greater in number.

2. The actual terms that make up the non-default names tend to be more consistent. For example, in the ‘dry / wet’ examples above, twenty of the ‘dry’ names all derive basically from the Sc term *dry* with the remaining eleven coming from five other roots. For ‘wet’ however the four terms derive from three separate roots, thus there is no *de facto* term for ‘wet’ in Scottish hydronymy.

3. Probably because of the greater frequency of non-default terms, the components of *geogscores* for non-default terms tend to be more spread out and overall greater, see figure 4.9 on page 144 for example.

4.3 Semantic Categories

This survey follows the same structure as that in the previous chapter, using the same analytical tools. For each of the adjectival terms, the default and non-default terms are sought for.

4.3.1 Colour

Colour[2] is one of the furthest ranging semantic types in that it seems to be productive in all strata, and at all times. The earliest examples seem to be Findhorn and Deveron, with names such as Black Burn being formed up until a few hundred years ago.

It appears that there are two ways in which colour is used within Celtic hydronymy. The first is literally to describe the water, i.e. G *dubh*, 'black' is utilised because the water is murky or muddy. The second is to use terms such as black and white as a binary pair, in order to distinguish the two. See section 4.3.25 on page 170 for the discussion about pairs.

These two different usages occur at all levels of Celtic hydronymy, and more weakly within Scots hydronymy. Since smaller watercourses are likely to flow less quickly and are more prone to becoming murky, it is unsurprising that amongst smaller watercourses terms for 'black' and 'dark' are much more common than terms for 'white' and 'clear'; in fact for watercourses between 0 and 5 km long 121 have an etymology of 'black' whilst only 24 have one for 'white'.

Figure 4.4 compares the qualities of RNs with colour terms, as opposed to all the RNs, ignoring, for the time being, linguistic stratum. The similarity between the two patterns is very strong; the two sets are virtually identical, and even the average points are very close. The chief deviation is for names with a *kmscore* of over 6.5. Essentially this means that any watercourse can have a colour term for a name, except the very longest rivers. It would be appropriate then to investigate this relationship. The ten longest watercourses with colour terms as their name are listed in table 4.5.

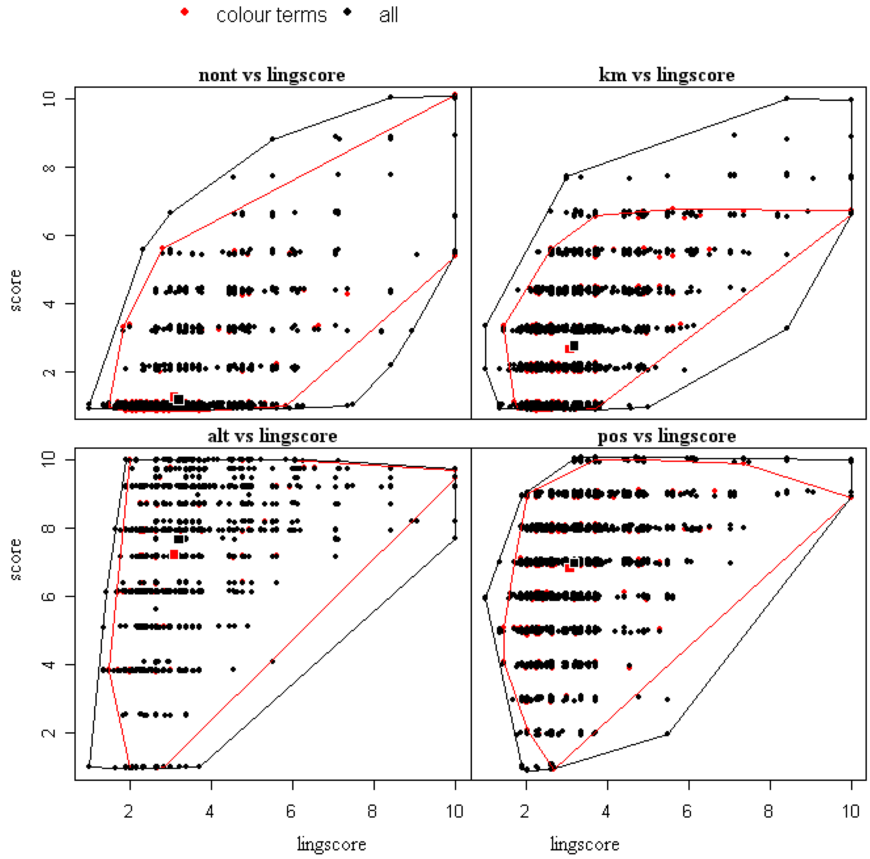
Table 4.5: Ten Longest Watercourses with Colour Terms for their Names

name	length in km
River Tummel	28.5
Loch Tummel	26.5
River Lochay	26.5
River Allan	25
Black Water	24
Lochter Burn	24
Lunan Burn	20.5
Water of Dye	20
Lunan Water	19
Grayburn	18

River and Loch Tummel, meaning dark or gloomy, probably relate to the physical aspects of the

[2] This category corresponds to Stewart's category B: 'naming by color' under his 'descriptive names'. See: Stewart, *Names on the Globe*, p. 91

Figure 4.4: Comparison of Colour Terms and All Terms



watercourse. The -l- ending is rare in Celtic hydronymy and as it is usually used as a locative suffix, it could mean 'dark place' rather than 'dark'.[3]

Since the Celtic colour classification includes not just spectral qualities but factors such as reflectivity and so on, words which are often translated as 'shining' have been included within this section. The derivation of the River Allan is uncertain, but if it does mean 'shining', it is the name with the highest score which relates to colour or reflexivity. It is possible that the Allan and the Devon are a binary pair, although it is equally possible the Devon relates to **domnona*, 'deep'.[4] It is possible the name could be a pair with the Gaelic word for the Forth, Abhainn Dubh as discussed above.

It is of note that of these ten, five represent pre-Gaelic terms for reflexivity. The Allan was mentioned above, and River Lochay and Lochter Burn mean 'black / shining'[5] and Lunan Burn and Water mean 'shining like the moon'.[6] This could mean either that names concerning reflexivity should be treated differently from general colour terms, or it could mean that these reflexivity terms were once

[3]See Padel, *Cornish Place-Name Elements*, p. 138 for a discussion of the -l suffix. The suffix can also be seen in names such as Oykel from OC **uxell-*, 'high'.

[4]John Wilkinson, 'Deep Thoughts on the Devon and a Fresh Look at the Nith', *Nomina* 25 (2002), p. 139-143.

[5]King, "Lochy' Names and Adomnan's Nigra Dea'.

[6]Jacob King, 'Endrick and Lunan', *The Journal of Scottish Name Studies* 1 (2007).

much more widespread and only the names associated with the longest watercourses have survived.

This still does not really explain why the longest watercourses are prohibited from having colour terms for their names. This is all the more baffling since nearly all the longest watercourses have an adjectival term of some sort as their name. It could be that the watercourses were considered as being so long, they could not have the same colour along its whole course, because of the changes in geographical conditions between its source and confluence (this would concur with the discussion of Abhainn Dubh above).

Table 4.6 shows, within a Gaelic context, the relative amounts of colour terms. A number of points stand out:

Table 4.6: Frequency of Gaelic Colour Terms

term	frequency
G *dubh*, 'black'	89
G *ruadh*, 'red'	22
G *glas*, 'grey-green'	21
G *geal*, 'white'	20
G *buidhe*, 'yellow'	17
G *dearg*, 'red'	17
G *odhar*, 'dun'	16
G *uaine*, 'green'	14
G *riabhach*, 'brown'	11
G *bàn*, 'white'	8
G *gorm*, 'green'	5
G *breac*, 'speckled'	3
G *fionn*, 'white'	3
G *ciar*, 'dark, gloomy'	2
G *donn*, 'brown'	2
G *grianach*, 'sunny, shining'	2

1. As stated above, *dubh* is by far the commonest term.

2. Whilst terms meaning 'black' occur with greater frequency, there are fewer terms for them. Disregarding forms and derivatives of *dubh* the only other term is G *ciar*, yet for terms meaning white or shining there are several, G *airgiod*, 'silver, G *bàn*, 'white', G *buidhe*, 'yellow', G *can* 'white', G *fionn* 'white', G *geal*, 'yellow', G *grianach*, 'sunny, G *leusach* 'blazing, G *soilleir*, 'bright'.

3. The term *fionn* is actually relatively rare within a Gaelic context in the AOS, and there is no single term which is the opposite of *dubh*. This fits in with the notion of defaults discussed above.

The reason for the lack of parity in the distribution of the various colour semantemes is probably due to two factors. Firstly, as mentioned above, the physical environment simply contains more smaller streams with dark muddy water in them. One of the main ways in ancient times for a watercourse to have been considered ‘white’ must have been the fact that it reflected the sun. For it to do this, the watercourse must be large enough to have a relatively even surface. It is a geographical fact that larger watercourses are rarer than smaller ones. Another way in which a watercourse could be considered ‘white’ is by its generation of white foam.[7]

The second phenomenon at work here is that once the above situation exists, watercourses meaning ‘black’ become more ‘fashionable’, and thus extend themselves over watercourses that do not have black qualities. This two-step process is a common process in nomenclature. It is thus clear that in terms of a default, ‘white’ is the default and ‘black’ the non-default term under the parameters listed above.

4.3.2 Age

Semantemes for ‘age’ exist throughout Scotland and countries with Celtic hydronymy and seem to go back to the earliest stages, with old names such as Abhainn Sin in Scotland and the Shannon in Ireland cognate with G *sean*, ‘old’, also, Holder cites Sēna as a river in Umbria.[8] Names meaning ‘young’ are somewhat rarer in Celtic hydronymy, and are maybe even absent in this earliest stratum.

In the AOS, a small number of names seem to be named for their apparent ‘age’, each with a semanteme meaning ‘old’ or ‘young’. In the case of the Celtic names meaning ‘new’, the root they all derive from is OC **now(i)jo-*, which may have a sense of ‘fresh’ in P-Celtic hydronymy which would be an appropriate quality.[9] The name Abernyte suggests the term was known in P-Celtic times, and contrasts well with other Aber- names such as Aberlady, which means ‘rotting, stagnant’. In later strata, the G *nuadh* and Sc *new* are used, which, whilst having a possible meaning of ‘fresh’ both have the productive semanteme of ‘new’ in the sense of ‘recent’.

Alongside this are the terms meaning ‘old’, predominantly from G *sean*, ‘old’ and in one case Auld Water. In the case of Auld Water, Alexander[10] says of it: ‘The old channel of the Mossat burn, before its diversion.’ The definition, then, seems to be a watercourse whose upper reaches have been diverted. For the remainder of the names, however, it is not intuitive in these cases to imagine how ‘old’ and ‘young’ could be applied to watercourses. It is of course possible the ‘new’ names in a handful of cases may be equivalent to name such as Newbiggings, i.e. in reference to a newly created features. In other cases the names could represent newly used burns, which previously were located in unused areas of land, whilst ‘old’ burns were the ones which were previously used for some function.

If these uses were the case most of the time however, one would perhaps expect a fair degree of name pairing. In fact there is only one known surviving name-pair in the AOS, that is Allt Nuadh (4967) and Sean Allt (5536). Conceptually, one might imagine the ‘old’ watercourse as being the

[7]Compare coastal Onich from G *omhanach*, ‘place of foam’.

[8]Holder, *Alt-Celtischer Sprachschatz* (Leipzig, 1904), pp. 1464 vol. 2.

[9]For further discussion about this element, see: Wilkinson, ‘Deep Thoughts on the Devon and a Fresh Look at the Nith’, p. 143-145

[10]Alexander, *The Place-names of Aberdeenshire*, p. 157.

main watercourse, with the new one, a ‘branch’ or ‘offshoot’; but the opposite occurs here: the new watercourse is the parent watercourse and the old watercourse the tributary. Of course, this is only one instance of this type of structure, and one could put forward a number of scenarios as to why this is so, ranging from OS errors, to positing G *sean* meaning something like ‘senior, distinguished, set apart, remote’.[11]

O. J. Padel[12] makes an identification between the semantemes for ‘old’ and ‘winter’ in the Cornish element *hendre*, consisting of *hen*, ‘old’ and *tre*, ‘homestead’. It is possible that *sean* in these cases has a similar meaning to the names meaning ‘winter’. A comparison of the components of *geogscore* is inconclusive, and evidence for such a connection is tenuous, mainly because the names do not refer to hydronyms but settlements. Another possible interpretation would be a metaphorical extension of ‘old’ >‘slow, sluggish’ and ‘young’ >‘lively, bubbly’.[13] This would suggest that all or some of these names could perhaps be placed in the ‘manner’ category.

It would of course be inadvisable to attempt to find an overarching set of meanings for these types of names. They exist across a number of linguistic strata, and it may be that the names are not supposed to be contrasted at all. This might well be the case for Early Gaelic and P-Celtic names in the area, where the OC **now(i)jo-* names may mean ‘fresh’ and the **sen-* names may mean ‘senior’ or ‘remote’; moreover, see section 4.3.10 on page 147 where the possibility is discussed that G *mòr* is in opposition to G *òg*, ‘young’. Put like this, the names do not seem like opposites at all.

Whilst particular terms for ‘old’ and ‘new’ do not suggest a default, figure 4.5 on page 138 suggests that ‘young’ is the default and ‘old’ the non-default because the ‘young’ semanteme is rare, and the polygon that represents these names is smaller in the *nontscore* and *kmscore* graphs. This is reinforced by two other points. Firstly in the discussion of colour terms on page 171 it is shown that in earlier strata it might be possible for a naming pair to have only the non-default term marked, with the default term not present. It is possible this structure exists for ‘age’ in the name pair mentioned in Ravenna which are Leuca and Leugosena[14] other interpretations are possible, but this could reflect the default RN **leucā*, ‘shining one’ and the non-default RN with the ‘age’ term *senā*.

4.3.3 Smell / Taste / Feel

Hydronyms are predominantly coined by their reference to appearance or sound, but watercourses coined from the other senses also exist. It appears that these names do not have any particular defaults. This may be because a number of different small classes are involved.[15] It is possible to discern two semantemes, ‘clean’ (e.g. Burn of Glansie perhaps from G *glan*, ‘clean’) and ‘dirty’ (such as Clattie Burn from Sc *clatty*, ‘muddy, dirty’ or Allt Salach from G *salach*, ‘dirty’), but neither of these appears to possess the qualities of a default. In some cases, the names such as Nethy Burn from PG **neiktodia*

[11] Outside the AOS, the Old Water of Cluden (NX889797) or Old Water as it is on modern OS maps, is a tributary of the Cluden in Kirkcudbrightshire, fitting with this structure.

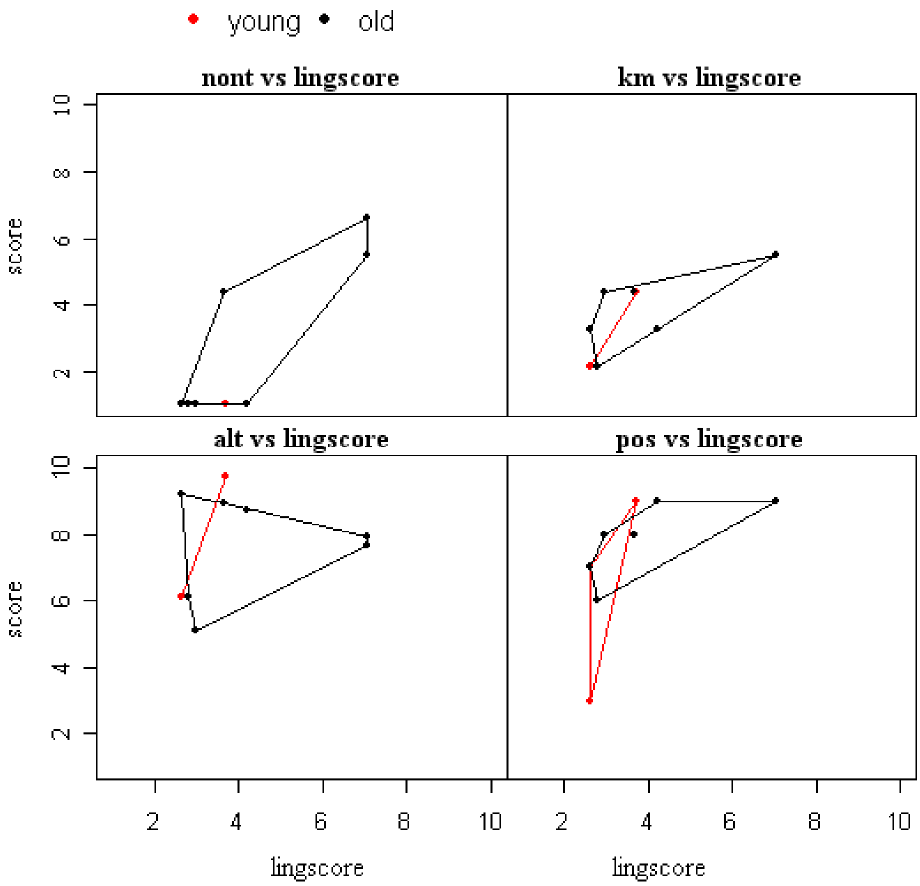
[12] Padel, *Cornish Place-Name Elements*, p. 129.

[13] Thanks to Professor Roibeart Ó Maolalaigh for this suggestion in personal communication.

[14] Rivet and Smith, *The Place-Names of Roman Britain*, p. 388.

[15] See: Stewart, *Names on the Globe*, p. 91 who first noticed the rarity of names of this type.

Figure 4.5: Comparison of 'young' and 'old'



OI *necht*, 'cleansing' may be better attributed to the effect / manner class rather than this one.

4.3.4 Manner

'Manner' refers specifically to the *manner* in which the water itself flows. It is divided into two groups, those which stress the animated, moving aspects, and those which stress the apathetic, still aspects of the water. The majority of the RNs in this group belong to the first category, with ninety-seven representing rough water, fifteen for still water and four for neither, as in table 4.7.

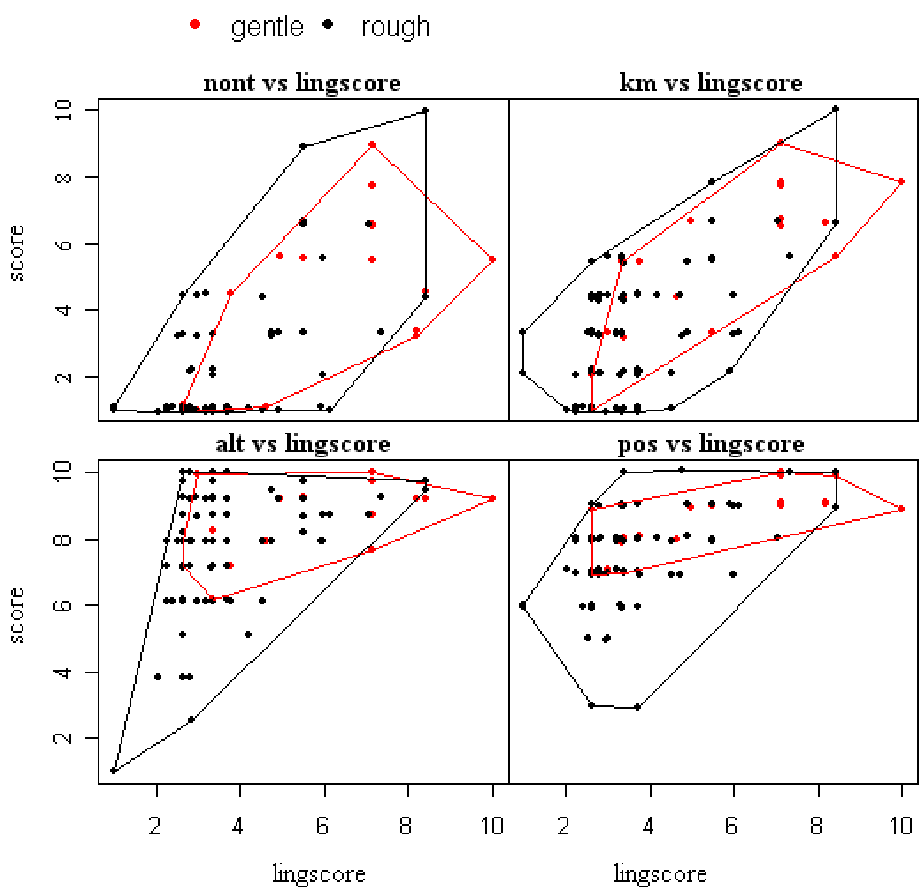
As can be seen, the commonest element is G *garbh*, 'rough', with the second being Sc *rough*. The AOS is basically an overlap zone between these two elements. Sc *rough* exists mainly in the Borders and somewhat North of the Forth, whilst *garbh* has a typical distribution for a G element. The *altave* is 419 for 'rough' water, and 326 for 'still' water, showing that 'rough' water flows through higher, more mountainous terrain and 'still' water is in flat land, as one would expect.

Concerning a default, figure 4.6 does not show a conclusive result. The near ubiquity of the terms *garbh* and *rough* however, suggest that these are the non-default terms, with the semanteme 'gentle' being the default.

Table 4.7: Commonest Derivations for 'rough' Water

etymology	frequency
G *garbh*, 'rough'	32
Sc *rough*, 'rough'	11
P **caleto*, 'hard'	8
G *fearg*, 'anger, rage'	5
G **garbhaid*, 'rough'	4
G *callaidh*, 'active quick'	3
PG *teine*, gen. *tened*, 'fire / firey one'	3
G *ceatharnach*, 'warriorlike'	2
G *criosda*, 'swift, active'	2
G *fiaclach*, 'jagged'	2

Figure 4.6: Comparison of 'gentle' and 'rough'



4.3.5 Sound

Watercourses named for ‘sound’ belong predominantly in two groups: those whose water is perceived as generating sound,[16] and those watercourses around which other noise occurs, usually provided by animals. In most cases it is very hard to discern between these two types. Since sound is transient, the method used here is not particularly suited to unravelling the nuances of the types of sound.

That said, certain correlations can be discerned when investigating soil classification or altitude. If one proposes the explanation that some watercourses whose names denote a degree of loudness and are thus named because they flow through rocky ground at a higher altitude, then it is no surprise the average altitude for ‘loud water’ is 415, whereas for ‘quiet water’ it is 350, much the same as for ‘rough’ and ‘still’ water above.

Concerning the existence of a default, figure 4.7 shows that this must surely be ‘quiet’, with the quality of loudness providing the cause for naming. In this figure it can be seen that there are only two names in the category of ‘quiet’ whilst table 4.8 shows four names. This is because the names *Bothrie[17] and *Haggernie are reconstructed names, and as such do not have generic elements, with the result that they do not have a *genelemscore*, and thus no *lingscore* and thus cannot be plotted.

Table 4.8: Hydronyms Meaning ‘quiet’

example	derivation	frequency
River Balvag	G *balbh*, ‘mute, silent’	1
Bothrie	PG *bodhar*, ‘deaf, quiet’	1
Haggernie	G *cagair*, ‘whisper’	1
The Shevock	G *sèamhag*, ‘little quiet one’	1

Table 4.9: Hydronyms Meaning ‘like a human voice’

example	etymology	frequency
Loch Ullachie	G *iolachdach*, ‘shouting’	2
Allt Eigheach	G *èigheachd*, ‘shouting’	2
Allt Bheadhair	G *beur*, ‘shrill, sonorous’[18]	1
Brerachan Water	G *briathrach*, ‘wordy’	1
Calair Burn	P **calar-*, ‘calling one’	1
Shouting Loch	Sc *shouting*, ‘shouting’	1
Skeugh Burn	G *sgiamhach*, ‘shrieking’ or Sc *skeugh*, ‘skewed’	1
River Ythan	OC **iektona*, ‘talkative one’	1

[16]See: Stewart, *Names on the Globe*, p. 91 who mentions that these types of name are rare except in hydronymy.

[17]This element has been said to derive from G *bodhar*, ‘deaf’ or from a P-Celtic cognate of the Welsh, *byddar*, ‘deaf’. The extended sense in both these linguistic strata is that it is quiet, that is, other people are deaf to it. An equivalent of this is G *caochan* ‘little blind one’, not meaning that the watercourse is blind, but that others are blind to it, i.e. it is hidden. See Watson, *The Celtic Place-Names of Scotland*, p. 435.

[18]*Allt Bheur* (1874 OS 6 inch 1st edn.); *Allt Bheur* (1912 Bartholomew: Survey Atlas of Scotland)

Figure 4.7: Comparison of ‘quiet’ and ‘loud’

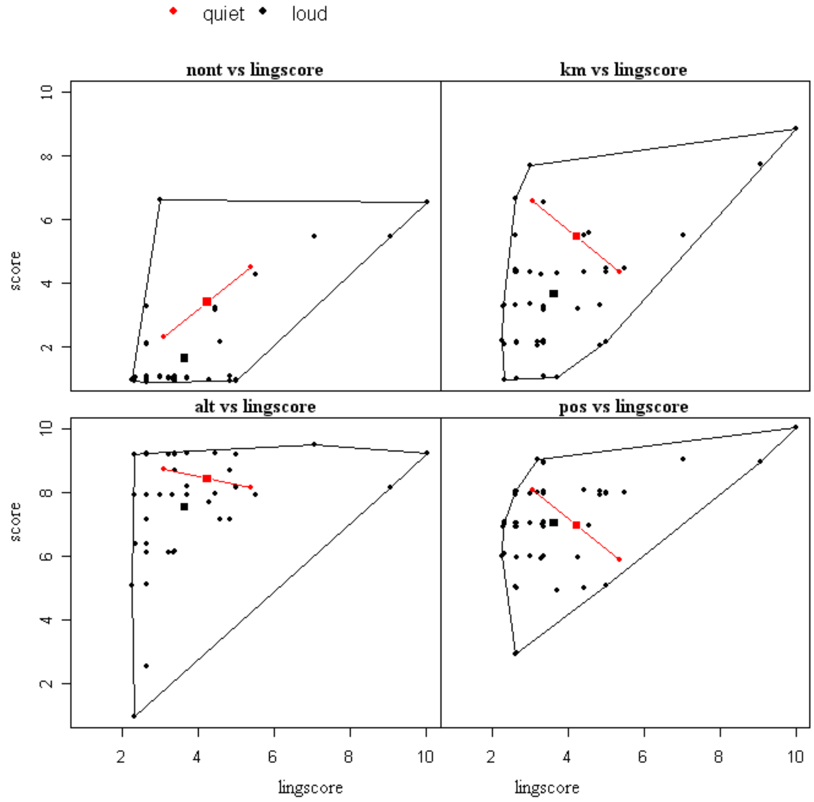


Table 4.10: Hydronyms Meaning ‘sounding like an animal or bird’

example	etymology	frequency
Gairney Burn	G *goirneag*, ‘little crier’, G *goir*	8
Allt Luineag	G *luinneag*, ‘song’	2
Allt a’ Bhuirich	G *bùrach*, ‘wailing’	2
Eas a’ Chaoinidh <	G *caoineadh*, ‘howling, wailing’	1
Cloak Burn	Sc *cloack*, clock, ‘clucking noise’	1
Caochan Rànaich	G *rànaich*, ‘roaring’	1
Allt Shallainn	G *sallan*, ‘singing’	1
Allt Sgairnich	G *sgàirneach*, ‘howling’	1

Table 4.11: Hydronyms Meaning ‘constant sound’

example	etymology	frequency
River Braan	PG **bremava, breamhainn*, ‘hum, buzz’	1
Altheyrenach	G *tàirneach*, ‘thundering’	1
Roar Burn	Sc *roar, roarie*, ‘roaring, loud’	1
Rumbling Burn	Sc *rumbling*, ‘rumbling’	1
Water of Tanar	OC **tanar-*, ‘thundering one’	1

Table 4.12: Hydronyms Meaning Miscellaneous or Generic Sounds

example	etymology	frequency
Allt Lowrie	PG/G *labhrach*, ‘noisy’	7
Rappla Burn	G *ràpalach*, ‘noisy’	2
Burn of Dararach	G *dararach*, ‘loud rattling noise’	2
Wharlish Burn	G *cairealach*, ‘noisy’	2
Allt na Galanaich	G *galanach*, ‘noisy’	1
Eas a’ Ghlaoidh	G *glaodh*, ‘cry, shot’	1
Glary Burn	G *glaothar*, ‘noisy’	1
Allt Clappy	G *clapach*, ‘clapping’	1
Fèith Thalain[19]	G *callan*, ‘noise, clamour’	1

4.3.6 Temperature

Figure 4.8 on page 143 shows the relationship between ‘hot’ and ‘cold’ burns. It weakly shows that the upper limits of names with the semanteme ‘cold’ are at a higher altitude than those with ‘hot’, which exist mainly at lower altitudes.

4.3.7 Course

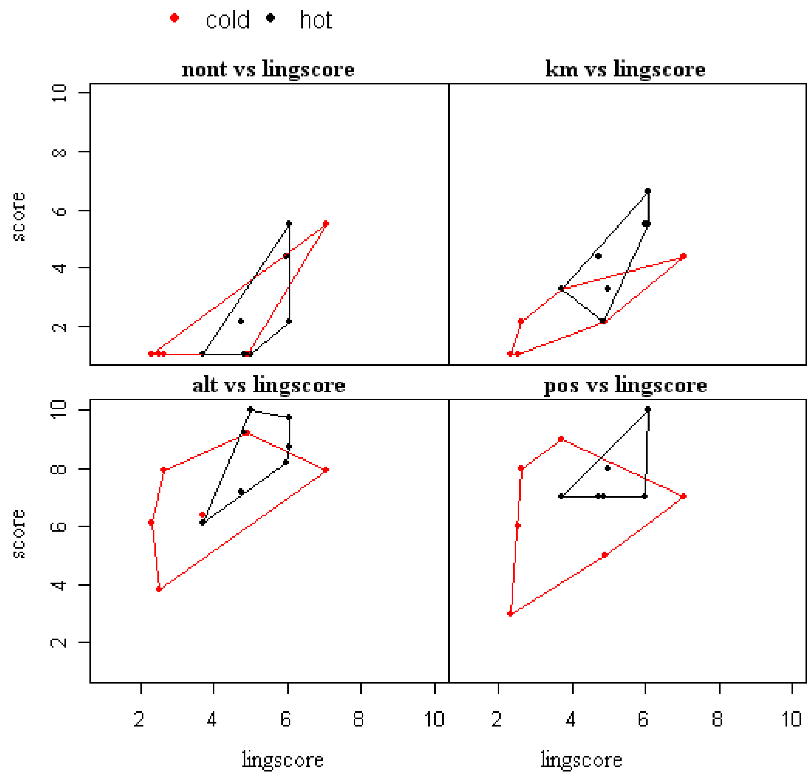
Figure 4.9 shows that a ‘straight’ watercourse is considered the default. This hypothesis is reinforced by the terminological structure of these semantemes. There are only six watercourses meaning ‘straight’, deriving from either G *dìreach* or G *dron*. There are nineteen terms for crooked; the top six are listed in table 4.13. The remaining terms occur only once each. Of this table, *cam* and *crom* and are clearly the most popular. Figure 4.10 on page 146 shows that the two terms represent similar watercourses. The difference between these two terms is unclear.

Table 4.13: Hydronyms Meaning ‘crooked’

etymology	frequency
G *crom*, ‘crooked’	18
G *cam*, ‘crooked’	12
G *carach*, ‘meandering, winding’	4
G *caim*, ‘loop or curve’	3
Sc *crooked*, ‘crooked’	3
G *cas, caise*, ‘steep, headlong’	2

[19] *Allt Fèith Chalan* 1912 Bartholomew: Survey Atlas of Scotland

Figure 4.8: Comparison of 'hot' and 'cold'



4.3.8 Effect / Character

Figure 4.11 on page 147 shows the difference between names meaning something broadly negative and names meaning something broadly positive. The most striking thing that can be seen from this is that the 'bad' names represent watercourses which are smaller than watercourses which 'good' names represent. This is probably for the following reason: A small watercourse which is considered dangerous, having a tradition of people drowning in it, would be named as such (e.g Allt Gauch from G *gàbhach*, 'perilous'). An equivalent small watercourse which is useful would likely be named for the way in which it was useful, if it did not drown anyone, it would be considered the default, it would not necessarily be considered 'good' for that. With the larger watercourses, a different attitude was taken. A large watercourse that drowned people was more likely to be treated with respect, since it was harder to avoid than a small mountain burn. Such a name may reflect this awe and respect rather than a name suggesting a desire for it to be avoided (perhaps such as the Dee or Don).

Conversely, the larger watercourses which were useful to the inhabitants would be called 'good' in the sense of 'useful' or 'abundant'(such as the Dichty and Mashie). It is possible of course that some of these names are euphemisms, i.e. that the river was in fact dreaded, and out of respect was called 'the good one' to avoid a perceived anger.[20]

[20]This phenomenon can be observed with the Greek Furies, vicious goddesses whose name *Eumenides* is a euphemism

Figure 4.9: Comparison of 'straight' and 'crooked'

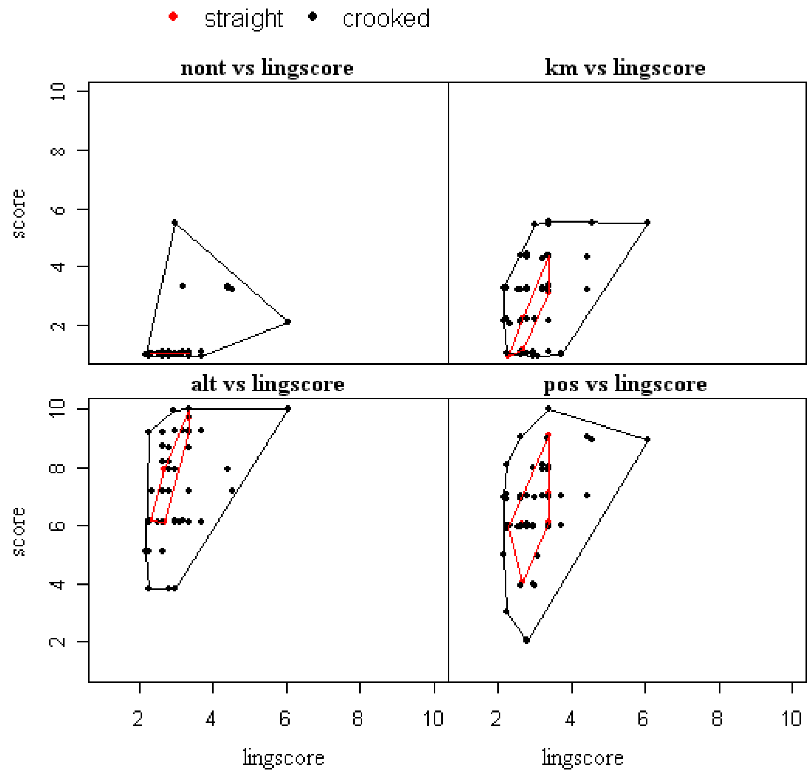


Table 4.14: Hydronyms with a Negative Meaning

example	etymology	frequency
Allt a' Chroin	G *cron*, 'harm'	5
Allt Gauch <	G *gàbhach*, 'perilous'	2
Allt Granda	G *grànda*, 'ugly'	2
Ess-growach	G *gruamach*, 'gloomy'[21]	2
Arity	G *earraideach*, 'quarrelsome'	1
Poll Bhat	G *bàidhte*, 'drowning'	1
Burn of Birse	G *bras*, 'rash, impetuous'	1
Bloody Burn	Sc *bloody*, 'bloody'	1
Bullie Burn <	G *boileach*, 'furious'	1
Allt Chiarlich	G *ciaralach*, 'perverse, quarrelsome'	1
Mossat Burn	PG *mosach*, 'filthy'	1
Peelie Burn	Sc *peelie*, 'ill looking'	1

meaning 'the well meaning ones'.

[21] This is comes from a single mention: 'burn cald Ess-growach, or wgly lynn' (c. 1591 Pont text 131v). It is unclear to me exactly what *growach* could relate to, unless one posits an orthographical error of *m* to *w*, which is possible in the style of the Pont MS.

Allt Sgionie	G *sgeunach*, 'fright, terror, dread' ?	1
Burn of Sorrow	Sc *sorrow*, 'sorrow'	1
Tifty Burn	Sc *tifty*, 'moody'	1

Table 4.15: Hydronyms with a Positive Meaning

example	etymology	frequency
Nethy Burn	PG **neiktodia OI necht*, 'cleansing'	5
Burn of Day	G *deagh*, 'good, excellent' or G *daigh*, 'fire'	4
Dichty Burn	PG *deagh*, 'good, excellent'[22]	4
Burn of Clearach	G *cliarach*, 'brave'	2
Condie	P *con*, 'wolf' or P *cun*, 'lovely'	2
Loch Mahaick	G *math*, 'good'	2
kunglas	P *cun*, 'lovely'	2
Bodería	OC **boud*, 'victory'[23]	1
Buthnott	PG *buadhnach*, 'healing one'	1
Comelyburn	Sc *comely*, 'comely'	1
Burn of Glansie	G *glan*, 'clean'	1
River Mashie	OP **mati-sia*, 'good'	1
Noran Water	OC **narona*, 'noble one'	1
Allt Slanaidh	G *slànadh* or G *slànuchadh*, 'healing'	1
Trusty Burn	Sc *trusty, trasty*, 'reliable'	1
Allt Unaig <	G *unach*, 'washing'	1
Glander	PG *glan*, 'clean'	1

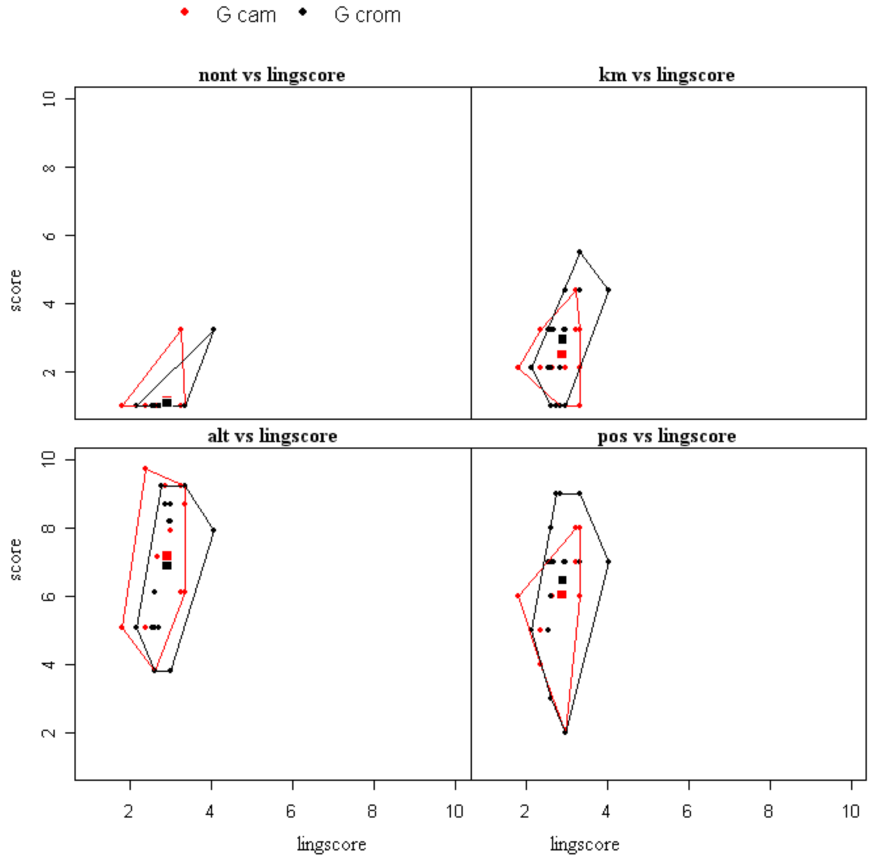
4.3.9 Visibility

In figure 4.12 on page 148 one might expect a difference of altitude, with perhaps exposed watercourses flowing at higher altitudes than hidden ones, however this is only weakly represented. It is probable that the factors contributing towards these semantemes are not present in the data gathered. That is, the difference between the altitude of the watercourse itself and the immediate land around, and local vegetation. For further discussion see the section on *caochan* in section 3.4.11 on page 80.

[22]See King, "Lochy' Names and Adomnan's Nigra Dea', p. 84-85 for discussion of the *-ty* ending, (in the case of Lochty). It is shown here that this ending reflects a phonological process as opposed to a morphologcal suffix.

[23]For debate on this name, see Rivet and Smith, *The Place-Names of Roman Britain*, p. 269-271. The derivation is not certain and it is possible this RN does not belong in this list.

Figure 4.10: Comparison of G *cam* and G *crom*



4.3.10 Dimensions

Table 4.16: Hydronyms Meaning 'large' and 'small'

large	
nontscore	*kmscore*
1.32	2.93
small	
nontscore	*kmscore*
1.07	2.40

Table 4.17: Hydronyms Meaning 'large' and 'small' as Part of a Pair

large	
nontscore	*kmscore*
1.45	1.90
small	
nontscore	*kmscore*
1.00	1.23

Figure 4.11: Comparison of 'good' and 'bad'

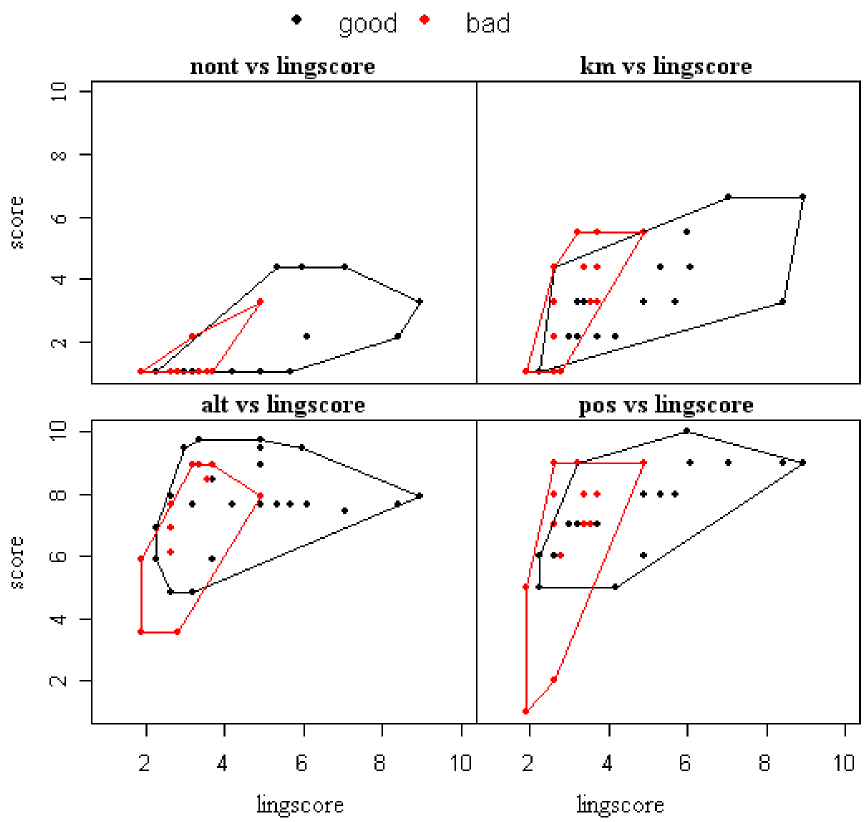


Table 4.16 shows the average *nontscore* and *kmscore* for watercourses with semantemes, 'large' and 'small'. If the comparison is restricted to only watercourses which form part of a pairing (such as *Allt Mòr* and *Allt Beag*) one sees the same correlation in table 4.17. The smaller difference between the largest and smallest names may seem counter-intuitive, but are due to two reasons. Firstly the number of entries is actually relatively small, there are only five pairs of names in this set. Secondly, names in pairs, where one is 'large' and the other 'small' tend to be river pairs which are of roughly similar size. Put another way, their names have been coined from their size because their size is not radically different. One would not expect, say, the Tay and the Dichty to form a RN pair, despite the fact one watercourse is larger than the other; they are in a different class of size. This was first noted by Stewart[24] who wrote: "In fact, rather few descriptions can be called absolute. Thus there is no absolute standard of bigness, and Big River may get its name only because it happens to be bigger than the streams near it."[25]

The distribution of *mòr* is not entirely typical of a Gaelic hydronym, it is predominantly restricted to the Highland Perthshire part of the Tay catchment area and the Spey catchment area. In terms

[24]Stewart, 'A Classification of Place Names', p. 3-4.

[25]This category corresponds approximately to Stewart's category A: 'namings from size' and partially C: 'configuration-names' under his 'descriptive names'. See: Stewart, *Names on the Globe*, p. 9-2

Figure 4.12: Comparison of ‘exposed’ and ‘hidden’

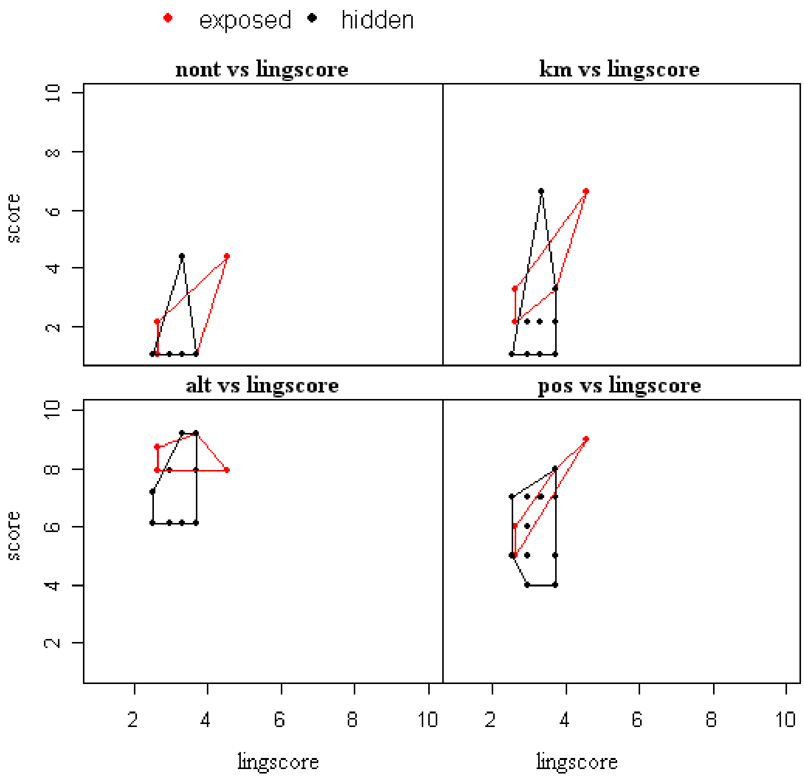


Table 4.18: Occurrences of G *mòr* by Catchment Area

RN	frequency
River Spey	17
River Tay	12
River Forth	3
Don to Spey	2
River Dee	1
Forth to Tay	1

of a default, it would seem that ‘small’ is the default. This is suggested by the higher number of watercourses meaning ‘large’ predominantly in the G term *mòr*. It is possible that *mòr* and *beag* are not to be considered opposites. In an article by Dodgshon[26] he notes that in certain Highland descriptive epithets, *mòr* is used in opposition with G *òg*, ‘young’. The G term *òg* is rare in Scottish hydronymy, which is what one would expect with a default term.

It is of note that whilst most defaults exist across the linguistic strata, there seems to be little or no evidence of ‘big’ and ‘small’ terms being used in this way in the Scots stratum. Indeed it seems that the Scots watercourses coined for their size are in fact well-named as table 5.1 on page 196 shows for

[26]Robert A. Dodgshon, ‘Symbolic Classification and the Development of Early Celtic Landscape’, *Cosmos: yearbook of the traditional cosmology society: Duality* (1985), p. 75.

Muckle Burn (4869).

4.3.11 Bed

This comprises terms denoting the quality of the bed of the stream, predominantly describing its hardness or stoniness. The top six are shown in table 4.19.

Table 4.19: Hydronyms Coined from the Quality of their Beds

etymology	frequency	example
P **cal*, 'hard'	11	Keltie Burn
G *gaineamhach*, 'sandy'	3	Allt Gaineamhach
Sc *stan, stane*, 'stone'	3	Stan Burn
G *clachach*, 'stony'	2	Allt Clachach
Sc *rocky*, 'rocky'	2	Rocky Burn
Sc *sand*, 'sand'	2	Sand Loch

The commonest term is P **cal*,**caled*. This is of uncertain meaning, whilst being cognate with W *caled*, 'hard' it is difficult to ascertain in what way water is 'hard'. Figure 4.13 compares these names with all P-Celtic names. Further work needs to be done with this root. It may be that the appropriate semantic class for **cal* is 'manner' as opposed to 'bed'. The semantemes 'hard water' and 'soft water' should not be confused with these phrases in English / SSE. The restricted usage of 'hard water' to describe water with mineral deposits in it is surely restricted to English and Scots, and it would be inadvisable to extend this meaning back to the P-Celtic stratum.

4.3.12 Number

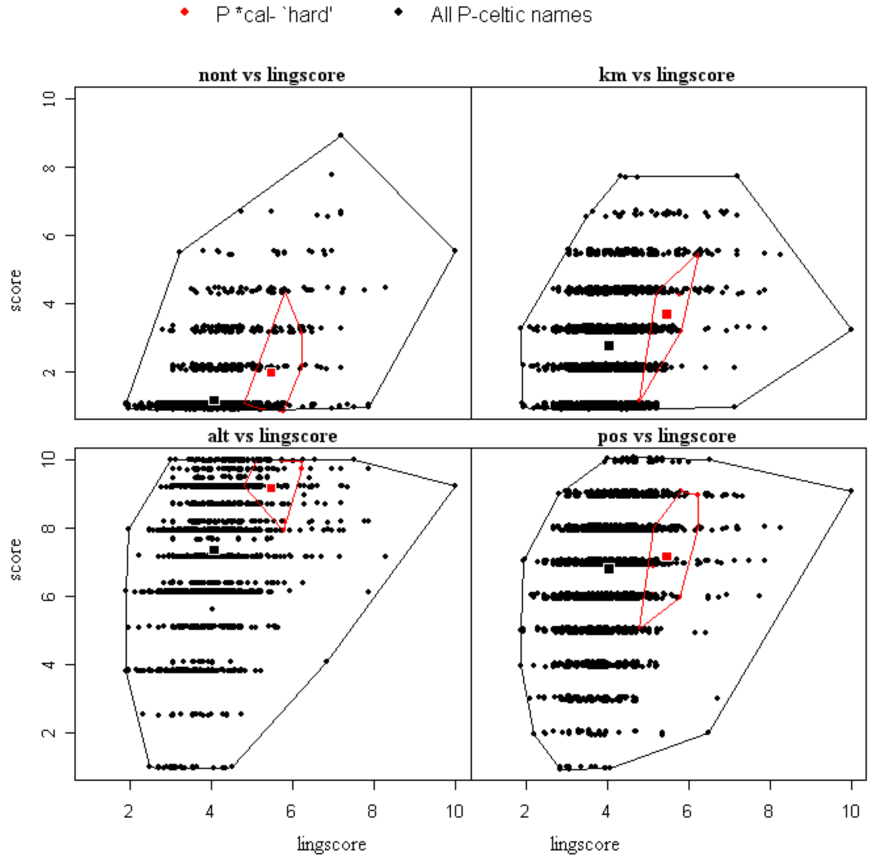
This is a small group which basically includes two types of names. The first are names denoted specifically as a pair, with three terms containing the term G *càraide*, 'couple'. These are *harity from Inverharity (Now on Allt na Beinne, but this was once called *Alt Herraty* c. 1591 Pont map 28) and two instances of Quharity Burn. Secondly, there are 'Three Burns' and 'Na Tri Chaochain' of identical meaning. It could be claimed that these terms are not hydronyms as such, but instead denote a general area, akin to G *alltach*, 'place of the burns'. They have been included here because they seem to denote a hydronymic term.

4.3.13 Dryness / Moistness

Needless to say, due to the inherent qualities of watercourses, they are all 'wet', so the semanteme 'dry' is much more common. The semanteme 'dry' probably signifies a watercourse which dries up or reduces in size at certain times of the year. This was Watson's view.[27] Of the names meaning 'wet', it is probable that the names have some other nuance, such as 'flowing water' (much like Nicolaisen's 'water-words'), or perhaps represent watercourses which have a deal of spray coming from them, so

[27] Watson, *The Celtic Place-Names of Scotland*, p. 447.

Figure 4.13: P **cal* vs All P-Celtic Names



that one gets wet standing near it. This is perhaps a case where visiting the actual watercourse will elucidate the underlying reasons for naming, in a way that this methodology will not.

In terms of defaults, it is self-evident that ‘wet’ is the default term. As one might expect, in the Scots hydronymicon at least, there is no particular term for ‘wet’ and the term ‘dry’ is much more common. This is because it is remarkable if a watercourse is dry, but not if it is wet.

4.3.14 Elevation

There appear to be three terms for ‘height’ within the AOS, these are G *àrd*, ‘height’, G *èirigh*, ‘rising’ (Allt na h-Eirghe) and P *uchel*, ‘high’. It is interesting to note there does not seem to be any obvious term for ‘low’ in opposition to any of these terms. One possible contender is G *moineiseach*, ‘low, inactive’, but this seems to be a derivative of G *moineis* with a primary meaning of ‘slowness’. Another possibility is G *domhainn*, ‘deep’; this has the meaning of ‘low’, and is the opposite of P *uchel*, ‘high’. This is perhaps the situation in the Ochil Hills, from P *uchel*, through which runs the Devon, cognate with G *domhainn*, but probably older.

In terms of discerning a default it would seem that G *domhainn* is the non default, since that is the consistent term for ‘low’ whilst the term for high is filled with a variety of terms. However, as with the issue of ‘age’ above, it is not certain that *domhainn* reflects an opposite of ‘high’, the term deep could possibly relate to a watercourse with steep sides, regardless of altitude. The evidence of Ochil and the Devon however suggests otherwise.

As with a comparison of *km* and names relating to ‘dimension’, a comparison between *alt* and ‘elevation’ is not particularly productive, suggesting the elevation is relative, not absolute. This is necessarily the case, since the namers of the watercourses could only consider the name as part of the landscape, as opposed to the system here, which considers the absolute distance from sea-level.

4.3.15 Flora

RNs relating to flora are rarer in larger watercourses than is generally accepted.[28] Two predominant traditional derivations are the Spey and the group based around the PG word *leamhann*, ‘elm’, in such names as Leven and Lyon. It is suggested here that neither of these names relate to flora.

Spey: On geographical terms if the Spey were to mean ‘hawthorn’, it would be by far the largest river in Scotland to relate to flora, and would not fit into the standard deviation model discussed in section 2.6.2.3 on page 45.

Leven / Lyon: If the Leven and Lyon names were considered to derive from G *leamhann*, ‘elm’, then they would be the five RNs with the highest *geogscore* in the semantic class ‘flora’. Figure 4.14 on page 153 shows that Leven / Lyon names are substantially different from other ‘flora’ names. Simon Taylor[29] and others[30] have suggested an alternate root for this name, from OC/P **lēm* / **lēb* + *(o)nā*, cognate with Welsh *llyfn*, ‘smooth’, in the sense of ‘smooth, slow flowing river’.[31] That the names now represent in Gaelic the word for ‘elm’ in the form of Leamhann suggest that many of these names have been reanalysed along these lines at a later date. In the case of Lyon in G Lìomhann or Lìobhann, it seems that these names have for some reason not been reanalysed into Leamhann and have preserved the long vowel. An alternate derivation, but with the same result, would be to propose a form of the Gaelic cognate of this root, *sleamhainn*, ‘slippery’, which has an inorganic *s-*. Either of these proposals would put the Leven names into the ‘manner’ category. Figure 4.15 on page 154 shows this interpretation of the names. This overlap clearly shows that ‘manner’ is a more fitting semanteme than ‘flora’, and as such has been used throughout this thesis.

Several points are notable:

1. There is a larger range of flora used in Gaelic nomenclature than in Scots.
2. Alder is curiously absent from Scots nomenclature. This is predominantly because the areas where these trees grow are mainly Gaelic speaking areas, i.e. the Highlands. This can be seen by the average altitude for these names.

[28] This type of name is discussed under Stewart’s ‘associative names’. See: Stewart, *Names on the Globe*, p. 99.
[29] Personal communication regarding material for forthcoming Place-names of Fife volume II
[30] Thanks to Alan James in personal communication for helping to formulate my thinking in this area.
[31] G. R. Isaac, ‘Place-Names in Ptolemy’s Geography’, CD-ROM (2004), p. 72.

Table 4.20: Comparison of Flora between Scots and Gaelic RNs

etymology	frequency	etymology	frequency
G *feàrna*, ‘alder’	20		
G *caorunn*, ‘mountain ash or rowan tree’	16	Sc *rowan tree*, ‘rowan (tree)’	6
G *seileach*, ‘willow’	16	Sc *sauch, sauchen / willow*	18
G *beith*, ‘birch’	15	Sc *birken*, ‘birch’	6
G *craobh*, ‘tree, bush’	11	Sc *bush*, ‘bush’	9
G *giubhas*, ‘fir’	9	Sc *fir*, ‘fir’	1
G *calltuinn*, G *call*, ‘hazel’	9	Sc *hazel*, ‘hazel’	2
G *luachar*, ‘rush’	9	Sc *rash*, ‘rush’	4
G *coille*, ‘wood’	8	Sc *wood*, ‘wood’	8
G *fraochach*, ‘heath covered’	8		
G *crann*, ‘tree’	7	Sc *tree*, ‘tree’	2
G *aitionn*, ‘juniper’	6		
G *cuilionn*, ‘holly’	6	Sc *hollin, hollen*, ‘holly’	2
G *dearchail*, ‘abounding in berries’	5	Sc *berry*, ‘berry’	2
G *raineach*, ‘fern’	4		
G *maide*, ‘wood, timber’	3		
G *broighleag*, ‘whortleberry’	2		
G *carran*, ‘spurrey weed’?	2		
G *iubhar*, ‘yew tree’	3		
G *feòrach, feur*, ‘grass(y)’	4		
G *meògach*, ‘abounding in whey’	2		
G *sgitheach*, ‘hawthorn’	2		
G *airne*, ‘sloe, wild plum’	1		
G *bealaidh*, ‘broom’	1	Sc *broom*, ‘broom’	5
G *biorag*, ‘dutch rushes’	1		
G *càilein*, ‘seedling, husk’	1		
G *seasganach*, ‘abounding in corn’	1		
G *coinnleach*, ‘of candles, i.e. fir roots’	1		
G *creamh*, ‘fern’ gen sg	1		
G *cuilc*, ‘reed, bullrush’	1		
G *curran*, ‘carrot’	1		
G *dair, darach*, ‘oak (tree)’	2	Sc *oak*, ‘oak’	1
G *droighnean, draighionn*, ‘blackthorn’	2		
G *eanach*, ‘type of plant’	1		
G *fàirneag*, ‘sloe plum bush’	1		
G *farradh*, ‘crop’	1		
G *fiùran*, ‘sapling’	1		

Table 4.20: Comparison of Flora between Scots and Gaelic RNs

etymology	frequency	etymology	frequency
G *freumhag*, 'root' or 'portion'	1		
G *gòinean*, 'couch-grass'	1		
G *leamhan*, 'elm'	1		
G *lurachan*, 'garlic'	1		
G *neanntag*, 'nettle'	1	Sc *nettly*, 'abounding in nettles'	1
G *slat*, 'rod, twig'	1		
G *barragan*, 'weed that floats on water'	1		

Figure 4.14: Relationship between Leven / Lyon Names and 'flora' Names

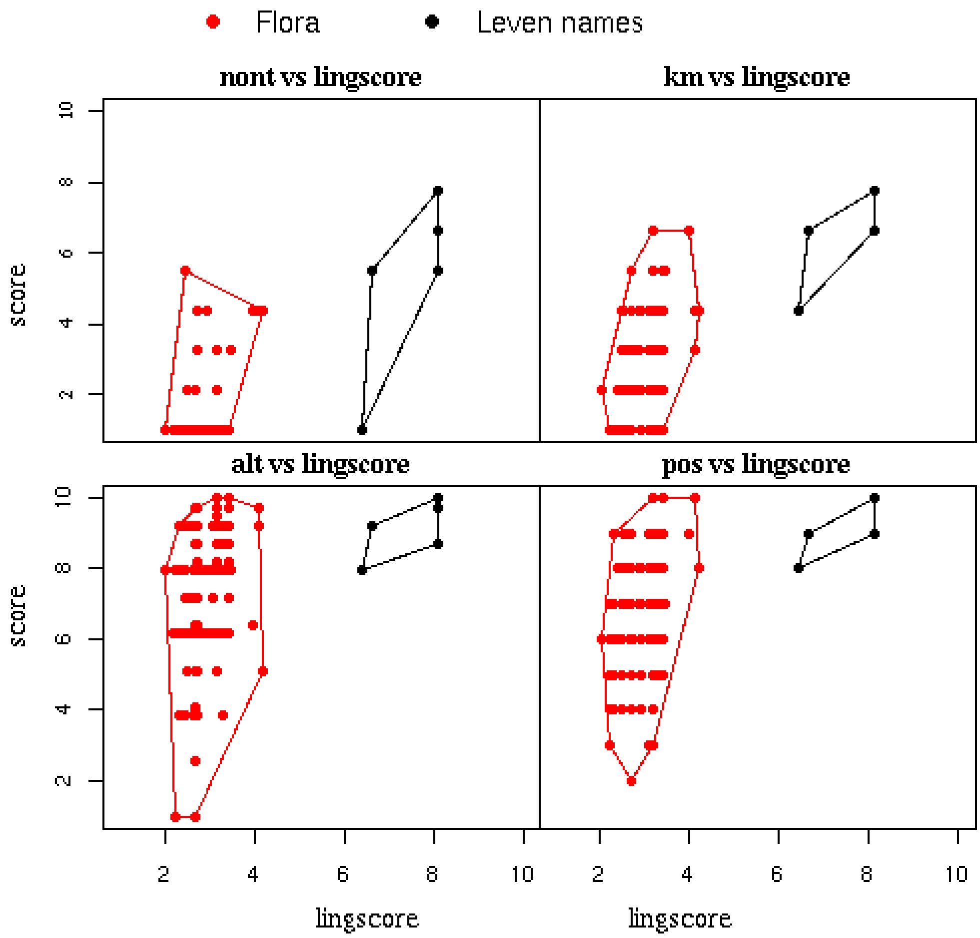
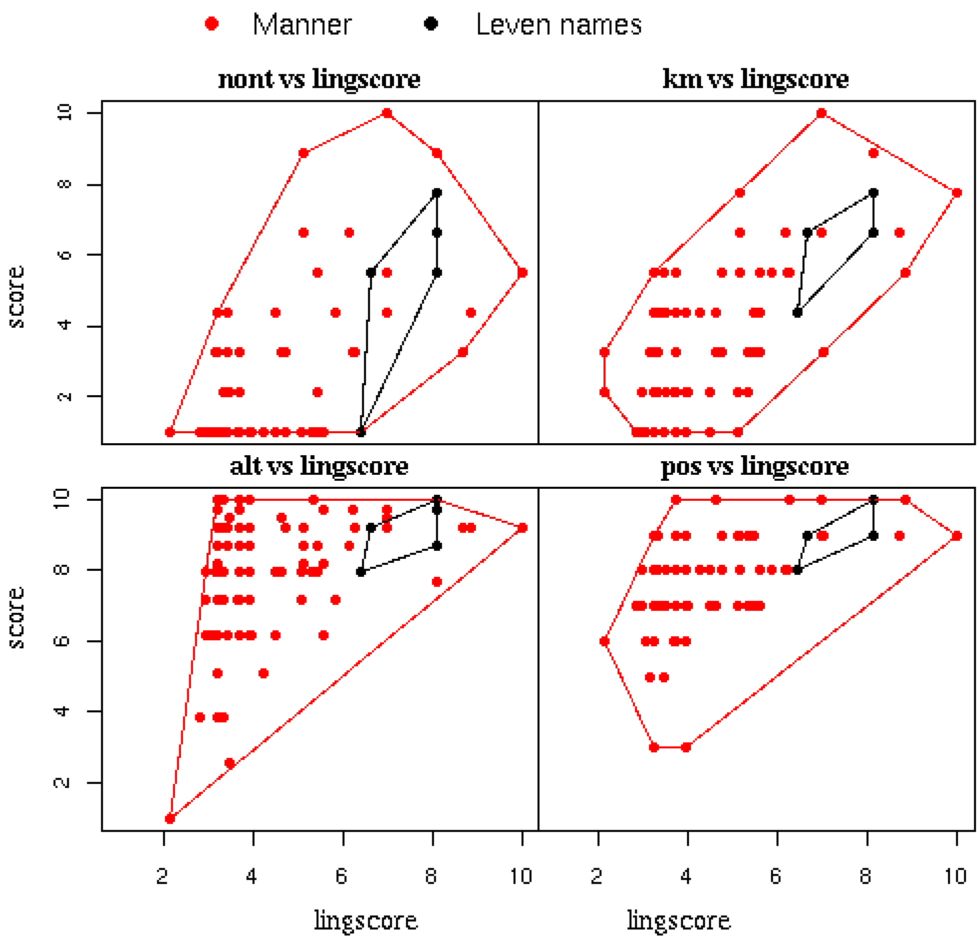


Figure 4.15: Relationship between Leven Names and 'manner' Names



4.3.16 Fauna

Table 4.3.16 on page 155 compares Scots and Gaelic terms for fauna in RNs. It is of course unwise to make too much comparison between the Scots and Gaelic occurrences of fauna and flora. In the case of fauna especially, it should not always be assumed that because an animal is referred to, it means the animal frequented the area banks of the watercourse. A number of other possibilities exist:

- With the earlier strata, the watercourses named after an animal can denote the fact that the qualities of the watercourse are similar to the qualities of the animal by a process of metaphorical extension. River Turk, for instance could represent a river that was conceived as flowing in a manner reminiscent of the way a boar moves, i.e. aggressively and quickly.[32] This may explain the relatively high number of watercourses named from such animals as bulls, wolves and maybe raven.

[32]See King, 'Endrick and Lunan', p. 151 for more discussion on how animals appear in RNs.

Table 4.21: Hydronyms Meaning ‘fauna’

animal	Gaelic	Scots
various birds	35	14
calf	21	1
goat	18	1
wolf, dog	18	6
cow	17	2
various fish	15	1
pig	14	
ox, stag	13	2
dog	13	1
horse	13	1
cat	10	1
bull	9	2
fox	9	2
deer	9	7
sheep	7	
kid	7	
badger	6	
heifer	4	
snake	3	
otter	3	
lamb	3	
insect	3	
boar	3	
kid	2	
hare	2	
foal, colt	2	
squirrel	2	
ewe	1	
snail, slug	1	
limpet	1	1
mastiff	1	
minnow, worm	1	
otter	1	
toad, frog	1	1
rabbit, coney	1	
horse leech	1	

- In some cases the watercourse could be coined from a natural feature which is perceived as looking like an animal. Several names in Scotland with the element *muc*, ‘pig’ probably do not denote the fact that pigs lived there, but that the particular hill looked like a pig’s back.

- If particular animals do frequent the area, it may not be the natural state of affairs but could represent a specific man-made use. This is a more restricted use than, say, transhumance, where sheep are kept in a particular place, but represents a situation such as Allt an Eich, ‘Burn of

> the Horse', which could represent any number of scenarios: a particular place where it was useful for people travelling by horse to water them, being easy to approach and at an equidistant point between two settlements for instance; a place where a particular horse drowned or where a ghostly horse was said to be seen. It does not necessarily mean that horses roamed the area in great numbers. This was first noted by Stewart who wrote: '...They do not mean that the animal was unusually plentiful at that spot or especially characteristic of it, but merely record a particular occasion upon which the animal was encountered'.[33]

The lack of parity between the amount of terms for flora and fauna between Gaelic and Scots is not entirely clear. There are two possible explanations. Firstly, it might be the case that areas of Scots settlement simply had less bio-diversity and as such the animals and plants were simply not present in these areas. This is hardly acceptable however, because, for instance, the most common absent fauna from the Scots hydronymicon are pigs and sheep with dogs, horses and goats also being rare. It is of course clearly wrong to suggest Scots speakers did not encounter these animals. The second explanation is that Scots speakers were less willing to elevate these terms to a toponymic status. The reasons for this is not clear, and may have its roots in the micro-management of agriculture which goes beyond the bounds of this thesis. To compare the Scots hydronymicon with that of the English, Ekwall states that for England "Derivations from names of animals are few. Some are probably of a late type".[34]

4.3.17 Hydronyms Coined from Places

Within this section, the various types of hydronyms named after places or areas nearby are discussed. These types of names divide approximately into two groups, names coined from natural features (e.g. Blackhill Burn), and names coined from man-made features (e.g. Burn of Oldtown). Another approximate division is between those named after a generic element alone (e.g. Glen Burn) and those named after a specific and generic element (e.g. Glendronach Burn). The boundaries between all four sets of names are fuzzy. In the first instance it is not always clear where an element represents a natural feature or a settlement coined from a natural feature. In the second case, it is not always clear where a hydronym is coined from a simplex non-hydronymic element (such as a place simply called 'Corrie').

These types of names correspond approximately to Stewart's *shift names* which he defines as: '...names places upon places by the mere shift of the specific form from one generic to another in the vicinity'.[35]

4.3.17.1 Hydronyms Coined from Other Hydronyms

There are three types of name which fall under this heading:

[33] Stewart, 'A Classification of Place Names', p. 4. See also this type of name as discussed under his 'associative names' in Stewart, *Names on the Globe*, p. 99-101

[34] Ekwall, *English River-Names*, p. li.

[35] Stewart, 'A Classification of Place Names', p. 9.

Pleonastic names These names are described as names where a generic element has been reanalysed as a specific element. An example would be Dour Burn, from the P-Celtic generic *dobhar*, now a specific element.

Hydronyms related to a neighbouring, but separate water feature This could represent a fully-fledged hydronym, such as Lochbroom Burn where Loch Broom is a toponymic entity in its own right. Alternatively the watercourse could relate to a small insignificant water-feature such as Snowgoat Burn from Sc *gote*, 'trench, watercourse'.

Hydronyms relating to settlement features In some cases hydronyms are coined from a settlement itself coined from the original name of the watercourse. Many of the names of the type '*Burn of* A(u)ld X' belong in this category. Take as an example Burn of Aultderg: Originally this name would have been what in modern Gaelic orthography would be Allt Dearg, 'Red Burn'. The settlement at the foot of the burn also took this name. At some point, probably with the decline of Gaelic in the area, the name of the burn was forgotten or became less important, and came to be given *burn* as a generic and be named after the settlement at its foot, hence Burn of Aultdearg. (In this case this must have happened relatively early, since Pont records 'B: of Alddeirg' in about 1591.) This process could be interpreted as metonymy.

To unravel which element is which in every case is often impossible, although there are several pointers in most situations. Table 4.22 on page 158 shows the frequency of these various elements.

Table 4.23 on page 159 counts the number of combinations of generic with specifics that denote watercourses or other bodies of water. This table shows all three types mentioned in the preceding section. In the instances, where one term is a watercourse and another a water-basin, clearly the name is of the second type. The 'Burn of Allt X' type was discussed above, and it is likely that most if not all these names relate to the third type mentioned above. Figure 4.16 on page 160 shows a distribution of these types of name, including not only *burn* and *allt*, but all combinations where the generic is Scots and the specific Gaelic, although the commonest type is that mentioned above. It is clear that these names lie along or near the 'Highland corridor', the area of greatest linguistic contact between Scots and Gaelic.

Figure 4.17 on page 161 shows two different sorts of relationships between watercourses and water-basins. The first type, in black on the map, represent pairs of names where a water-basin and watercourse have the same specific but different generic elements. There are forty-two names of this type. An example of this would be Gelly Burn and Loch Gelly. This is a common phenomenon throughout Scotland, and probably throughout the world. The second type represents names where the basin name is treated as a specific in the name of the watercourse. An example would be Lochgelly Burn and Loch Gelly. One striking thing about this second type is that eleven of the forty names are situated in the River Tummel, and ten of these relate to G *lochan* rather than *loch*.

Another cluster worthy of discussion is that in South Fife as represented in table 4.24 on page 160. It appears that, whilst the default naming style between a water-basin and a watercourse in most of the

Table 4.22: Hydronyms Coined from other Hydronyms

etymology	frequency
G *allt*, ‘burn’	54
G *loch*, ‘lake’	26
G *lochan*, ‘small loch’	19
G *linne*, ‘pool’	17
G *eas*, ‘waterfall / steep burn’	16
Sc *well*, ‘well’	15
G *slugan*, ‘gulf, whirlpool’	10
G *lùb*, ‘loop’	8
G *caochan*, ‘little blind one’	7
G *poll*, ‘pond’	7
PG **dubhag*, ‘black one’ / ‘black pool’	6
G *slug*, ‘puddle’	6
G *uar*, ‘waterspout’	5
G *uisge*, ‘water’	5
G *easan*, ‘little waterfall’	4
G *feadan*, ‘flute, spout’	4
G *fuaran*, ‘well’	4
PG *dobhar*, ‘water’	3
Sc *pot*, ‘hole in the ground;	3
Sc *spout*, ‘natural spring’	3
G *fuaran*, ‘well, spring’	3
Sc *grain*, ‘stream’	3
G *lòin*, ‘rivulet’	3
Sc *slouch*, ‘cataract’	3
G *tobar*, ‘well’	3
G *uarach*, ‘abounding in waterfalls’	3

AOS is to name the *burn* and *loch* (for example X Burn and Loch X) with the same specific element, in South Fife, the style was to name the watercourse after the basin (for example Loch X and Loch-X Burn).

4.3.17.2 Hydronyms Coined from Natural Features

Gaelic names are by far the commonest in this section as in table 4.25 on page 161.

Table 4.26 on page 162 shows the twenty commonest natural features used as specific elements in the AOS. The commonest element, G *coire* comprises half of the specific elements of the top ten. Table 4.27 on page 162 shows the twenty commonest man-made features used as specific elements in the AOS. For an explanation of ‘obscure PN element’, see the etymology entry on page 260. See page 126 for the definition of ‘man-made feature’.

Table 4.23: Combinations of Generic and Specific Element Coined from Another Hydronym

Generic	Pleonastic element	frequency
Burn	*Allt*	53
Allt	*Lochan*	16
Burn	*Well*	14
Allt	*Loch*	14
Burn	*Loch*	11
Allt	*Eas*	11
Burn	*Linne*	8
Burn	*Slugan*	6
Allt	*Fuaran*	6
Allt	**Dubhag*	5
Burn	*Eas*	4
Allt	*Poll*	4
Burn	*Caochan*	4
Allt	*Easan*	4
Allt	*Linne*	4
Allt	*Slugan*	4
Allt	*Feadan*	4
Burn	*Slug*	4
Burn	*Pot*	3
Loch	*Linne*	3
Allt	*Uisge*	3
Burn	*Uar*	3
Loch	*Lùb*	2
Burn	*Linn*	2
Lochan	*Uisge*	2
Caochan	*Lùb*	2
Burn	*Pow*	2
Burn	*Strand*	2
Allt	*Lòin*	2
Burn	*Stream*	2
Allt	*Tobar*	2
Burn	*Slouch*	2
Burn	*Uarach*	2
Burn	*Dobhar*	2
(The) *X* (only)	*Grain*	2
Other	*Spout*	2
Burn	*Lùb*	2
Allt	*Spùt*	2
Allt	*Coileach*	2
Burn	*Loch*	2

4.3.17.3 Hydronyms Coined from Settlements

To analyse these names in terms of the components of *geogscore*, caution needs to be exercised. Whilst correlations do exist, one must be sure whether one is measuring a quality of a watercourse or

Figure 4.16: Names with Scots Generic Elements and Gaelic Generic Elements as Part of the Specific Element

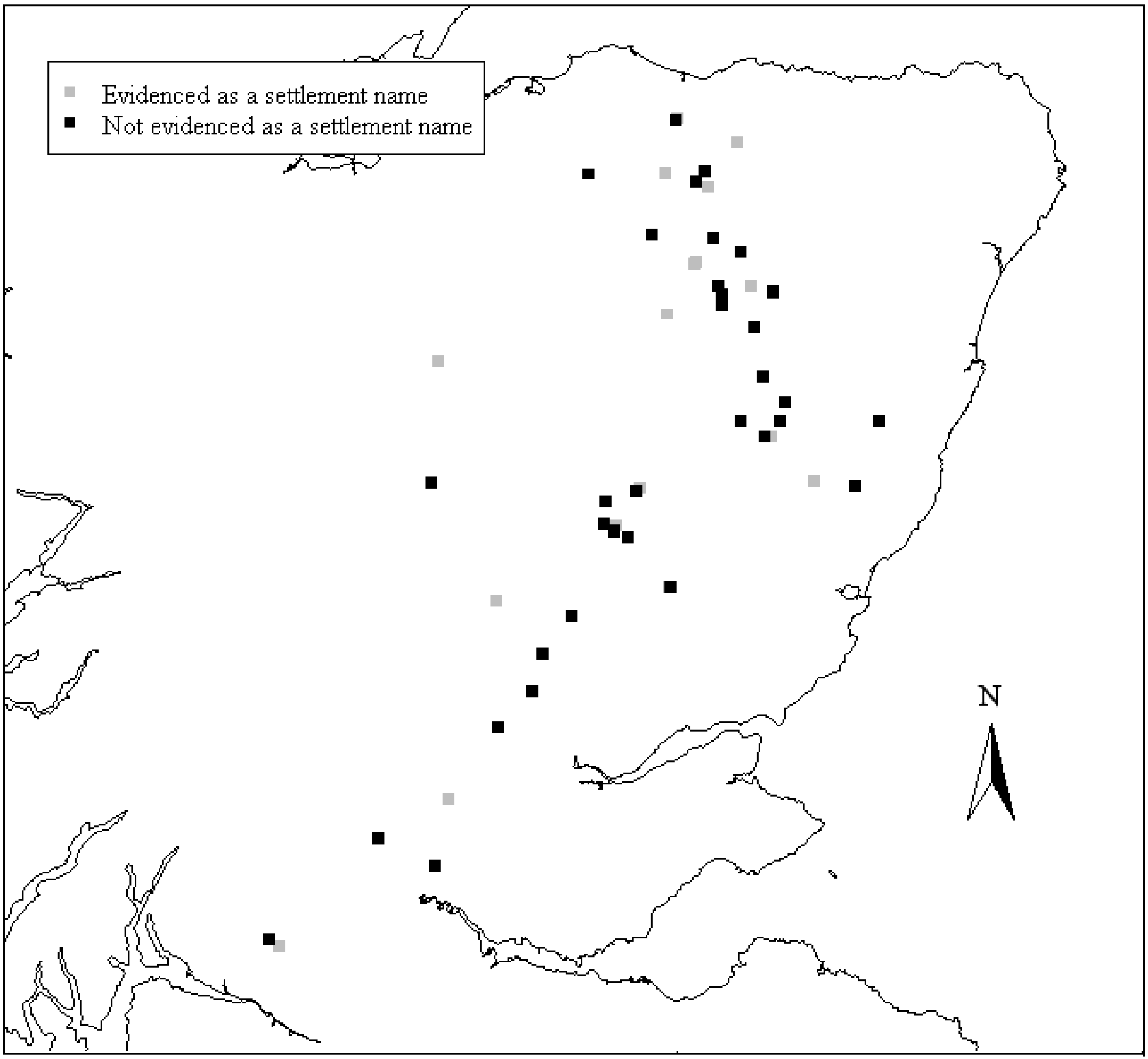


Table 4.24: Naming Style of lochs in South Fife

id	RN	basin
75	Aldlochglo	Loch Glow
4326	Lochfitty Burn	Loch Fitty
4327	Lochgelly Burn	Loch Gelly
4329	Lochmalony Burn	Loch Malony
4334	Lochorisburne	Loch Ore
4335	Lochornie Burn <	*Loch Ornie

measuring a quality of the settlement at second hand. For instance, to obtain the average altitude for all RNs with *achadh* as a specific may produce a correlation, but what is really being measured is the

Figure 4.17: Distribution of Basins and Watercourses

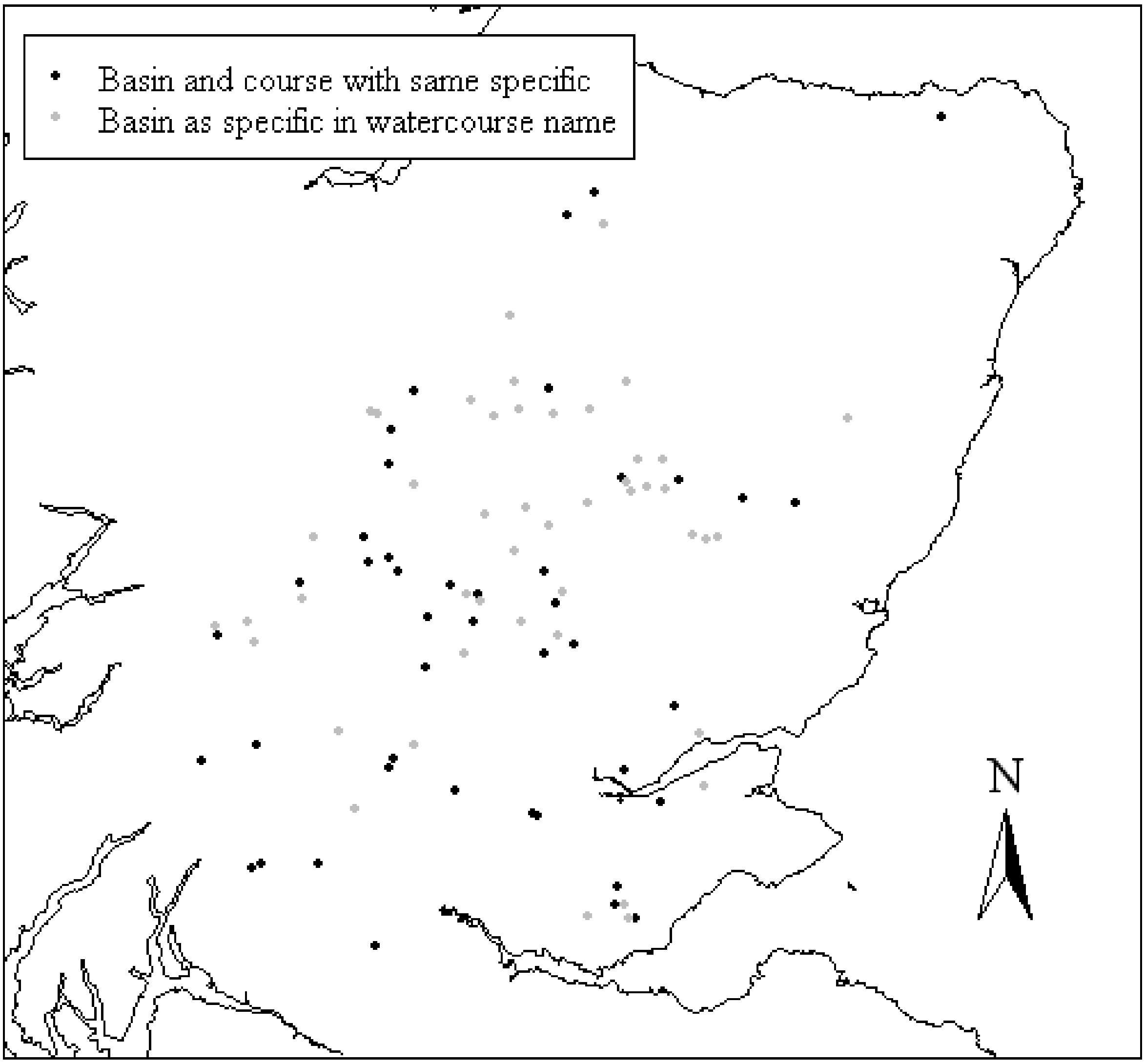


Table 4.25: Hydronyms Coined from Natural Features by Linguistic Stratum

specelem	frequency
OC / P-Celtic	7
P-Celtic	14
P-Celtic / Gaelic	22
Gaelic	2237
Scots	584

altitude of settlements with *achadh* as a specific. To be able to compare the altitude of the settlement and the altitude range of the associated watercourse would be appropriate, but goes beyond the bounds of this thesis.

Table 4.26: Twenty Commonest Natural Features in Hydronyms

etymology	frequency
G *coire*, ‘circular hollow’	407
G *creag*, ‘crag’	57
G *allt*, ‘burn’	56
Sc *glen*, ‘glen’	50
Sc *hill*, ‘hill’	44
Sc *den*, ‘narrow valley’	44
G *gleann*, ‘glen’	43
G *clais*, ‘furrow’	40
G *dùn*, ‘hill, hillfort’	40
G *bad*, ‘thicket’	38
G *tulach*, ‘hill’	36
G *beinn*, ‘hill’	33
G *druim*, ‘ridge’	32
G *ceann*, ‘head, end’	31
G *loch*, ‘lake’	28
G *bealach*, ‘pass’	28
G *fèith*, ‘bog’	28
Sc *lea, ley*, ‘untilled ground’	26
G *àird*, ‘height’	25
G *tom*, ‘knoll’	25

Table 4.27: Twenty Commonest Man-Made Features in Hydronyms

etymology	frequency
Sc *tun*, ‘town’	93
G *baile*, ‘township’	74
G *ruighe*, ‘slope’	65
G *achadh*, ‘shieling’	62
Sc *mill*, ‘mill’	51
G obscure PN element	34
G *pit*, ‘unit of land’	33
Sc *shiel*, ‘summer farm’	21
G *cill*, ‘church’	21
Obscure place-name element	17
G *àth*, ‘ford’	15
Sc *hall*, ‘hall’	15
G *inbhir*, ‘confluence’	15
Sc *side*, ‘side’	14
G *both*, ‘hut, dwelling’	14
G *muileann*, ‘mill’	14
Sc *ford*, ‘ford’	13
G *cùl*, ‘back’	13
Sc *house*, ‘house’	10
Sc *fold, fauld*, ‘fold, pen’	9

Table 4.28: Commonest Types of Natural Features Associated with Hydronyms

example	name	frequency
Altowybreck	Concavity	693
Addabing	Convexity	645
Burn of Aldachuie	Body of water	299
Aber Burn	Area not used for agriculture	264
Burn of Allachampit	Area used for agriculture	158
Lochan Beanaidh	Riparian area	68
Allt Bunbruach	None	59
Allt a’ Bhraonaich	Other	2

The *km* and *nont* of watercourses related to settlements is appropriate to analyse, however, since this has no analogy with the settlement. Table 4.29 shows the average, minimum and maximum values for *nontscore* and *kmscore*.

It is clear that nearly all, if not every single, watercourse named from a settlement is a small watercourse. This clearly fits the hypothesis that watercourses were used or at least encountered by people within an agricultural / economic context. This concurs with Stewart’s comment: “Small streams, especially, are likely to remain unnamed in the early stages of development, and later take the name of the village.”[36] Moreover, the larger the watercourse, the more likely the situation is to be

[36]Stewart, *Names on the Globe*, p. 158.

Table 4.29: Size of Watercourses Associated with Settlements

nontscore		
minimum	average	maximum
1	1.2	7.8
kmscore		
minimum	average	maximum
1	2.9	7.8

reversed: that the settlement will be named after the watercourse, rather than *vice versa* (e.g. Aberdeen from the Dee, Inverness from the Ness, Dundee from the Tay). Table 4.30 lists the fifteen longest watercourses which are coined from a settlement or man-made area of some sort.

Table 4.30: Fifteen Longest Watercourses Coined from a Settlement or Man-Made Area

km	RN	primary place-name	etymology	source	q
30.5	Burn of Garnesmiln	Garden's Mill	Sc *mill*, 'mill'	F1	
26.5	Tarland Burn	Tarland	Sc *land*, 'land, area'	OS	q
25.5	Water of Deer	Deer	PG obscure place-name element	F1	
24.5	Greenmire Burn	Greenmyre	Sc *mire*, 'mire"	OS	
22.5	Leochel Burn	Leochel-Cushnie	PG *lòchial*, 'bright place'	OS	
20.0	Monikie Burn	Monikie	G *mòine*, 'moss' / *monadh*, 'mountain'	OS	
19.0	Gallangad Burn	Gallangad	G obscure place-name element	OS	
18.5	Ernan Water	*Ernan	G *earrann*, 'division of land'	OS	q
16.0	Feardar Burn	*Feardar	PG *dobhar*, 'water'	OS	
15.5	Baddoch Burn	Baddoch	G *badach*, 'abounding in thickets'	OS	
15.5	aquas de Fetteresso	*Fetteresso	P **uotir*, 'district'	F1	
15.0	Lumphanan Burn	Lumphanan	G *lann fionnain*, 'church'	OS	
15.0	Pollagach Burn	The Pollach	G *bad*, 'thicket'	OS	
14.5	Newmill Burn	Mill of Fintray	Sc *mill*, 'mill'	OS	
14.0	Kirkney Water	Kirkney	G *cearc*, 'hen'	OS	

As can be seen, the five longest watercourses all have a degree of uncertainty about them; either they are poorly documented or are of uncertain derivation. The longest watercourse that definitely derives from a man-made settlement is in my opinion Leochel Burn, deriving in all likelihood from a parish name of P-Celtic origin. Conversely if one looks at the fifteen longest watercourses in table 4.31 on page 164, ten of the RNs have settlement-names coined directly from them (such as Dundee from the Tay), none of the RNs, however, are coined from a settlement.

4.3.18 Specific Person / Occupation

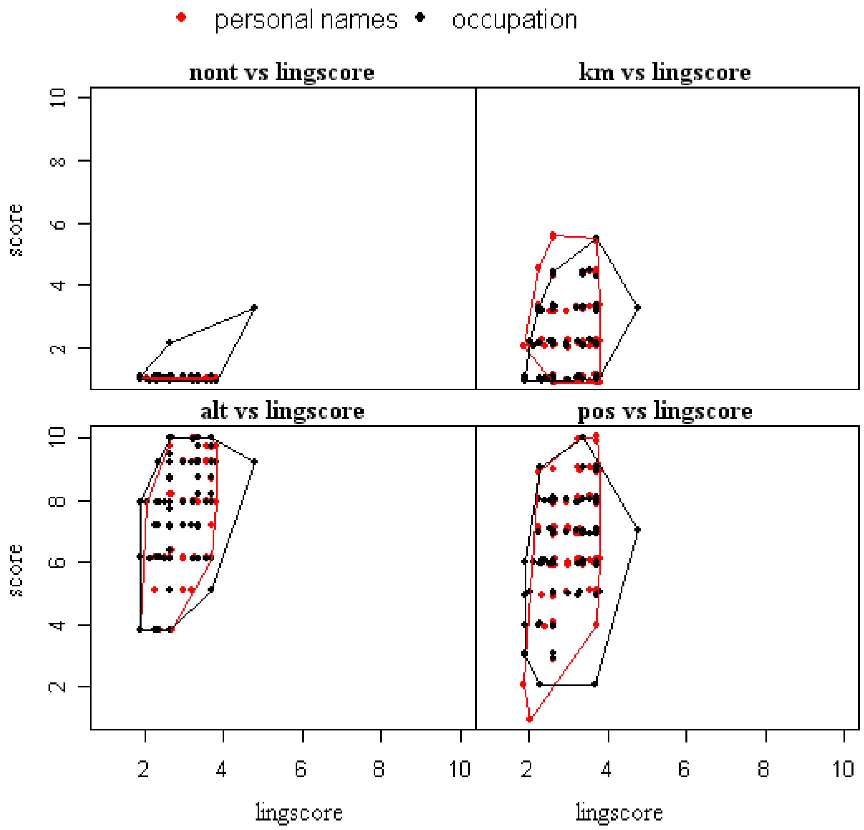
This section includes elements comprising both personal names, and occupations or trades.[37] Figure 4.18 shows how similar the two elements are, and that they should effectively be treated as one. The frequency of personal names broadly correlates with the frequency of Scottish names from the late Medieval era up to the present day. For instance, the commonest name is *Eòghainn*, in SSE Euan, obviously a common name. Following with three of each are: Aillig, Matthew, Padruic and Bhaltair. Of the non-personal names, G *cailleach*, 'woman' is by far the commonest, with twelve examples.

[37] This class corresponds to 2c *Mythological name* in Stewart's classes: Stewart, 'A Classification of Place Names', p. 4.

Table 4.31: Fifteen Longest Watercourses and the Settlements associated with them

km	RN	secpn	etymology	source	q
159.5	River Spey	Inverspey	OP *sk^wei*, 'vomit'	OS	
129.5	River Dee	Aberdeen, Inverdee	OC **deva*, 'goddess'	OS	
116.5	River Don	Inverdon	OC **devona*, 'goddess'	OS	
98.0	River Deveron	Deveron	OC *earn*, 'river'	OS	qq
85.5	Loch Tay	Kinmore	OC **tava*, obscure	OS	
85.5	River Tay	Dundee	OC **tava*, obscure	OS	
77.0	River Forth		PG *fo-rith*, 'slow running one'	OS	
69.5	River Ythan	Inverythan	OC **iektona*, 'talkative one'	OS	
66.5	River Earn		OC **isarona*, 'fast flowing river'	OS	
56.5	Loch Ericht		G *eireachd*, 'assembly'	OS	
51.5	River Avon	Inveraven	PG **abona*, G *abhainn*, 'river'	OS	
50.0	River Devon	Aberdona	OC **dumona*, 'deep'	OS	
46.0	River Almond	Inveralmond	OC **ambona*, 'damp one'	OS	
46.0	River Eden		OC **ituna*, '?'	OS	q
45.0	River Isla		OC **ila(f)-*, '?'	OS	q

Figure 4.18: Comparison of Personal Names and Occupational Names



All the names are Gaelic or Scots, and relate generally to smaller watercourses. Personal names as hydronyms do not seem to occur in older linguistic strata, although there is evidence for Pictish personal names in glen names, such as Glentarken for *Talorgan* a Pictish name. Of course, the name *Dobur Artbranani*, the lost name from the Life of St Columba as discussed on page 55, shows that such names did exist, but have been eradicated.

4.3.19 Supernatural entity

Figure 4.19: Comparison of Dee and Don with other RNs Meaning ‘supernatural entity’

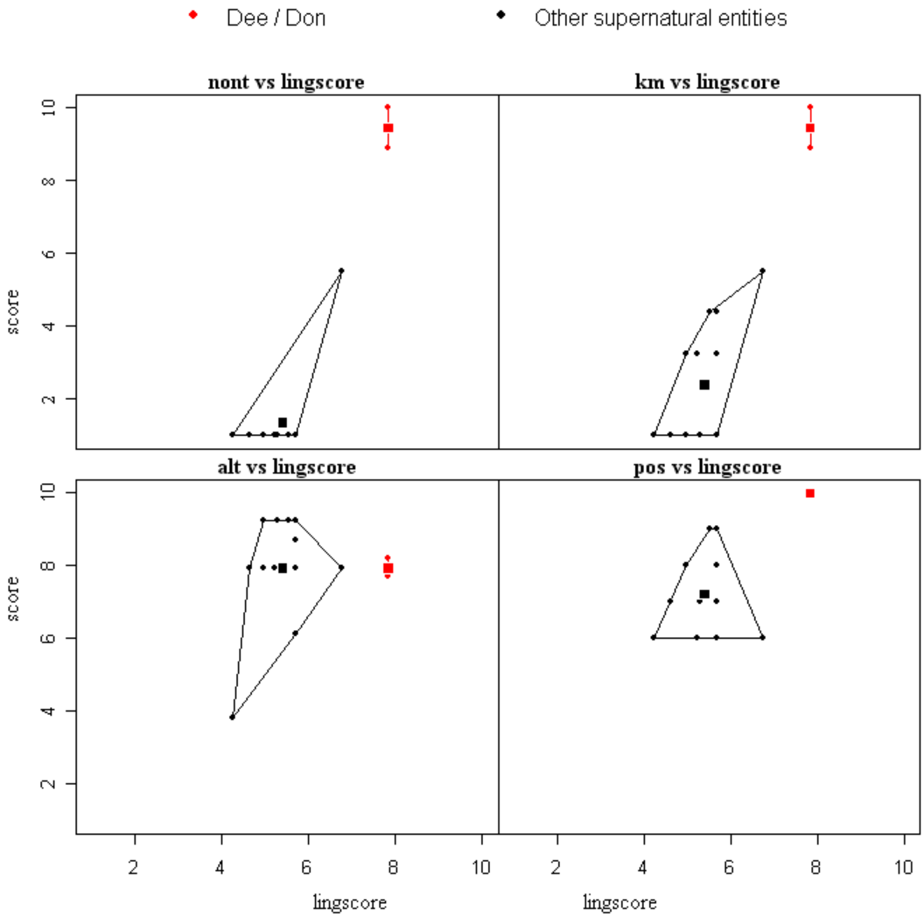


Figure 4.19 compares the Dee and Don (which both derive from an Early Celtic root meaning ‘goddess’) with other names relating to supernatural entities. These two watercourses are far larger than the others in this group. This variation presents a methodological problem. On the one hand, it seems clear that the Dee and the Don as a group, and the other names, represent two different semantic groups somehow. On the other hand, it is difficult to find a way of semantically dividing the names in a satisfactory way so that the Dee and the Don are separated.

This issue is not a problem in general toponymy; it almost certainly reflects the way data have been interpreted within this methodology. That said, if one looks at the semantic content of the ten largest watercourses in the AOS where the meaning is known as in table 4.32, it can be seen that apart from the Dee and Don, all the other names represent qualities of the water (as expected). The term 'goddess' can hardly be said to be conceptually situated in this semantic area.

Although a full discussion of the names relating to Dee and Don goes outside the remit of this thesis, it is worthy of note that this pattern exists throughout the British Isles. The idea that these names represented a taboo divine name that was too holy to speak may offer an acceptable explanation here. For instance in Mainland Europe, where records go much further back, the names of the largest rivers are much more evenly distributed between names that represent general adjectival terms such as the Seine[38] and divine names. The divine names, however, are specific names, rather than the perhaps titular 'goddess', such as the Marne from Matrona, 'mother goddess'.[39] This may have been the situation once too in the British Isles. For instance, it has been demonstrated that the Dee in Cheshire was Aeruen (=Aerfen) 'a goddess of war' in early Welsh poetry.[40]

It may be that the division between divine names and adjectival names is largely one of modern scholars' creation; that is, some of the names which modern scholars regard as representing the manner of a watercourse, relate also to a divine epithet or name. For instance names such as the Ythan, mentioned above, could be construed as the 'talking one' or perhaps, 'the talking divine being', with no distinction between these two names being made in the minds of the coiners. Evidence for these 'adjectival-divine names' is stronger in a name such as the Tanner which has divine cognates such as Tanaros, a Celtic thunder deity, but also is linked to the Celtic word for 'thunder'.

Table 4.32: Semantic Content of Ten Longest Watercourses

RN	etymology	*semtype*
River Spey	OP *sk^wei*, 'vomit'	4
River Forth	PG *fo-rith*, 'slow running one'	4
River Dee	OC **deva*, 'goddess'	28
River Deveron	OC *earn*, 'river'	22
River North Esk	OP **esc*, 'water'	21
River Don	OC **devona*, 'goddess'	28
River Tummel	OC **temelo-*, 'dark'	2
River Ythan	OC **iektona*, 'talkative one'	5
River South Esk	OP **esc*, 'water'	21
River Leven	P *llyfn*, 'smooth, flowing gently'	4

The remaining names generally denote various supernatural creatures which were probably thought to inhabit the local area. The G term *sìdhean* for a fairy-hill has not been included in this section, since the term does not denote the watercourse, but a neighbouring hill. The names are all Gaelic or Scots and relate to a non-Christian belief in invisible beings haunting particular areas. The two exceptions are 'Burn of Angels', which appears to be Christian, although this could equally be G *aingeal*, which has the meaning of both 'fire' and 'angel', reanalysed as Sc *angel*.

[38]This is OC *Se:koána* from the root **secu-*, 'cutting' Isaac, 'Place-Names in Ptolemy's Geography'.
[39]Holder, *Alt-Celtischer Sprachschatz*, p. 468-470.
[40]Skene, *The Four Ancient Books of Wales*, p. 341.

4.3.20 Event

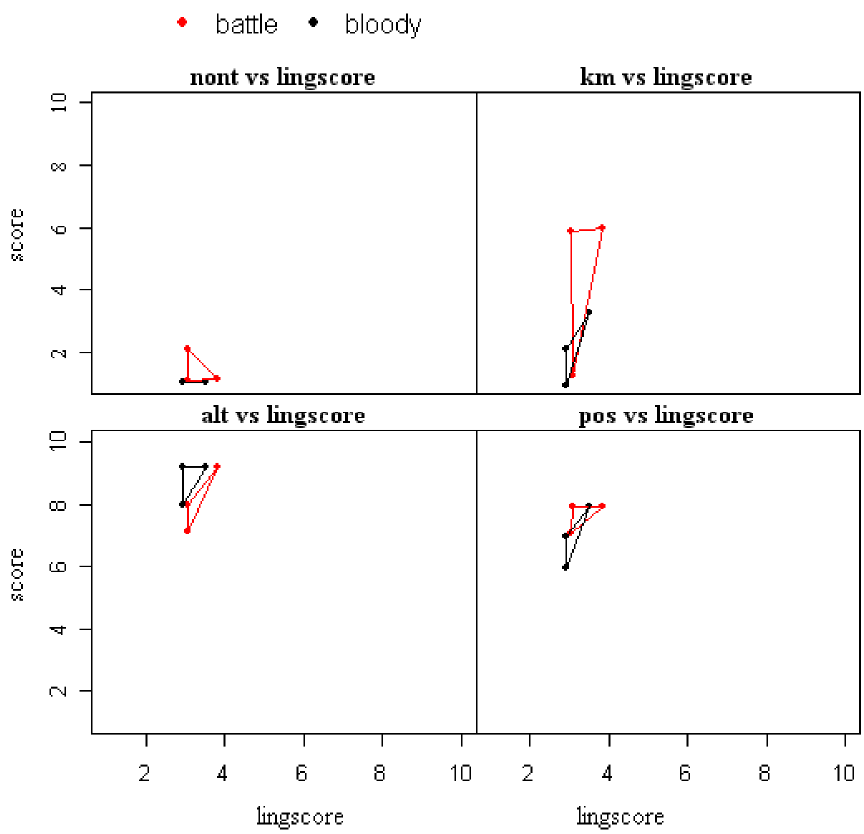
Antiquarians were very fond of positing events as the derivation of various place-names. In fact there are only five names in the database explicitly relating to an event.[41] It may be that other names were so coined after an event, particularly watercourses named after people, could be so called due to some once known event involving the person and the watercourse, but now forgotten. Three of the names relate to battles:

Table 4.33: Hydronyms Meaning 'battle'

id	RN	etymology
519	Battle Burn <	Sc *battle*, 'battle'
1291	Allt Catha	G *cath* 'battle' gen sg or G *càth*, 'seeds'
1815	Allt Chomhraig	G *còmhrag*, 'battle, fight'

It may be that names with 'bloody' represent sites where battles had taken place.

Figure 4.20: Comparison of 'bloody' and 'battle'



[41]This class corresponds to Stewarts class 3: Incident names identified 'by means of some incident which has occurred at or near it': Stewart, 'A Classification of Place Names', p. 4.

4.3.21 Relation to other features

The usage of these directional terms is largely straightforward. Table 4.34 on page 168 shows the two commonest terms being *cùl* and *back* suggest that 'fore' or 'front' is the default term here. In this case, this is expected, a nearby burn could be named anything, but one further from oneself or one's settlement would be worthy of being named such.[42]

Table 4.34: Common Directional Semantemes

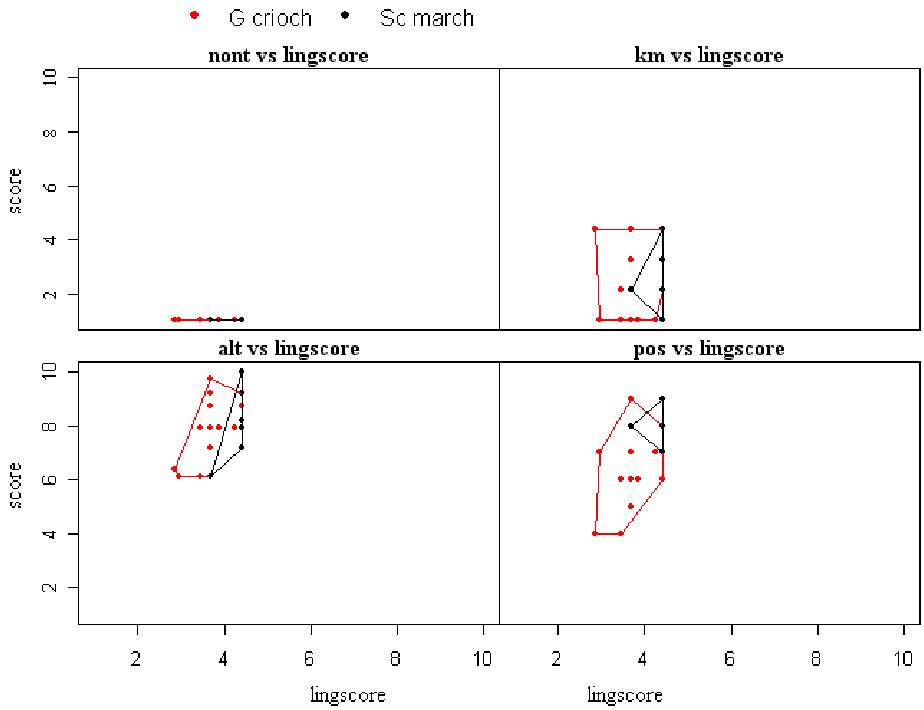
etymology	frequency
Sc *back*, 'back'	36
G *cùl*, 'back'	24
Sc *mid*, 'middle'	17
Sc *west*, 'west'	16
Sc *east*, 'east'	14
Sc *wester*, 'west'	8
Sc *easter*, 'east'	7
G *earas*, 'back part'	5
Sc *fore*, 'fore'	4
Sc *north*, 'north'	4
G *cùlaibh*, 'back parts'	3
G *deas*, 'south'	3
PG *fo*, 'sub'	3
G *cùlaidh*, 'back place'	3
G *suas*, 'upstream'	3
G *bac*, 'hindrance'	2
Sc *bottom*, 'bottom'	2
G *cùileach*, 'angular'	2
Sc *inner*, 'inner'	2
Sc *outer*, 'outer'	2
G *sìos*, 'downstream'	2
Sc *south*, 'south'	2
G *tòn*, 'anus'	2
G *clì*, 'left'	1
Sc *far*, 'far, distant'	1
Sc *inmost*, 'innermost'	1
Sc *foot*, 'foot of a burn'	1
G *màs*, 'bottom'	1
G *meadhan*, 'middle	1
G *meadhanach*, 'intermediate, central'	1
Sc *outmost*, 'outermost'	1
Sc *over*, 'over'	1
Sc *upper*, 'upper'	1
Sc *nether*, 'nether'	1

[42]This class corresponds to Stewart's *Relative Description*: Stewart, 'A Classification of Place Names', p. 3.

4.3.22 Boundary

Names in this section predominantly belong to two terms: G *crìoch*, 'boundary' and Sc *march*, 'boundary'. Figure 4.21 shows that the two names largely inhabit the same conceptual space. One interpretation of this graph is that *march* is a translation of the previous *crìoch* name. Another interpretation is that when Scots became the dominant language in the area, the agricultural landscape essentially remained the same, so the same watercourses still acted as boundaries. The answer undoubtedly lies somewhere between the two of these scenarios. Certainly, there is evidence for interchange between these two terms, as in Allt Chriochaidh (2119) mentioned as *Alt-Chriachie or the marching-burne* (c.1591 Pont Text 137r) and March Burn (747) as *Easter Altcrich, Wester Altcrich or March Burn* (1810 Knox).[43]

Figure 4.21: Comparison of G *crìoch* and Sc *march*



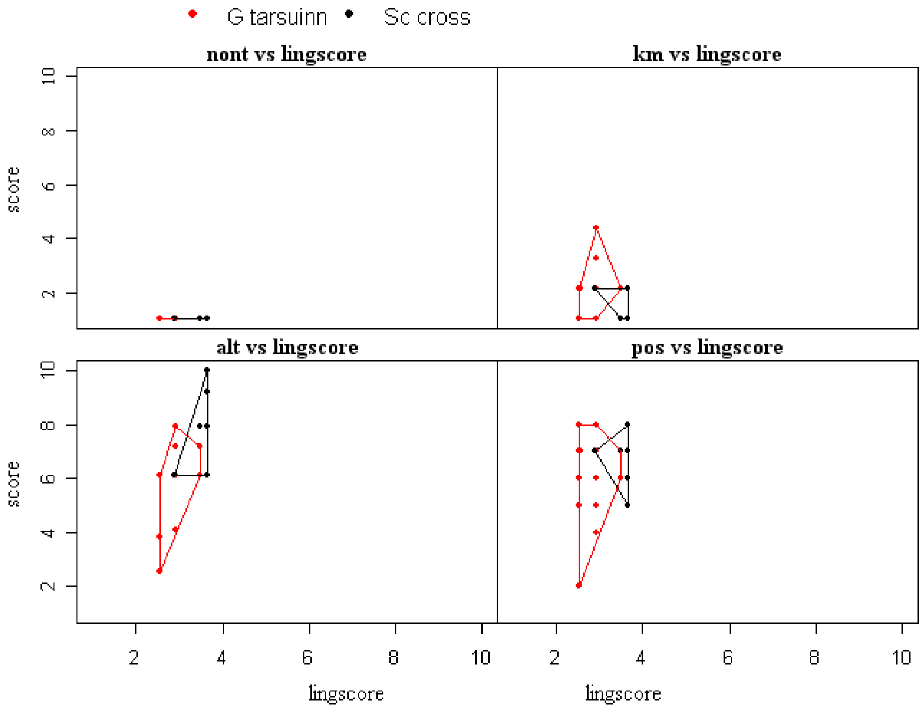
4.3.23 Crossing

Names in this group are coined from their usage as crossing points. The two commonest terms are Sc *cross*, 'cross' and G *tarsainn*, 'cross', both meaning in this context, 'a point between A and B'. Names denoting specific fords are under the section of 'specific man-made areas: riparian area'. Figure 4.22 shows clearly that the two terms are equivalent and take the same space in the hydronymicon allowing

[43]This class corresponds to Stewart's class of *Associative Description* in which he gives Boundary Creek as an example: Stewart, 'A Classification of Place Names', p. 3.

for some locational differences between the Gaelic and Scots strata.

Figure 4.22: Comparison of G *tarsuinn* and Sc *cross*



4.3.24 Water Words

Nicolaisen posited a semantic class called 'water word',[44] essentially representing a class of rivers simply meaning, 'flowing one, or wet one'. A decision has been made not to include this class, for the following reasons:

- The number of watercourses within the AOS which Nicolaisen posits for this group is very small.
- Nicolaisen himself says that these names probably had nuances lost to us now,[45] hence many of these names can be ascribed to the *semtype* 'manner'.
- Several of the etyma for these RNs have been reanalysed along other lines.

4.3.25 RN pairs

When discussing RNs, the most commonly stated fact is that some neighbouring pairs of rivers contain colour terms, and this colour is only used to differentiate one river from another, i.e it does not

[44]Nicolaisen, 'The Semantic Structure of Scottish Hydronymy', p. 232-33.
[45]Ibid.

physically represent the river.[46] This tenet, although broadly correct, only tells a part of the story. RN pairing is notable since it seems to occur at all linguistic strata and with a broad range of syntactical constructions. For the purposes of this discussion, the names have been divided into certain categories.

4.3.25.1 Early Celtic Names with Identifiers from a Later Stratum

These names generally represent RNs of considerable antiquity which have had usually Black and White terms tagged on as identifiers. Such names are: Black and White Cart, Blackadder and Whiteadder, Black Esk and White Esk and Deveron and Findhorn. The last pair are relatively unusual since the identifiers are Gaelic or P-Celtic as opposed to Scots. Other not so obvious names are: Devon and Black Devon. This name is idiosyncratic since there is no evidence to suggest that the Devon was ever called *White Devon, one might reasonably expect to have found this name in old forms, since that area has been relatively well documented since Scots-speaking times.

Another name, perhaps rather of the Findhorn / Deveron type, is the Ugie, which has two main tributaries, the North and South Ugie Water. In the Book of Deer, the South Ugie seems to be referred to as Dubuci, that is Dubh Uige, 'the Black Ugie'.[47] It would seem reasonable to infer that the North Ugie was called *Finnuci, 'the White Ugie' (although see discussion below).

The River South Esk, which is in opposition to River North Esk has at its confluence the name Inveriscandye, and in old forms has the name Escandye W(ater). See 3.3.1 on page 53 . Two points can be made here. Firstly, there are a number of colour pairings where there is only direct evidence for one of the colours. These are Black Devon, Dubuci and Escandye. In all three cases it is the 'black' colour that exists, whilst the 'white' name is absent. This could perhaps be put down to coincidence were it not for the fact that this phenomenon exists in later strata too, where 'black' continues to outnumber occurrences of 'white'. Given the concept of defaults, it could be said that a viable pairing system for names at this stratum is to have one name in the pair denoted the non-default 'black' alongside a specific element, with the other denoted simply with the specific. This would fit the evidence of occurrences of the terms and also would fit the evidence of these earliest names. This is not to say that all pairing operates thus at this stratum: there is of course plenty of evidence for the traditional system such as Black and White Adder.

Secondly, in all the cases mentioned above, the South or Eastern watercourse is the 'black' one, and the North or West name is the 'white' one. The South / North opposition is the predominant one, and the only other one that makes its way into the lexicon of oppositions i.e. in this section no rivers are opposed with the terms East and West. Why this should be is not certain, but it also occurs in other linguistic strata suggesting its cause is something suggested by the environment. Since in Scotland the Sun traverses the sky in the Southern section of the sky, one might have thought that the Southern watercourse would be the 'white' lighter one, with the Northern one conceptually further from the Sun, but the opposite situation occurs.

---

[46] For instance under the entry for the River Adder in Ross, *Scottish Place-names*, p. 5.

[47] Simon Taylor, personal communication and Simon Taylor, *Studies in the Book of Deer* (2008: Forthcoming), chap. Place-names in the Gaelic Notes in the Book of Deer.

4.3.25.2 Transparent Scots or Gaelic names

The opposition of RNs with colour is not as common in the AOS as one might think. There are many RNs with the semanteme 'black' and some, though not as many, meaning 'white'; there is, however, only a handful of examples where they can be definitely shown to be in pairs. The predominant opposition terms used are cardinal points for these two strata. There are three examples of White Burn and Black Burn, and one Black Burn vs White Stripe. As for Gaelic, there is a notable lack of these types of names, barring one dubious instance: Caochan Riabhach and Caochan Bàn. G *fionn*, 'white, fair', is conspicuous by its absence.

This phenomenon probably reflects some conceptual view of the world that once existed. It is known in Indo-European cosmology that North is considered 'left' and South as 'right'; this concept exists, for example, as far afield as India, where a region of India known as the Deccan is the right or South (Deccan is distantly cognate with G *deas*). This is still the case in Celtic languages where in Gaelic there is the following situation according to Dwelly as seen in diagram 4.35:

Table 4.35: Relationship between Cardinal Points and Relative Directions

	tuath 'left' North	
iar 'end' West	⇒	*ear* 'front' East
	deas 'right' South	

Since 'white' and 'North' are equivalent as are 'black' and 'South' one might posit that left may be equivalent to white, and black with right. Some evidence for this is also in Dwelly where G *bán*, 'white' also means left hand side of ploughed land in contrast to G *dearg*, 'dark red' which means the right hand side. What is interesting here is that although *bàn* and *dearg* mean light and dark, these terms are not used in opposition in toponymy, suggesting that the relationship between light and dark on the one hand and north and south on the other comes from outside the toponymicon.

It would seem then, that in some cases, the condition for colour pairing is dependent on cardinal direction. It would be interesting to investigate whether colour names not in pairs also represent cardinals, i.e. Was a simple 'Black Burn', not in a pair, coined because of its relation to other features?

This information is somewhat at odds with Indo-European symbolism, whilst facing East is the conceptual default (for instance people were buried along an East-West axis, with the head pointing Eastward). North and West are associated with death, and South and East with life, and left with 'wrong' and right with 'right'. If this interacts with our system, it would imply that black is associated with life and white with death, a counter-intuitive identification in modern western culture, but in the Orient and at times in Mediaeval Europe, white was in fact associated with death and mourning.

4.4 Semantics vs Components of Geogscore

Each of the graphs in this section plots a particular component of *geogscore* grouped by each *semtype*, a jitter has been added so that individual points can be discerned. This is the same type of graph as in figure 4.3 on page 131, but instead of using *semtypescore* as the y-axis, the components of *geogscore* have been used. A number of points are noticeable when the data are presented in this way.

As before, *alt* and *pos* concur with each other, as do *km* and *nont*. The following semantemes have a wide *alt / pos* range: Colour, Manner, Effect / Character, Flora, Fauna, Specific Natural Feature. Within the *km / nont* ranges, the hydronyms coined from external features, Manner and Sound have a wide range. A number of outliers can be observed; these have been discussed above. In the *km / nont* graphs these are:

1. The Dee and Don in the Supernatural Entity axis.

2. The Water of May in the 'boundary' axis.

3. The River Tummel in the 'colour' axis.

4. Bodería in the 'effect / character' axis

In the *alt / pos* these are:

1. The outlier for 'flora' is Allt an Dà Chraobh Bheath, that is: 'Burn of the Two Birch Trees'. This is at a high altitude on the slopes of Lochnagar and it sounds as if the existence of two birch trees was remarkable enough to be commented upon.

2. The outlier for 'age' in the *pos* graph is Allt Ogline, treated in this case as from *ògail*, 'young, youthful' + masculine diminutive *an*, 'little young one'. This name may be better interpreted as 'young person' and thus be moved to the 'specific person' category where it would not be an outlier. Alternatively, the name may relate to names of the Ogle / Ochil type, ultimately from OC **uxellos*, 'high'.[48]

It is interesting to note that the outliers in the *km / nont* group are all higher than the average, whilst the outliers in the *alt / pos* are all lower than the average.

[48] Both these possibilities are derived from Angus Watson, 'Place-Names Land and Lordship in the Medieval Earldom of Strathearn', Ph. D thesis, St Andrews (2002), p. 28.

Figure 4.23: *Semtype* by *altscore*

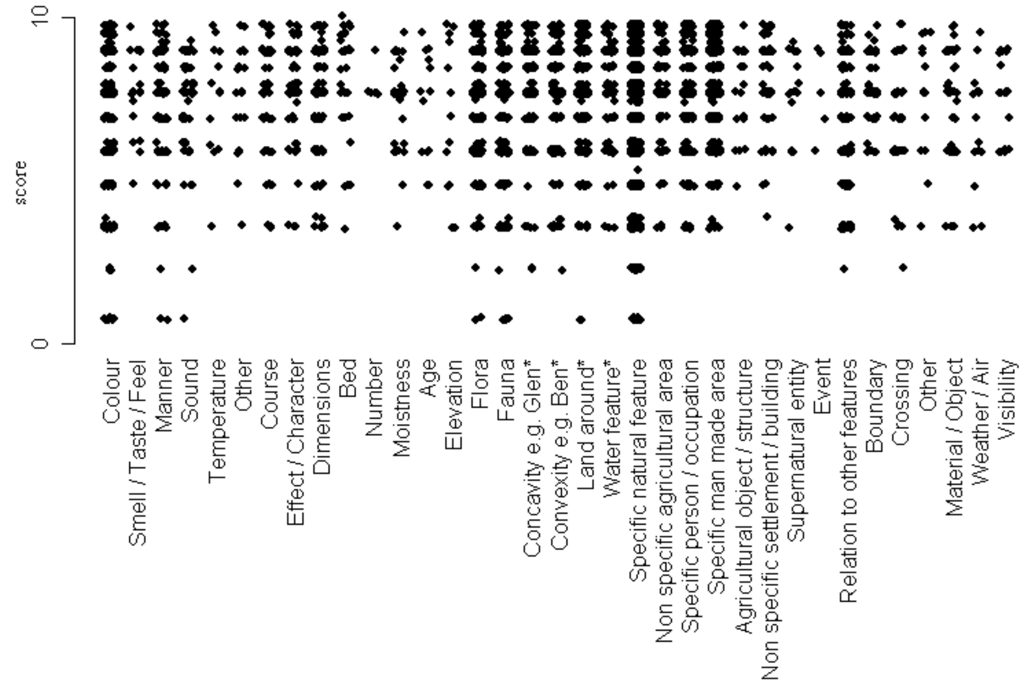


Figure 4.24: *Semtype* by *posscore*

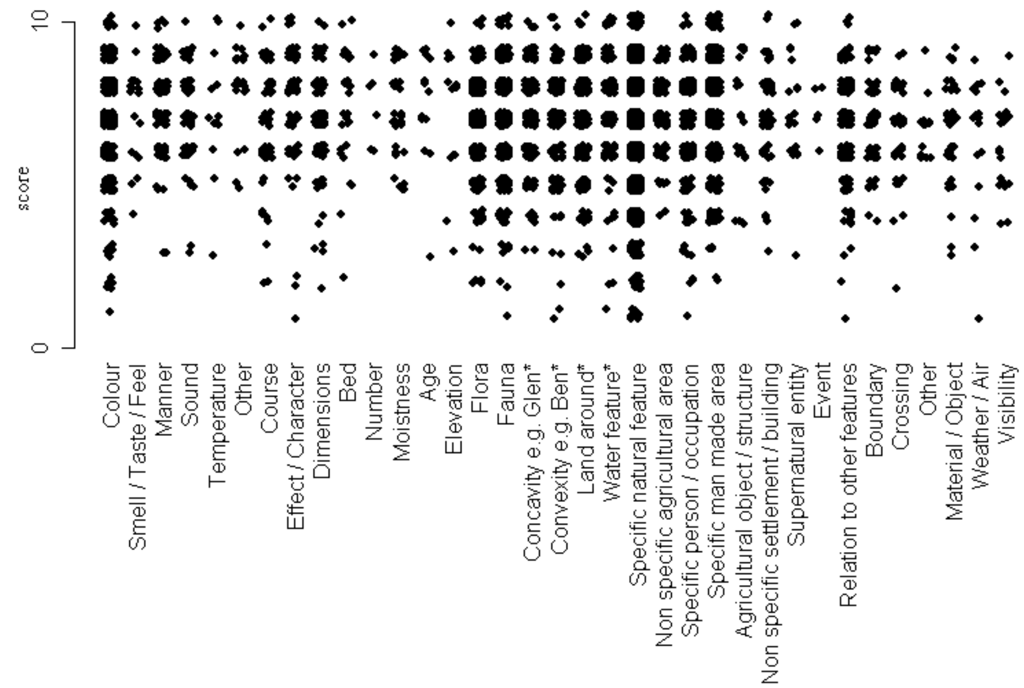


Figure 4.25: *Semtype* by *nontscore*

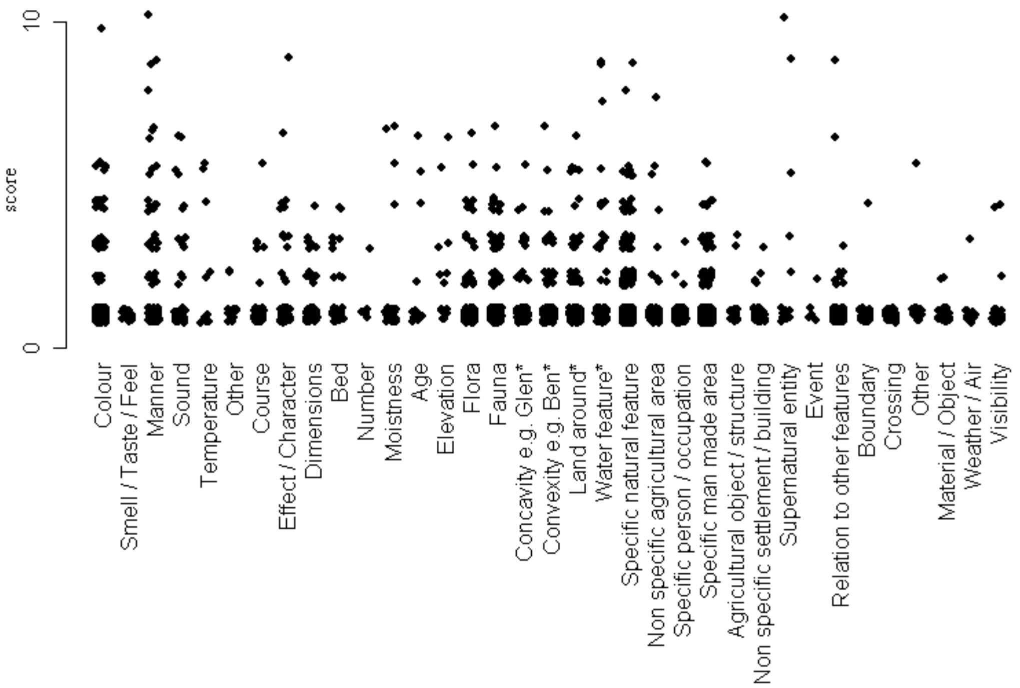
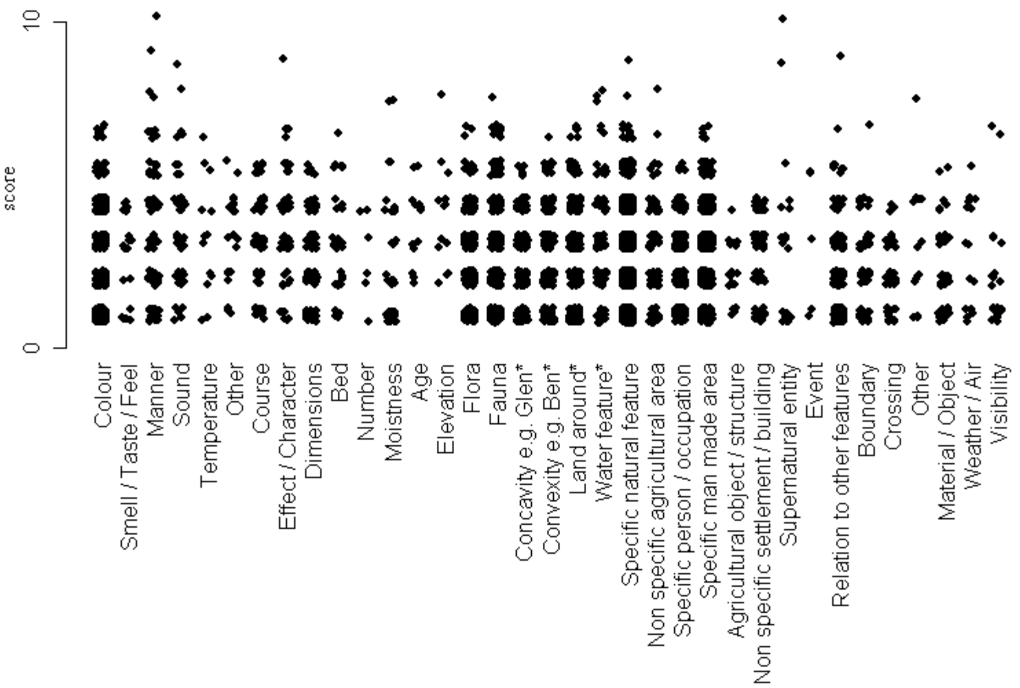


Figure 4.26: *Semtype* by *kmscore*



4.5 Semtype and the Statistical Method

There are two main approaches to the statistical method outlined in the second chapter. The first is *semtypefinder* which works in a similar way to *genelemfinder*. The second was described under the section for Frandy Burn in section 2.6.1.1 on page 34.

4.5.1 Semtypefinder

The equivalent of *genelemfinder* as explained on page 112 can be used as *semtypefinder*. This process involves getting each component of *geogscore* for a specific RN and discovering which *semtypescore* ranges it fits. Unfortunately it does not work very well for the majority of RNs. This is because out of the thirty-five classes of *semtype*, the majority of entries in the database will lie between the majority of these names which are not at the extremities of the data ranges. Moreover, the function only works in cases where only the *semtype* is not known. This means that if the *specelem* (or linguistic stratum) is not known either, the function will not work.

Here are a few examples:

Table 4.36: Possible derivations for the River Tay

name	minkm	maxkm	minnont	maxnont	minalt	maxalt	minpos	maxpos
Character	1	8.88	1	8.88	3.83	10	1	10
Manner	1	10	1	10	1	10	3	10

Table 4.36 shows that the only possible meaning for the River Tay, if it were not to alter the *semtypescore*, are 'manner' or 'effect / character'. This is hardly surprising, since this is the class into which the traditional 'water-words' have been placed, and is also the class where names such as the Spey, the Earn and the Almond belong. This echoes a sentiment found amongst place-name scholars, that although the meaning of Tay may not be known, that there is nevertheless a restricted set of meanings it could have, e.g Tay could relate to the way in which its water flows. To say, for instance, that the Tay was coined from a nearby field, would 'feel wrong'. With *semtypefinder*, it can be suggested that this is indeed the case.

The tool used above ignores the linguistic stratum. This is not hugely problematic, since the semantic content of a name is more likely to be influenced by the physical qualities of a watercourse rather than the linguistic stratum from which the name derives. In some cases, however, one may wish to look only at the semantic classes derived from the specific linguistic stratum from which the name derives. Needless to say, this will only work with names where the meaning is not known, but the stratum is: e.g. with a Gaelic name where the Gaelic specific is an obscure term.

An example here is Allt Ainndeir, a Gaelic name of uncertain meaning. Running normal *semtypefinder* on table 4.38 on page 178. This table gives twenty-four options. Running the same query, but only counting Gaelic RNs table 4.37 on page 177 is produced with fourteen suggestions. This shows a refining of the semantic classes. Since the meaning of the name is obscure, one cannot take the methodology any further, but someone researching the name could use the refined list as a guide to the sorts of semantemes that the name is likely to have.

Table 4.37: *Semtypefinder* for Allt Ainndir with only Gaelic Names

name	minkm	maxkm	minnont	maxnont	minalt	maxalt	minpos	maxpos
Crossing	1	4.38	1	1	2.54	10	2	9
Relation to other features	1	5.5	1	6.63	3.83	10	1	10
Weather / Air	1	5.5	1	3.25	3.83	9.74	1	9
Material / Object	1	5.5	1	2.13	3.83	10	3	9
Agricultural object / structure	1	4.38	1	3.25	5.11	10	4	9
Colour	1	6.63	1	10	1	10	1	10
Flora	1	6.63	1	6.63	1	10	2	10
Specific natural feature	1	8.88	1	8.88	1	10	1	10
Specific person / occupation	1	5.5	1	3.25	3.83	10	1	10
Fauna	1	7.75	1	6.63	1	10	1	10
Concavity e.g. Glen	1	5.5	1	4.38	2.54	10	3	10
Dimensions	1	5.5	1	3.25	3.83	10	2	10
Course	1	5.5	1	5.5	3.83	10	2	10
Moistness	1	5.5	1	6.63	3.83	9.74	5	9
Convexity e.g. Ben	1	6.63	1	6.63	2.54	10	1	10
Other	1	7.75	1	5.5	3.83	9.74	6	9
Non specific settlement / building	1	4.38	1	3.25	4.09	10	3	10
Land around	1	6.63	1	6.63	1	10	3	10
Non specific agricultural area	1	7.75	1	7.75	3.83	10	4	9
Other	1	5.5	1	2.13	3.83	10	5	10
Specific man-made area	1	6.63	1	4.38	3.83	10	2	10
Water feature	1	7.75	1	8.88	3.83	10	1	10
Bed	1	6.63	1	4.38	3.83	10	2	10
Sound	1	8.88	1	6.63	1	9.49	3	10
Effect / Character	1	8.88	1	8.88	3.83	10	1	10
Temperature	1	6.63	1	5.5	3.83	10	3	10
Elevation	2.13	7.75	1	6.63	3.83	10	3	10
Manner	1	10	1	10	1	10	3	10
Age	2.13	5.5	1	6.63	5.11	10	3	9
Supernatural entity	1	10	1	10	3.83	9.23	6	10

Table 4.38: *Semtypefinder* for Allt Ainndir

name	minkm	avgkm	maxkm	minnont	avgnont	maxnont	minalt	avgalt	maxalt	minpos	avgpos	maxpos
Relation to other features	1	2.17	5.5	1	1.09	6.63	3.83	7.01	10	1	6.75	10
Visibility	1	2.08	6.63	1	1.34	4.38	6.14	7.62	10	4	6.26	9
Colour	1	2.65	6.63	1	1.26	10	1	7.23	10	1	6.85	10
Flora	1	2.68	6.63	1	1.17	6.63	1	7.41	10	2	6.8	10
Specific natural feature	1	2.74	8.88	1	1.12	8.88	1	7.44	10	1	6.85	10
Fauna	1	2.76	7.75	1	1.28	6.63	1	7.3	10	1	6.84	10
Concavity e.g. Glen	1	2.5	5.5	1	1.18	4.38	2.54	7.92	10	3	6.97	10
Moistness	1	2.58	5.5	1	1.34	6.63	3.83	7.48	9.74	5	7.18	9
Course	1	2.93	5.5	1	1.25	5.5	3.83	7.68	10	2	6.75	10
Boundary	1	2.56	6.63	1	1.07	4.38	6.14	8.06	10	4	6.96	9
Convexity e.g. Ben	1	2.81	6.63	1	1.22	6.63	2.54	7.7	10	1	6.93	10
Other	1	3.17	7.75	1	1.32	5.5	3.83	7.61	9.74	6	6.71	9
Non spec. agric. area	1	2.58	7.75	1	1.26	7.75	3.83	8.06	10	4	7.19	9
Land around	1	2.9	6.63	1	1.22	6.63	1	7.94	10	3	7.11	10
Water feature	1	3.01	7.75	1	1.53	8.88	3.83	8.21	10	1	7.32	10
Specific man-made area	1	2.96	6.63	1	1.12	4.38	3.83	8.58	10	2	7.43	10
Bed	1	3.25	6.63	1	1.56	4.38	3.83	8.06	10	2	7.28	10
Sound	1	3.66	8.88	1	1.75	6.63	1	7.66	9.49	3	7.13	10
Effect / Character	1	3.47	8.88	1	1.51	8.88	3.83	8.05	10	1	7.25	10
Temperature	1	3.43	6.63	1	2.13	5.5	3.83	7.74	10	3	7.23	10
Elevation	2.13	4.03	7.75	1	2.39	6.63	3.83	7.19	10	3	7.23	10
Manner	1	3.46	10	1	2.22	10	1	7.79	10	3	7.59	10
Age	2.13	3.82	5.5	1	2.22	6.63	5.11	7.72	10	3	7.42	9
Supernatural entity	1	3.32	10	1	2.41	10	3.83	8.02	9.23	6	7.69	10
Colour	1	2.3	6.63	1	1.11	5.5	1	6.68	10	1	6.54	9
Flora	1	2.77	6.63	1	1.2	6.63	1	7.17	10	2	6.64	10
Fauna	1	2.78	7.75	1	1.3	5.5	1	7.06	10	1	6.71	10
Specific nat. feature	1	2.76	6.63	1	1.11	5.5	1	7.24	10	1	6.76	10
Concavity e.g. Glen	1	2.51	5.5	1	1.26	4.38	2.54	7.42	10	3	6.79	9
Convexity e.g. Ben	1	2.85	6.63	1	1.23	6.63	2.54	7.62	10	1	6.89	10
Course	1	3.02	5.5	1	1.25	5.5	3.83	7.73	10	2	6.77	10
Water feature	1	3.01	5.5	1	1.29	4.38	3.83	7.71	10	1	6.87	10
Supernatural entity	1	2.5	5.5	1	1.75	5.5	3.83	7.47	9.23	6	7.17	8
Non spec. agric. area	1	2.71	7.75	1	1.47	7.75	3.83	7.78	10	4	7	9
Land around	1	2.96	6.63	1	1.26	6.63	1	7.72	10	3	7.09	10
Manner	1	2.95	7.75	1	1.71	8.88	1	7.48	10	3	7.25	10
Sound	1	3.55	7.75	1	1.58	6.63	1	7.37	9.23	3	6.98	9
Spec. man-made area	1	2.98	6.63	1	1.12	4.38	3.83	8.38	10	2	7.27	10
Visibility	1	3.25	6.63	1	2.13	4.38	6.14	7.91	10	5	6.57	9

As an example, where the *semtype* is actually known, I have chosen Caochan na Gaibhre (3277), from G *gobhar*, 'goat', which is in the semantic class, 'Fauna' as in table 4.40.

Table 4.40: *Semtypefinder* for Caochan na Gaibhre

Colour
Smell / Taste / Feel
Manner
Sound
Other
Course
Effect / Character
Dimensions
Number
Flora
Fauna
Concavity e.g. Glen
Convexity e.g. Ben
Land around
Water feature
Specific natural feature
Non specific agricultural area
Non specific settlement / building
Specific man-made area
Agricultural object / structure
Person / Occupation
Supernatural entity
Event
Relation to other features
Boundary
Crossing
Weather / Air
Visibility

This shows that this watercourse is suitable for thirty of the thirty-five semantic classes. This would not seem to be very useful, but at least shows that Fauna is a viable semanteme for this watercourse under the methodology.

4.5.2 Semtypefinder2

Another tool can be used which was outlined in a simplified form under Frandy Burn in section 2.6.1.1 on page 34. Rather than looking at the individual components of *geogscore* this system looks only at other entries in the database with similar *geogscores* to a given name, and counts the amount of semantic classes possessed by this subset. As an example Allt Ainndir may be used again as in table 4.41.

Again, this list can be used as a guide alongside table 4.37 on page 177. Caution should be

Table 4.41: *Semtypefinder2* on Allt Ainndir

meaning	frequency	percentage
Specific natural feature	23	34.85%
Specific man-made area	16	24.24%
Land around	4	6.06%
Flora	3	4.55%
Convexity e.g. Ben	2	3.03%
Concavity e.g. Glen	2	3.03%
Manner	2	3.03%
Water feature	2	3.03%
Specific person / occupation	2	3.03%
Non specific agricultural area	1	1.52%
Non specific settlement / building	1	1.52%
Dimensions	1	1.52%
Colour	1	1.52%
Age	1	1.52%
Fauna	1	1.52%
Material / Object	1	1.52%
Other	1	1.52%
Bed	1	1.52%
Effect / Character	1	1.52%

exercised here because the two most likely outcomes in the table are ‘specific natural feature’ and ‘specific man-made area’. These are some of the commonest names in any case and as such will always appear near the top of any list.

Two further refinements can be made with this function. Firstly, it was mentioned above that this system gathers *geogscores similar* to the *geogscore* of the queried RN. It is possible to widen the range if the results are too few (or none at all). This is especially useful for the larger watercourses, where fewer names exist in the area. An example of this is *Toúaisis*, the old name for the Spey. Running the normal function produces table 4.42.

Table 4.42: *Semtypefinder2* on Toúaisis

meaning	frequency	percentage
Manner	1	100%

One would expect this, since the geographical qualities of the Spey and Toúaisis are identical of course. In this sense, this result is misleading. Therefore if the tolerance of the *geogscore* were increased, to search for anything $\pm$ 2 of the *geogscore* of Toúaisis, the results are as in table 4.43 which shows the common semantic types for watercourses of this size or RNs at this stratum. The Dee and Don are referred to in the second line.

Table 4.43: *Semtypefinder2* on Toúaisis with a tolerance of ± 2

meaning	frequency	percentage
Manner	3	30%
Supernatural entity	2	20%
Sound	1	10%
Specific natural feature	1	10%
Water feature	1	10%
Land around	1	10%
Relation to other features	1	10%

4.6 Semtype and the Hierarchical Network

Table 4.44 shows the number of occurrences of the various combinations of semantic classes within the network (ignoring cases where the main and tributary are the same). For instance the top entry represents six hundred and fifty occurrences of a watercourse with an adjectival name having a watercourse flowing into it which has a name relating to topography. This is on an analogy with section 3.11 on page 117 in the chapter on generic elements. A few points are worthy of mention concerning table 4.44:

1. The top entry is expected, since nearly all the large rivers contain an adjectival term, and many of the smaller rivers have names related to topography. (For example Touch Burn from G *tulach*, 'hill' flowing into the River Forth, an adjectival name discussed on page 40).

2. The second entry represents the Dee and Don, since they are in the 'Human' category, as discussed on page 165.

3. The third and forth entries both show names in the 'Human' category are common names for tributaries. This relates to the low *posscore* of many of these names; the same can be seen for 'Situation' lower down the list.

This table corroborates broadly the concept of semantic distance. If one puts this into a hierarchical order by looking only at the commonest combination of any given two types (e.g. counting only Adjective : Topography and not *vice versa*, on the grounds that the former occurs 650 times and the latter only 290) and then placing them in the order of decreasing occurrence, the following order is shown, which is clearly a reflection of the semantic distance spectrum as explained in section 4.2.2 on page 128. The implications of this are discussed in more depth in the conclusion.

$$\text{Adjective} < \text{Ecosystem} < \text{Topography} < \text{Human} < \text{Situation}$$

There are many cases in Scotland where a single watercourse has different names for its upper and lower sections. It has been observed that in this particular region of the network one of the names relates to a settlement or natural feature of some sort. Table 4.45 shows the occurrences of this pattern. It is similar to the one above, except that names relating to topography and human activity feature more strongly. Only these two semantic classes are present in the top five most popular combinations. This essentially shows that in the majority of cases, where a section of a watercourse has been renamed, it has generally been renamed after some external feature, such as a settlement or natural feature. In a few cases there is evidence that a watercourse has been renamed to an adjectival structure. This suggests adjectival terms originate from a slightly earlier stratum within the Scots and Gaelic nomenclature than names for external features, or at least that adjectival naming styles may have ceased to be productive after a point in time.

Table 4.44: Number of Occurrences of the Various Combinations of Semantic Classes

main	tributary	frequency
Adjective	Topography	650
Human	Topography	436
Topography	Human	432
Adjective	Human	309
Topography	Adjective	290
Ecosystem	Topography	249
Topography	Ecosystem	225
Adjective	Ecosystem	155
Human	Adjective	138
Topography	Situation	102
Ecosystem	Human	91
Human	Ecosystem	83
Ecosystem	Adjective	81
Adjective	Situation	46
Situation	Topography	40
Ecosystem	Situation	35
Human	Situation	34
Situation	Human	19
Situation	Adjective	11
Situation	Ecosystem	7

Table 4.45: Number of Occurrences of the Various Combinations of Semantic Classes where one Watercourse has a Different Name for its Different Sections

main	tributary	frequency
Topography	Topography	57
Human	Topography	35
Human	Human	31
Topography	Human	30
Adjective	Topography	18
Adjective	Adjective	15
Adjective	Human	10
Topography	Ecosystem	10
Ecosystem	Topography	10
Topography	Adjective	9
Ecosystem	Human	9
Human	Adjective	8
Human	Ecosystem	6
Adjective	Ecosystem	6
Situation	Topography	5
Topography	Situation	4
Ecosystem	Adjective	4
Adjective	Situation	2
Ecosystem	Ecosystem	2
Human	Situation	2
Ecosystem	Situation	1
Situation	Ecosystem	1
Situation	Adjective	1
Situation	Situation	1
Situation	Human	1

4.7 Conclusion

In this chapter, I have attempted to demonstrate three things:

1. To show a framework for the construction of the concept of semantic distance.

2. To show how tools can be used to suggest meanings for terms in obscure hydronyms.
3. To discover what terms are considered default for hydronyms, and what qualities these defaults and non-defaults have.

4.7.1 Semantic Distance

It would be a mistake to create too rigid a hierarchy of distance. Different interpretations and different methods give varying results. Rather, a general set of principles can be gleaned from the preceding discussion. A number of trends can nonetheless be discerned:

- The original hierarchy mentioned in the introduction for this section above reflects a possible range. That is, if 'adjective' reflects the closest relationship between specific element and watercourse, then this reflects the widest score range (i.e. RNs with adjectives can exist for watercourses of all sizes). At the other end of the spectrum, watercourses named after features somewhat removed, reflect a more restricted range (i.e. RNs coined after, for instance, a crossing, can exist only for a relatively specific range of watercourses).
- Within this range, the difference between the maximum and minimum varies greatly. Whilst adjectival RNs have a wide range, the other end of the spectrum shows a more homogeneous picture, with names coined from situation or topography being somewhat similar in range. It may be suggested that for this range, the difference between the maximum and minimum is how the notion of semantic distance should be judged.

4.7.2 Tools

The tools *semtypefinder* and *semtypefinder2* as described in section 4.5 starting on page 176 are not as useful as the equivalent tools for other linguistic components, but they nevertheless show correlations. In no cases was it found that the tools gave inaccurate or improbable results.

4.7.3 Defaults

It appears that the default qualities of a watercourse are as follows: 'wet', 'quiet', 'light-coloured', 'straight', 'young', 'calm', 'shallow', 'good', 'small' and 'front'. Defaults do not exist for 'temperature', 'smell' or 'taste'. This concept of defaults is discussed at greater length in the conclusion.

Chapter 5

Linguistic Strata

5.1 Introduction

In this chapter, the principles which influence the attribution of a watercourse to a specific linguistic stratum will be investigated. This relationship between a linguistic stratum and a particular name has always been perceived as an important area of research in toponymy. Scholars have traditionally preferred to be able to say a name is 'Gaelic' or 'Pictish' as a starting point for discussion. Within the approach of this thesis, I have attempted to move away somewhat from this position, treating the notion of 'linguistic stratum' as one of a number of linguistic qualities. In the preceding two chapters, linguistic stratum has largely been ignored, in favour of individual terms (such as *burn* or *allt*, or *rough* and *garbh*). In some countries, the boundaries between linguistic strata are clearly delineated. Studying American hydronymy for instance, would show a clear division between European languages and Native American languages. In Scotland however, as explained in the next section, the situation is more blurred.

Another more significant issue also arises with the concept of 'linguistic stratum'. It is perhaps easy enough to divide Scotland's linguistic strata into Gaelic, P-Celtic and so on, but how should these be applied to an actual name? As mentioned before, the stratum is considered by looking at only the original form of the specific element. Aside from the fact that this often cannot be known, many other factors need to be taken into account, such as the stratum, identity and position of the generic element, the surface phonology of the element, the stratum of any external feature denoted where relevant, and even the location of the watercourse itself.

5.1.1 Definitions of Linguistic Strata

Old or Early Celtic This stratum represents two types of names: Firstly it represents RNs which are of sufficient antiquity that a differentiation between P- and Q-Celtic cannot be made. Secondly, where the name is obscure, and clearly ancient, it is considered Old or Early Celtic. Nicolaisen's Old European hydronyms are also in this section; this is mainly because there are too few names which Nicolaisen considers pre-Celtic to form a coherent class of their own, and also because the whole theory of 'Old European' as a theory has recently come into question, and this thesis is not the place to further this debate.[1] In terms of chronology, these names relate to names of unknown, but considerable antiquity (first recorded by Ptolemy in c. 150 AD) up to around the end of the Roman occupation of the British Isles.[2]

P-Celtic This stratum generally represents pre-Gaelic Pictish names, although in some cases, especially in Clackmannanshire and around Loch Lomond, a different P-Celtic dialect / language other than Pictish may have been spoken. Thus the term P-Celtic is used. The Picts as an identifiable people seem to have come into being roughly around 300 AD[3] and seem to have been subsumed into Gaelic culture over a period ending roughly in 900.[4] These are necessarily approximate dates, but it would not seem unreasonable to say that the majority of names attributed to this stratum in the database were coined during this time.[5]

Gaelic This stratum specifically represents Scottish Gaelic. Under the traditional paradigm, Dalriadic Gaelic speakers encroached from Ireland from around the fifth and sixth centuries.[6] Some more recent work, however, has argued that parts of Argyll have been Gaelic speaking at least since pre-Roman times, and that no mass migration from Ireland to Scotland took place in the Dark Ages.[7] Gaelic's decline began around 1100,[8] and has continued to the present day. The earliest coined Gaelic RNs appearing in this database were probably coined somewhat after the sixth century, since the AOs does not include Argyleshire, the area of earliest Gaelic settlement. Naming probably ceased to be productive around the time of the clearances and other upheavals in Scotland in the eighteenth and nineteenth centuries.[9]

---

[1]The only work known to the author to directly address this issue is Eric Hamp, 'Varia', *Scottish Gaelic Studies* XVI (1990), whilst a number of other more recent works implicitly reject the 'Old European' theory, such as G. R. Isaac, 'Place-Names in Ptolemy's Geography', CD-ROM (2004) and G. R. Isaac, 'The Antonine Itinerary Land Routes: Place-Names of Ancient Europe and Asia Minor. An electronic database with etymological analysis of the Celtic name-elements', CD-ROM (2002)

[2]The following works are pertinent to the study of this stratum: Richard Coates and Andrew Breeze, *Celtic Voices English Places* (Stamford, 2000); Eilert Ekwall, *English River-Names* (Oxford, 1928); T. V. Gamkrelidze and V. V. Ivanov, *Trends in Linguistics: Indo-European and the Indo-Europeans* (New York, 1995); H. Krahe, *Unsere ältesten Flussnamen* (Wiesbaden, 1964); Pokorny, *Indogermanisches Etymologisches Wörterbuch* (Munich, 1959); A. L. F. Rivet and C. Smith, *The Place-Names of Roman Britain* (London, 1979).

[3]E. Sutherland, *In Search of The Picts* (Britain, 2000), p. 40.

[4]Ibid., p. 238.

[5]The following works are pertinent to the study of this stratum: Simon Taylor, *Studies in the Book of Deer* (2008: Forthcoming), chap. Place-names in the Gaelic Notes in the Book of Deer; E. Sutherland, *In Search of The Picts* (Britain, 2000); E. H. Nicoll, ed., *Pictish Panorama: The Story of The Picts And a Pictish Bibliography* (Balgavies, 1995); David Henry, ed., *The Worm, The Germ And The Thorn: Pictish And Related Studies Presented To Isabel Henderson* (Balgavies, 1997); W. A. Cummins, *The Age of The Picts* (Frome, 1995); W. A. Cummins, *The Lost Language of the Picts* (Trowbridge, 2001); W. F. Skene, *Chronicles of The Picts, Chronicles of the Scots* (Edinburgh, 1867); A. Small, ed., *The Picts: A New Look at Old Problems* (Dundee, 1987); F. T. Wainwright, ed., *The Problem of The Picts* (Edinburgh, 1955).

[6]MacNeill and MacQueen, *Atlas of Scottish History to 1707*, p. 58.

[7]See Ewan Campbell, 'Were the Scots Irish?', *Antiquity* 75 (2001) and Prof. David Dumville, *Rannsachadh na Gàidhlig* (Aberdeen, 2002), chap. Ireland and North Britain in the Earlier Middle Ages: Contexts for the Míniugud Senchasa Fher nAlban for more discussion of this paradigm.

[8]MacNeill and MacQueen, *Atlas of Scottish History to 1707*, p. 426.

[9]The following works are pertinent to the study of this stratum: E. Dwelly, *The Illustrated Gaelic-English Dictionary* (Edinburgh, 1901-11); William J. Watson, *The Celtic Place-Names of Scotland* (Edinburgh, 1993); William J. Watson,

Scots This stratum encompasses Scots, SSE and English itself. Scots broke off from its Anglian neighbours and encroached into southern Scotland, where Gaelic and / or a P-Celtic languages where presumably spoken. As a distinctive language it dates to around the seventh century.[10] By the fourteenth century it was the dominant tongue in the lowlands. Whilst it survives to the present day, it has to a certain extent been subsumed by both English and SSE and is now largely considered a dialect. It seems reasonable that names from this stratum were coined from the earliest stages of Scots settlement up to the nineteenth century as was also the case with Gaelic.[11]

Latin No RNs within the AOS can be derived from Latin. In certain cases, a name such as Burn of King Edward has an old form such as: *aqua regia* lie *King's water* (1668 Retours (Aberdeen) no. 125); this is clearly simply a translation as opposed to a *bone fide* Latin name. The stratum is useful however, for tagging generic elements such as Latin *aqua*, 'water' and *torrens*, 'torrent' as they appear in old manuscripts and maps. A possible field of study would be to compare the various Latin generics to answer questions such as: What size does a watercourse have to be to be considered *aqua* in the Retours? A question such as this is only of peripheral value to the thesis however, and should be considered carefully in the context of the various manuscripts, something which goes outside the bounds of this thesis.[12]

Two other intermediate strata have been used in this thesis, 'Old Celtic / P-Celtic' and 'P-Celtic / Gaelic'. These are an attempt to avoid over-simplification of the strata. Within the score system methodology discussed in chapter two, it occasionally became apparent that some names would not fit neatly into either category. An example of such a name would be the P-Celtic / Gaelic name Burn of Brown, which can be derived from a P-Celtic root **brutona*, 'boiling', but can also be more immediately derived from G *bruthainn*, 'heat'. It could be said that a 'Scots / Gaelic' stratum should have been used, but there are very few names which would fit into this category, mainly only names with similar terms, such as *croft* and *croit* or *miln* and *muileann* etc. Moreover, since these are the only two linguistic strata which are in any way productive, in the vast majority of cases it is clear whether some element derives from Gaelic or Scots.

Scottish Place-Name Papers (Edinburgh, 2002); William J. Watson, *Place-Names of Ross And Cromarty* (Edinburgh, 1904); Roy Wentworth, *Gaelic place-names of Beinn Eighe National Nature Reserve* (Perth, 1999); C. W. J. Withers, *Gaelic In Scotland 1698-1981* (Edinburgh, 1984); E. Quin, ed., *Dictionary of The Irish Language* (Dublin, 1913-79); Alexander MacBain, *An Etymological Dictionary of the Gaelic Language* (Inverness, 1896); M. Maclennan, *Gaelic Dictionary* (Aberdeen, 1979); Charles Ferguson, 'Gaelic Names of Birds', *Transactions of the Gaelic Society of Inverness* 12 (1885-1886); T. S. Ó Máille, 'Irish Place-Names in -as, -es, -is, -os, -us', *Ainm* 4 (1989-1990); T. S. Ó Máille, 'Place-Name Elements -ar', *Ainm* 2 (1987); Gregory Toner, 'The Backward Nook: *Cúil* and *Cúl* in Irish Place-Names', *Ainm* 7 (1996); D. S. Wodtko, *Sekundäradjektive in den altirischen Glossen* (Innsbruck, 1995).

[10] Robinson, *The Concise Scots Dictionary*, p. ix.

[11] The following works are pertinent to the study of this stratum: M. Robinson, *The Concise Scots Dictionary* (Aberdeen, 1987); 'Dictionary of The Scots Language' ⟨URL: `www.dsl.ac.uk`⟩; J. R. Clark Hall, *A Concise Anglo-Saxon Dictionary* (London, 1984); C. Kenneth, *English Place-Names* (London, 1961); P. H. Reany, *English Place Names* (London, 1977).

[12] The following work is pertinent to the study of this stratum: C. T. Lewis, *Elementary Latin Dictionary* (Oxford, 1956).

5.2 Specelem and Components of Geogscore

In each of these sections two graphs are used, which are available as R functions as *specelem1* and *specelem2*. These are explained in more detail in the following section.

5.2.1 Km

The set of graphs in figure 5.1 plots the percentage occurrence of each *kmscore* for each linguistic stratum. The peak of each graph shows what the majority of watercourses have as their length. The further right in the graph the peak is, the longer the majority of watercourses are. One can see that the older the stratum the longer the watercourses are.

Figure 5.1: Occurrence of *kmscore* for each Stratum

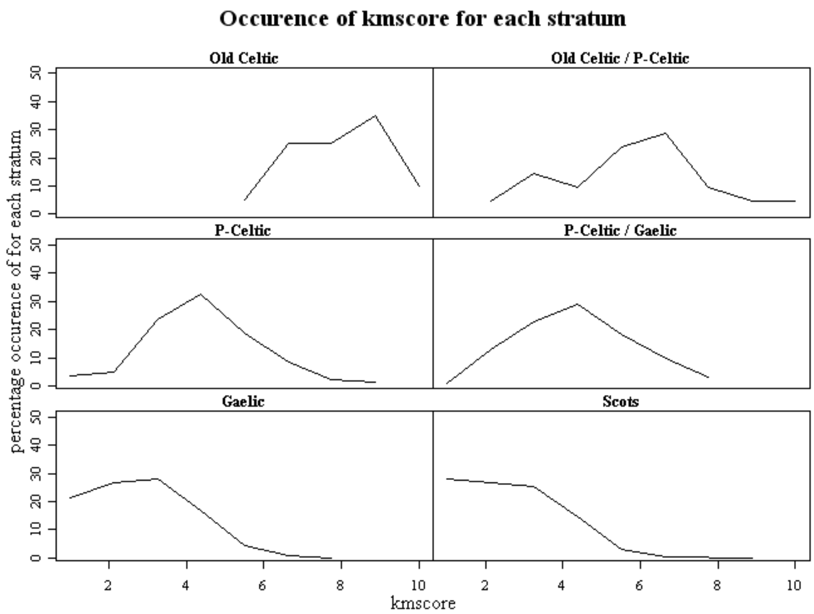
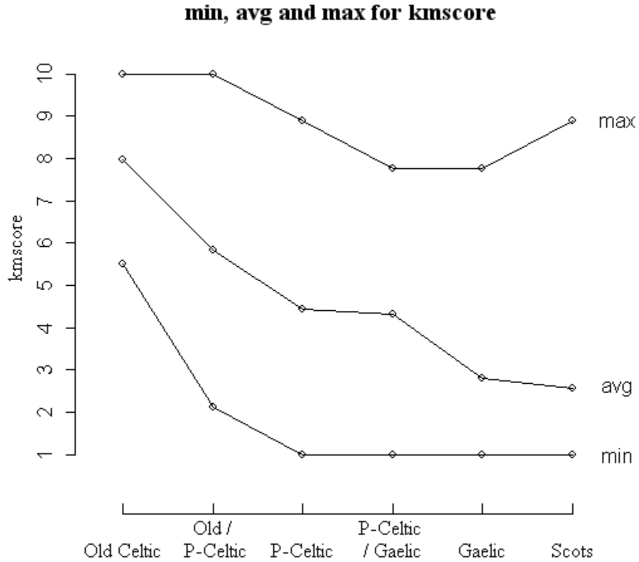


Figure 5.2 shows the same data but presented in a different way. The middle line is the most important here, showing the average *kmscore* for each stratum. One can see that in general the *kmscore* decreases as the strata become later. The rise in the maximum line is not particularly problematic, because whilst the average line is based on all the data, the maximum and minimum values are plotted based on only one RN, and thus some 'noise' is expected to enter (this is why averages are generally used throughout this methodology).

5.2.2 Nont

Starting with the Gaelic and Scots names in figure 5.3, the data for *nontscore* of 1 goes off the scale. This was not corrected since it is desirable to have the same scale for all these graphs, and to change the y-axis scale would have created undesirable graphs in most instances. Essentially this shows that

Figure 5.2: Minimum, Maximum and Average *kmscore* for each Linguistic Stratum



nearly all watercourses with Scots and Gaelic names have no or very few named tributaries. This situation is less true for P-Celtic or Early Gaelic names, although the pattern is partially preserved. The lack of a curve in the graph for Old Celtic is a result of the low numbers of names in this group (20). The dip can be effectively ignored, showing an opposite pattern to the other names.

Figure 5.3: Occurrence of *nontscore* for each Stratum

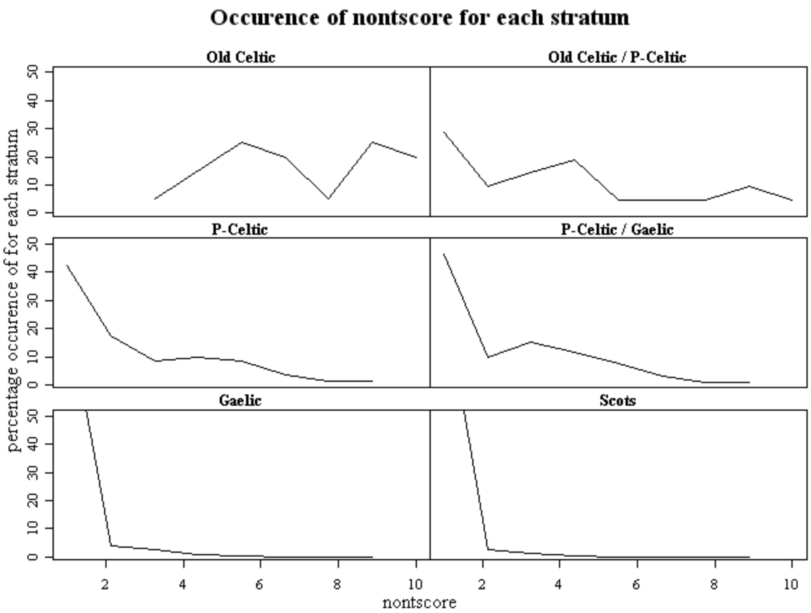
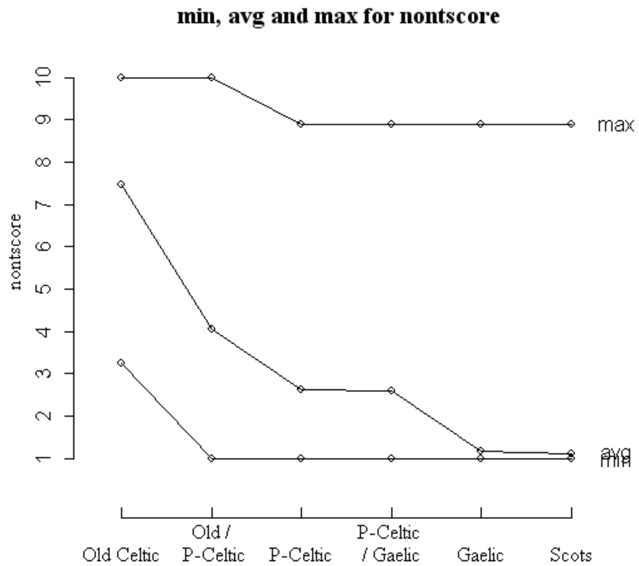


Figure 5.4 shows a broad correlation with figure 5.2 for *kmscore* on page 189. Figure 5.3 on page 189, however, shows a stronger correlation with figure 5.1 on page 188, that the larger the watercourse, the earlier the stratum.

Figure 5.4: Minimum, Maximum and Average *nontscore* for each Linguistic Stratum



5.2.3 Pos

Concerning figure 5.5 on page 191, as noted above in figure 2.9 on page 24, one encounters a particular phenomenon when investigating *specelem* and *pos* in that the older strata behave in a markedly different way to the later strata. This is a different visual interpretation of Nicolaisen's comments that: "Proof of stream-names as markers of relative chronology within a certain catchment area... could be provided... from many parts of Scotland. ...In practically all instances, if the river chosen is only large enough, one would ultimately end up with an Early Celtic or even pre-Celtic name...".[13] Figure 5.6 on page 191 shows a 'kink' in the average plot line in the region of Gaelic and Scots which is also visible in figure 5.8. This essentially shows that the Gaelic and Scots strata are not in the expected order. The implication here is that incoming speakers of successive languages name successively smaller watercourses, and adopt the names of successively larger already-named watercourses with the apparent exception of the Scots stratum. The implications of this in relation to relative chronology is discussed in greater length in the conclusion, in light of other discoveries outside this chapter.

[13]Nicolaisen, *Scottish Place-Names*, p. 225-226.

Figure 5.5: Occurrence of *posscore* for each Stratum

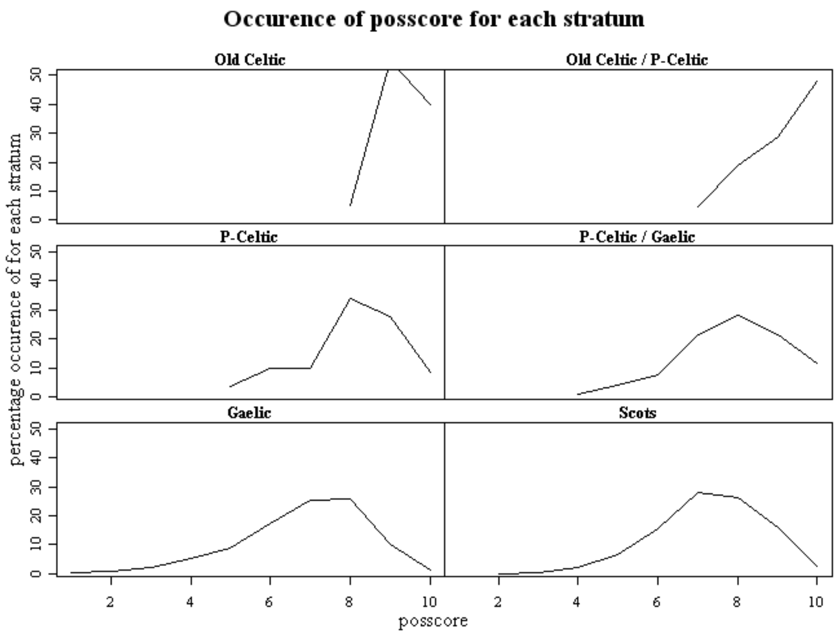
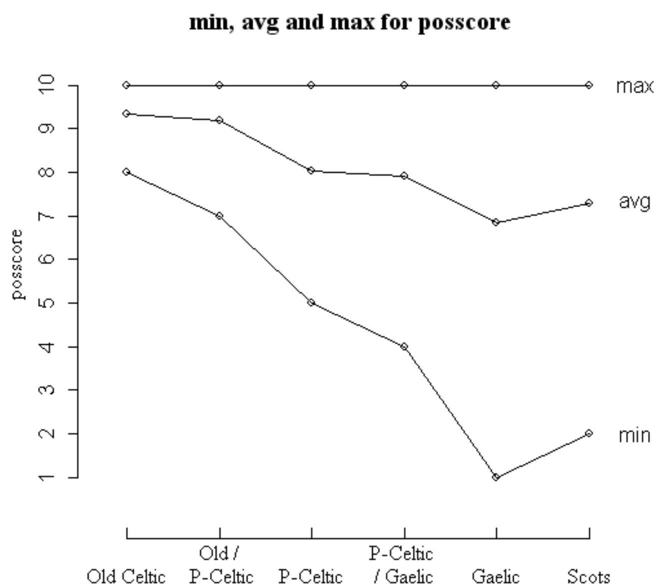


Figure 5.6: Minimum, Maximum and Average *posscore* for each Linguistic Stratum



5.2.4 Alt

Figure 5.7 appears ‘choppy’ because there is no data for some of the *altscore*, due to gaps in the data, which registers as zero. Essentially it shows a similar situation as for *posscore* above; this is corroborated by figure 5.8 showing a very similar result to 5.6.

Figure 5.7: Occurrence of *altscore* for each Stratum

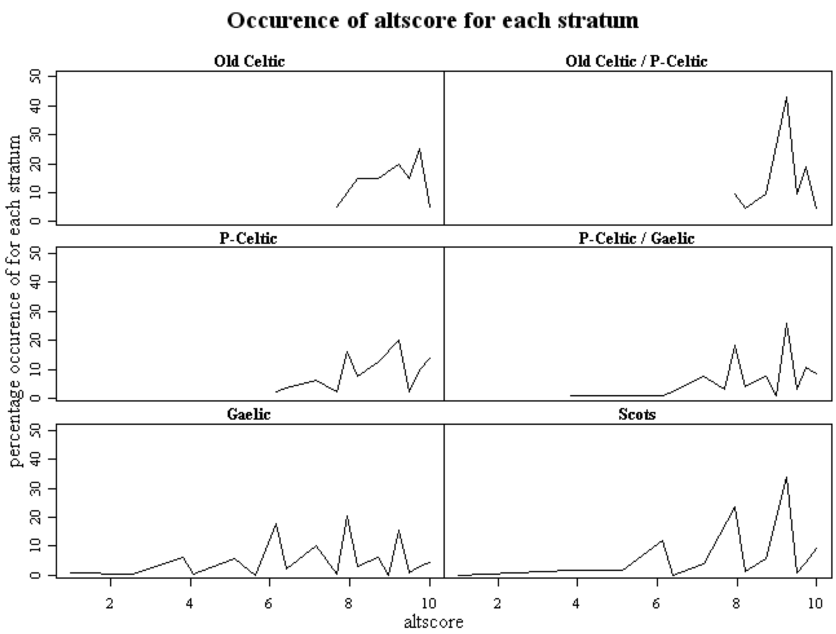
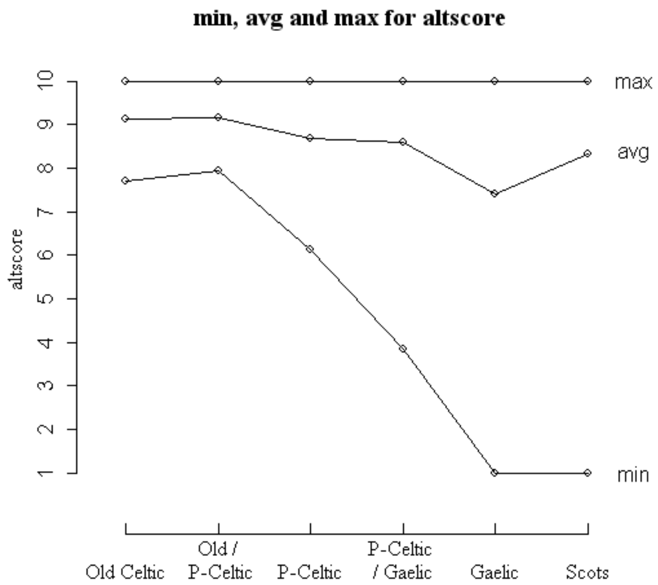


Figure 5.8: Minimum, Maximum and Average *altscore* for each Linguistic Stratum



5.3 Specelem and Components of Lingscore

Elsewhere in this thesis, elements of *geogscore* have been compared to components of *lingscore*. It is also the case, however, that the three components of *lingscore* may be compared with each other. It should be said that a similar comparison of *geogscores* could be made; however, this would have nothing to do with names *per se*, and would only be relevant in a study of the purely geological aspects of a watercourse, i.e. a study in *hydronomy* not *hydronymy*.

A comparison of *genelem* and *specelem* has not been made here. This is because the principles underlying the relationship between the two are governed by *specelem*. That is, a comparison between generic elements and the meaning of a name, whilst ignoring the linguistic strata from which the names come, would be inappropriate.

5.3.1 Specelem and Genelem

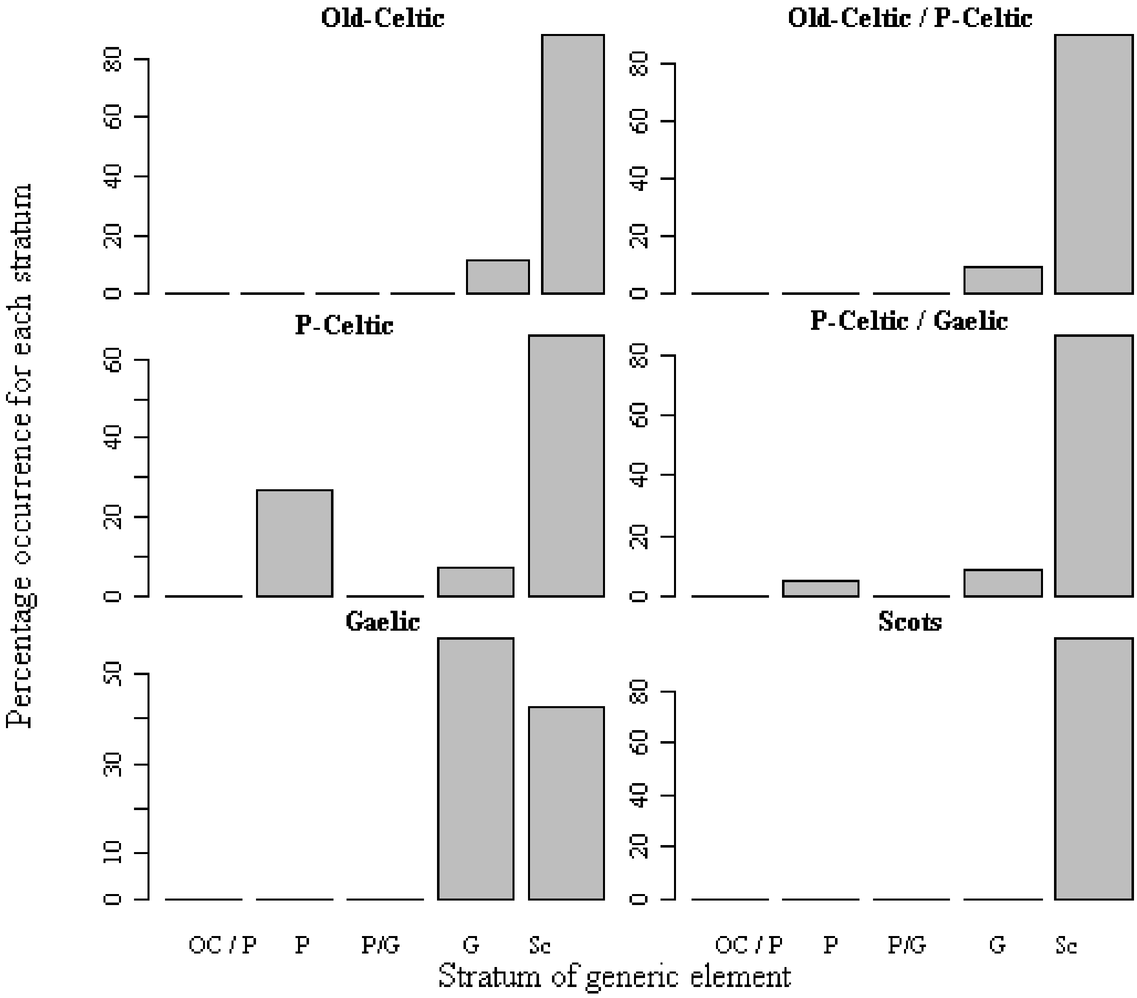
Figure 5.9 on page 194 shows, for each linguistic stratum, the percentage amount of names with generic elements belonging to each linguistic stratum. It shows that the predominant generic for all the strata is Scots, with the exception of Gaelic, which has more Gaelic generic elements, but also with many Scots generic elements (e.g. Linn Burn vs Allt na Linne Mòire). The fact that the majority of Celtic RNs have Scots generic elements is of course due to the fact that as an epexegetic element, the generic is less closely bound to the 'identity' of the name as other features of the name. The fact that Gaelic names largely buck this trend is due to two factors: Firstly, the way in which Gaelic and Scots speakers interacted was largely peaceful, with Gaelic speakers in general becoming bilingual then Scots / SSE monolingual. This is unlike the situation with the relationship between pre-Norse and Norse settlement in some part of Scotland such as the Western Isles, where it is generally accepted that no pre-Norse Gaelic settlement names survive. The second factor involved in the relative survival of Gaelic generic elements is the fact of the recentness of Gaelic as a living language. In certain areas of the AOS it is still used, and in many places was certainly used at the time of the first OS data gathering. If one could imagine a similar study at the time of the gaelicisation of the Picts, one could imagine a similar set of graphs, with each language moving 'one-up' with Scots removed, and Gaelic in dominant position, with P-Celtic names having similar amounts of Gaelic and P-Celtic generic elements. The general rule here is:

> Where the stratum is not productive, the generic elements from the latest stratum are used, where the stratum is productive, that stratum's generic elements are used. If the stratum is productive or semi-productive and not the latest stratum, that stratum's generic elements are used along with generic elements from all later productive strata.

5.3.2 Specelem and Semtype

Figure 5.10 on page 195 shows the predominant semantic types for each stratum. Although it is not particularly easily interpretable by the eye, a number of things can be seen from the graph.

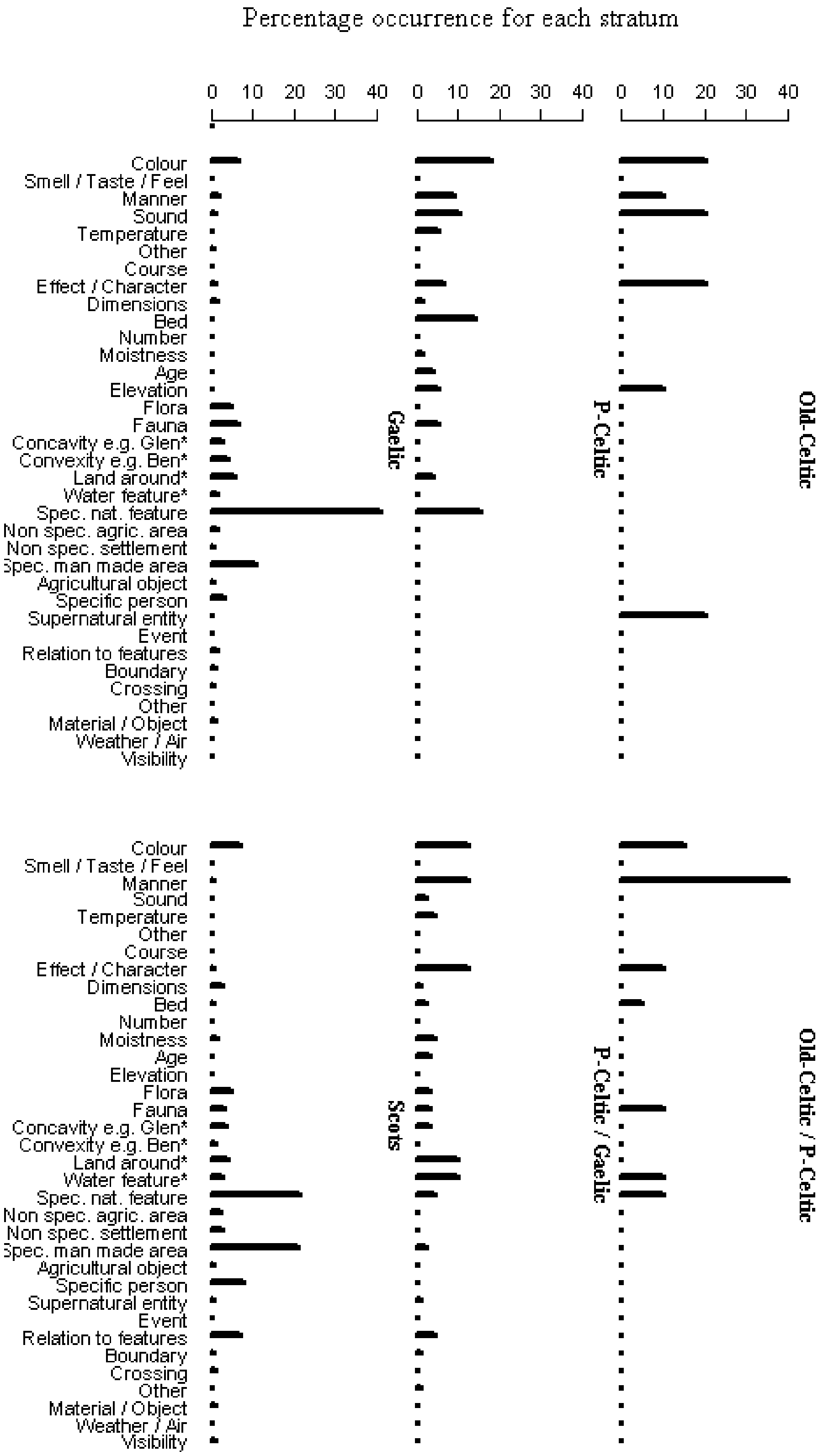
Figure 5.9: Comparison of Linguistic Stratum of a Generic and Specific Element



- In the Old Celtic graph, the two names Don and Dee fit into the supernatural entity slot, and as such can be seen to be semantic outliers. This is discussed above in section 4.3.19 on page 165.
- By far the commonest type of name for Gaelic is ‘specific natural feature’.
- It can be noticed that the later the stratum, the further to the left the peaks appear, with a peak appearing far to the left for Old Celtic, and being at the right for Scots. This is because the semantic classes are ordered by superordinate types, Adjective, Ecosystem, Topography, Human and Situation.

If one makes the same set of graphs, but groups the strata by semantic type as opposed to the class, graph 5.11 on page 197 is shown. This shows the semantic outliers of the Dee and the Don even more acutely, with all other Old Celtic names relating to adjectives. ‘Old-Celtic / P-Celtic’ and

Figure 5.10: Comparison of Linguistic Stratum and a Semantic Type



‘P-Celtic’ names show a similar but less strong correlation. As expected ‘P-Celtic / Gaelic’ names show a halfway situation between the two strata. Gaelic and Scots names show a predominance of names concerned with Topography and Human activity, whilst having several names in all categories. This of course ties in with what one expects of names of this type.

As a general principle, there are certain semantemes which are only applied to hydronyms from later linguistic strata (such as those pertaining to agriculture and human activity). The semantemes applied to the older linguistic strata however, are not solely the domain of those older strata, they also exist for the later names (for example, names pertaining to ‘manner’ can appear in RNs at all strata as discussed on page 176). There are no semantic diagnostic tools for detecting antiquity. Because of the close relationship between *geogscore* and *semtype*, the same can be stated of *geogscore* and *specelem*, that whilst smaller watercourses have names with particular semantemes which cannot be applied to larger watercourses, the reverse is true of larger watercourses.

There is inconclusive evidence that some semantic types should be judged purely within a stratum, and some across all strata. In the chapter on semantics, a homogeneous approach was taken, whereby individual languages were often ignored. It is probably the case however, that in some, if not most, cases semantics should be studied on a language by language basis, ignoring all other RNs. A prime example of this is Muckle Burn (4869), a tributary of the Allan. Whilst this watercourse is far from being the longest watercourse in the Allan catchment area, it is the longest watercourse with a Scots name, as table 5.1 on page 196 shows. It should be stated, however, that this phenomenon is not sustainable in many other situations.

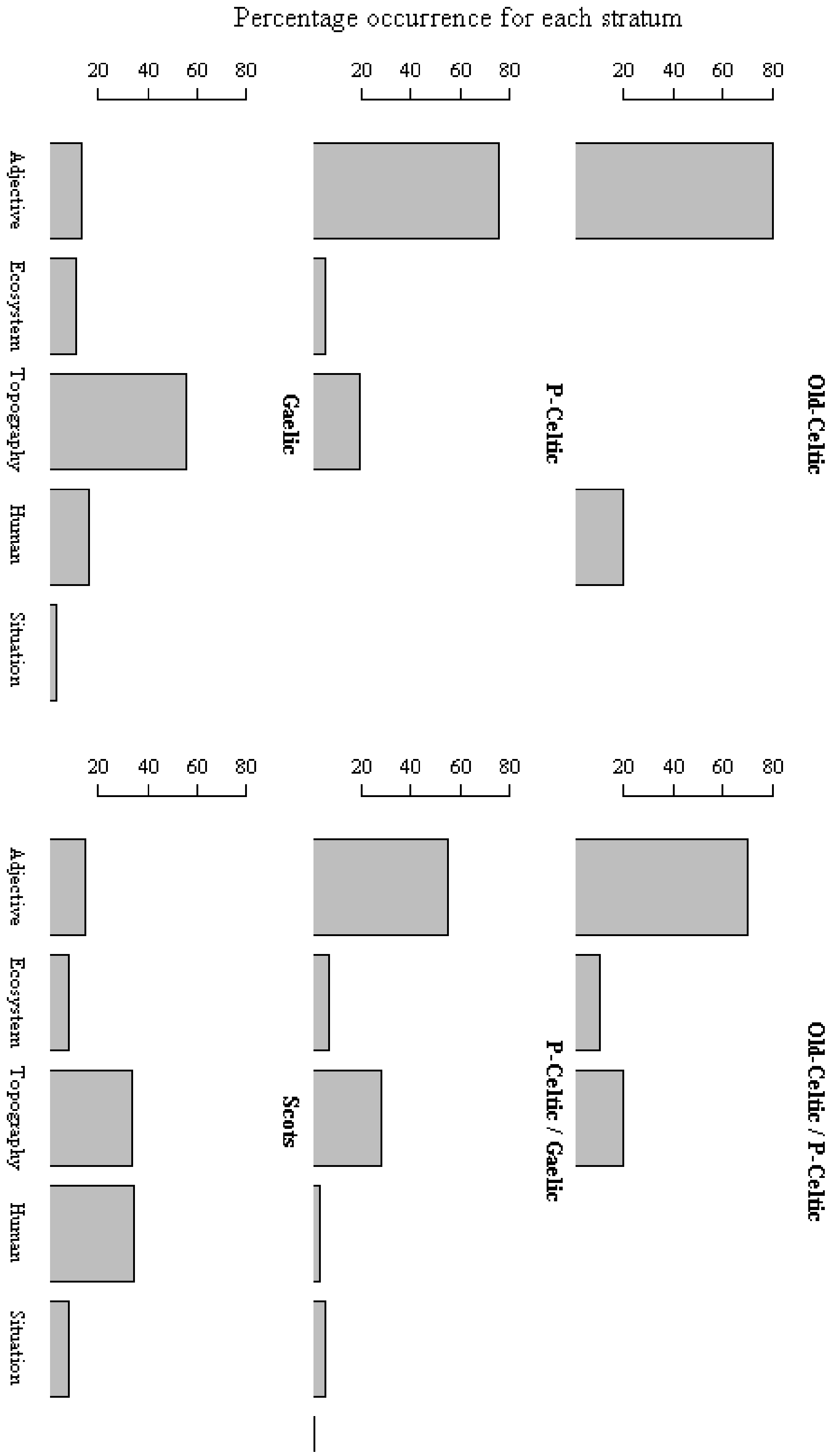
Table 5.1: Ten Longest Watercourses with Scots Names in the Allan Catchment Area

id	rn	km
4869	Muckle Burn	9.0
4991	Burn of Ogilvie	6.0
4345	Lodge Burn	6.0
6273	Whiteburn	5.5
4709	Millstone Burn	5.5
305	Back Burn	5.0
5999	Todhill Burn	3.5
3712	Green Burn	3.0
3590	Burn of Glenbank	2.5
1479	Cock’s Burn	2.5

5.4 Specelem and the Statistical Method

The tools used in the previous chapters are not quite so readily applicable to linguistic strata as to other linguistic factors. This is mainly because, in cases where the linguistic stratum is unknown in a name, it is unlikely that the meaning will be known. Thus it is rare for *specelem* to be the only missing bit of data in the components of *lingscore*.

Figure 5.11: Comparison of Linguistic Stratum and a Semantic Group



Whilst it is then not appropriate to search for *specelem* by investigating the other components of *lingscore*, it is of course possible to look at the components of *geogscore*. To use the equivalent of *genelemfinder* technically works, but in effect gives little information. This is because whilst *genelem* has forty-four distinct classes, *semtype* has only six. Thus a search for most names simply returns the four latest strata, hardly a useful conclusion.

The equivalent of *genelemfinder2*, i.e. *specelemfinder2*, was briefly touched upon in chapter 2 above. In this system, other names with a similar or identical *geogscore* are grouped and counted according to their linguistic stratum. As mentioned above, one can adjust the tolerance of the *geogscore*, and the geographical range. This is a more effective tool, as shown in the following case study.

5.4.1 Case Study: Hadydarn Burn, Hodyclach Burn and Hoodiemart Burn

To show the system in action, I have chosen three burns which are all close to each other in the Ochil Hills, whose linguistic strata and meanings are unclear, although it would seem possible their specific elements contain a similar initial element, these are: Hadydarn Burn (3761), Hodyclach Burn (3812), Hoodiemart Burn (3830).[14] To start, the normal function is applied in table 5.2.

Table 5.2: *Specelemfinder* on Three Related Hydronyms

	Hadydarn Burn		Hodyclach Burn		Hoodiemart Burn	
stratum	frequency	percentage	frequency	percentage	frequency	percentage
Gaelic	57	85.07%	57	85.07%	57	85.07%
Scots	10	14.93%	10	14.93%	10	14.93%

Since the burns all have similar geographical qualities, and are all in the immediate vicinity of each other, they all give the same results. From this view, it would seem the names are fairly clearly Gaelic. However, if one wished to use the tool to only look at names in the vicinity of the Ochil Hills, which has a distinctive nomenclature, one could do so by setting the range to 10000, this takes in only those watercourses with a source within 100000 m or 10 km in section 5.3.

Table 5.3: *Specelemfinder* on Three Related Hydronyms within a Range of 10 km

	Hadydarn Burn		Hodyclach Burn		Hoodiemart Burn	
stratum	frequency	percentage	frequency	percentage	frequency	percentage
Scots	4	80%	4	80%	4	80%
Gaelic	1	20%	1	20%	1	20%

From this it seems that the situation has been reversed, but the amount of data are far too small, since the function has only taken in five watercourses. In order to remedy this, it is possible to slightly expand the tolerance of the *geogscore*, to accept more watercourses. A small increase in the tolerance

[14] These three RNs were first discussed in Angus Watson, *The Ochils: Placenames History Tradition* (Perth, 1995), p. 83-84.

tends to accept many names when dealing with small watercourses. An increase in tolerance of $\pm$ 0.3 accepts thirty-one names and yields a similar result as in table 5.4. To increase the tolerance to $\pm$ 0.4 of the *geogscore* starts to include P-Celtic names into the equation as in table 5.5.

Table 5.4: *Specelemfinder* on Three Related Hydronyms with a tolerance of $\pm$ 0.3

	Hadydarn Burn		Hodyclach Burn		Hoodiemart Burn	
stratum	frequency	percentage	frequency	percentage	frequency	percentage
Scots	26	83.37%	26	83.37%	26	83.37%
Gaelic	5	16.13%	5	16.13%	5	16.13%

Table 5.5: *Specelemfinder* on Three Related Hydronyms with a tolerance of $\pm$ 0.4

	Hadydarn Burn		Hodyclach Burn		Hoodiemart Burn	
stratum	frequency	percentage	frequency	percentage	frequency	percentage
Scots	32	84.21%	32	84.21%	32	84.21%
Gaelic	5	2.63%	5	2.63%	5	2.63%
P-Celtic	1	2.63%	1	2.63%	1	2.78%

The conclusion that can be gained from this is that if one looks across the whole data-set, these names are most likely to be Gaelic, but when looking only at names in or near the vicinity, the watercourses' 'peer group' so to speak, Scots names are more likely. In fact it looks like these names are indeed some sort of combination between Gaelic and Scots, whilst the first Hoodie / Hady / Hody elements is obscure, the other elements could perhaps reflect Sc *darn*, 'hidden', G *mart*, 'ox or cow' (or Sc *mart*, a loanword from Gaelic of the same meaning) and G *clach*, 'stone'.

5.4.2 Data-mining and Specelem

Another method of statistical analysis can be performed. This is of the 'data-mining' type, whereby the whole data-set is investigated for particular anomalies of correlations. The column *posid* denotes the *id* of the watercourse which the watercourse for a given entry flows into. This information enables us to determine all the instances, for example, where a watercourse with a Scots name has a watercourse with a Gaelic name flowing into it.

Table 5.6 shows the top fifteen watercourses with Scots names, ordered by the number of tributaries they have with watercourses with Gaelic names. What is clear from these names is that although they are all Scots, they are also suggestive of names from earlier strata. The two examples of Black Water are clearly reminiscent of *Allt Dubh or *Abhainn Dubh, as are the Glen names reminiscent of *gleann*. The existence of Invernuide suggests that Milton Burn (4731) originally had the Gaelic name of *Nuide. West Water lies between North and South River Esk. If one posits from this evidence that West Water was at some point considered as an **esc* name, this would remove this name as an anomaly from the table. Likewise, Water of Saughs is the upper part of West Water.

Table 5.6: Fifteen Watercourses with Scots Names, having the most Tributaries with Gaelic Names

RN	number of tributaries	other names
Black Water	13	
West Water	11	
Littleglen Burn	8	
Water of Saughs <	7	
Glencally Burn	7	
Pow Water	6	Pow of Inchaffray, Polpefferie
Milton Burn <	6	*Nuide (from Invernuide)
Glenbrighty Burn	6	
Black Water	6	
Menstrie Burn	5	
White Water	5	
Milton Burn	4	
Little Burn	4	
Water of Philorth	4	
Milltown Burn	4	

Table 5.7: Pre-Gaelic Hydronyms which Flow into Watercourses with Gaelic or Scots Names

example	meaning	frequency
Keillor Burn	‘wood’	5
Callater Burn	‘hard water’	5
Nettie Burn	‘cleansing’	2
Strathie Burn	‘valley’	2
Tennen Burn	‘fire’	2

Table 5.7 shows the number of pre-Gaelic hydronyms which flow into watercourses with Gaelic or Scots names. The amount is necessarily fewer, but it can be seen that names deriving from two elements stand out as anomalous: PG **coille* + *ar*, ‘wood’ such as Keillor, and names from P **cal* / *caled*- ‘hard’. The first element is poorly understood, but seems to relate to a wooded area, although the relationship between this and names deriving from P *coet* is uncertain. It may be that the forms in Keillor are not in fact P-Celtic at all, but early Gaelic. After all, Inverkeillor is attested with a number of these names but never *Aberkeillor. This would explain the situation above and the relationship with **coet*. This is only a suggestion however, and more work needs to be done on this term.

Similar things can be said of names deriving from **cal*-; its precise meaning and usage are not clear. In this case however it may be analogous to the term Gaelic term *garbh*. As a P-Celtic term it represents relatively steep, rough, watercourses. Looking at all the names of this type, they are predominantly in the P-Celtic / Gaelic category, (such as Strathie Burn or Tennen Burn) few if any are definitively P-Celtic.

Table 5.8: Number of Combinations of Linguistic Strata and their Tributaries

main	tributary	frequency	direction
Gaelic	Gaelic	2342	normal
Gaelic	Scots	741	normal
Scots	Scots	418	normal
Scots	Gaelic	385	contrary
Old Celt	Gaelic	380	normal
P-Celtic / Gaelic	Gaelic	316	normal
P-Celtic	Gaelic	243	normal
OC / P-Celtic	Gaelic	196	normal
Old Celt	Scots	185	normal
P-Celtic	Scots	107	normal
OC / P-Celtic	Scots	94	normal
P-Celtic / Gaelic	Scots	79	normal
Gaelic	P-Celtic / Gaelic	24	contrary
Old Celt	P-Celtic / Gaelic	16	normal
Old Celt	P-Celtic	11	normal
Gaelic	P-Celtic	10	contrary
P-Celtic	P-Celtic	8	normal
P-Celtic / Gaelic	P-Celtic / Gaelic	8	normal
OC / P-Celtic	P-Celtic / Gaelic	7	normal
OC / P-Celtic	P-Celtic	6	normal
Old Celt	Old Celt	6	normal
Scots	P-Celtic	5	contrary
OC / P-Celtic	OC / P-Celtic	4	normal
P-Celtic / Gaelic	P-Celtic	3	contrary
Gaelic	OC / P-Celtic	3	contrary
P-Celtic	Old Celt	3	contrary
Scots	P-Celtic / Gaelic	3	contrary
P-Celtic	P-Celtic / Gaelic	3	normal
OC / P-Celtic	Old Celt	2	contrary
Old Celt	OC / P-Celtic	2	normal
P-Celtic / Gaelic	OC / P-Celtic	1	contrary
P-Celtic	OC / P-Celtic	1	contrary

5.5 Specelem and the Hierarchical Network

Table 5.8 shows the number of combinations of linguistic strata and their tributaries. It can be seen that the majority of combinations show watercourses with a later linguistic stratum flowing into a watercourse of an earlier stratum, these are marked ‘normal’. The few marked ‘contrary’ relate to the opposite situation. Most of these examples are where, for instance, a watercourse with a P-Celtic / Gaelic name flows into a watercourse with a Gaelic name, and is probably a matter of interpretation. The main exception is the Gaelic / Scots divide, where 385 watercourses with a Gaelic name flow into a watercourse with a Scots name. A large proportion of these names have as their main watercourse one of the names discussed in the previous section, e.g. West Water. Nonetheless there is a significant group of names which are certainly Scots, with Gaelic-named tributaries. These largely relate to the situation where a watercourse with a Scots name, or at least with Scots phonology, runs through a mountainous area with Gaelic mountain burns flowing into it. This phenomenon will be taken up in the chapter on phonology, whereby this phenomenon can be observed more clearly. In this sense, the process of watercourse renaming can be seen as one similar to that of anglicisation of Gaelic names.

It is also possible to investigate this pattern but only regarding the cases where the upper part of a watercourse has a different name from the lower as opposed to tributaries. It can be seen in table 5.9 that the situation is the same as that in table 5.8.

Table 5.9: Number of Combinations of Linguistic Strata and their Tributaries where Upper Part of a Watercourse has a Different Name from the Lower Part

stratum	stratum	frequency
Gaelic	Gaelic	114
Gaelic	Scots	58
Scots	Scots	49
Scots	Gaelic	40
P-Celtic / Gaelic	Gaelic	9
P-Celtic	Gaelic	9
Old Celt	Gaelic	5
P-Celtic / Gaelic	P-Celtic / Gaelic	2
OC / P-Celtic	Scots	2
Old Celt	P-Celtic / Gaelic	2
P-Celtic / Gaelic	Scots	1
Old Celt	Scots	1
P-Celtic	Scots	1

5.5.1 Old / Early Celtic Names

The statistical analysis performed above concurs with what is already known about the oldest names within the Scottish hydronymicon: The rivers with such names are longer, and thus tend to flow into the sea or watercourses flowing into the sea, tend to flow through low lying areas and have a large number of tributaries.

In terms of position, the *pos* section of figure 5.12 on page 204 shows that most of the names of this type flow directly into the sea, for reasons discussed above. Table 5.10 shows the cases where a watercourse with an Old / Early Celtic name does not flow into the sea, and into which river it does flow, with its stratum.

It shows that in the majority of cases the stratum of the river into which the watercourse in question flows is of the same stratum. The Forth contradicts this, although this anomaly is ameliorated by the fact that the original name for the Forth was of course OC Bodería as discussed on page 40. The remaining cases, of the Deveron and River South Esk can be put down largely to a matter of interpretation and the fact that no toponymic system is precise.

Table 5.10: Watercourses with Old / Early Celtic Names and the Watercourses into Which They Flow

stratum of main	main	tributary
Old Celt	River Dee	Water of Tanar
OC / P-Celtic	River Deveron	River Isla
P-Celtic	River Forth	River Allan
P-Celtic	River Forth	River Devon
P-Celtic	River Forth	River Teith
Old Celt	River Tay	River Almond
Old Celt	River Earn	Loch Earn
Old Celt	River Tay	River Isla
Old Celt	River Tay	Loch Tay
Old Celt	River Tay	River Tummel
OC / P-Celtic	River South Esk	Noran Water

5.5.2 P-Celtic Names

P-Celtic names are of a similar size as Old Celtic names, but in terms of altitude and position, they more closely resemble Gaelic names. This essentially means that P-Celtic names have some qualities of older names and some qualities of later names. Applying this to the historical situation, it would seem that names of watercourses at lower altitudes are more resistant to renaming than those at higher altitudes. This can be easily seen in figure 5.12.

5.5.3 Gaelic and Scots Names

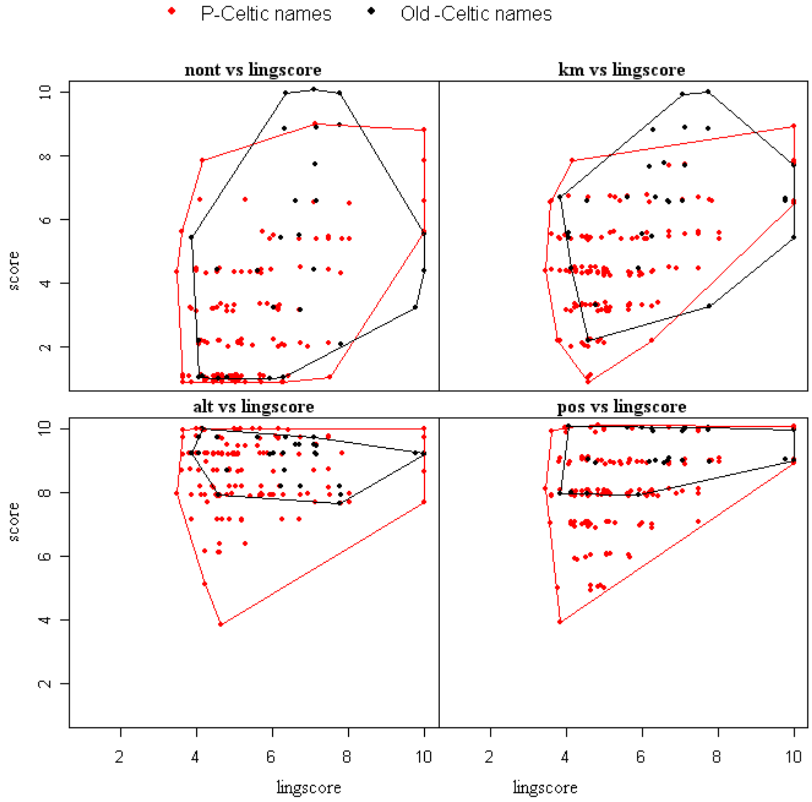
Watercourses with Gaelic and Scots names are physically very similar. Ignoring the location for a moment, any watercourse with a Scots name could also have a Gaelic name, and not change any minimum or maximum values in the database. A comparison of the two in figure 5.13 shows that like the case of G *loch* and G *lochan* as discussed in section 3.4.10.1 on page 77, any watercourse which has a Scots name could also have a Gaelic name.

5.5.4 Hierarchical Order of Semtype

Another use of *posid* is analogous to that used in section 4.6 on page 181 in chapter 4. If one counts the number of combinations of watercourses with a name from a given stratum flowing into another watercourse with a name of a given stratum, and then puts this into a hierarchical order by only looking at the commonest combination of any given two types (e.g. counting only Gaelic : Scots and not *vice versa* because the former occurs 2342 times and the latter only 385) and then placing them in the order of decreasing occurrence, the expected chronological order is given:

$$\text{Old Celtic} < \text{OC / P-Celtic} < \text{P-Celtic / Gaelic} < \text{P-Celtic} < \text{Gaelic} < \text{Scots}$$

Figure 5.12: Comparison of P-Celtic names and Old Celtic Names



5.6 Relationship between Pnstrat and Semtype

Hydronyms coined from place-names were discussed in the chapter on semantics. In this section I wish to look at the relationship between the linguistic stratum of a hydronym derived from a place-name and the linguistic stratum of the place-name itself, ignoring in this chapter the physical type of place denoted.

Linguistic stratum as used in this thesis reflects the stratum from which the term occurring in the specific element derives. The only exception to this is in hydronyms derived from place-names; in this case the stratum denotes when the place-name was taken over as a specific element in a hydronym. The linguistic stratum of the place-name is stored in *pnstrat*. Thus a name such as Loch of Auchlossan is a Scots hydronym deriving from a settlement with a Gaelic name. In many cases, this is a matter of interpretation, since in the majority of instances it is unknown whether a derived name was generated in the same stratum as the place-name the new name derives from, or from a later stratum. (For instance, was the name Balnacoul Burn coined in the Scots era, from the settlement Balnacoul, or does it derive from a Gaelic **Allt Baile nan Cùl*?) In general hydronyms have been taken to be derived from the same stratum as the place-name, although it is possible, and even probable, that some names of this type were coined in the Scots era.

Table 5.11 shows that nearly all place-names are taken from the previous or same linguistic strata

Figure 5.13: Comparison of Gaelic Names and Scots Names

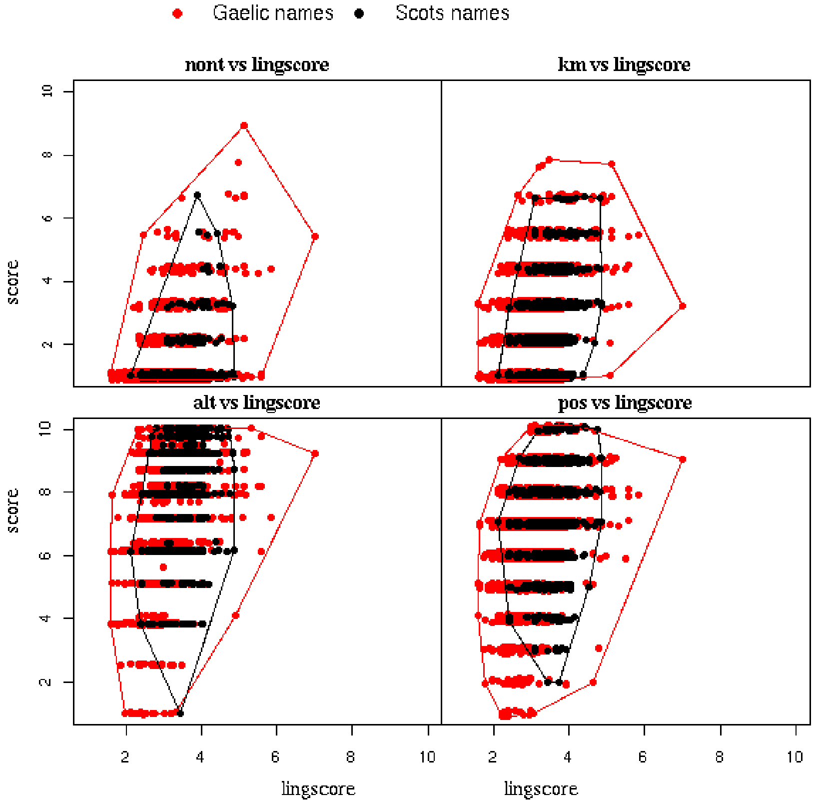


Table 5.11: Combinations of Linguistic Stratum and Stratum of Place-Name where the Watercourse is Coined From a Specific Place

specelem	*pnstrat*	frequency
Gaelic	Gaelic	2112
Scots	Scots	647
Scots	Gaelic	59
Gaelic	P-Celtic	32
Scots	P-Celtic	28
Gaelic	P-Celtic / Gaelic	13
P-Celtic / Gaelic	P-Celtic	3
P-Celtic	P-Celtic	3
Gaelic	Scots	1
Scots	Old Celtic	1

when naming hydronyms. This may seem like an obvious point, but it is not necessarily self evident. Names such as Allt Galloper, a Gaelic name with a Scots place-name generic could exist but are in fact very rare. In other areas of Scotland where the relationship between Gaelic and Scots / SSE is different, more of these names exist. The only example in this AOS is Allt Croft Stochdanan, in this case the word ‘croft’, rather than being considered a Scots word appearing in a Gaelic RN, should probably be considered an orthographic variant of G *croit*, the Gaelic loan of the Scots term. The spoken form of this name must surely have been Allt Croit Stochdanan, rather than the name being a unique Scots-Gaelic hybrid. The example above of a Scots name denoting an OC hydronym is 'Pools of Dee'.

5.7 Conclusion

The analysis of linguistic strata using this system is not as fruitful as it is for other types of data; it may be that under this methodology the original linguistic stratum of the name is not as significant as the linguistic stratum of the generic element, or the surface phonology. Indeed, the correlations between things such as meaning and length are often more significant than simply the linguistic stratum and length. Nonetheless, in the few cases where *specelem* is not known, the method is shown to work according to expected principles. The *posid* data-mining analysis was a valuable exercise, since it is able to show a list of RNs likely to be originally from an earlier linguistic stratum. Moreover, the expected relative chronology of the various linguistic strata was shown. A 'kink' was noticed in some of the graphs, and this is discussed in much more detail in the final conclusion.

Chapter 6

Phonological and Orthographical Overlay

6.1 Introduction

As discussed above on page 32, the categorising of a name by its phonological or orthographic qualities is somewhat different to the other factors employed in the calculus used in this thesis. The other features were all present at the creation of the RN as a linguistic entity; whilst every name clearly had phonology at its inception, this is something that can be altered. To use an analogy, the linguistic stratum or meaning of a RN is equivalent to a person's genetic code, whereas the phonology and orthography of the name equates to a person's socio-economic status.

In keeping with the philosophy and methodology of this thesis, this section will attempt to create a quantifiable range of 'gaelicisation to anglicisation'. I am aware the term 'anglicisation' strictly relates to English as a language, and Scots is the language in question here, but I have used the term since 'scotticisation' in the present context is a clumsy term in my opinion.

6.2 Factors Involved in Anglicisation and Gaelicisation

If one imagines the process of anglicisation from a Gaelic name to a Scots name as a path, it is fallacious to believe that every name begins at one end and travels the whole course. Not all names are created in full Gaelic phonology and orthography, many names having a mixture of Gaelic and Scots at their inception. For instance, a name may be bilingually Gaelic and Scots, or a name may be a Gaelic RN named after a Scots settlement name. This point about settlement names is important, since a settlement name may have a long history of linguistic change before it becomes part of a RN. Extending the analogy further, it is also possible for a name to travel in the reverse direction along this path. The famous example of this in British toponymy is the case of Prestatyn[1] in Denbighshire in Wales. This name derives from Old English *prēosta-tūn*, 'priests' farm'. Elsewhere in England,

[1]B. G. Charles, *Non-Celtic Place-Names in Wales* (London, 1938), p. 230-231.

this name becomes Preston, but in this case the name was borrowed into Welsh and the penultimate syllable was preserved in accordance with Welsh phonological developments. Nowadays Prestatyn is in a chiefly English speaking area again. The name thus shows an ebb and flow from English to Welsh and back to English again.

Phonological / Orthographical Overlay of the Linguistic Stratum of the Specific Element This is deemed to have changed if the spelling is considered formally 'incorrect' for the linguistic stratum of the name. This concerns the specific element only, since generic elements react differently to these pressures. An example where the phonology is Gaelic but the orthography is Scots would be Loch Dhu, from G *dubh*, 'black'. In this example a typical pronunciation of the name may not tell one whether the name was phonologically Gaelic or Scots, but the spelling definitely suggests a Scots flavour. Of course, it could be said (much like *Burn of* X distributions) that the variations in the orthography are a result of the vagaries of the Ordnance Survey cartographers, but with the evidence shown below, I endeavour to prove that this is a less important factor than may have been supposed. This also includes cases where the pronunciation of the specific element is deemed to have changed from the original element from which it was coined. An example is Quhomery Burn (pronounced 'Fummery' by the locals), from G *comar*, 'confluence' (with the Scots form implying the G form *chomair*). It is a Gaelic name, borrowed by Scots speakers with subsequent changes in phonology. This is stored within the database in the *phono* column.

Linguistic Stratum of the Generic Element As also discussed in chapter 3 on Generic Elements, a name can have a specific element from one stratum and a generic element from another (usually later). This of course should be taken into account, especially considering names such as Burn of Coire nan Dùn, which is a Scots generic element, but with a specific element in correct Gaelic orthography.

Linguistic Stratum of the Specific Element 'Linguistic Stratum' is of course the starting point for judging this phenomenon, but there is also the additional issue here of translated names. For instance a name such as Black Burn could hypothetically be a translation of a Gaelic Allt Dubh, but in 99% of cases one can only speculate, since Black Burn is also a productive RN within Scots, regardless of Gaelic. Thus it should be borne in mind that some names could have undergone this process before any old forms were recorded. The date of recording is largely an arbitrary date which can only prove the earliest date at which a name existed.

In the description above linguistic change has not been discussed within a particular stratum; as any toponymist notices, the phonology of place-names often change without ever changing any of the factors above, except perhaps orthography. Whilst this is a fascinating topic of study, in this section, only change from one stratum to another has been taken into account, since it is much more easily quantifiable.

It may also be said that where a RN relates to an external feature, then the linguistic feature of that place-name should be taken into account, however, where a modern name relates to an old feature, then it is most appropriate that the name of the external feature is considered in the same stratum. For instance, the Dee is an ancient name, yet there is probably nothing ancient about the name 'Pools of Dee'. This brings into focus a factor not often discussed, that the linguistic stratum of a specific element of a name is generally quoted as the earliest stratum, whereas the name is actually in the latest stratum. Take for example the name Don, there are spellings for this which show it as an Old Celtic name: *Deouana*, British: *Doen*, Gaelic: *Deon* and now a Scots / SSE spelling and pronunciation: The Don. At each stage it was nonetheless considered a name in that language, as, in a sense, the name Don is considered a Scots / SSE name now, insofar as it is pronounced and written in Scots / SSE, regardless of its ancient origins[2].

The data-set used here is only those data derived from current OS maps. This is done because one needs names gathered at the same time as a 'snapshot' to study the fluidity of phonological / orthographical overlay. For instance some names shown on Pont maps are categorised under the old forms for a modern name, but where the modern name is no longer extant, that name has its own entry. This policy seemingly distorts the results, since it will seem that some names have an orthography from Pont's time, whilst some have a modern orthography. Choosing the OS names as a 'snapshot' is a conscious choice, but other choices could have been made. A native Gaelic speaker, for instance, might view the head-forms differently, for instance by considering a name such as Abhainn Tatha the default name to investigate, rather than River Tay. Table 6.1 shows different types of names, with frequency and examples. The string represents a concatenation of *genelem1*, *phono* and *specelem*.

Table 6.1: Grades of Anglicisation from Gaelic to Scots

string	frequency	example	*genelem1*	*phono*	*specelem*
777	1698	Elfhouse Burn	Scots	Scots	Scots
776	1667	Aber Burn	Scots	Scots	Gaelic
766	43	Leuchar Burn	Scots	Gaelic	Gaelic
677	1	Allt Galloper	Gaelic	Scots	Scots
676	77	Loch Achray	Gaelic	Scots	Gaelic
666	2130	Allt a' Choire Odhair	Gaelic	Gaelic	Gaelic

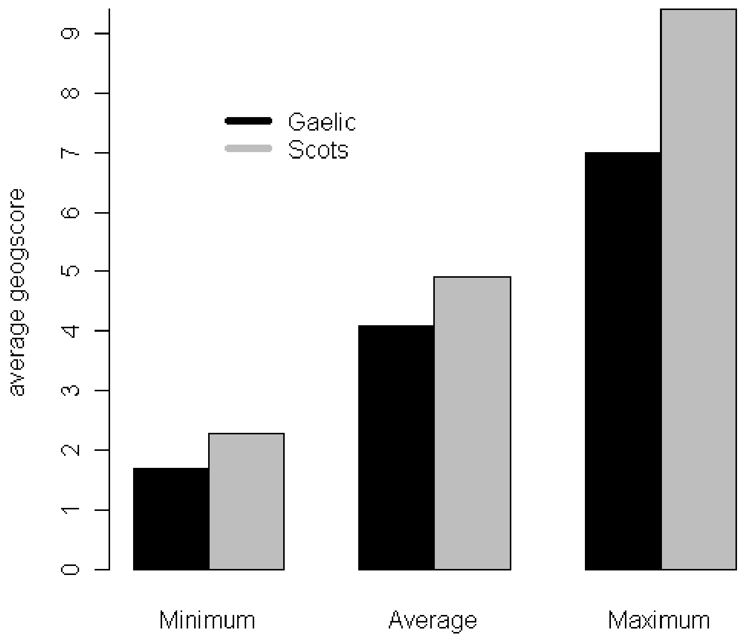
6.2.1 Anglicisation and Geogscore

Geogscore or physical attributes do not influence anglicisation in the way it influences generic element. For example, one cannot on the whole deduce the degree of anglicisation from the physical features of a watercourse, which is mainly determined by location, of course. The main exception to this is that larger watercourses are less resistant to anglicisation than smaller ones. This can be seen in the fact that all the large watercourses, or 'rivers' as they may be called, are all generally anglicised, even

[2]This mirrors an issue in historical linguistics, where linguists say Welsh is a Celtic language, but few ethnologists would call the Welsh people 'Celtic' as such.

in generally Gaelic speaking areas the names tend to have two names, a Gaelic one with *abhainn*, and a SSE one in *river* (such as Lettie River/Abhainn Deataidh from OS gazetteer). This is shown in figure 6.1.[3]

Figure 6.1: Various *geogscores* for RNs with Gaelic and Scots Phonology / Orthography



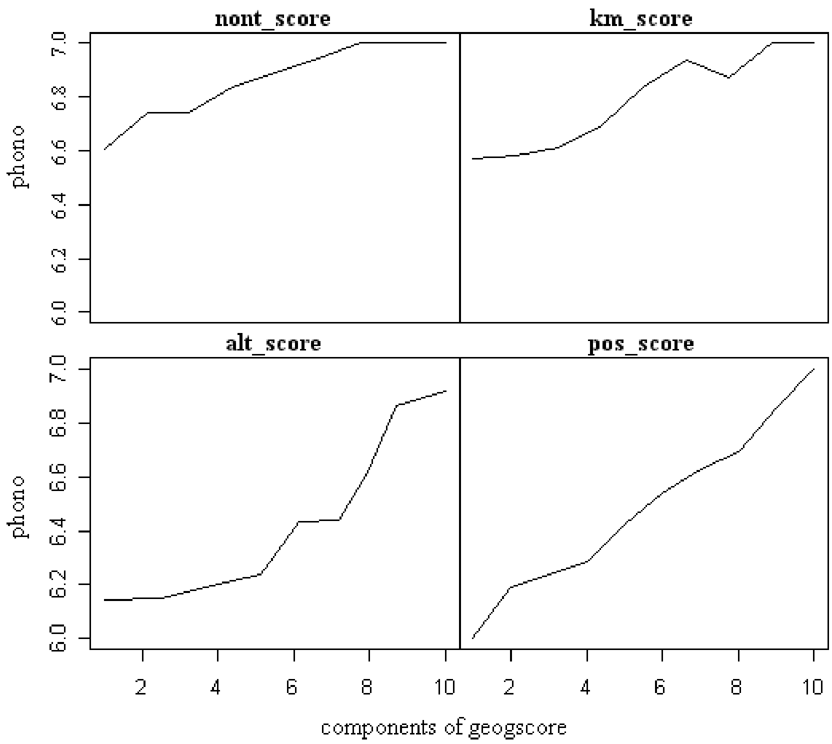
Of course, RNs are only one type of place-name, and this phenomenon occurs throughout all echelons of the toponymicon, but it is observable most clearly in smaller features, and names with generic elements, so RNs are as good a data-set as any other, although hill names could also yield comparable results.

Figure 6.2 breaks down the individual scores that comprise *geogscore*, showing that essentially the same phenomenon is at work with each score that comprises *geogscore*, the larger the watercourse, the more likely the RN is to have a phonology or orthography from a later linguistic stratum. *Pos* has the most direct correlation, which is surprising, since this was the most controversial score.

When these are compared to the same data, however, except with *specelem* used instead of *alt*, an interesting picture emerges, as in figure 6.3 on page 212. The lower the altitude, the more likely the phonology or orthography of a name will be Scots. This is the same as with *alt*, in that the lower

[3]It is interesting to note that this phenomenon does not seem to occur with mountains and hills, for instance Ben Nevis and Glen Nevis are spoken in common parlance, but River Nevis, never Abhainn Nevis, or Ben Lawers and Lawers Burn. Of course, it is true that Gaelic has more currency in mountainous areas, but this does not explain why this situation exists in the Highlands.

Figure 6.2: Comparison of Phonology / Orthography and Various Components of *Geogscore*



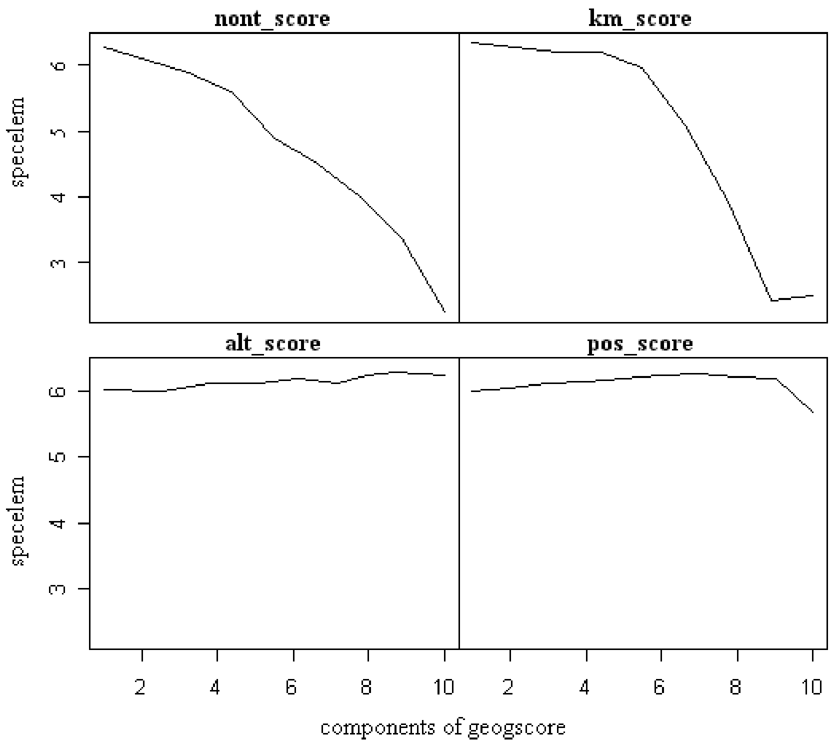
a watercourse is, the more likely it is that it will derive from a Scots name. The remaining scores, however, are radically different, if not opposite. The longer a watercourse or the more tributaries a watercourse has, the more likely it will be from an earlier linguistic stratum. This is to be expected, and is a bi-product of the phenomenon that "the older a name is, the larger the watercourse is", however, it seems to contradict the way *phono* works.

The relationship between these two factors shows the dual phenomena at work in *pos*. In numbers of low position (or high *posscore* as shown here), the graph has a lower average *phono*, since it includes names from all linguistic strata, but the higher the position, the more likely the name is to be Gaelic, hence the approach to the number 6, i.e. Gaelic.

In order to discern a Highland line, or corridor, one should first look at the encroachment of Scots names West, and the Gaelic names East. For this, each different combination of names from Gaelic to Scots as shown in table 6.1 on page 209 is plotted on the distribution map in figure 6.4 on page 213. A number of points can be seen from this distribution:

1. The areas of the incidence of most names not purely one language or the other (i.e. between 666 and 777) are also the areas of the greatest incidence of Y *of* X names discussed in section 3.7.2 on page 96.

Figure 6.3: Comparison of Linguistic Strata and Various Components of *Geogscore*



2. There are more Scots names in the Gaelic area than *vice versa*. This situation is to be expected as Scots was the encroaching, later language. The Scots names in question predominantly belong to watercourses which are tributaries of very large rivers running through glens, such as the Spey Valley, the Tay catchment area and Upper Deeside. This is presumably because the lines of linguistic communication travelled more freely through the glens than over mountainous terrain.

3. The Highland line is a corridor or zone, more than a line. In large parts of Perthshire the corridor is thinner, but the two 'intermediate' areas show a more gradual process of anglicisation. Another way of showing this is with contour plots, showing the average *phono* for small areas, as in figure 6.5 on page 214.

In terms of semantics, the names are typical of Scots names in other areas with two exceptions. Firstly colour names are absent except for Red Burn (5265) which as a point of note, does not occur on the first OS maps. Secondly, the name Milton Burn appears three times on the Spey (the same three Milton Burns that appear in table 5.6 on page 200).

Figure 6.4: The Highland Corridor according to the Orthography and Phonology of Hydronyms

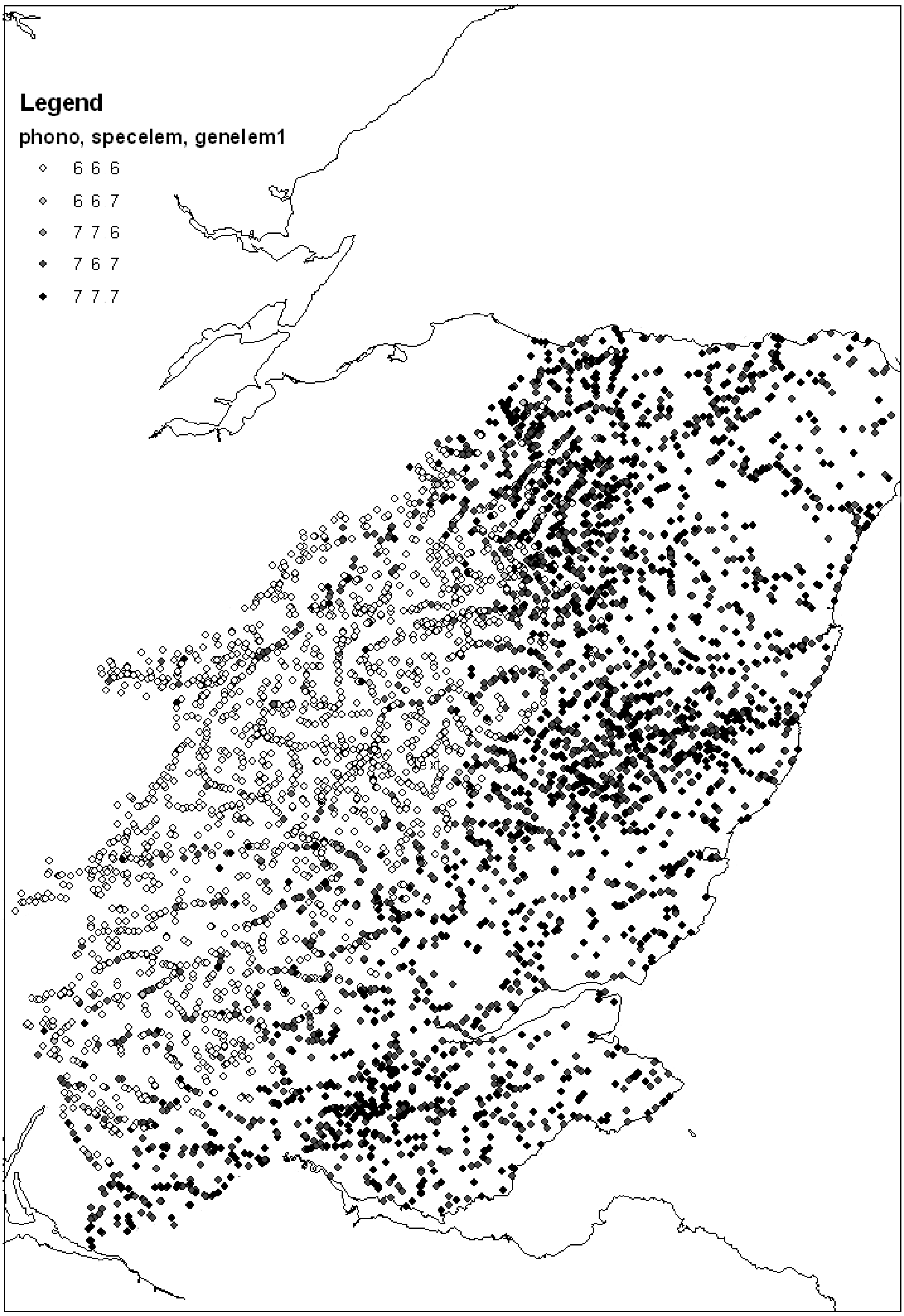
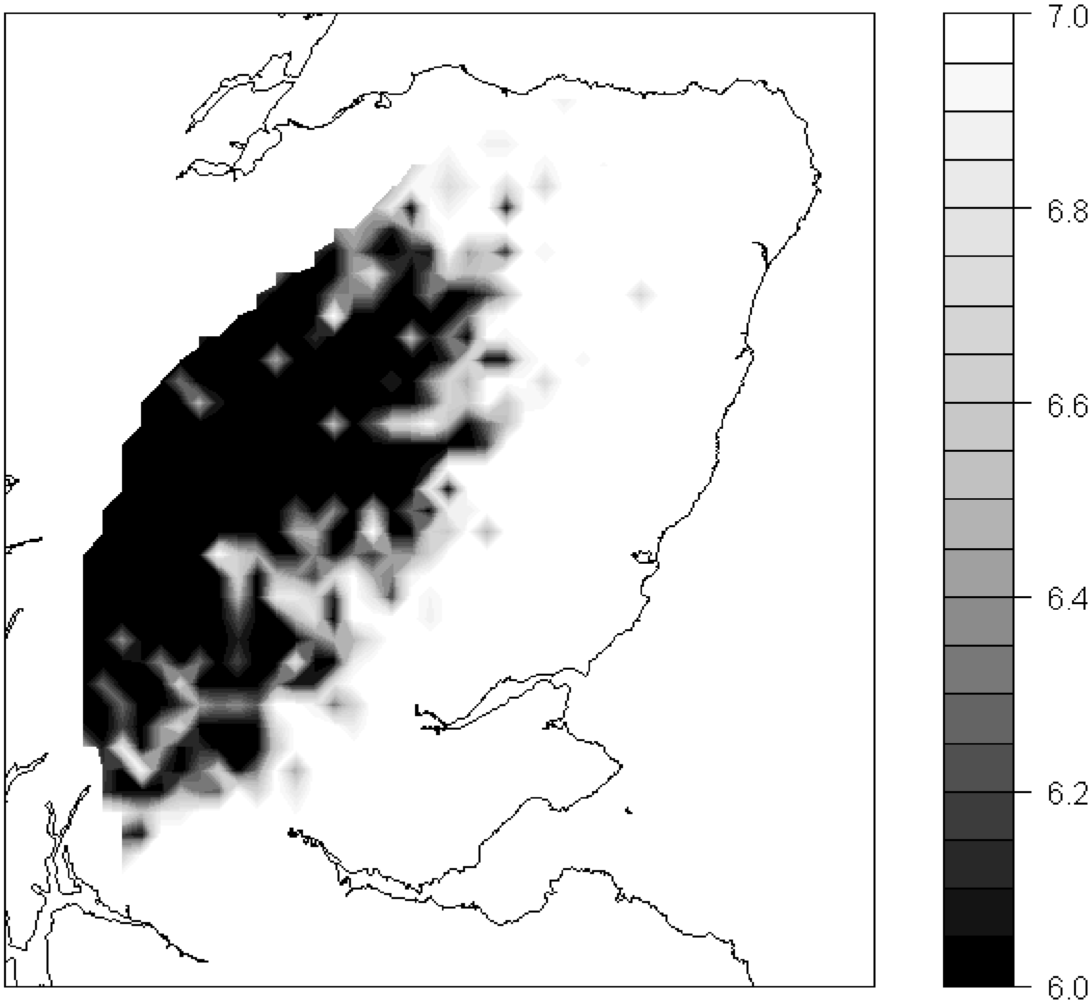


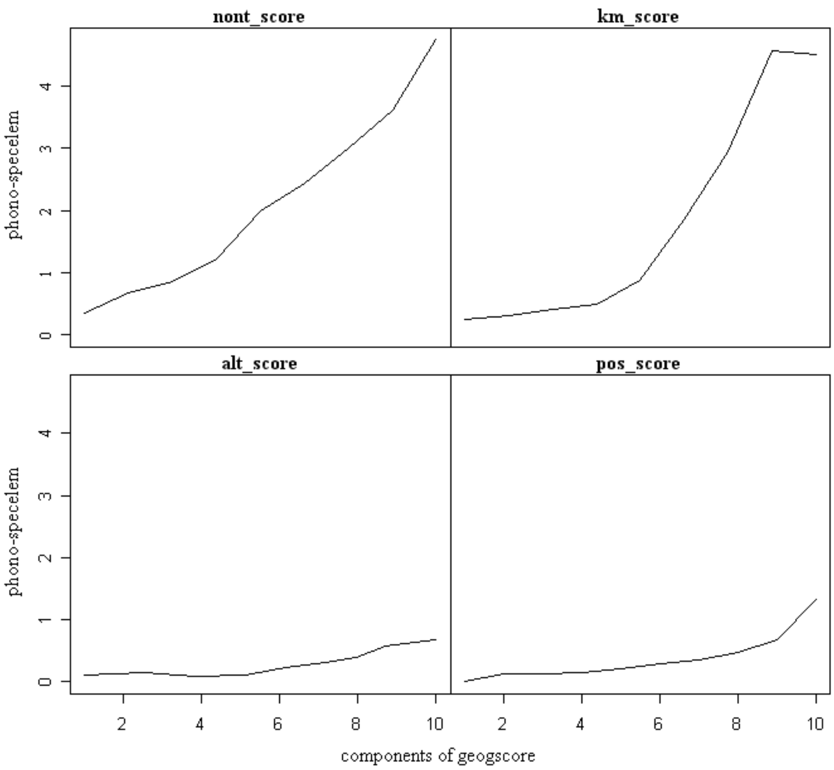
Figure 6.5: Contour Map of Surface Orthography and Phonology of Watercourses



6.2.2 Anglicisation and Components of Geogscore

A comparison between figure 6.6 on page 215 and 6.3 on page 212 shows that a name such as Allt Dubh has a *specelem* and *phono* of 6, so the difference is zero, but a name such as Tuach Burn (from G *tulach*, 'hill') has a *phono* of 7 (i.e. Scots) but a *specelem* of 6 (i.e. Gaelic). It is possible that where the difference is 1 the name is Scots with a Gaelic phonology, but no names such as this exist (such names have generally been interpreted as containing a Gaelic term that is a loan word from Scots, such as Allt a' Bhùirn, 'burn of the burn').

Figure 6.6: Comparison of *phono* and *specelem* with Components of *geogscore*



For *nont* and *km* in this graph, the trend is that the larger the watercourse, the greater the difference between original linguistic strata and surface phonology / orthography. This can be seen in that small watercourses such as Allt Dubh or Ballinloan Burn have either an identical *specelem* and *alt*, or a difference of 1 (between Scots and Gaelic), yet a large watercourse such as the Tay has a surface phonology / orthography of Scots, whilst belonging to an Old / Early Celtic stratum (having a difference of 5).

For *alt* and *pos*, the situation is essentially the same, but the scale is much smaller. The greatest difference is 1. Watercourses at the lowest altitudes, closest to the sea, all have a Scots phonology or orthography but can have any linguistic stratum for their origin. Watercourses at high altitudes, conversely, are all Gaelic in origin with Gaelic phonology and orthography.

6.3 Relationship between Phono and Specelem

Table 6.2 represents the number of occurrences of combinations of phonology / orthography and linguistic stratum. One can clearly see that the differences in phonology / orthography are only between neighbouring linguistic strata. This table was taken only from OS data, and there are other names which show a different situation with old forms and so on. These 'accidents of survival' generally muddy the picture. What is interesting in this pattern is that the current dominant language, Scots / SSE becomes the *de facto* phonology / orthography for dead linguistic strata. This means that whilst names from dead languages, such as the Tay and so on existed throughout the entire Gaelic speaking era of the area, the name is now spoken of always in a Scots / SSE phonology. Little effort is made to speak of it as a Gaelic name (such as Abhainn Tatha), yet a name such a Allt a' Mhuilinn has retained its phonology and orthography.

Table 6.2: Number of Occurrences of Combinations of Phonology / Orthography and Linguistic Stratum

specelem	*phono*					
	Old Celt	OC / P-Celtic	P-Celtic	P-Celtic / Gaelic	Gaelic	Scots
Old Celt	0	0	0	0	0	17
OC / P-Celtic	0	0	0	0	0	21
P-Celtic	0	0	0	0	0	51
P-Celtic / Gaelic	0	0	0	0	2	68
Gaelic	0	0	0	0	2173	1741
Scots	0	0	0	0	0	1686

6.4 Anglicisation and the Statistical Method

Since *phono* is substantially different from the components of *lingscore*, to apply similar methods would be nonsensical. Phonology / orthography is largely dependent on location, the only situation in which a statistical method would be useful would be to guess the location of a watercourse where its name, orthography and phonology was known. There are no situations such as this, since any given RN whose location is unknown is from old records, which necessarily have a different phonology and orthography from that of the present day.

Despite this, the phonological qualities of hydronyms can be investigated within the context of the hierarchical network along similar lines to that done in the chapter on linguistic strata. Table 6.3 is an equivalent of 5.8 on page 201 but counting the phonology and orthography of hydronyms rather than the linguistic stratum.

Table 6.3 shows that the second most popular combination is for watercourses with Gaelic names to flow into watercourses with Scots names. This is the same phenomenon as that mentioned in the previous chapter. The configuration of the Gaelic names in this case is typical of Gaelic names in general.

Table 6.3: Number of Combinations of Phonological / Orthographical Overlay and their Tributaries

main	tributary	frequency	direction
Scots	Scots	3359	normal
Scots	Gaelic	1233	contrary
Gaelic	Gaelic	940	normal
Gaelic	Scots	125	normal

The amount of these names as opposed to the reverse surely shows that the hierarchical network sequence is different from the chronological sequence. That is, one is used to thinking of Scots / SSE as the most recent linguistic stratum, which is true, but that does not mean it occurs at the end of the hierarchical sequence.

6.5 Conclusion

This chapter is substantially different from the others, and in a sense does not directly build onto the methodology for discovery shown in other sections. It does, however, show how analytical tools can uncover patterns such as the Highland corridor and the hierarchical network.

Chapter 7

Conclusion

The first section of this chapter draws together certain strands in the thesis in order to reveal conclusions not necessarily evident in the individual chapters. In the second section the strengths of the approaches taken in this thesis and the resolution of potential drawbacks mentioned in the introduction are discussed. Finally, some future uses of the methodology are envisaged.

7.1 Patterns Revealed by the Methodology

There are a number of emergent patterns which are not fully evident in the individual chapters. These are discussed in the sections below.

7.1.1 The Hierarchical Network

Throughout each preceding chapter about the individual linguistic qualities Scottish hydronymy was investigated not by studying a specific linguistic quality as it relates to a specific RN, but instead by studying the combinations of linguistic qualities where two watercourses intersect. In each case an attempt has been made to discern a hierarchy of discrete units judged by their significance in the network. The resulting correlations are stronger in some cases than in others:

Generic elements display a fairly clear order, as seen in 3.26 on page 119. This was mainly due to two facts: firstly that generic elements are generally attached to RNs unambiguously, that is, one will nearly always know precisely what a generic element is in a given name. Secondly, the relatively large number of generic elements means that there is plenty of scope for a hierarchy.

In **Semantics**, the hierarchy was less easily distinguishable. This was mainly due to the fact that whilst a hydronym clearly has a particular generic element, the classification of its meaning is generally open to interpretation. On page 129 the question was raised as to what level of specificity it is best to classify a semanteme. As mentioned above in section 4.1 on page 130, should G *sean*, 'old' be interpreted merely as an adjective, as an adjective concerning age, or as an adjective specifically representing old age? The answer is that the larger these classifications, the more data they necessarily contain and thus the more statistically meaningful the results. The highest-level semantic class showed the most sense, whilst the lower-level more specific types showed less meaningful results. Nonetheless, a distance can clearly be observed amongst the larger semantic groups as shown on page 181. This is:

$$\text{Adjective} < \text{Ecosystem} < \text{Topography} < \text{Human} < \text{Situation.}$$

This hierarchy is a representation of the conceptual distance between the semantic group of the RN and the semanteme 'water' or 'flowing water'. One can imagine this as the distance from a watercourse a viewer would need to be to notice the relevant semanteme. Thus, one must be close to a watercourse to detect an adjectival quality such as colour.[1] A viewer needs to step back from the river to see the flora and fauna that surround it, and must look further afield for nearby settlements or natural features. This is a concrete description of the notion of 'importance' of RNs, in a semantic context.

Linguistic Strata, having only six distinct groups, show clear correlations in many situations. Throughout this thesis the linguistic strata as in table B.7 on page 248 has been used. This order reflects the chronological situation. In the majority of instances this chronological order has been reflected in the research; in several situations, however, it would seem that Gaelic and Scots are in a reverse order from the temporal order. These are in the following situations:

- In the hierarchy of the watercourse network, where Gaelic occurs more commonly as the uppermost watercourse at the watershed, than do watercourses with Scots names, as discussed in section 5.5.4 on page 203.
- As in figure 2.9 on page 24 which shows shows the percentage occurrence of each position within each linguistic stratum.
- As in figure 2.6 on page 22 which shows the average altitude according to linguistic stratum.
- As in figure 5.8 on page 192 which shows the average altitude according to linguistic stratum.
- As in figure 5.6 on page 191 which shows the average position according to linguistic stratum.
- As in figure 6.3 on page 217 which shows the number of combinations of phonological overlay for watercourses and their tributaries.

In each of the graphs mentioned above, Gaelic and Scots names are reversed when analysing altitude or position, creating a visual 'kink' in these graphs. This is the result of the way a particular phenomenon has been actuated: a new stratum tends to name 'lesser' watercourses rather than replacing the names of the 'larger' watercourses. This results in the situation that generally 'the older the name, the bigger the river'. In this particular situation the term 'lesser' here represents the higher, more mountainous watercourses (i.e. those with lower *posscores* and *altscores*). As with any linguistic process, its progress can be arrested, slowed or even reversed at certain points. It seems that in this case, the renaming of the highest watercourses would be expected to have been undertaken by the

[1] This is primarily on a conceptual level; one cannot see 'age' for instance, and in some situations colour terms are not representative of the physical colour of the water.

latest stratum, the Scots speakers, but this has not happened as thoroughly as one might have expected; it is perhaps a process still in flux, resulting in the kink. This image is consistent in cases where there is relatively rapid language change, as was the case with the decline of the Gaelic language and the rise of Scots and SSE. It is also consistent with a situation of bilingualism and language replacement, as opposed to the replacement of actual people of one tongue by another. Whether the process will continue or has been halted remains to be seen.

The hierarchy of linguistic strata is quantitatively different from that of generic elements or semantics. In these two cases the hierarchy reflects a cognitive process within the mind of a (hypothetical) speaker who coined a name. Linguistic strata reflects the process over a period of time; in a sense it is diachronic, whilst the other two hierarchies are synchronic.

7.1.2 Peripherality and Centrality

One trend that has occurred repeatedly throughout the analysis is that given a certain correlation between two qualities, the names which best fit this correlation tend to be better documented than names which are at the extremities of the clustering. An example of this was shown in the methodology chapter; names which were documented only once tended to be further from the line of best fit in figure 2.11 on page 40. Table 4.30 on page 163 and the following discussion also shows a list where the longest watercourses which derive from settlements are also some of the most uncertain.

RNs that are peripheral in these analyses tend to have one or more of the following qualities:

- They are not well documented as RNs. Within the database, these names are marked as having a source of F1 (see table B.6 on page 248), meaning the name derives from only one independent source, with no corroboration.
- It is not clear whether a name is a primary or secondary hydronym (Such as Loch Lomond and Ben Lomond).
- The derivation is uncertain, or the term might be ascertained, but its meaning may not be clear (as with **cal*).

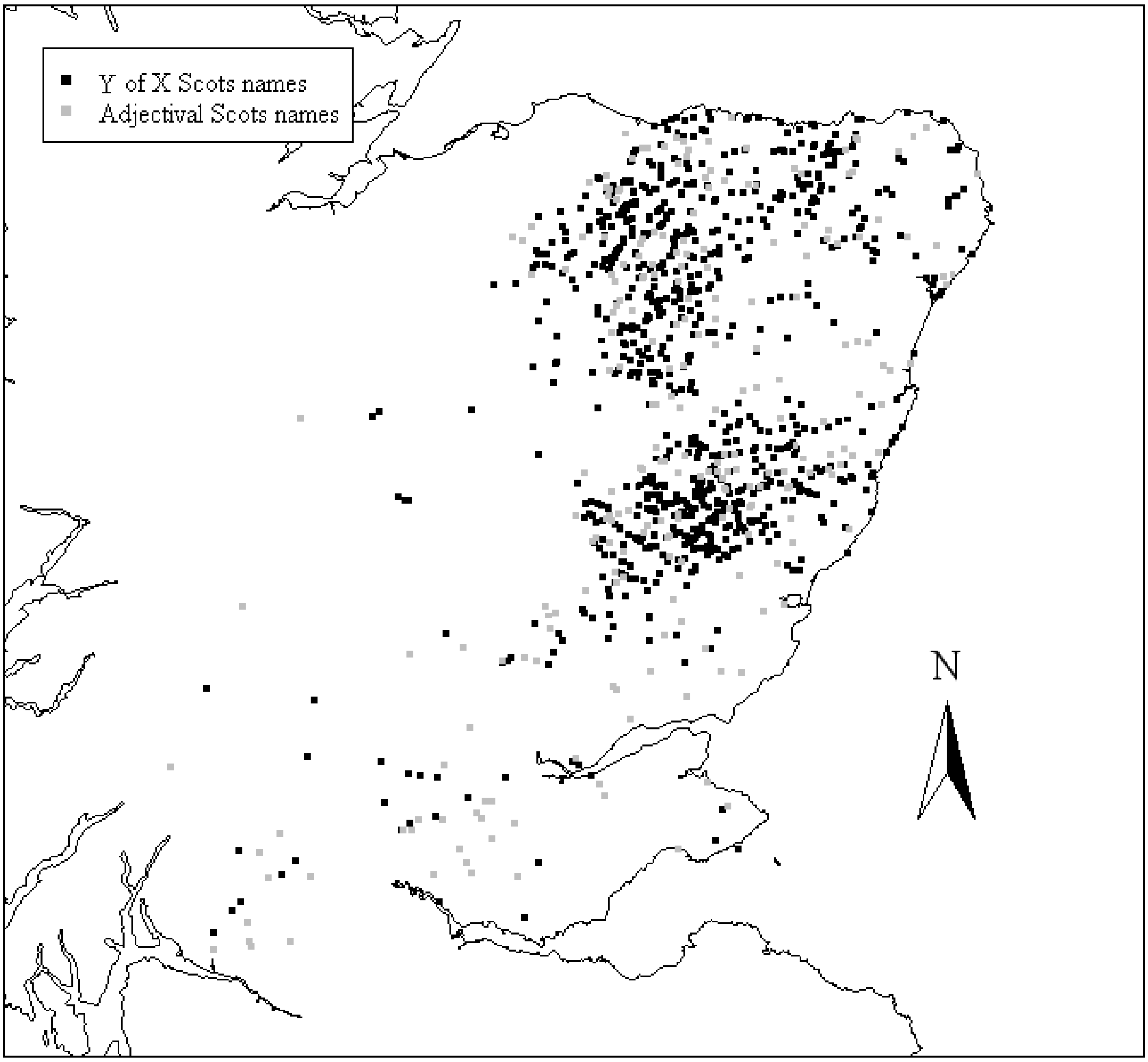
Unfortunately, evidence for this trend is not particularly forthcoming when looking at the data. This is due to a number of external complicating factors. For instance, larger watercourses tend to be better documented in old forms because they are larger and therefore have more impact on material culture and so on.

7.1.3 Distribution Patterns across Linguistic Features

Figure 3.47 on page 123 showed a number of generic elements with similar distributions, these contain the following elements:

Y *of* X, *burn* X, *den*, *stripe*, *slack*, *meur*, *grain*, Adjectival Scots names can also be added to the above list. A comparison of the Y *of* X names mentioned above and Scots adjectival terms show a similar distribution as in figure 7.1.

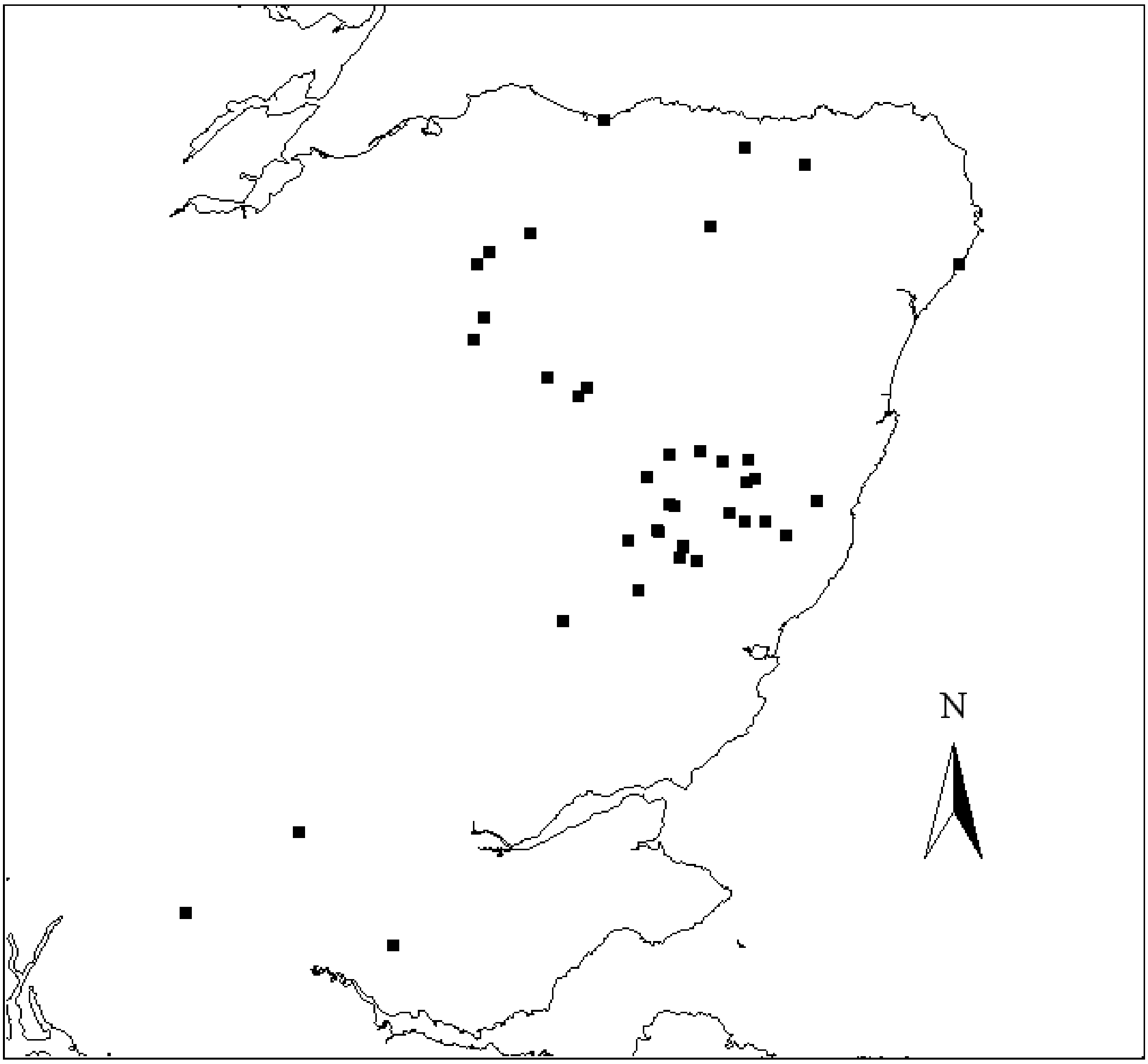
Figure 7.1: Various Similar Distributions



Scots names not falling in these categories do exist in these areas, but are also fairly evenly distributed throughout the other areas of the AOS. It seems that within the Scots hydronymicon X Y structures tend to be associated with adjectives, and Y *of* X structures with place-name / noun elements. Despite this, thirty-six names fulfil both criteria as in 7.2 on page 223, although it should be said all bar two of these names have Gaelic specific elements (the two exceptions are Burn of Sorrow, an odd hydronym in any case, and the water of Michall, only attested once.)

On structural grounds, the G term *meur* stands out as being the only non-Scots term included

Figure 7.2: Y *of* X Adjectival Scots Names



here. This may mean that in the AOS at least, *meur* is a translation of *grain*. This would suggest a gaelicisation of Scots names.

Broadly speaking, there are three areas where these names are located. A small patch exists in the Ochil Hills, and two larger areas exist. The more Southerly is approximately situated in the Eastern part of the Grampian Mountains, and the Northerly section is the Banff / Strathbogie area. There is a divide between the two larger sections which runs along the Dee / Don watershed. Why this should be is not precisely clear, beyond the fact that these areas denote the greatest concentration of Scots in the AOS. Much of these areas can be described as 'upland rural areas' where human subsistence activity took place. This is reflected by the number of Scots terms for small watercourses with various nuances. It may be significant that there are also overlaps with where Gaelic was once spoken but has been superseded, it could be interpreted as a somewhat wider version of the Highland Corridor. Although this would not explain its distribution completely, such as the lack of these terms in Angus and between the Dee and Don, it would explain why the *Burn* X and adjectival Y *of* X names largely have specific elements that are Gaelic in origin.

7.1.4 Default Qualities of Watercourses

Under section 4.7.3 on page 183 the concept of semantic defaults was discussed. These are broadly: 'wet', 'quiet', 'light-coloured', 'young', 'calm', 'shallow', 'good', 'straight', 'small' and 'front'. The watercourses these names represent tend to have certain qualities. Some of these qualities were outlined in 4.2.3 on page 132, others have come to light in the following discussion. In general these are:

1. The non-default names tend to be greater in number.
2. The actual terms that make up the non-default names tend to be less varied in number.
3. The components of *geogscores* for non-default terms tend to be more spread out and overall greater. This is shown in figure 4.2 on page 131 where many of these qualities have the highest *geogscores* of all the semantemes.
4. In name-pairs, there is evidence that the default term need not appear in the coupling. This would go some way to explaining point 1 above.

These qualities can be partially explained under the semantic notion of *markedness*[2] as discussed on page 132. This occurs in pairs of opposites, where one term is designated marked and the other unmarked.

The reason or reasons why these specific qualities emerge from the data as 'default' or 'unmarked' is not entirely clear; they must have originally derived from the landscape itself. It would be a useful exercise to investigate whether these terms are universally used to describe watercourses, or if it is restricted to Scotland, the British Isles or the Indo-European world. The remarkable consistency of these terms, spanning from the earliest Celtic strata up to the latest Germanic strata, suggest that this phenomenon exists at least on the Indo-European stage, and may even be a linguistic / toponymic universal.

Within generic elements, it seems clear that *burn* and *river* in Scots and *allt* and maybe *abhainn* in Gaelic are default terms for a watercourse. This means that when asked what a watercourse was, they would reply with this term. There is a correlation here between the *geogscore* ranges of these generic elements and those discussed in point three above, that is, the components of *geogscore* are spread out over a large conceptual area.

These correlations are manifestations of psycholinguistic processes which reflect how the brain processes language and handles perception. It goes somewhat beyond the bounds of this thesis, but the

[2]Cruse, *Meaning in Language*, p. 172.

data gathered here may prove a basis for a further research into historical cognitive linguistics. Work in this field has recently been carried out by Carole Hough of the University of Glasgow. In an article[3] she discusses the prototype theory, which is:

> "the notion that some members of a (semantic) class are more typical examples than others... This may offer an alternative to the unlikely hypothesis that... each 'cold stream' entailed the nearby presence of warmer streams... These and similar formations may rather have been regarded as the prototypical, or 'best examples' of their kind."[4]

The notion of prototypes[5] is clearly equivalent, if not identical, to the notion of defaults as discussed in this thesis. The application of the notion of prototypes in toponymy could be a valuable resource for cognitive linguistics on the whole, since (with the exception of the study of personal names) no other application gives concrete examples of the referent to which the prototypes are being attached. In other words, it is possible to investigate the phenomenon more clearly than otherwise, because one can compare the prototypes to physical objects (in this case watercourses) which themselves can be quantified. It is hoped that in the future the methodology in this thesis will discover more correlations and concrete examples for this fascinating sub-discipline of toponymy. This in turn may give us a unique window into the minds of the original coiners of the RNs.

7.1.4.1 A Syntax for the Hydronymicon

Throughout this thesis, the term hydronymicon has been used as a conceptual part of the larger toponymicon, this term is itself considered a part of the lexicon. The lexicon is a vocabulary of a given language.[6] A concrete example of a lexicon is a dictionary. A language, however, is made of both lexical items (words) and syntax. The conclusions drawn here are a reflection of the syntax of the hydronymicon. In other words, they reflect the acceptable ways in which the various generic elements, semantemes and specific elements can be combined to make RNs, in the same way the syntax of a language describes the way in which the various lexical items can be combined to make phrases, clauses and sentences.

7.1.5 Challenging or Confirming Currently Accepted Meanings

Throughout this thesis a number of case studies have been done to show the practical applications of the tools. These earlier studies tested well-established concepts and ascertained they were correct and also quantified them. Here is a short list of the types of problems that have been solved:

1. The methodology has been able to solve specific issues concerning the identity of a particular watercourse or RN. Such examples were shown in Abhainn Dubh and the Forth in section 2.6.2.1 on page 40 or Cart Burn in section 2.6.2.4 on page 45.

[3]Carole Hough, 'Commonplace Place-Names', *Nomina* 30 (2007).
[4]Ibid., p. 106.
[5]See also: Anderson, *The Grammar of Names*, p. 7-9.
[6]David Crystal, *An Encyclopedic Dictionary of Language and Languages* (St Ives, 1994), p. 227-228.

2. The methodology was also able to fill in ‘missing gaps’ in information. For example in the case study of Abercairny in section 3.9.1 on page 112, where **bothrie* is thought to be a Pictish RN, it was shown that the element *glas* is the likeliest generic to be attached to this name, despite the fact this is not directly in evidence anywhere.

3. These tools work not only with discrete units such as generic elements, but also with more nebulous semantic concepts. The use of the tools in section 4.3.15 on page 151 challenge proposed meanings for the Spey on the one hand and the Leven and Lomond on the other.

4. Even for obscure names, where the meaning and / or linguistic stratum may never to be known, the tools can make a suggestion as to the most likely outcome, as with the names Hadydarn Burn, Hodyclach Burn and Hoodiemart Burn in section 5.4.1 on page 198.

As discussed in chapter 2, this tool does not only suggest answers to specific problems, it can also carry out data-mining whereby a large number of names can be investigated, and any anomalies can be detected. Some examples are:

1. Section 2.6.2.2 on page 43 shows that the name ‘aqua de North’ representing the Don was probably not a name that was used in common speech.

2. Graph 4.3 on page 131 and 4.19 on page 165 show that the Dee and the Don should not be considered as ‘supernatural entities’ in the same sense as other RNs.

3. Section 5.4.2 on page 199 shows West Water and its upper reaches, called Water of Saughs, are likely to be Scots translations of older, Gaelic or P-Celtic names. This can be seen by the anomalously high amount of watercourses with Gaelic and earlier names which flow into them.

The case studies presented here are of course only a small selection of those that could be done. Whilst none of the case studies is able to make statements to an absolute degree of certainty, it is possible to say, for instance, Cart Burn could be an Old / Early Celtic name, but if it were, it was by far the shortest watercourse with an Old / Early Celtic name, and as such it is a very unlikely scenario.

7.2 Strengths and Resolution of Drawbacks of this Approach

7.2.1 Strengths

In section 1.2.5 on page 15 some strengths of this approach were discussed. The evidence presented in this thesis has shown that these strengths are valid. Further to this, other strengths can now be assessed in light of the conclusions which could have been unsatisfactory in two ways. Firstly, the results given in the functions such as *lingscorecomparer* could have been too vague to draw conclusions from. Secondly, the results could have been contrary to what could be reasonably explained.

In the first case, where the experiment would have been a failure, this has largely not occurred. Time and time again, viable results have arisen from the data. Whilst the data have rarely given very clear cut results, the trend has usually suggested a meaningful result. In some cases, the presentation of the data, especially from a visual standpoint, had to be altered to show the strongest correlations, but an effort has always been made to be consistent across data sets in this way, and I have also made an effort to be open about how the data have been presented, as can be seen from the source code in appendix D on page 265.

In the second case, a baffling result has been surprisingly rare. In most cases the direction or trend has been expected (for instance in a correlation between linguistic strata and a component of *geogscore*), but the rate of the trend has been unexpected. In one or two cases the opposite results have been presented. The most important cases were discussed above in section 7.1.1 on page 219.

Other strengths are the speed and relative ease of data visualisation and analysis. In the past, distribution maps did exist, but took a long time to create, were rarely consistent and often uncheckable. With R and other GIS applications one can quickly create consistent maps. The same can be said of the concept of 'the visualisation of conceptual space'. In the past this phrase was used in a fuzzy way, but in this thesis the conceptual space was mapped visually in the *lingscorecomparer* graphs.

7.2.2 Resolution of Potential Drawbacks

These issues were first flagged in section 1.2.6 on page 16.

1. The issue of indeterminacy discussed above is mitigated somewhat by the concept of the statistical approach. The nature of the statistical approach is such that, by gathering a large amount of data, even if some bits of data are not representative or incorrect, the amount of data will create an overall correct picture. Within the thesis here, even if some bits of information have been 'rounded off' to a slightly simplified state (such as assigning a *specelem* to each name), the existence, for instance, of a sine-wave in figure 2.3 on page 20 proves that the choices made to do this rounding were nevertheless meaningful.

2. The issue, mentioned above, of the difficulty of recording multiple generic element variation in old forms has been circumvented by a judicious use of these data at the appropriate points. It should also be stated that no other toponymic work has studied this phenomenon in this way before.

3. Potential drawbacks in the macro approach have been avoided. Issues of circular reasoning, too little data and issues with micro data which underpins the macro data have not occurred. In cases where these issues could have arisen, a warning has been given to this effect. As mentioned before, an undertaking such as this would be much easier for England, where the English Place Name Society Volumes have authoritatively covered nearly all RNs in the whole country. Within the Celtic linguistic landscape, works such as that of Simon Taylor[7], Richard Cox[8] and George Broderick[9] may one day become the data basis for taking this approach to a smaller area, with

[7]Taylor and Márkus, *The Place-Names of Fife*.
[8]Cox, *The Gaelic Place-names of Carloway*.
[9]Broderick, *Placenames of the Isle of Man*.

more toponymic data.

7.3 Ideas for the Future

7.3.1 Areas for Further Improvement

The improvement of this approach is dependent on two factors, the data and the methodology itself.

An attempt here has been made to gather a large amount of data. This could be further improved by gathering more data, e.g. expanding the AOS, and also by gathering more accurate data about the watercourses and RNs. Due to time constraints this was not possible, but a number of the factors mentioned in the methodology chapter above would further refine the approach. Particularly more precise historico-geological information about the condition of the landscape in the past would be especially useful. Despite this the AOS has proven to be a well-chosen area, encompassing an area where all the most important linguistic strata (excluding Norse) interact, including areas of particular interest to hydronymy such as: areas covered by Ptolemy's Geography, 'Pictland' and the Highland Corridor. The AOS also only encompasses the mainland; no islands have been taken into account. It is possible that the phenomena occurring here are actuated differently in the Western and Northern Isles, where it is not possible to have such large tributary networks. This was a conscious choice.

The algorithms and methodology used in this thesis, whilst clearly meaningful, could always be refined and improved. In fact, there are a number of free data-mining programs[10] which, with the necessary processing power, would be able to analyse the data here using 'machine learning' whereby rules and patterns can be discerned from large data sets. A version of that has been done in this thesis, but with the advances of computing power, it will soon be possible to perform much more complex tasks in far less time.

7.3.2 Applying the Methodology to Other Data Sets

As mentioned in the introduction, the most obvious expansion of this thesis would be to include RN data from the rest of Scotland, and even the rest of the British Isles. Aside from this, it would also be possible to apply the same or a similar methodology to other types of toponymic features. This section discusses a few possibilities.

7.3.2.1 Settlement Features

Settlement features are difficult to quantify by the fact that an important physical attribute of a settlement is its population, which is a very unstable statistic. The same issues exist with natural features but to a much lesser degree. The phenomenon of transferred names, where a name moves from its initial location to another site is present to a far lesser degree for natural features. Information such as altitude and aspect, however, could be quantifiable for names of small places. Whilst initial area of a

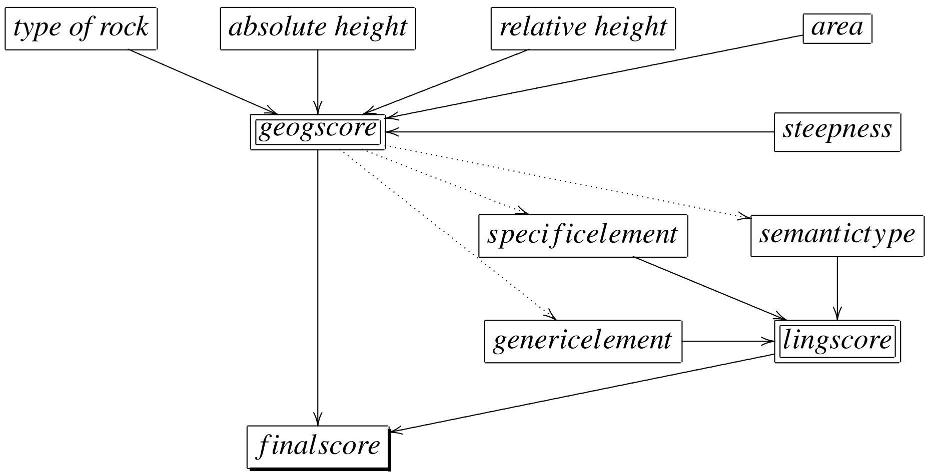
[10] Such as RapidMiner: `www.rapidminer.com`

settlement is difficult to ascertain, the distance between one settlement of a given type (or with a given generic element such as **pit*-) could possibly be gathered as a marker of initial size.

7.3.2.2 Natural Features

Natural features are far less susceptible to the issues mentioned above. Aside from watercourses, the other obvious type of feature would be hills. It would be relatively easy to gather information such as: height, height relative to surrounding hills, area, steepness, colour and type of stone. Much research could be done along similar lines into the relationship between, for example: the generic element and physical features, meaning of the specific element and physical features, and so on. Moreover, it would doubtless be easier to gather much of this information, since it is common for maps and books about hiking and so on to contain much of this information in a way that is lacking for watercourses. Table 7.3 is a version of 2.10 on page 33 but with these factors included.

Figure 7.3: Proposed Structure for a System for Hill-Names



An application of these analytical tools in this way would be an exciting one, not only to the field of hill-name study, but also as an exercise to test the methodology in this thesis in fields beyond hydronymy. The author looks forward to such work in the future.

Part II

Bibliography and Appendices

Bibliography

'National Library of Scotland: Map Collection' ⟨URL: `http://www.nls.uk/maps/`⟩.

'Charting the Nation: maps of Scotland and associated archives 1550-1740' ⟨URL: `www.chartingthenation.lib.ed.ac.uk`⟩.

Abraham, Ortelius, 'Scotiae Tablua', Map (1527-1598).

Adair, John, 'The East Part Of Fife', Map (1684).

Adair, John, 'The Turnings of the River Forth / Clackmanan', Map (1688).

Adair, John, 'Map Of Clackmannanshire / Map Of Strathdevon', Map (c. 1681).

Adomnan of Iona, *Adomnan's Life of St Columba* (Edinburgh, 1961).

Adomnan of Iona, *Life of St Columba* (London, 1995).

Ahlqvist, A., 'Two Ethnic Names in Ptolemy', *The Bulletin of The Board of Celtic Studies* 26 (1975), pp. 28–93.

Ainslie, John, 'Map of Fife', Map (1775).

Ainslie, John, 'Map of the County of Forfar or Shire of Angus', Map (1794).

Alexander, William, *The Place-names of Aberdeenshire* (Aberdeen, 1952).

Allard, Carel, 'Novissima Regni Scotiae septentrionalis et meridionalis tabula', Map (1697).

Anderson, John M., *The Grammar of Names* (Oxford, 2008).

Anderson, Marjorie O., ed., *Kings and Kingship in Scotland* (1973).

Bannatyne Club, ed., *Registrum Episcopatus Moraviensis* (Edinburgh, 1837).

Bannatyne Club, ed., *Registrum Episcopatus Brechensis* (Aberdeen, 1856).

Barrow, G. W. S., *Uses of Place-Names* (St Andrews, 1998), chap. The Uses of Place-names and Scottish History - Pointers and Pitfalls.

Bartholomew, J. G., *The Survey Atlas of Scotland* (Edinburgh, 1912).

Beveridge, Erskine, *The 'Abers' And 'Invers' of Scotland* (Edinburgh, 1923).

Birse, E. L., *Soil Survey of Scotland* (Aberdeen, 1970).

Black, George F., *The Surnames of Scotland* (Chippenham, 1946 reprinted 2004).

Blaeu, J., *The Blaeu Atlas of Scotland* (Edinburgh, 1654 (2006)).

Broderick, George, *Placenames of the Isle of Man* (Tübingen, 2005).

Caerwyn, Williams, 'WYSG (river-name), WYSG, HWYSGYNT, RHWYSG', *Studia Celtica* 21 (1990), pp. 670–678.

Campbell, Ewan, 'Were the Scots Irish?', *Antiquity* 75 (2001), pp. 285–292.

Charles, B. G., *Non-Celtic Place-Names in Wales* (London, 1938).

Clarke, Robert, ed., *The Book of Deer* (Edinburgh, 1869).

Coates, Richard and Breeze, Andrew, *Celtic Voices English Places* (Stamford, 2000).

Coronelli, Vincenzo, 'Le Royaume d'Escosse divisé en deux parties', Map (1689).

Coronelli, Vincenzo, 'Scotia: parte settentrionale', Map (1696).

Cox, Richard A. V., *The Gaelic Place-names of Carloway* (Isle of Lewis, 2002).

Cruse, Alan, *Meaning in Language* (Oxford, 2000).

Crystal, David, *An Encyclopedic Dictionary of Language and Languages* (St Ives, 1994).

Cummins, W. A., *The Age of The Picts* (Frome, 1995).

Cummins, W. A., *The Lost Language of the Picts* (Trowbridge, 2001).

Cunningham, Ian, *The nation survey'd: essays on late sixteenth-century Scotland as depicted by Timothy Pont* (East Linton, 2001).

Dodgshon, Robert A., 'Symbolic Classification and the Development of Early Celtic Landscape', *Cosmos: yearbook of the traditional cosmology society: Duality* (1985), pp. 61–83.

Dumville, Prof. David, *Rannsachadh na Gàidhlig* (Aberdeen, 2002), chap. Ireland and North Britain in the Earlier Middle Ages: Contexts for the Míniugud Senchasa Fher nAlban, pp. 185–211.

Dwelly, E., *The Illustrated Gaelic-English Dictionary* (Edinburgh, 1901-11).

Ekwall, Eilert, *Introduction to the Survey of English Place-Names* (Cambridge, 1925), chap. The Celtic Element, pp. 15–35.

Ekwall, Eilert, *English River-Names* (Oxford, 1928).

Ekwall, Eilert, *The Concise Oxford Dictionary of English Place-Names* (Oxford, 1940).

Evans, D. E., *Gaulish Personal Names* (Oxford, 1967).

Ferguson, Charles, 'Gaelic Names of Birds', *Transactions of the Gaelic Society of Inverness* 12 (1885-1886), pp. 327–333.

Forsyth, Kate, *Language in Pictland* (Utrecht, 1997).

Gamkrelidze, T. V. and Ivanov, V. V., *Trends in Linguistics: Indo-European and the Indo-Europeans* (New York, 1995).

Gordon, Robert, 'A detailed map including Glenmore, Findhorn', Map (c. 1636-52).

Gordon, Robert, 'Scotiae provinciae mediterraneae inter Taum flumen et Vararis aestuarium', Map (c. 1636-52).

Hall, J. R. Clark, *A Concise Anglo-Saxon Dictionary* (London, 1984).

Hamp, Eric, 'DOUBS', *Études Celtiques* XXV (1988), pp. 128–128.

Hamp, Eric, 'Varia', *Scottish Gaelic Studies* XVI (1990), pp. 191–195.

Hamp, Eric, 'Varia IV: Goidelic *alt* and *allt*', *Ériu* 43 (1992), pp. 209–211.

Henry, David, ed., *The Worm, The Germ And The Thorn: Pictish And Related Studies Presented To Isabel Henderson* (Balgavies, 1997).

Holder, *Alt-Celtischer Sprachschatz* (Leipzig, 1904).

Homann, Johann Baptist, 'Magnae Britannia : pars septentrionalis qua regnum Scotiae in suas partes et subja centes insulas divisum', Map (1663-1724).

Hough, Carole, 'Commonplace Place-Names', *Nomina* 30 (2007), pp. 101–120.

Isaac, G. R., 'The Antonine Itinerary Land Routes: Place-Names of Ancient Europe and Asia Minor. An electronic database with etymological analysis of the Celtic name-elements', CD-ROM (2002).

Isaac, G. R., 'Place-Names in Ptolemy's Geography', CD-ROM (2004).

Jackson, Kenneth, *Language And History In Early Britain* (Edinburgh, 1953).

Jackson, Kenneth, *The Problem of the Picts* (1955), chap. The Pictish Language, pp. 129–166.

Jackson, Kenneth, *The Gaelic Notes in the Book of Deer* (Cambridge, 1972).

Jansson, Jan, 'Scotia Provinciae intra Flumen Taum', Map (1659).

Johnston, J. B., *The Place Names of Stirlingshire* (Stirling, 1904).

Johnston, J. B., *The Place Names of Scotland* (London, 1934).

Johnston, J. G., *Come Fish With Me* (London, 1948).

Joyce, P. W., *The Origin and History of Irish Names of Places*, vol. 2 (London, 1973).

Kadmon, N., *Toponymy: The Lore, Laws and Language of Geographical Names* (New York, 2000).

Kenneth, C., *English Place-Names* (London, 1961).

King, Jacob, "Lochy' Names and Adomnan's Nigra Dea', *Nomina* 28 (2005), pp. 69–91.

King, Jacob, 'Endrick and Lunan', *The Journal of Scottish Name Studies* 1 (2007), pp. 150–156.

Kneen, John Joseph, *The place-names of the Isle of Man : with their origin and history* (Douglas, 1925-6).

Krahe, H., *Unsere ältesten Flussnamen* (Wiesbaden, 1964).

Lazar-Meyn, Heidi Ann, *Language Contact in the British Isles: Proceedings of the Eighth International Symposium on Language Contact in Europe* (Tübingen, 1991), chap. The Colour Systems of the Modern Celtic Languages: Effects of Language Contact, pp. 227–242.

Lewis, C. T., *Elementary Latin Dictionary* (Oxford, 1956).

Livingstone, M., ed., *Registrum Secreti Sigilli Regum Scottorum* (Edinburgh, 1908).

MacAirt, S., ed., *The Annals of Ulster* (Dublin, 1983).

MacAuly, Donald, 'Studying the Place Names of Bernera', *Transactions of the Gaelic Society of Inverness* 47 (1970-72), pp. 313–337.

MacBain, Alexander, *An Etymological Dictionary of the Gaelic Language* (Inverness, 1896).

MacDonald, A., *The Place-Names of West Lothian* (Edinburgh and London, 1941).

MacDonald, J, *Place Names of West Aberdeenshire* (Aberdeen, 1899).

MacFarlane, Walter, *Geographical collections relating to Scotland made by Walter Macfarlane* (1906).

Mackenzie, W. C., *Scottish Place-Names* (London, 1931).

Maclennan, M., *Gaelic Dictionary* (Aberdeen, 1979).

MacNeill, Peter B. G. and MacQueen, Hector L., *Atlas of Scottish History to 1707* (Edinburgh, 1996).

Margaret, Gelling and Cole, Anne, *The Landscape of Place-names* (Stamford, 2000).

Matheson, D., *Place Names of Elginshire* (Stirling, 1905).

Mercator, Gerhard, 'Scotiae Regnum south and north sheet', Map (1595).

Monier-Williams, Monier, *A Sanskrit English Dictionary* (Delhi, 1993).

Morgan, T., *The Place-Names of Wales* (Newport, 1912).

Nicolaisen, W. F. H., *Studies in Scottish Hydronymy unpublished B.Litt. Thesis* (Aberdeen, 1956).

Nicolaisen, W. F. H., 'Die alteuropäischen Gewässernamen der britischen Hauptinsel', *Beiträge Zur Namenforschung* 8 (1957), pp. 211–268.

Nicolaisen, W. F. H., 'The Semantic Structure of Scottish Hydronymy', *Scottish Studies* 1 (1957), pp. 211–240.

Nicolaisen, W. F. H., *Proceedings of The Eighth International Congress of Onomastic Sciences* (Netherlands, 1966), chap. Scottish Water-courses As Boundaries, pp. 211–240.

Nicolaisen, W. F. H., *The Picts And Their Place Names* (Inverness, 1996).

Nicolaisen, W. F. H., *Names Places and People* (Stamford, 1997), chap. The Dee at Chester and Aberdeen, pp. 117–125.

Nicolaisen, W. F. H., *Scottish Place-Names*, 2nd edition (Edinburgh, 2001).

Nicoll, E. H., ed., *Pictish Panorama: The Story of The Picts And a Pictish Bibliography* (Balgavies, 1995).

Ó Máille, T. S., 'Place-Name Elements -ar', *Ainm* 2 (1987), pp. 27–36.

Ó Máille, T. S., 'Irish Place-Names in -as, -es, -is, -os, -us', *Ainm* 4 (1989-1990), pp. 124–143.

Ó Murchú, Máirtín, *East Perthshire Gaelic* (Dublin, 1989).

Ordnance Survey, *1:25 000 Scale Explorer Series* (Southampton, 2003).

Ordnance Survey, *1:50 000 Scale Landranger Series* (Southampton, 2003).

Padel, O. J., *Cornish Place-Name Elements* (Cambridge, 1985).

Pokorny, *Indogermanisches Etymologisches Wörterbuch* (Munich, 1959).

Pont, Timothy, *Atlas Novus* (Amsterdam, 1662).

'Dictionary of The Irish Language' ⟨URL: `www.dil.ie`⟩.

Quin, E., ed., *Dictionary of The Irish Language* (Dublin, 1913-79).

Reany, P. H., *English Place Names* (London, 1977).

Rivet, A. L. F. and Smith, C., *The Place-Names of Roman Britain* (London, 1979).

Robertson, James, 'Topographical and military map of the counties of Aberdeen, Banff and Kincardine', Map (1822).

Robertson, Joseph, ed., *Illustrations of the Topography and Antiquities of the Shires of Aberdeen and Banff* (1847).

'Dictionary of The Scots Language' ⟨URL: `www.dsl.ac.uk`⟩.

Robinson, M., *The Concise Scots Dictionary* (Aberdeen, 1987).

Ross, D., *Scottish Place-names* (Edinburgh, 2002).

Rumsey, Deborah, *Statistics Workbook for Dummies* (Indiana, 2005).

Sanson, Nicolas, 'L' Escosse delà le Tay', Map.

Sinclair, John, ed., *The Statistical Account of Scotland* (Edinburgh, 1791-1799).

Sinclair, John, ed., *The New Statistical Account of Scotland* (1845).

Skene, W. F., *Chronicles of The Picts, Chronicles of the Scots* (Edinburgh, 1867).

Skene, W. F., *The Four Ancient Books of Wales* (Edinburgh, 1868).

Small, A., ed., *The Picts: A New Look at Old Problems* (Dundee, 1987).

Smith, Albert, *English Place-Name Elements*, vol. 25-26 (Cambridge, 1956).

Spittal, J. and Field, J., *A Reader's Guide To The Place-Names of The United Kingdom* (Stamford, 1990).

St. Andrews Priory, *Liber Cartarum Prioratus Sancti Andree In Scotia* (Edinburgh, 1841).

Stewart, George R., 'A Classification of Place Names', *Names* II no i (1954), pp. 1–13.

Stewart, George R., *Names on the Globe* (New York, 1975).

Stobie, James, 'South East Part of Perthshire', Map (1783).

Stone, J. C., *The Pont Manuscript Maps of Scotland: Sixteenth century origins of a Blaeu atlas* (Tring, 1989).

Stuart, J., ed., *The Exchequer Rolls of Scotland* (Edinburgh, 1878-1908).

Sutherland, E., *In Search of The Picts* (Britain, 2000).

Taylor, Allen, *SQL for Dummies* (Indiana, 2003).

Taylor, Simon, *Settlement Names of Fife PhD Thesis* (Edinburgh, 1995).

Taylor, Simon, 'Generic Element Variation, with Special Reference to Eastern Scotland', *Nomina* 20 (1997), pp. 5–22.

Taylor, Simon, ed., *The Uses of Place-Names* (St Andrews, 1998).

Taylor, Simon, *The Fife Book* (Edinburgh, 2000), chap. Place-Names of Fife.

Taylor, Simon, 'Place-names and Archaeology', *History Scotland* 3 no. 6 (2003), pp. 50–53.

Taylor, Simon, *Studies in the Book of Deer* (2008: Forthcoming), chap. Place-names in the Gaelic Notes in the Book of Deer.

Taylor, Simon and Márkus, Gilbert, *The Place-Names of Fife*, vol. 1 (Donington, 2006).

Thomas, Colin, 'Place-name Analysis in the Geographical Study of the Rural Landscape of Wales', *Studia Celtica* 9 (1973-74), pp. 299–318.

Thomas, R. J., *Enwau Afonydd a Nentydd Cymru* (Caerdydd, 1938).

Thomson, John Maitland, ed., *Registrum Magni Sigilli Regum Scottorum* (Edinburgh, 1882).

Thomson, T., ed., *Inquisitionum ad capellam domini regis retornatarum* (Edinburgh, 1811-16).

Thurneysen, R., *A Grammar of Old Irish* (Dublin, 1946).

Toner, Gregory, 'The Backward Nook: *Cúil* and *Cúl* in Irish Place-Names', *Ainm* 7 (1996), pp. 113–117.

Upton, Graham and Cook, Ian, *Oxford Dictionary of Statistics* (Oxford, 2002).

Various, *Geiriadur Prifysgol Cymru* (Aberystwyth, 1950-2002).

Vendryes, J., *Lexique Etymologie De L'Irlandais Ancien Dublin* (Paris, 1987).

Wainwright, F. T., ed., *The Problem of The Picts* (Edinburgh, 1955).

Watson, Adam and Allan, Elizabeth, *The Place Names of Upper Deeside* (Aberdeen, 1984).

Watson, Angus, *The Ochils: Placenames History Tradition* (Perth, 1995).

Watson, Angus, 'Place-Names Land and Lordship in the Medieval Earldom of Strathearn', Ph. D thesis, St Andrews (2002).

Watson, William J., *Place-Names of Ross And Cromarty* (Edinburgh, 1904).

Watson, William J., *The Celtic Place-Names of Scotland* (Edinburgh, 1993).

Watson, William J., *Scottish Place-Name Papers* (Edinburgh, 2002).

Wentworth, Roy, *Gaelic place-names of Beinn Eighe National Nature Reserve* (Perth, 1999).

Whittington, G. and Soulsby, J. A., 'Preliminary Report on Pit Place-Names', *Scottish Geographical Magazine* 84 (1968), pp. 205–220.

Wilkinson, John, 'Deep Thoughts on the Devon and a Fresh Look at the Nith', *Nomina* 25 (2002), pp. 139–145.

Withers, C. W. J., *Gaelic In Scotland 1698-1981* (Edinburgh, 1984).

Wodtko, D. S., *Sekundäradjektive in den altirischen Glossen* (Innsbruck, 1995).

Appendix A

Survey of Computing Resources

A.1 MySQL

MySQL[1] is a powerful database engine which can store a large amount of data and manipulate them in very complex ways. It also has the advantage of being free. MySQL and other equivalent database engines operate by a language called SQL, which stands for Structured Query Language. As an example, if one wished to retrieve a list of all the watercourses between 1 and 5 km in length, which had an etymology of 'black', one would enter:

```
select rn from list2 where km between 1 and 5 and etymology like ''%black%''
```

This is how the vast majority of tables have been created in this thesis. It also underpins nearly all the graphs and maps.

An expert in database management will probably notice that the database does not conform to many of the best practice concepts appropriate for websites, such as 'third normal form'[2] or ACID[3]. There are several reasons for this:

1. Joining tables to prevent the reduplication of common data (such as the word 'Gaelic') is best practice for fast servers which contain databases with millions of records. Since this database is relatively small, however, the issue of hard drive space does not exist, whilst the creation of joins slows down a normal computer. That said, many columns, such as *specelem* and *catch* were created with join tables.

2. Many of the best practices are used with an assumption that the database is a back end to a website or exists on a network with many users. In this case this assumption is not true.

[1] http://www.mysql.com/
[2] Allen Taylor, *SQL for Dummies* (Indiana, 2003), p. 116-117.
[3] Ibid., p. 278-278.

A.2 R

R[4] is a language and environment for statistical computing and graphics. It can connect with the MySQL database easily. In this thesis it was primarily used in three ways.

GIS I wrote a script called Onymagic which operates as a basic replacement for ArcGIS, tailored specifically creating distribution maps for British toponymy.

Data visualisation Some one-off graphs were tailor-made, but many are very similar with just slightly different queries entered such as *lingscorecomparer*. For these I wrote a number of functions. This is the easiest way of running chunks of code.

Tables Whilst MySQL is very versatile, sometimes a coding environment is necessary to get to more complex representations of the data, such as the finder function *semtypefinder*, which runs a series of complex queries.

Like MySQL, R is free and open source. Please see the code in section D below for specific source code of these scripts.

A.3 ArcGIS

ArcGIS[5] is an integrated collection of software products for building a complete Geographical Information System (GIS). The program has many uses; for the purposes here it enables the user to rapidly create distribution maps and other choropleth maps. It can access MySQL databases and spreadsheet files. The main drawback is the cost of ArcGIS which is over a thousand pounds. Because of this, I wrote an R script which performed many of the same functions called Onymagic (see section on R above).

A.4 Server2Go

The CD-ROM accompanying this thesis was created with a program called Server2Go[6] which allows database driven websites to run on a CD-ROM or other removable media. The site allows the user to browse the MySQL database in a user friendly way. The script manager used is PERL[7]. This product is free but not open source.

[4] `http://www.r-project.org/`
[5] `http://www.esri.com/software/arcgis/`
[6] `http://www.server2go-web.de/`
[7] `http://www.perl.org/`

A.5 EDINA Digimap

EDINA Digimap[8] is a group of cartographic resources. It is available only through subscription. Most of the OS data were gathered using Digimap Carto, a Java Web-based mapping application. It is also possible to download the OS Landranger data from this site, as discussed on page 12.

[8] `http://edina.ac.uk/digimap/`

Appendix B

Tables

B.1 Primary tables

This section contains the tables which appear in the database directly. They do not themselves contain information but are a key to the data in the database itself .

B.1.1 Geographical tables

Table B.1: *Catch*

catch	id	RN
Unknown	1	Unknown
DEE	2	River Dee
DON	3	River Don
DTS	4	Don to Spey
FOR	5	River Forth
FTT	6	Forth to Tay
LOM	7	Loch Lomond
SPE	8	River Spey
TAY	9	River Tay
TTD	10	Tay to Dee

Table B.2: *Altitude*

Range	Description
0-200m	Lowland
200-400m	Foothill
400-800m	Upland
>800m	Mountain

Table B.3: *Catch2*

catch2	id	RN
	1	Unknown
ALL	2	River Allan
ALM	3	River Almond
AVO	4	River Avon
BER	5	Bervie Water
BLA	6	Black Devon
BRA	7	River Braan
BUC	8	Water of Buchat
CAR	9	Carron Water
CLU	10	Clunie Water
COW	11	Cowie Water
CUL	12	Burn of Cullen
DES	13	Deskry Water
DEV	14	River Devon
DOC	15	River Dochart
DRU	16	River Druie
DUC	17	Duchray Water
DUL	18	River Dulnain
EAR	19	River Earn
END	20	Endrick Water
ERN	21	Ernan Water
EYB	22	Ey Burn
FAL	23	River Falloch
FES	24	River Feshie
FEU	25	Water of Feugh
FID	26	River Fiddich
GAI	27	River Gairn
GEL	28	Geldie Burn
ISL	29	River Isla
LEV	30	River Leven
LIV	31	River Livet
LUI	32	Lui Water
LYO	33	River Lyon
MUI	34	River Muick
MUL	35	Burn of Mulben
NET	36	River Nethy
NOC	37	Water of Nochty
NOR	38	River North Esk
ORE	39	River Ore
QUO	40	Quoich Water
SOU	41	River South Esk
TAN	42	Water of Tanar
TEI	43	River Teith
TRO	44	River Tromie
TRU	45	River Truim
TUM	46	River Tummel
UGI	47	River Ugie
URI	48	River Urie
YTH	49	River Ythan

Table B.4: Soil Classification

Climatic conditions	class	class2
Warm dry lowland	EE	1E
Warm rather dry lowland	EH	1H
Warm moist lowland	EM	1M
Warm rather wet lowland	ER	1R
Warm wet lowland	EV	1V
Fairly warm dry lowland	LE	2E
Fairly warm rather dry lowland	LH	2H
Fairly warm moist lowland and foothill	LM	2M
Fairly warm rather wet lowland and foothill	LR	2R
Fairly warm wet lowland and foothill	LV	2V
Cool rather dry lowland	MH	3H
Cool moist lowland and foothill	MM	3M
Cool rather wet lowland, foothill and upland	MR	3R
Cool wet foothill and upland	MV	3V
Cold rather wet foothill and upland	SR	4R
Cold wet upland	SV	4V
Very cold wet upland and mountain	VV	5V
Extremely cold wet mountain	ZV	6V

Table B.5: *Class*

class	id	class	id	class	id
	1	ER LR LV	57	LR MR MV	113
EE	2	ER LR LV MV	58	LR MR MV SV	114
EE EH	3	ER LR LV MV SR SV	59	LR MR MV SV VV	115
EE EH EM LM	4	ER LR LV MV SV	60	LR MR SR	116
EE EH LH	5	ER LR LV MV SV VV	61	LR MR SR SV	117
EE EH LH LM	6	ER LR MV	62	LR MR SV	118
EE EH LH LM LR MR	7	ER LR MV LV SV	63	LR MV	119
EE EH LH LM MM MR MV	8	ER LR MV SV	64	LR MV SV	120
EE EH LM	9	ER LV	65	LR MV SV VV	121
EE EH LR	10	ER LV MV	66	LR MV SV VV ZV	122
EE LE	11	EV	67	LR SR	123
EE LE LH	12	EV LV	68	LR SV	124
EE LE LH LM	13	EV LV MV	69	LR SV VV	125
EE LE LH LM LR MR SV	14	EV LV MV SV	70	LV	126
EE LE LH LM MM	15	LE	71	LV MV	127
EE LH	16	LE LH	72	LV MV SV	128
EE LH LM	17	LE LH LM	73	LV MV SV VV	129
EE LM LR	18	LH	74	LV MV VV	130
EH	19	LH LM	75	LV SV	131
EH EM	20	LH LM MM MR SR SV VV	76	MM	132
EH EM ER	21	LH LM MM SR	77	MM LM	133
EH EM ER EV LV	22	LH LM MR	78	MM MR	134
EH EM ER LR	23	LM	79	MM MR SR	135
EH EM LM	24	LM EH	80	MM MR SR SV	136
EH EM LM LR	25	LM EM	81	MM MR SR SV VV	137
EH EM LM MR SV VV ZV	26	LM LH	82	MM MR SV VV	138
EH EM LR	27	LM LR	83	MM MV SV	139
EH EM LR LV MV SV	28	LM LR MR	84	MM SR	140
EH EM LR MR	29	LM LR MV	85	MM SR SV	141
EH EM LR MV	30	LM MM	86	MM SV	142
EH LH	31	LM MM LR	87	MR	143
EH LH LM	32	LM MM MR	88	MR MV	144
EH LM	33	LM MM MR MV	89	MR MV SV	145
EH LM LR	34	LM MM MR SR	90	MR MV SV VV	146
EH LR	35	LM MM MR SR SV	91	MR SR	147
EH MM	36	LM MM MR SR SV VV	92	MR SR SV	148
EM	37	LM MM MR SV	93	MR SR SV VV	149
EM ER	38	LM MM MR SV VV	94	MR SV	150
EM ER LR	39	LM MM SR	95	MR SV VV	151
EM ER LR LV	40	LM MM SR SV	96	MV	152
EM ER LR MV	41	LM MM SV	97	MV SR	153
EM ER LV	42	LM MR	98	MV SV	154
EM LM	43	LM MR MV	99	MV SV VV	155
EM LR	44	LM MR MV SV VV	100	MV SV VV ZV	156
EM LR LV	45	LM MR SR	101	MV VV	157
EM LR LV MV	46	LM MR SR SV	102	SR	158
EM LR MR	47	LM MR SV	103	SR MR	159
EM LR MR MV	48	LM MV	104	SR SV	160
EM LR MR MV SV	49	LM MV SR	105	SR SV VV	161
EM LR MV	50	LM MV SV	106	SR VV	162
EM LR MV SV	51	LR	107	SV	163
EM MR MV	52	LR LV	108	SV VV	164
ER	53	LR LV MV	109	SV VV ZV	165
ER EV	54	LR LV MV SV	110	VV	166
ER EV LV	55	LR LV SV	111	VV ZV	167
ER LR	56	LR MR	112	ZV	168

B.1.2 Linguistic tables

Table B.6: *Source*

source	description	id
F1	Old forms attested from only one / dubious source	2
F2	Old forms attested from more than one / trustworthy source	3
OS	Ordnance Survey form	4
PL	Pleonastic	5
PN	Place name	6

Table B.7: *Specelem* (*et al*)

id	stratum
1	Obscure / None
2	Old Celt
3	OC / P-Celtic
4	P-Celtic
5	P-Celtic / Gaelic
6	Gaelic
7	Scots
8	Latin

Note on table B.8 on page 249: Similar lists of generic elements have been constructed before, but this one was done from scratch. It will be seen that there are several equivalents from one stratum to another, e.g. *Allt na* X and *Burn of* X are analogous to each other, further down the list these analogies break down, but they are there for as long as is useful. That is, if one wanted to search for a list of all RNs with *Allt na* X and *Burn of* X, one would search for *genelem2* as 1 and not specify *genelem1*, however if one wanted to search for only *Allt na* X, one would then specify *genelem2* as 3.

Table B.8: *Genelem2*

id	Gaelic	Scots	P-Celtic	Latin
0				
1	Allt X	Burn X		fluvium
2	Allt na X	Burn of X		torrens
3	X Allt	X Burn		
4	Alltan			
5	Uisge	X Water	esk	aqua
6		Water of X		
7	Abhainn	River	dobhar	
8	An X (only)	(The) X (only)		
9	Gleann	Glen		
10	Glas	Slack	glas	
11	Fèith	Bog		
12	Caochan	Stank		
13	Eas(an)	Falls		
14	Àth / Cul-àth	Den		
15	Lag	Latch		
16	Meur	Grain		
17		Stripe		
18	Other	Other		
19	Loch X	Loch X		
20	X Loch	X Loch		
21	Loch a’ X	Loch of X		lacus
22	Lochan	Pond		
23	Ailnaig	Lake		
24	Poll	Pool	pol	
25	Linn	Pow		
26		Pow of X		

Table B.9: *Genelem3*

id	order
0	Simplex
1	Specific + Generic
2	Specific + Generic + Genitive + Specific
3	Specific + Generic + Specific
4	Specific + Specific + Generic
5	Generic + Specific
6	Generic + Genitive + Specific
7	Generic + Specific + Specific
8	Generic + Specific + Genitive + Specific

B.2 Derived tables for scores

The layout of some of these tables differ slightly from that in the database, for instance, some of the columns have been altered slightly to make them more easily appreciable to the eye. The data are

Table B.10: *Semtype*

1	Water Word	Water Word	1
2	Adjective	Colour	2
3	Adjective	Smell / Taste / Feel	2
4	Adjective	Manner	2
5	Adjective	Sound	2
6	Adjective	Temperature	2
7	Adjective	Other	2
8	Adjective	Course	2
9	Adjective	Effect / Character	2
10	Adjective	Dimensions	2
11	Adjective	Bed	2
12	Adjective	Number	2
13	Adjective	Moistness	2
14	Adjective	Age	2
15	Adjective	Elevation	2
16	Ecosystem	Flora	3
17	Ecosystem	Fauna	3
18	Topography	Concavity e.g. Glen	4
19	Topography	Convexity e.g. Ben	4
20	Topography	Land around	4
21	Topography	Water feature	4
22	Topography	Specific natural feature	4
23	Human	Non specific agricultural area	5
24	Human	Non specific settlement / building	5
25	Human	Specific man-made area	5
26	Human	Agricultural object / structure	5
27	Human	Specific person / occupation	5
28	Human	Supernatural entity	5
29	Human	Event	5
30	Situation	Relation to other features	6
31	Situation	Boundary	6
32	Situation	Crossing	6
33	Other	Other	7
34	Topography	Material / Object	4
35	Topography	Weather / Air	4
36	Adjective	Visibility	2

Table B.11: *Adjtype*

id	*semtype2*	*semtype4*	*semtype5*	*semtype6*	*semtype8*
1	light	gentle	quiet	cold	straight
2	dark	rough	loud	hot	crooked
id	*semtype10*	*semtype13*	*semtype14*	*semtype36*	
1	big	dry	young	exposed	
2	small	wet	old	hidden	

Table B.12: *Pntype*

Meanings of elements contained in places after which watercourse is named	
Concavity	1
Convexity	2
Body of water	3
Land not used for agriculture	4
Land used for agriculture	5
Riparian area	6
Other	7

identical.

Table B.13: *Genelemscore*

average *genelem*	element	order
1.000	Uisge	Specific + Generic + Specific
1.713	Àth / Cul-àth	Generic + Specific
1.835	Lochan	Generic + Specific + Genitive + Specific
1.932	Fèith	Specific + Generic + Specific
2.186	Àth / Cul-àth	Simplex
2.257	Loch	Specific + Generic + Specific
2.344	Caochan	Generic + Specific + Genitive + Specific
2.408	Caochan	Generic + Specific + Specific
2.626	Other	Specific + Generic
2.661	Allt	Generic + Specific + Genitive + Specific
2.695	Àth / Cul-àth	Specific + Generic + Genitive + Specific
2.695	Uisge	Generic + Specific + Genitive + Specific
2.748	Caochan	Specific + Generic + Specific
2.772	Meur	Generic + Specific
2.792	Lochan	Generic + Genitive + Specific
2.978	Latch	Generic + Genitive + Specific
2.994	Lochan	Specific + Generic
3.064	Lag	Generic + Genitive + Specific
3.091	Lochan	Generic + Specific
3.204	Àth / Cul-àth	Specific + Generic

Table B.13: *Genelemscore*

average *genelem*	element	order
3.204	Pool	Generic + Genitive + Specific
3.268	Loch	Specific + Generic
3.283	Uisge	Generic + Specific
3.333	Alltan	Specific + Generic
3.347	Alltan	Specific + Generic + Specific
3.359	Grain	Specific + Generic
3.362	Grain	Generic + Genitive + Specific
3.362	Stripe	Specific + Generic + Genitive + Specific
3.551	Loch	Generic + Genitive + Specific
3.574	Caochan	Generic + Specific
3.580	Fèith	Generic + Specific
3.612	Alltan	Generic + Specific
3.617	Ailnaig	Generic + Specific
3.639	Allt	Generic + Specific + Specific
3.664	Other	Generic + Genitive + Specific
3.709	Meur	Generic + Genitive + Specific
3.819	Other	Generic + Specific
3.820	Caochan	Specific + Generic
3.840	Other	Simplex
3.871	Bog	Generic + Genitive + Specific
3.908	Other	Generic + Genitive + Specific
3.973	Grain	Specific + Generic + Genitive + Specific
4.175	Eas(an)	Generic + Specific
4.247	Stripe	Specific + Generic
4.267	Stripe	Generic + Genitive + Specific
4.286	Allt	Generic + Specific
4.288	Allt	Generic + Genitive + Specific
4.292	Allt	Specific + Generic + Specific
4.298	Slack	Specific + Generic
4.355	Gleann	Specific + Generic
4.462	Eas(an)	Specific + Generic
4.533	Caochan	Generic + Genitive + Specific
4.558	Glen	Generic + Specific
4.570	Eas(an)	Generic + Genitive + Specific
4.599	Poll	Generic + Specific
4.602	Loch	Simplex
4.607	Stank	Specific + Generic

Table B.13: *Genelemscore*

average *genelem*	element	order
4.635	Allt	Specific + Generic
4.650	Àth / Cul-àth	Generic + Genitive + Specific
4.686	Uisge	Specific + Generic
4.784	Other	Specific + Generic
4.788	Lag	Generic + Specific
4.788	Pool	Specific + Generic
4.871	Burn	Specific + Generic + Genitive + Specific
4.874	Meur	Specific + Generic
4.929	Loch	Specific + Generic
4.958	Allt	Specific + Specific + Generic
4.977	Other	Simplex
4.982	Loch	Generic + Specific
5.100	Latch	Specific + Generic
5.200	An X (only)	Simplex
5.297	Poll	Generic + Genitive + Specific
5.376	Loch	Generic + Specific
5.394	Burn	Generic + Genitive + Specific
5.431	(The) X (only)	Simplex
5.555	Burn	Specific + Specific + Generic
5.619	Other	Generic + Specific
5.699	(The) X (only)	Specific + Generic
5.747	Burn	Specific + Generic
5.827	Burn	Generic + Specific
5.836	Pond	Specific + Generic
5.918	Den	Specific + Generic
5.953	Slack	Generic + Genitive + Specific
5.958	Gleann	Generic + Specific
5.982	Lake	Specific + Generic
6.042	Loch	Generic + Genitive + Specific
6.086	Uisge	Simplex
6.345	Pond	Generic + Genitive + Specific
6.369	glas	Specific + Generic
6.371	Glen	Specific + Generic
6.462	Den	Generic + Specific
6.462	Den	Generic + Genitive + Specific
6.529	Abhainn	Generic + Specific
6.577	Abhainn	Specific + Generic

Table B.13: *Genelemscore*

average *genelem*	element	order
6.707	Water	Generic + Specific
6.854	Lake	Generic + Genitive + Specific
7.047	dobhar	Specific + Generic
7.272	Pow	Specific + Generic
7.363	pol	Specific + Generic
7.862	Water	Specific + Generic
8.141	Water	Generic + Genitive + Specific
8.506	Pow	Generic + Genitive + Specific
8.773	pol	Generic + Specific
9.008	esk	Specific + Generic
9.517	Water	Specific + Specific + Generic
10.000	River	Generic + Specific

Table B.14: *Specelemscore*

specelemscore	stratum
1.000	Gaelic
1.726	Scots
2.458	Obscure / None
4.290	Pictish / Gaelic
4.466	Pictish
7.135	OC / Pictish
10.000	Old Celt

Table B.15: *Semtypescore*

semtype-score	*semtype*	type	name	frequency
1.00	3602	Adjective	Visibility: hidden	19
1.66	2210	Topography	Specific natural feature: Concavity	664
1.67	200	Adjective	Colour: other	96
1.72	3200	Situation	Crossing:	44
1.84	3500	Topography	Weather / Air:	17
1.93	3000	Situation	Relation to other features:	198
2.10	601	Adjective	Temperature: cold	6
2.92	3400	Topography	Material / Object:	47
3.03	1600	Ecosystem	Flora:	288
3.04	801	Adjective	Course: straight	6

Table B.15: *Semtypescore*

semtype-score	*semtype*	type	name	frequency
3.06	2230	Topography	Specific natural feature: Body of water	318
3.20	1700	Ecosystem	Fauna:	334
3.21	202	Adjective	Colour: dark	222
3.28	2600	Human	Agricultural object / structure:	23
3.30	1301	Adjective	Moistness: dry	31
3.32	2400	Human	Specific person / occupation:	268
3.41	1002	Adjective	Dimensions: small	54
3.45	1000	Adjective	Dimensions: other	7
3.66	802	Adjective	Course: crooked	55
3.72	1401	Adjective	Age: young	7
3.73	3100	Situation	Boundary:	53
3.81	1900	Topography	Convexity e.g. Ben:	187
3.81	1502	Adjective	Elevation: low	8
3.87	2200	Topography	Specific natural feature:	56
4.01	2220	Topography	Specific natural feature: Convexity	604
4.06	201	Adjective	Colour: light	101
4.09	2260	Topography	Specific natural feature: Riparian area	63
4.11	3300	Other	Other:	14
4.11	300	Adjective	Smell / Taste / Feel:	18
4.12	1001	Adjective	Dimensions: big	61
4.15	2270	Topography	Specific natural feature: Other	7
4.15	2250	Topography	Specific natural feature: Land used for agriculture	179
4.15	1800	Topography	Concavity e.g. Glen:	174
4.22	2700	Human	Non specific settlement / building:	71
4.30	2000	Topography	Land around:	319
4.30	2530	Human	Specific man-made area: Body of water	5
4.45	2240	Topography	Specific natural feature: Land not used for agriculture	188
4.49	2300	Human	Non specific agricultural area:	94
4.78	402	Adjective	Manner: rough	92
4.81	2900	Human	Event:	5
4.98	1200	Adjective	Number:	6
5.21	2500	Human	Specific man-made area:	48
5.24	502	Adjective	Sound: loud	54
5.39	2350	Human	Non specific agricultural area	18

Table B.15: *Semtypescore*

semtype-score	*semtype*	type	name	frequency
5.42	2550	Human	Specific man-made area: Land used for agriculture	492
5.46	700	Adjective	Other:	23
5.49	2100	Topography	Water feature:	106
5.75	2540	Human	Specific man-made area: Land not used for agriculture	225
5.79	1100	Adjective	Bed:	33
5.96	2560	Human	Specific man-made area: Riparian area	54
6.04	2800	Human	Supernatural entity:	19
6.07	900	Adjective	Effect / Character:	68
6.59	501	Adjective	Sound: quiet	4
6.74	3601	Adjective	Visibility: exposed	5
7.52	1402	Adjective	Age: old	8
8.57	1501	Adjective	Elevation: high	7
8.87	602	Adjective	Temperature: hot	7
9.01	1302	Adjective	Moistness: wet	7
10.00	401	Adjective	Manner: gentle	16

Appendix C

Gazetteer Information

The database can be accessed from the CD-ROM which accompanies this thesis.

C.1 The Main Table

The primary table is called *list2*. In the introductory section, the types of information catalogued were outlined in general terms, in the following section the specifics are given. The terminology in brackets is the data type for the column using MySQL terminology. In many column refers to a key which contains the actual relevant data. For example, in the column *specelem* 'Gaelic', 'P-Celtic' and so on are not repeated hundreds of times, but instead Gaelic represents 6, P-Celtic 4 and so on as listed in table B.7 on page 248. In another table the specific strata are listed. This saved disk space (since the word 'Gaelic' is not repeated hundreds of times) and thus speeds up queries. "Links to:" explains which tables the column links to.

id (int(4)) Identification number: unsigned. An incremental number to differentiate each entry. The order of the id roughly follows that of the alphabetic order of the specific element but there are many exceptions.

RN (varchar(40)) River name: The name of the watercourse.

catch (int(2) unsigned) Primary catchment area: The overall catchment area to which the watercourse belongs. Note that strictly some areas are not catchment areas as such, such as Forth to Tay, but represent a number of watercourses flowing into the sea. Links to table B.1.

catch2 (int(2) unsigned) Secondary catchment area: Within the table B.1 there are smaller watercourse, which are still of considerable length, for instance within the Tay catchment area are rivers such as the Isla, the Garry and the Tummel. Links to table B.3.

coord (varchar(8)) Coordinates: This is a 6 digit OS grid reference accurate to 100m. In cases where the river is larger than this area, or the area is not exactly known, a 4 digit reference is given.

LAT (int(6)) Latitude: OS easting. This is essentially the same data as the coord above, but ArcGIS and a number of other programs prefer various formats of the data, so both are present.

LON (int(6)) Longitude: OS northing. See LAT above

class (varchar(20)) Soil class: This lists all the types of soil through which the river flows according to the Soil Classification. The list in each cell represents all the different soil classifications through which the watercourses run, starting with the lowest point. Links to table B.4

class2 (longtext) Soil class: This lists the same information as above but corresponds to column *class2* in table B.4. It is technically redundant.

firstclass (longtext) First soil class: This lists the soil classification of the lowest point on the watercourse. It is technically redundant.

lastclass (longtext) Last soil class: This lists the soil classification of the highest point on the watercourse. It is technically redundant.

altmin (int(3)) Minimum altitude: This and the following fields represent the minimum, average and maximum altitude of the watercourse. This is derived from the soil classification system as shown in table B.2 on page 245.

altave (int(3)) Average altitude: See *altmin* entry.

altmax (int(3)) Maximum altitude: See *altmin* entry.

km (decimal(4,1)) Kilometres: The general extent of the watercourse as discussed in the introduction above. The length is measured in kilometres, to the nearest .5, except for watercourses shorter than that length, which are all rounded to 0.5.

pos (int(11)) Position: Position of the watercourse. The number here represents which column the name is in, in the hierarchically arranged RNs section of the database. E.g. if the watercourse flows into the sea, the RN has a pos of 1, etc.

nont (int(11)) Number of named tributaries: An integer also derived from the hierarchically arranged RNs section of the database.

parbound (varchar(20)) Parish boundaries: This is a text explaining whether any part of the watercourse runs (or ran) along a parish boundary.

countbound (varchar(20)) County boundaries: This is a text explaining whether any part of the watercourse runs (or ran) along a county boundary.

primpn (varchar(65) Primary place-name: If the RN takes its specific element directly from a place name, the name should go here. E.g. ‘Aboyne’ and ‘Burn of Aboyne’.

secpn (varchar(65) Secondary : If there are any place-names which take their names directly from the watercourse, the name should go here. E.g. ‘Corrie of Allt nan Aighean’ and ‘Allt nan Aighean’.

relatpn (varchar(65) Related place-name: If there are any place-names which derive from a feature or element from which the RN also derives, or, if there are any other names of relevance, the name should go here. E.g. ‘Coire nan Aighean’ and ‘Eas nan Aighean’.

genelem1 (int(2)) Generic element: An integer corresponding to the linguistic stratum to which the generic element belongs, i.e. Gaelic or Scots. Links to table B.7.

genelem2 (int(2)) Generic element: An integer corresponding to the individual generic element to contained in the RN, i.e. *allt* or *burn*. Links to table B.8.

genelem3 (int(2)) Generic element: An integer corresponding to the syntactic structure of the generic element in relationship to the specific element, i.e. Allt Dubh vs. Dubh Allt, or Burn of Corrie vs. Corrie Burn. Links to table as in table B.9.

specelem (int(2)) Specific element: An integer corresponding to the linguistic stratum of the specific element. Links to table B.7.

phono (int(2)) Phonology: An integer corresponding to the phonology of the RN according to table B.7.

orthog (int(2)Orthography: An integer corresponding to the orthography of the specific element. Links to table B.7.

pleon1 (int(2)) Pleonastic element: If the RN is ‘pleonastic’, then pleon1, pleon2 and pleon3 are the equivalent of *genelem1*, *genelem2* and *genelem3*. For instance, Dhualt Burn has for its pleonastic entries what Dubh Allt has for its *genelem* entries. Links to table B.7.

pleon2 (int(2)) Pleonastic element: See above. Links to table B.8.

pleon3 (int(2)) Pleonastic element: See above. Links to table as in table B.9.

semtype (int(2)) Semantic type: This integer relates to the semantic group from which the name derives. Links to table B.10.

semtype2 (int(2)) Secondary semantic type: This integer relates to the semantic group from which the name derives, if there is a secondary meaning, or more than one element. For example, East Burn of Builg has a primary meaning deriving from *builg* and a secondary from ‘East’. (The field etym2 refers to the actual secondary element). Links to table B.10

adjtype (int(1)) Adjectival type: If the semtype of the RN pertains to an adjective, this field separates binary oppositions. For example, for RNs named from colour, *adjtype* will distinguish between ‘light’ and ‘dark’, if the RN is coined for its age, *adjtype* will distinguish between ‘young/new’ and ‘old’. Links to table B.11.

adjtype2 (int(1)) Secondary adjectival type: This integer relates to the specific type of adjective. E.g. if the adjective relates to temperature, is it hot or cold.

pntype (int(2)) Place-name type: If the semtype of the RN pertains to an external place-name in some way, this field specifies the type of place. For instance, Dunino Burn relates to the place-name, Dunino, deriving from G *dùn*, 'hill, hillfort', which is classified as 'convexity'. The specific place-name is listed in the *primpn* field. Links to table B.12.

pnstrat (int(1)) Linguistic strata of the place-name: If the semtype of the RN pertains to an external place-name in some way, this field specifies the linguistic stratum from which the place-name (*not* the RN) derives. This is useful in RN such as Cupar Burn, whilst Cupar as a place-name is Pictish, the RN Cupar Burn has been judged to be coined in a Scots context. Links to table B.7.

2genelem1 (int(2)) Generic element: This and the following two entries correspond to the same tables as *genelem1*, *genelem2* and *genelem3* except they are given values only if there is an old form which has a different generic element, but same specific element compared to the RN entry. For example, Banvie Burn has as old forms: Auld Banavie (Stobie); Ald Banowy (c. 1591 Pont map 19). Links to table B.7

2genelem2 (int(2)) Generic element: See above. Links to table B.8

2genelem3 (int(2)) Generic element: See above. Links to table B.9

q (longtext) Query: This field contains a 'q' if the etymology given is uncertain, if it is even more uncertain 'qq' is given. This is similar to a simple question mark after the etymology in most place-name dictionaries.

spec (varchar(255)) Specific: The specific element of the RN is given here. If the name is Gaelic and lenition has occurred, the lenition has been removed in this field.

etymology (longtext) Etymology: This gives the specific element in the following format: Abbreviated linguistic stratum, dictionary form, meaning. e.g. G *achadh*, 'shieling'. For names requiring further discussion, see *comments*. Note that for hydronyms deriving from toponyms, only the specific element of the toponym is given, for instance, a name such as Allt a' Choire Buidhe, has an etymology of G *coire*, 'circular hollow', but the second element is not discussed, since it refers only to the *coire*, not the watercourse itself. If further etymologisation of a toponym is desired, this is found in comments. In the cases where the derivation is uncertain, as much information as possible is entered. For instance, if the RN is utterly obscure, 'obscure' is entered; if the name is demonstrably Gaelic, but nonetheless obscure 'G obscure' is entered. If the RN derives from a nearby place-name, which is obscure, then 'obscure PN element' is added. If this place-name element is obviously Gaelic then 'obscure G PN element' is entered, and so on.

etym1a (varchar(46)) The first section of etymology i.e. if etymology is G *achadh*, 'shieling', this is G achadh.

etym1b (varchar(64)) The second section of etymology i.e. if etymology is G *achadh*, 'shieling', this is shieling.

etym2 (longtext) Secondary etymology: If semtype2 above is used, the particular element goes here, in the same format as etymology.

oldforms (longtext) Old forms: Old forms are listed here, in a general text format. See section on sources for the accepted format.

othernames (varchar(90)) Other names: If the watercourse has at any time had another name this goes here. The name could have come from the old forms, or could be inferred from a settlement or pleonastic element, such as 'hadden' from Inverhadden Burn.

otherid (int(4)) Other ID number: This relates to the id of the name in the field othernames.

comments (longtext) Comments: This is a text area for any other comments or discussion on the name. See below.

source (longtext) Source: This column explains how the RN name is known. In the majority of cases, the name is "OS", that is, known from current Ordnance Survey data, but others are known only from old forms or settlements etc. Links to table B.6.

hiera (int(4))ID in hiera: This integer corresponds to the id in the hierarchical section of the database.

county (varchar(45)) County: The county in which the watercourse is situated (currently empty).

id2 (int(4)) Secondary ID: If two or more RNs are considered to be in pairs, the same id number is in this field for all relevant entries. For instance, North Ugie Water and South Ugie Water both have a unique id2 of 8.

posid (int(4)) Position ID: The ID of the watercourse into which this one flows. Only entries with a source of OS have a *posid* value. The number of occurrences of this number will be equal to the *nont* of the watercourse with the posid.

C.1.1 Old forms and Comments

The oldforms for the names follow the accepted format as seen in Taylor and Márkus, *The Place-Names of Fife* and so on, with a few exceptions: since the data are stored in a MySQL database it is not possible to italicise the oldforms. Also, the date information about the source is placed in brackets. A new line is marked by a semi-colon, which is rendered into a carriage return in the CD-ROM.

The comments relate to any discussion or, in the case where a secondary source has been used to derive the old forms, the appropriate reference is made here. The main titles mentioned here are: Simon Taylor and Gilbert Márkus, *The Place-Names of Fife*, vol. 1 (Donington, 2006); William J. Watson, *The Celtic Place-Names of Scotland* (Edinburgh, 1993); J MacDonald, *Place Names of West*

Aberdeenshire (Aberdeen, 1899); William Alexander, *The Place-names of Aberdeenshire* (Aberdeen, 1952); Angus Watson, 'Place-Names Land and Lordship in the Medieval Earldom of Strathearn', Ph. D thesis, St Andrews (2002); Adam Watson and Elizabeth Allan, *The Place Names of Upper Deeside* (Aberdeen, 1984) and Angus Watson, *The Ochils: Placenames History Tradition* (Perth, 1995).

The abbreviations on page 263 are used in the database for certain common manuscripts and maps. The dates appearing in the list are those of publication, they are not necessarily the same as the recorded dates as listed in the old forms.

C.2 The River Hierarchy

This table, called *hiera2* in the database, contains all the RNs whose source is from the Ordnance Survey listed as shown in table 1.1 on page 12. Some names in this are not in the database itself. These are predominantly the 'east side', 'west side' markers for large lochs. Also occurring are the odd small feature not included in the database, (such as Rooking Linn) or an unnamed burn (marked UNB) which are needed to explain the hierarchy but are not linguistically significant. This table is not used much in specific calculations but it is a useful visualisation of the hierarchical network of rivers.

Table C.1: List of Abbreviations for Old Forms

Adair: Clackmannanshire = John Adair, 'Map Of Clackmannanshire / Map Of Strathdevon', Map (c. 1681)
Adair: Fife = John Adair, 'The East Part Of Fife', Map (1684)
Adair: Forth = John Adair, 'The Turnings of the River Forth / Clackmanan', Map (1688)
Ainslie: Angus = John Ainslie, 'Map of the County of Forfar or Shire of Angus', Map (1794)
Ainslie: Fife = John Ainslie, 'Map of Fife', Map (1775)
Allard = Carel Allard, 'Novissima Regni Scotiae septentrionalis et meridionalis tabula', Map (1697)
Antiquities = Joseph Robertson, ed., *Illustrations of the Topography and Antiquities of the Shires of Aberdeen and Banff* (1847)
AU = S. MacAirt, ed., *The Annals of Ulster* (Dublin, 1983)
Bartholomew = J. G. Bartholomew, *The Survey Atlas of Scotland* (Edinburgh, 1912)
Beveridge = Beveridge, *The 'Abers' And 'Invers' of Scotland*
Blaeu map = J. Blaeu, *The Blaeu Atlas of Scotland* (Edinburgh, 1654 (2006))
Blaeu text available from 'National Library of Scotland: Map Collection' ⟨URL: `http://www.nls.uk/maps/`⟩
Coronelli map 1 = Vincenzo Coronelli, 'Le Royaume d'Escosse divisé en deux parties', Map (1689)
Coronelli map 2 = Vincenzo Coronelli, 'Scotia: parte settentrionale', Map (1696)
RMS = John Maitland Thomson, ed., *Registrum Magni Sigilli Regum Scottorum* (Edinburgh, 1882)
REM = Bannatyne Club, ed., *Registrum Episcopatus Moraviensis* (Edinburgh, 1837)
REB = Bannatyne Club, ed., *Registrum Episcopatus Brechensis* (Aberdeen, 1856)
Dwelly = Dwelly, *The Illustrated Gaelic-English Dictionary*
Exch. Rolls = J. Stuart, ed., *The Exchequer Rolls of Scotland* (Edinburgh, 1878-1908)
Homann = Johann Baptist Homann, 'Magnae Britannia : pars septentrionalis qua regnum Scotiae in suas partes et subja centes insulas divisum', Map (1663-1724)
Jansson = Jan Jansson, 'Scotia Provinciae intra Flumen Taum', Map (1659)
MacDonald = MacDonald, *Place Names of West Aberdeenshire*
Macfarlane Geog. Coll. MacFarlane, *Geographical collections relating to Scotland made by Walter Macfarlane*
Mercator = Gerhard Mercator, 'Scotiae Regnum south and north sheet', Map (1595)
Ortelius = Ortelius Abraham, 'Scotiae Tablua', Map (1527-1598)
Pont map = Timothy Pont, *Atlas Novus* (Amsterdam, 1662)
Ptol. = Ptolemy's map derived from Isaac, 'Place-Names in Ptolemy's Geography'
Retours = T. Thomson, ed., *Inquisitionum ad capellam domini regis retornatarum* (Edinburgh, 1811-16)
Robert Gordon map 1 = Robert Gordon, 'A detailed map including Glenmore, Findhorn', Map (c. 1636-52)
Gordon Braid-Allaban = Robert Gordon, 'Scotiae provinciae mediterraneae inter Taum flumen et Vararis aestuarium', Map (c. 1636-52)
Robertson map= James Robertson, 'Topographical and military map of the counties of Aberdeen, Banff and Kincardine', Map (1822)
RSS = M. Livingstone, ed., *Registrum Secreti Sigilli Regum Scottorum* (Edinburgh, 1908)
Robert Gordon map 1 = Gordon, 'A detailed map including Glenmore, Findhorn'
Robert Gordon: Braid Allaban = Gordon, 'Scotiae provinciae mediterraneae inter Taum flumen et Vararis aestuarium'
Stobie = James Stobie, 'South East Part of Perthshire', Map (1783)
Stat. Acc. = John Sinclair, ed., *The Statistical Account of Scotland* (Edinburgh, 1791-1799)
Stat. Acc. = John Sinclair, ed., *The New Statistical Account of Scotland* (1845)

Appendix D

Source code

The code used here to generate the graphs and tables will largely give identical results to the graphs and tables in this thesis. In certain situations, however, there may be some discrepancies. This is partially because the graphs or tables were created at different times, and the data may have changed slightly since then. In other cases, I have occasionally tweaked the presentation of the final output to make it more visually understandable. In neither case has the data been significantly changed to undermine any argument.

It should be said that the code here worked specifically on the author's computer. The R code will probably not work without some modifications. The MySQL code, however, will work if copied and pasted into the relevant window on the CD-ROM.

D.1 MySQL views

The following code represents the MySQL views This is a MySQL table derived completely from another table or tables.

geogscore

```
select (round(((
if((‘list2‘.‘KM‘ between 0 and 1),1,
if((‘list2‘.‘KM‘ between 1.5 and 2),2,
if((‘list2‘.‘KM‘ between 2.5 and 4),3,
if((‘list2‘.‘KM‘ between 4.5 and 8),4,
if((‘list2‘.‘KM‘ between 8.5 and 16),5,
if((‘list2‘.‘KM‘ between 16.5 and 32),6,
if((‘list2‘.‘KM‘ between 32.5 and 64),7,
if((‘list2‘.‘KM‘ between 64.5 and 128),8,
if((‘list2‘.‘KM‘ > 128.5),9,NULL))))))))) +
if((‘list2‘.‘nont‘ between 0 and 4),1,
if((‘list2‘.‘nont‘ between 5 and 8),2,
if((‘list2‘.‘nont‘ between 9 and 16),3,
```

```
if((‘list2‘.‘nont‘ between 17 and 32),4,
if((‘list2‘.‘nont‘ between 33 and 64),5,
if((‘list2‘.‘nont‘ between 65 and 128),6,
if((‘list2‘.‘nont‘ between 129 and 256),7,
if((‘list2‘.‘nont‘ between 257 and 512),8,
if((‘list2‘.‘nont‘ > 513),9,NULL))))))))) +
abs((‘list2‘.‘pos‘ - 11))) + abs((((((left(‘list2‘.‘firstclass‘,1) *
 left(‘list2‘.‘lastclass‘,1)) - 1) / 35) * 9) + 1) - 11))),2) / 4) AS
  ‘score‘,‘list2‘.‘id‘ AS ‘id‘,‘list2‘.‘RN‘ AS ‘RN‘,  ‘list2‘.‘catch‘ AS ‘catch‘,
  ‘list2‘.‘catch2‘ AS ‘catch2‘,  ‘list2‘.‘coord‘ AS ‘coord‘,
  ‘list2‘.‘class‘ AS ‘class‘,  ‘list2‘.‘KM‘ AS ‘KM‘,
  ‘list2‘.‘parbound‘ AS ‘parbound‘,  ‘list2‘.‘countbound‘ AS ‘countbound‘,
  ‘list2‘.‘primpn‘ AS ‘primpn‘,‘list2‘.‘secpn‘ AS ‘secpn‘,
  ‘list2‘.‘relatpn‘ AS ‘relatpn‘,  ‘list2‘.‘etymology‘ AS ‘etymology‘,
  ‘list2‘.‘q‘ AS ‘q‘,  ‘list2‘.‘genelem1‘ AS ‘genelem1‘,
  ‘list2‘.‘genelem2‘ AS ‘genelem2‘,  ‘list2‘.‘genelem3‘ AS ‘genelem3‘,
  ‘list2‘.‘specelem‘ AS ‘specelem‘,  ‘list2‘.‘phono‘ AS ‘phono‘,
  ‘list2‘.‘orthog‘ AS ‘orthog‘,  ‘list2‘.‘pleon1‘ AS ‘pleon1‘,
  ‘list2‘.‘pleon2‘ AS ‘pleon2‘,  ‘list2‘.‘semtype‘ AS ‘semtype‘,
  ‘list2‘.‘semtype2‘ AS ‘semtype2‘,‘list2‘.‘pntype‘ AS ‘pntype‘,
  ‘list2‘.‘pnstrat‘ AS ‘pnstrat‘,‘list2‘.‘2genelem1‘ AS ‘2genelem1‘,
  ‘list2‘.‘2genelem2‘ AS ‘2genelem2‘,‘list2‘.‘spec‘ AS ‘spec‘,
  ‘list2‘.‘oldforms‘ AS ‘oldforms‘,‘list2‘.‘othernames‘ AS ‘othernames‘,
  ‘list2‘.‘comments‘ AS ‘comments‘,‘list2‘.‘class2‘ AS ‘class2‘,
  ‘list2‘.‘firstclass‘ AS ‘firstclass‘,‘list2‘.‘lastclass‘ AS ‘lastclass‘,
  ‘list2‘.‘altmin‘ AS ‘altmin‘,‘list2‘.‘altmax‘ AS ‘altmax‘,
  ‘list2‘.‘altave‘ AS ‘altave‘,‘list2‘.‘LAT‘ AS ‘LAT‘,  ‘list2‘.‘LON‘ AS ‘LON‘,
  ‘list2‘.‘hiera‘ AS ‘hiera‘,  ‘list2‘.‘source‘ AS ‘source‘,  ‘list2‘.‘county‘ AS ‘county‘,
  ‘list2‘.‘nont‘ AS ‘nont‘,  ‘list2‘.‘id2‘ AS ‘id2‘,‘list2‘.‘pos‘ AS ‘pos‘,
  if((‘list2‘.‘nont‘ between 0 and 4),1,
  if((‘list2‘.‘nont‘ between 5 and 8),2,
  if((‘list2‘.‘nont‘ between 9 and 16),3,
  if((‘list2‘.‘nont‘ between 17 and 32),4,
  if((‘list2‘.‘nont‘ between 33 and 64),5,
  if((‘list2‘.‘nont‘ between 65 and 128),6,
  if((‘list2‘.‘nont‘ between 129 and 256),7,
  if((‘list2‘.‘nont‘ between 257 and 512),8,
  if((‘list2‘.‘nont‘ > 513),9,NULL))))))))) AS ‘nont_score‘,
  abs((‘list2‘.‘pos‘ - 11)) AS ‘pos_score‘,
  if((‘list2‘.‘KM‘ between 0 and 1),1,
  if((‘list2‘.‘KM‘ between 1.5 and 2),2,
  if((‘list2‘.‘KM‘ between 2.5 and 4),3,
  if((‘list2‘.‘KM‘ between 4.5 and 8),4,
  if((‘list2‘.‘KM‘ between 8.5 and 16),5,
  if((‘list2‘.‘KM‘ between 16.5 and 32),6,
  if((‘list2‘.‘KM‘ between 32.5 and 64),7,
  if((‘list2‘.‘KM‘ between 64.5 and 128),8,
  if((‘list2‘.‘KM‘ > 128.5),9,NULL))))))))) AS ‘km_score‘,
```

```
  round(abs(((((((left(‘list2‘.‘firstclass‘,1) *
  left(‘list2‘.‘lastclass‘,1)) - 1) / 35) * 9) + 1) - 11)),2)
   AS ‘alt_score‘ from ‘list2‘;
```

lingscore

```
select (((9 * (avg(‘geogscore‘.‘score‘) - (select avg(‘geogscore‘.‘score‘)
AS ‘non_adj_sem‘
from (‘geogscore‘ join ‘specelem‘)
 where (‘geogscore‘.‘specelem‘ = ‘specelem‘.‘id‘)
group by ‘geogscore‘.‘specelem‘
order by avg(‘geogscore‘.‘score‘) limit 1))) /
 ((select avg(‘geogscore‘.‘score‘) AS ‘non_adj_sem‘
 from (‘geogscore‘ join ‘specelem‘)
where (‘geogscore‘.‘specelem‘ = ‘specelem‘.‘id‘)
group by ‘geogscore‘.‘specelem‘
order by avg(‘geogscore‘.‘score‘) desc limit 1) -
 (select avg(‘geogscore‘.‘score‘) AS ‘non_adj_sem‘
from (‘geogscore‘ join ‘specelem‘)
where (‘geogscore‘.‘specelem‘ = ‘specelem‘.‘id‘)
group by ‘geogscore‘.‘specelem‘
order by avg(‘geogscore‘.‘score‘) limit 1))) + 1) AS ‘avg_specelem_score‘,
‘geogscore‘.‘specelem‘ AS ‘specelem‘,
‘specelem‘.‘stratum‘ AS ‘stratum‘,
avg(‘geogscore‘.‘score‘) AS ‘non_adj_sem‘,
count(0) AS ‘count‘
from (‘geogscore‘ join ‘specelem‘)
where (‘geogscore‘.‘specelem‘ = ‘specelem‘.‘id‘)
group by ‘geogscore‘.‘specelem‘
```

finalscore

```
select ‘list2‘.‘RN‘ AS ‘rn‘,‘list2‘.‘id‘ AS ‘id‘,
round((‘lingscore‘.‘score‘ * ‘geogscore‘.‘score‘),2) AS ‘final_score‘,
‘lingscore‘.‘score‘ AS ‘lingscore‘,‘geogscore‘.‘score‘ AS ‘geogscore‘,
‘lingscore‘.‘avg_specelem_score‘ AS ‘avg_specelem_score‘,
‘lingscore‘.‘avg_semtype_score‘ AS ‘avg_semtype_score‘,
‘lingscore‘.‘avg_genelem_score‘ AS ‘avg_genelem_score‘,
‘list2‘.‘catch‘ AS ‘catch‘,‘list2‘.‘catch2‘ AS ‘catch2‘,
‘geogscore‘.‘nont_score‘ AS ‘nont_score‘,
‘geogscore‘.‘pos_score‘ AS ‘pos_score‘,
‘geogscore‘.‘km_score‘ AS ‘km_score‘,
‘geogscore‘.‘alt_score‘ AS ‘alt_score‘
from ((‘list2‘ join ‘lingscore‘) join ‘geogscore‘)
where ((‘geogscore‘.‘id‘ = ‘list2‘.‘id‘) and
(‘list2‘.‘id‘ = ‘lingscore‘.‘id‘))
order by round(((‘lingscore‘.‘score‘ * ‘geogscore‘.‘score‘) / 2),2) desc;
```

genelem

```
select ‘list2‘.‘genelem1‘ AS ‘genelem‘,
if((‘list2‘.‘genelem1‘ = 3),‘genelem2‘.‘gaelic‘,
if((‘list2‘.‘genelem1‘ = 4),‘genelem2‘.‘scots‘,
if((‘list2‘.‘genelem1‘ = 2),‘genelem2‘.‘pceltic‘,NULL))) AS ‘element‘,
‘genelem2‘.‘id‘ AS ‘id‘,‘genelem2‘.‘gaelic‘ AS ‘gaelic‘,
‘genelem2‘.‘scots‘ AS ‘scots‘,‘genelem2‘.‘pceltic‘ AS ‘pceltic‘
from ((‘list2‘ join ‘genelem2‘) join ‘specelem‘)
where (‘list2‘.‘genelem2‘ = ‘genelem2‘.‘id‘) group by ‘genelem2‘.‘id‘;
```

genelem2table

```
select concat(‘list2‘.‘genelem1‘,‘list2‘.‘genelem2‘) AS ‘code‘,
if((‘list2‘.‘genelem1‘ = 7),‘genelem2‘.‘scots‘,
if((‘list2‘.‘genelem1‘ = 6),‘genelem2‘.‘gaelic‘,
if((‘list2‘.‘genelem1‘ = 8),‘genelem2‘.‘latin‘,
if((‘list2‘.‘genelem1‘ = 4),‘genelem2‘.‘pceltic‘,
_latin1’none’))))
AS ‘element‘,
count(0) AS ‘count‘
from (‘genelem2‘ join ‘list2‘)
where (‘genelem2‘.‘id‘ = ‘list2‘.‘genelem2‘)
group by concat(‘list2‘.‘genelem1‘,‘list2‘.‘genelem2‘)
```

genelem3table

```
select concat(‘list2‘.‘genelem1‘,‘list2‘.‘genelem2‘,‘list2‘.‘genelem3‘)
AS ‘code‘,
if((‘list2‘.‘genelem1‘ = 7),‘genelem2‘.‘scots‘,
if((‘list2‘.‘genelem1‘ = 6),‘genelem2‘.‘gaelic‘,
if((‘list2‘.‘genelem1‘ = 8),‘genelem2‘.‘latin‘,
if((‘list2‘.‘genelem1‘ = 4),‘genelem2‘.‘pceltic‘,_latin1
’none’)))) AS ‘element‘,‘genelem3‘.‘order‘
 AS ‘order‘,count(0)
AS ‘count‘,
if((count(0) between 1 and 2),1,
if((count(0) between 3 and 4),2,
if((count(0) between 5 and 8),3,
if((count(0) between 9 and 16),4,
if((count(0) between 17 and 32),5,
if((count(0) between 33 and 64),6,
if((count(0) between 65 and 128),7,
if((count(0) between 129 and 256),8,
if((count(0) between 257 and 512),9,
if((count(0) between 513 and 1024),10,
if((count(0) between 1025 and 2048),11,
if((count(0) > 2049),12,NULL)))))))))))) AS ‘adjcount‘,
```

```
if((`genelem3`.`order` = _utf8'Simplex'),
convert(if((`list2`.`genelem1` = 7),`genelem2`.`scots`,
if((`list2`.`genelem1` = 6),`genelem2`.`gaelic`,
if((`list2`.`genelem1` = 8),`genelem2`.`latin`,
if((`list2`.`genelem1` = 4),`genelem2`.`pceltic`,
_latin1'none')))) using utf8)
,replace(`genelem3`.`order`,_utf8'Generic',
convert(`genelem2table`.`element` using utf8))) AS `order2`
from (((`genelem2` join `list2`) join `genelem3`) join `genelem2table`)
where ((`genelem2`.`id` = `list2`.`genelem2`)
and (`genelem3`.`id` = `list2`.`genelem3`)
and (`genelem2table`.`code` = concat(`list2`.`genelem1`,`list2`.`genelem2`)))
group by concat(`list2`.`genelem1`,`list2`.`genelem2`,`list2`.`genelem3`)
```

genelemscore

```
select (((9 * (avg(`geogscore`.`score`) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)`
from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) limit 1))) / ((select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)`
from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) desc limit 1) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) limit 1))) + 1) AS `avg_genelem_score`,
(((9 * (min(`geogscore`.`score`) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
 group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) limit 1))) / ((select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
 group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) desc limit 1) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
 order by avg(`geogscore`.`score`) limit 1))) + 1) AS `min_genelem_score`,
(((9 * (max(`geogscore`.`score`) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) limit 1))) / ((select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) desc limit 1) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) limit 1))) + 1) AS `max_genelem_score`,
```

```
concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
AS `genelem2`,avg(`geogscore`.`score`) AS `non_adj_gen`,min(`geogscore`.`score`)
AS `non_adj_gen_min`,max(`geogscore`.`score`) AS `non_adj_gen_max` from `geogscore`
where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by (((9 * (avg(`geogscore`.`score`) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) limit 1))) / ((select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) desc limit 1) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore` where (`geogscore`.`genelem1` <> 8)
group by concat(`geogscore`.`genelem1`,`geogscore`.`genelem2`,`geogscore`.`genelem3`)
order by avg(`geogscore`.`score`) limit 1))) + 1);
```

semtypescore

```
select (((9 * (avg(`geogscore`.`score`) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore`
group by concat(`geogscore`.`semtype`,`geogscore`.`pntype`,`geogscore`.`adjtype`)
order by avg(`geogscore`.`score`) limit 1))) / ((select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore`
group by concat(`geogscore`.`semtype`,`geogscore`.`pntype`,`geogscore`.`adjtype`)
order by avg(`geogscore`.`score`) desc limit 1) - (select avg(`geogscore`.`score`)
AS `avg(``geogscore``.``score``)` from `geogscore`
group by concat(`geogscore`.`semtype`,`geogscore`.`pntype`,`geogscore`.`adjtype`)
order by avg(`geogscore`.`score`) limit 1))) + 1) AS `avg_semtype_score`,avg(`geogscore`.`score`)
AS `non_adj_sem`,min(`geogscore`.`score`) AS `min_semtype_score`,max(`geogscore`.`score`)
AS `max_semtype_score`,concat(`geogscore`.`semtype`,`geogscore`.`pntype`,`geogscore`.`adjtype`)
AS `semtype` from `geogscore`
group by concat(`geogscore`.`semtype`,`geogscore`.`pntype`,`geogscore`.`adjtype`)
order by avg(`geogscore`.`score`);
```

semtypetable

```
select concat(`list2`.`semtype`,`list2`.`pntype`,`list2`.`adjtype`) AS `code`,
`semtype`.`name` AS `sem`,`pntype`.`name` AS `pn`,`semtype`.`Type` AS `type`,
if((`list2`.`semtype` = 2),`adjtype`.`semtype2`,
if((`list2`.`semtype` = 4),`adjtype`.`semtype4`,
if((`list2`.`semtype` = 5),`adjtype`.`semtype5`,
if((`list2`.`semtype` = 6),`adjtype`.`semtype6`,
if((`list2`.`semtype` = 8),`adjtype`.`semtype8`,
if((`list2`.`semtype` = 9),`adjtype`.`semtype9`,
if((`list2`.`semtype` = 10),`adjtype`.`semtype10`,
if((`list2`.`semtype` = 13),`adjtype`.`semtype13`,
if((`list2`.`semtype` = 14),`adjtype`.`semtype14`,
if((`list2`.`semtype` = 15),`adjtype`.`semtype15`,
if((`list2`.`semtype` = 36),`adjtype`.`semtype36`,_utf8'None'))))))))))) AS `adj`,
```

```
count(0) AS `amount` from (((`pntype` join `list2`) join `semtype`) join `adjtype`)
where ((`pntype`.`id` = `list2`.`pntype`) and (`semtype`.`id` = `list2`.`semtype`) and
 (`adjtype`.`id` = `list2`.`adjtype`))
group by concat(`list2`.`semtype`,`list2`.`pntype`,`list2`.`adjtype`)
order by count(0) desc;
```

specelemscore

```
select (((9 * (avg(`geogscore`.`score`) - (select avg(`geogscore`.`score`) AS `non_adj_sem`
from (`geogscore` join `specelem`)
 where (`geogscore`.`specelem` = `specelem`.`id`)
group by `geogscore`.`specelem`
order by avg(`geogscore`.`score`) limit 1))) / ((select avg(`geogscore`.`score`) AS `non_adj_sem`
from (`geogscore` join `specelem`)
where (`geogscore`.`specelem` = `specelem`.`id`)
group by `geogscore`.`specelem`
order by avg(`geogscore`.`score`) desc limit 1) - (select avg(`geogscore`.`score`) AS `non_adj_sem`
from (`geogscore` join `specelem`)
 where (`geogscore`.`specelem` = `specelem`.`id`)
group by `geogscore`.`specelem`
order by avg(`geogscore`.`score`) limit 1))) + 1) AS `avg_specelem_score`,
`geogscore`.`specelem` AS `specelem`,`specelem`.
`stratum` AS `stratum`,avg(`geogscore`.`score`) AS `non_adj_sem`,
count(0) AS `count`
from (`geogscore` join `specelem`)
where (`geogscore`.`specelem` = `specelem`.`id`)
group by `geogscore`.`specelem`;
```

D.2 MySQL Queries for Tables

The following code is for tables generated by MySQL directly, not through one of the R functions below. In some cases it was not possible to get MySQL to exactly generate the required tables, in these cases either a combination fo queries or an explanatory note have been given.

Table 2.1

```
SELECT specelem, round(min(km),1) as 'minimum km',
round(avg(km),1) as 'average km',
round(max(km),1) as 'max km' FROM list2
where source != "F1" and specelem !=1 group by specelem
```

Tables 2.2 and 2.3

See geogscore above.

Table 2.5

```
SELECT specelem.stratum as stratum, count(*) as amount,
concat(round((count(*)/70)*100,2),"\%") as percentage FROM geogscore
join specelem where geogscore.score  = 5.235 and specelem.id =
geogscore.specelem group by geogscore.specelem
```

Table 2.6

```
SELECT specelem.stratum as stratum, count(*) as amount,
concat(round((count(*)/1681)*100,2),"%") as percentage FROM geogscore
join specelem where geogscore.score  between 4.835 and 5.635 and
specelem.id = geogscore.specelem group by geogscore.specelem
```

Table 2.7

```
SELECT semtypetable.sem as meaning, count(*) as amount
FROM finalscore join list2 join semtypetable
where geogscore between 4.835 and 5.635 and list2.id = finalscore.id and lat between
(select lat-40000 from list2 where id = 3244) and (select lat+40000 from list2 where id = 3244)
and lon between (select lon-40000 from list2 where id = 3244) and
(select lon+40000 from list2 where id = 3244)
and concat(list2.specelem,list2.pntype) = semtypetable.code
group by semtypetable.code order by count(*) desc
```

Table 2.8

```
SELECT allsem.meaning, concat(frandy.percentage-allsem.percentage,"\%") as difference,
 concat(allsem.percentage,"\%") as "all Gaelic names", concat(frandy.percentage,"\%") as "Frandy"
 FROM allsem left outer join frandy on (frandy.id = allsem.id)
order by frandy.percentage-allsem.percentage desc
```

Table 2.9

```
select lat between 285000 and 325300 and
lon < 720300 and altmax > 600 and altmin > 100 and lat
between 285000 and 325300 and lon < 720300 and left(firstclass,1) > 2}
```

Table 2.10

```
SELECT RN, id, lingscore, geogscore,
abs(((geogscore-2.8586)/lingscore)-.8126)FROM finalscore
order by abs(((geogscore-2.8586)/lingscore)-.8126) desc
```

Table 2.12

Table 2.12

```
SELECT std(lingscore), geogscore
FROM finalscore, list2 where finalscore.id = list2.id and
list2.semtype != 1 group by geogscore
order by std(lingscore) desc
```

Table 3.1

```
select etymology, count(*) as amount from list2
where genelem1 = 4 and genelem2 = 4
group by etymology order by count(*) desc
```

Table 3.3

```
SELECT rn, s2.description, s1.description, altave
FROM list2, class, soilclass as s1, soilclass as s2
where class.id = list2.class and right(class.class,2) = s2.class and
left(class.class,2) = s1.class and (genelem2 = 17 or
etymology like "%pow%") and genelem1 != 6 order by spec
```

Table 3.4

This table is a combination of these two tables:

```
SELECT count(*) as amnt, catch.rn FROM list2 join catch where
genelem1 = 6 and genelem2 = 1 and source != "OS" and list2.catch != 0
and list2.catch = catch.id group by list2.catch
order by count(*) desc

SELECT count(*) as amnt, catch.rn FROM list2 join catch where
genelem1 = 6 and genelem2 = 1 and source = "OS" and list2.catch != 0
and list2.catch = catch.id group by list2.catch
order by count(*) desc
```

Table 3.5

```
select count(*) as number, county from allos.osscot
where name like "%abhainn%" group by county order by county
```

Table 3.6

```
SELECT semtype.name, count(*) FROM list2, semtype
where genelem2 = 8 and genelem1 = 6 and
semtype.id = list2.semtype group by semtype order by count(*) desc
```

Table 3.7

```
SELECT RN as name, null as variant FROM hieralist2
where (genelem2 = 13) and genelem1 = 6
```

Table 3.8

This table is a combination of two queries.

```
select "loch" as element, min(score) as minimum, avg(score) as average,
max(score) as maximum from geogscore where genelem1 = 6 and genelem2 = 14

select "lochan" as element, min(score) as minimum, avg(score) as average,
max(score) as maximum from geogscore where genelem1 = 6 and genelem2 = 15
```

Table 3.9

This table is a combination of two queries.

```
select "alltan" as element, min(score) as minimum, avg(score) as average,
max(score) as maximum from geogscore where genelem1 = 6 and genelem2 = 2
select "allt" as element, min(score) as minimum, avg(score) as average,
max(score) as maximum from geogscore where genelem1 = 6 and genelem2 = 1
```

Table 3.10

```
select nont, count(*) as amount from list2
where genelem1 = 6 and genelem2 = 2 group by nont
```

Table 3.11

```
select nont, count(*) as amount from list2
where genelem1 = 6 and genelem2 = 9 group by nont
```

Table 3.12

```
SELECT round(avg(score),2) as average score, RN, genelem3
from geogscore join genelem3 where genelem3.id = geogscore.genelem3 and
 genelem1 = 6 and (genelem3 = 1 or genelem3 = 5 or genelem3 = 6)
group by genelem3
```

Table 3.13

```
SELECT RN as 'typical name', etymology, count(*) as amount
FROM list2 where (genelem1 = 6 and genelem3 = 1 )
group by left(etymology,8) order by count(*) desc
```

Table 3.14

```
SELECT RN as 'typical name', etymology, count(*) as amount
FROM list2 where (genelem1 = 6 and genelem3 = 6) group by
left(etymology,8) order by count(*) desc limit 20
```

Table 3.15

```
select RN, oldforms, count(*), g.order2, h.order2
from list2 join genelem3table g join genelem3table h
where 2genelem1 is not null and g.code = concat(genelem1,genelem2,genelem3)
and h.code = concat(2genelem1,2genelem2,2genelem3)
group by concat(genelem1,genelem2,genelem3,2genelem1,2genelem2,2genelem3)
order by count(*) desc
```

Table 3.16

```
select id, RN, oldforms from list2 where
2genelem1 = 6 and 2genelem2 = 4 and genelem1 = 7 and genelem2 = 3
```

Table 3.17

```
SELECT RN, etymology FROM list2 where
concat(genelem1,genelem2,genelem3) = 715
```

Table 3.18

```
SELECT count(*) FROM allos.osscot
where name like "% of %"
```

Table 3.19

```
SELECT genelem2.scots, count(*) as number from list2 join genelem2
where genelem2.id = list2.genelem2 and genelem3 = 6 and genelem1 = 7
group by genelem2 order by count(*) desc
```

Table 3.20

The database is not specifically set up to detect plurals as such, but the following query gives some of the results:

```
SELECT rn from list2 where rn like "%s of %"
```

Table 3.21

```
SELECT  RN as name, count(*) as amount FROM list2 l where genelem1 = 7
and genelem2 = 5 and specelem != 6 group by etymology order by count(*) desc
```

Table 3.22

```
SELECT  RN as name, etymology FROM list2 l where genelem1 = 7
and genelem2 = 5 and specelem = 6 order by specelem
```

Table 3.23

```
SELECT id, RN, etymology FROM rndb.list2 l where genelem1 = 6 and genelem2 = 5
```

Table 3.44

```
select genelem2table.element, max(pos) as maximum, avg(pos) as average,
min(pos) as minimum from list2 join genelem2table
where genelem2table.code =  concat(genelem1,genelem2)
group by concat(genelem1,genelem2) order by avg(pos)
```

Table 3.25

```
select g.element as main, h.element as tributary, count(*) as number
FROM list2 l join list2 u join genelem2table g join genelem2table h
where  l.posid = u.id and concat(u.genelem1,u.genelem2) = g.code and
concat(l.genelem1,l.genelem2) = h.code group by concat(g.code,h.code)
order by count(*) desc limit 50
```

Table 3.26

```
select g.element as main FROM list2 l join list2 u join genelem2table g
join genelem2table h where  l.posid = u.id and concat(u.genelem1,u.genelem2) = g.code
and concat(l.genelem1,l.genelem2) = h.code and h.code != g.code
group by concat(g.code) order by count(*) desc limit 50
```

Table 3.27

```
SELECT g.element, h.element, count(*) FROM list2 l join list2 u join
genelem2table g join genelem2table h where u.posid = l.id and
u.spec = l.spec and concat(u.genelem1,u.genelem2) = g.code and
concat(l.genelem1,l.genelem2) = h.code
group by concat(g.code,h.code) order by count(*) desc
```

Table 3.28

```
select l.rn, u.rn, count(*) as amount FROM list2 l join list2 u
where concat(left(l.lat,3),left(l.lon,3)) = concat(left(u.lat,3),left(u.lon,3)) and
l.id != u.id and l.spec = u.spec and concat(u.genelem1,u.genelem2) != concat(l.genelem1,l.genelem2)
group by l.spec order by amount desc
```

Table 3.29

```
SELECT order2, round(avg_genelem_score,2) FROM genelem3table join genelemscore
where genelem3table.code = genelemscore.genelem2 and code like "7%" and count > 5
order by avg_genelem_score desc
```

Table 3.29

```
SELECT order2, round(avg_genelem_score,2) FROM genelem3table join genelemscore
where genelem3table.code = genelemscore.genelem2 and code like "4%"
order by avg_genelem_score desc
```

Table 3.12.3

```
SELECT order2, round(avg_genelem_score,2) FROM genelem3table join genelemscore
where genelem3table.code = genelemscore.genelem2 and code like "6%"
and count > 5 order by avg_genelem_score desc
```

Table 4.5

```
SELECT rn as name, km as "length in km"
from list2 where semtype = 2
order by km desc limit 10
```

Table 4.6

```
SELECT etymology as term, count(*) as amount
FROM list2 where semtype  = 2 and specelem = 6
group by etymology order by count(*) desc limit 18
```

Table 4.7

```
SELECT id, count(*) as amount, etymology
FROM list2 l where semtype = 4 and adjtype = 2
group by etymology order by count(*) desc limit 10
```

Table 4.13

```
SELECT etymology, count(*) FROM list2
where semtype = 8  and adjtype = 2 group by etymology
order by count(*) desc, specelem limit 6
```

Table 4.3.8

```
SELECT RN as example, etymology, count(*) as amount
FROM list2 where semtype = 9 and adjtype = 2 group by etym1a
order by count(*) desc
```

Table 4.34

```
SELECT count(*), etymology FROM list2 l
where semtype = 30 group by etymology order by count(*) desc
```

Table 4.44

```
select   s.type as ‘main’, t.type as ‘tributary’, count(*) as amount
FROM list2 l join list2 u join semtype s join semtype t
where  u.posid = l.id  and  l.semtype != 33  and  u.semtype != 33
and  l.semtype = s.id and  u.semtype = t.id  and  u.RN like "%<%"
group by  concat(s.type, t.type )  order by  count(*) desc
```

Table 4.45

```
select s.type as 'main', t.type as 'tributary', count(*) as amount
FROM list2 l join list2 u join semtype s join semtype t
where u.posid = l.id and l.semtype != 33 and u.semtype != 33
and l.semtype = s.id and u.semtype = t.id and s.type != t.type
group by concat(s.type, t.type ) order by count(*) desc
```

Table 5.1

```
SELECT id, rn, km  FROM list2 l where concat(catch,catch2) = 52
and specelem = 7 order by km desc limit 10
```

Table 5.6

```
SELECT l.RN, count(*) as amount, l.othernames FROM list2 l join list2 u
where u.posid = l.id and l.specelem = 7 and u.specelem= 6 group by l.id
order by amount desc limit 15
```

Table 5.7

```
SELECT u.RN as example, u.etym1b as meaning, count(*) as amount FROM list2 l join list2 u
 where u.posid = l.id and l.specelem > 5 and u.specelem  between 4 and 5
group by left(u.etym1b,5) order by amount desc limit 5
```

Table 5.8

```
select s.stratum as 'main', t.stratum as 'tributary', count(*) as amount,
if(l.specelem > u.specelem, "contrary", "normal") FROM list2 l join list2 u
join specelem s join specelem t where u.posid = l.id and l.specelem != 1
and u.specelem != 1  and l.specelem = s.id and u.specelem = t.id
group by concat(l.specelem, u.specelem ) order by count(*) desc
```

Table 5.9

```
select   s.stratum,   t.stratum, count(*)#l.RN as 'main', u.RN as 'tributary'
 FROM list2 l join list2 u join specelem s join specelem t
where  u.posid = l.id  and  l.specelem != 1 and  u.specelem != 1
and  u.RN like "%<%"  and  l.specelem = s.id  and
 u.specelem = t.id group by concat(l.specelem, u.specelem) order by count(*) desc
```

Table 5.10

```
select  s.stratum as 'stratum of main',l.RN as 'main', u.RN as 'tributary'
 FROM list2 l join list2 u join specelem s join specelem t
 where  u.posid = l.id  and  l.specelem != 1  and  u.specelem = 2
 and  l.specelem = s.id  and  u.specelem = t.id  order by  l.catch
```

Table 5.11

```
select specelem, pnstrat, count(*) as amount from list2 where
pnstrat != 1 and pnstrat is not null and specelem != 1 and
specelem is not null group by concat(specelem, pnstrat) order by count(*) des
```

Table 6.1

```
select concat(genelem1,phono,orthog,specelem) as string,
 round((`list2`.`phono`+`list2`.`orthog`+`list2`.`specelem`+`list2`.`genelem1`)/4,2) as avg,
 count(*) as amount, RN as example, g.stratum as genelem1, p.stratum as phono,
 o.stratum as orthog, s.stratum as specelem from list2, specelem g, specelem p,
specelem o, specelem s where source = "os" and specelem > 5 and g.id = list2.genelem1 and
p.id = list2.phono and o.id = list2.orthog and s.id = list2.specelem
group by concat(genelem1,phono,orthog,specelem)
order by concat(genelem1,phono,orthog,specelem) des
```

Table 6.2

```
SELECT phono, specelem, count(*) FROM list2 where source = "OS"
 group by concat(phono, specelem) order by count(*) desc
```

Table 6.3

```
select s.stratum as 'main', t.stratum as 'tributary', count(*) as amount,
if(l.phono > u.phono, "contrary", "normal") as direction FROM list2 l join
list2 u join specelem s join specelem t where u.posid = l.id and l.phono != 1
 and u.phono != 1 #and l.specelem > u.specelem and l.phono = s.id and u.phono = t.id
 group by concat(l.phono, u.phono ) order by count(*) desc
```

D.3 R source Code for Graphs

R source code are here only for 'one-off graphs', i.e. those which were not derived from functions.

Figure 2.2

```
library(RMySQL)
MySQL(max.con = 1, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("select avg(km) as km, length(replace(rn, ' ','')) as len
from list2 where source = 'os' group by len"))
dat <- fetch(res, n=7000)
plot(dat$len, dat$km, xlab = "length of river name in letters",
ylab = "average length of river in km")
lines(dat$len, dat$km)
#myline.fit <- lm(dat$len ~ dat$km)
myline.fit <- lm(dat$km~dat$len)
abline(myline.fit, col = "black", lwd = 1, lty = 1)
title(main = "Relationship between a length of a river name and
 the length of its watercourse",  cex.main = .8)
```

```
dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 2.3

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("select specelem, avg(km) as km from list2
where specelem != 1
group by specelem"))
dat <- fetch(res, n=7000)

plot(dat$specelem, dat$km, type = "b", ylab ="average length",
xlab ="specelem",  xlim = c(2,7), ylim=c(1,60), axes= FALSE)
axis(1, 2:7, labels = c("Old Celtic", "Old /\n P-Celtic",
"P-Celtic", "P-Celtic \n/ Gaelic", "Gaelic", "Scots"),
line = 0, tck = .02)
axis(2)

dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 2.5

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("select avg(km) as km, altave from list2
where altave is not null
group by altave"))
dat <- fetch(res, n=7000)

plot(dat$altave, dat$km, type = "b", ylab ="average km",
xlab ="altitude", xlim =c(100,900), ylim=c(2,4.5))

#dbClearResult(res)
#mysqlCloseConnection(con)
```

Figure 2.6

```
dbClearResult(res)
mysqlCloseConnection(con)
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("select specelem, avg(altave) as alt from list2
where specelem != 1
group by specelem"))
dat <- fetch(res, n=7000)
```

```
plot(dat$specelem, dat$alt, type = "b", ylab ="average altitude",
xlab ="specelem",  xlim = c(2,7), ylim=c(250,450), axes= FALSE)
axis(1, 2:7, labels = c("Old Celtic", "Old /\n P-Celtic",
"P-Celtic", "P-Celtic \n/ Gaelic", "Gaelic", "Scots"),
line = 0, tck = .02)
axis(2)
#axis(2, 250:450,  labels = c(250, 300, 350, 400, 450))
```

Figure 2.7

```
library(RMySQL)
MySQL(max.con = 1, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("select min(pos) as min,
avg(pos) as pos, max(pos) as max, specelem
from list2 where specelem !=1 group by specelem"))
dat <- fetch(res, n=7000)
plot(dat$specelem, dat$pos, type = "l", ylab = "",
xlab = "", ylim =c(0,10), xlim=c(1,8), axes = FALSE)
lines(dat$specelem, dat$pos)
points(dat$specelem, dat$pos)
lines(dat$specelem, dat$max)
points(dat$specelem, dat$max)
lines(dat$specelem, dat$min)
points(dat$specelem, dat$min)
axis(1, 2:7, labels = c("Old Celtic", "Old /\n P-Celtic",
"P-Celtic", "P-Celtic \n/ Gaelic", "Gaelic", "Scots"),
line = 0, tck = .02)
axis(2, 1:10, labels = 1:10, tck = -.02, line = -3)
tick = FALSE, pos = 6.9, padj = 0)
text(7.3,9, "max")
text(7.3,3.5, "avg")
text(7.3,1, "min")
text(0.8, 6,"position", srt = 90)
dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 2.8

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb",
user = "safe", host = "localhost", password = "changeme")
res <- dbSendQuery(con, paste("select avg(km) as avg,
pos from list2 where pos is not null group by pos"))
dat <- fetch(res, n=7000)
plot(dat$pos, dat$avg, ylab ="average km",
xlab ="pos", xlim =c(1,10), ylim=c(1,15))
lines(dat$avg)
dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 2.9

```
sc = c(2.8, 16.6, 28.1, 27.9, 15.3, 6.4, 2.1, 0.6, 0.1, 0)
g = c(1.2, 11, 28, 25.6, 16.2, 8.7, 5.6, 2.3, 0.9, 0.4)
pg = c(13.4, 25.6, 29.3, 22, 4.9, 4.9, 0, 0, 0, 0)
p = c(10.3, 32.8, 34.5, 12.1, 6.9, 1.7, 1.7, 0, 0, 0)
op = c(50, 31.3, 12.5, 0, 6.3, 0, 0, 0, 0, 0)
oc = c(42.3, 46.2, 7.7, 3.8, 0, 0, 0, 0, 0, 0)
q = NA
plot(q, xlim = c(10,-1), ylim = c(0,60), xlab = "position",
ylab = "percentage occurence of position within stratum")
lines(sc)
lines(g)
lines(pg)
lines(p)
lines(op)
lines(oc)
axis(1, 1:10, labels = c(1:10))
text(0.7, 2.8, "Scots", adj =c(0,0))
text(0.7, 1.2, "Gaelic", adj =c(0,1))
text(0.7, 13.4, "P-Celtic / Gaelic", adj =c(0,1))
text(0.7, 10.3, "P-Celtic", adj =c(0,1))
text(0.7, 50, "Old / P- Celtic", adj =c(0,1))
text(0.7, 42.3, "Old Celtic", adj =c(0,1))

title(x, main = "Embeddedness by stratum by percentile")
```

Figure 2.10 and 2.11

```
library(RMySQL)
MySQL(max.con = 2, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb",
user = "root", host = "localhost", password = "")
res <- dbSendQuery(con, paste("SELECT rn, geogscore, lingscore, final_score from finalscore
where geogscore*lingscore is not null")) #
dat <- fetch(res, n=7000)

#hdr.boxplot.2d(dat$lingscore, dat$geogscore)
plot(dat$lingscore, dat$geogscore, xlab ="lingscore", ylab ="geogscore",
xlim =c(1,10), ylim=c(1,10))
myline.fit <- lm(dat$geogscore ~  dat$lingscore)
abline(myline.fit, col = "gray", lwd = 1, lty = 1)

myline.fit1 <- lm(dat$geogscore-2.2 ~  dat$lingscore)
abline(myline.fit1, col = "gray", lwd = 1, lty = 1)
myline.fit2 <- lm(dat$geogscore+2.2 ~  dat$lingscore)
abline(myline.fit2, col = "gray", lwd = 1, lty = 1)


dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 2.13

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("SELECT (nont_score * pos_score * km_score * alt_score)
 as geogscore,
(avg_specelem_score * avg_semtype_score * avg_genelem_score)
as lingscore,
final_score from finalscore")) #
dat <- fetch(res, n=7000)

plot(dat$lingscore, dat$geogscore, xlab ="lingscore", ylab ="geogscore",
#xlim =c(1,80), ylim=c(1,2000)
)
myline.fit <- lm(dat$geogscore-2.2 ~  dat$lingscore)
abline(myline.fit, col = "gray", lwd = 1, lty = 1)
myline.fit <- lm(dat$geogscore+2.2 ~  dat$lingscore)
abline(myline.fit, col = "gray", lwd = 1, lty = 1)
myline.fit <- lm(dat$geogscore ~  dat$lingscore)
abline(myline.fit, col = "gray", lwd = 1, lty = 1)

dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 3.11

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), username = "root", dbname = "rndb", port = 3306)
res1<- dbSendQuery(con, paste("select altave, round(count(*)/17.48,2)
as percentage from list2
where (genelem1 = 6 and genelem2 between 1 and 4) or
(2genelem1 = 6 and 2genelem2 between 1 and 4)
group by altave")) #
dat1 <- fetch(res1, n=7000)
dbClearResult(res1)
res2 <- dbSendQuery(con, paste("select altave, round(count(*)/63.5,2)
as percentage from list2 group by altave")) #
dat2 <- fetch(res2, n=7000)
dbClearResult(res2)
plot(dat1$altave, dat1$percentage, ylab = "percentage occurence",
xlab = "average altitude", xlim =c(0,1000), ylim=c(0,30))
lines(dat1$altave, dat1$percentage)
points(dat2$altave, dat2$percentage, col = 2)
lines(dat2$altave, dat2$percentage, xlim =c(0,1000), ylim=c(0,30), col = 2)
legend(500, 30, c("Allt", "All generic elements"), col = c("black", "red"),
       text.col = "black", lty = 1, lwd = 2, bty = "n")

dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 3.12

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)

#res1 = allt names in black
res1<- dbSendQuery(con, paste("select avg(altave) as altave,
min(altmin) as altmin, max(altmax) as altmax, left(lat,2) as lat from list2
where ((genelem1 = 6 and genelem2 between 1 and 4) or
(2genelem1 = 6 and 2genelem2 between 1 and 4))
#and specelem = 6
and left(lat,2) != 10
group by left(lat,2)")) #
dat1 <- fetch(res1, n=7000)
dbClearResult(res1)
#res2 = all gaelic names in red
res2 <- dbSendQuery(con, paste("select avg(altave) as altave,
 min(altmin) as altmin, max(altmax) as altmax, left(lat,2) as lat from list2
where specelem = 6
and concat(genelem1,genelem2) not between 61 and 64
and left(lat,2) != 10
group by left(lat,2)")) #
dat2 <- fetch(res2, n=7000)
dbClearResult(res2)

plot(dat1$lat, dat1$altave, xlab = "<-- west latitude east -->",
ylab = "average altitude", xlim =c(22,42), ylim=c(0,650))
lines(dat1$lat, dat1$altave, lwd = 2)
points(dat2$lat, dat2$altave, col = "grey")
lines(dat2$lat, dat2$altave, col = "grey", lwd = 2)

dbClearResult(res1)
dbClearResult(res2)
mysqlCloseConnection(con)

legend(23, 200, c("RNs with allt as a generic",
"All other Gaelic names"), col = c("black", "grey"),
       text.col = "black", lty = 1, lwd = 2, bty = "n")
```

Figure 3.44

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "safe",
host = "localhost", password = "mellon")
res <- dbSendQuery(con, paste("select element, min(pos_score) as minpos,
avg(pos_score) as avgpos, max(pos_score) as maxpos, genelem1
from geogscore join genelem2table
where concat(geogscore.genelem1,geogscore.genelem2) = genelem2table.code
 and altave is not null
 and count > 11
group by concat(geogscore.genelem1,geogscore.genelem2)
```

```
order by avg(pos_score)")) #
dat <- fetch(res, n=7000)

rowno = as.numeric(rownames(dat))
ave = dat$avgpos
min = dat$minpos
max = dat$maxpos

plot(NULL,xlim = c(1,max(rowno)),ylim =c(0,10), ylab = "pos range",
xlab = "", axes = FALSE)
newcol = ifelse(dat$genelem1==7,1,ifelse(dat$genelem1==4,2,8))
segments(rowno,max,rowno,min, col = newcol, lwd = 3)

axis(rowno, at = rowno, side = 1, labels = dat$element, pos =.3,
las = 2, tick = FALSE, line = NA)
axis(c(0,10), side = 2)

dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 3.43

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), username = "root", dbname = "rndb", port = 3306)
res <- dbSendQuery(con, paste("select element, avg(altmin) as altmin,
avg(altmax) as altmax, avg(altave) as altave, genelem1
from list2 join genelem2table
where concat(list2.genelem1,list2.genelem2) = genelem2table.code
 and altave is not null
 and count > 11
group by concat(list2.genelem1,list2.genelem2)
order by avg(altave)")) #
dat <- fetch(res, n=7000)

rowno = as.numeric(rownames(dat))
ave = dat$altave
min = dat$altmin
max = dat$altmax

plot(NULL,xlim = c(1,max(rowno)),ylim =c(0,1000),
ylab = "altitude range", xlab = "", axes = FALSE)
newcol = ifelse(dat$genelem1==7,1,ifelse(dat$genelem1==4,2,8))
segments(rowno,max,rowno,min, col = newcol, lwd = 3)

axis(rowno, at = rowno, side = 1, labels = dat$element,
pos =25, las = 2, tick = FALSE, line = NA)
axis(c(0,1000), side = 2)

text(7,900, "Gaelic elements", pos = 2)
segments(8,900,13,900, col = 8, lwd = 3)
text(7,850, "Scots elements", pos = 2)
segments(8,850,13,850, col = 1, lwd = 3)
```

```
dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 3.45

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), username = "root", dbname = "rndb", port = 3306)
res <- dbSendQuery(con, paste("select element, min(km_score) as minkm,
avg(km_score) as avgkm, max(km_score) as maxkm, genelem1
from geogscore join genelem2table
where concat(geogscore.genelem1,geogscore.genelem2) = genelem2table.code
 and altave is not null
 and count > 11
group by concat(geogscore.genelem1,geogscore.genelem2)
order by avg(km)")) #
dat <- fetch(res, n=7000)

rowno = as.numeric(rownames(dat))
ave = dat$avgkm
min = dat$minkm
max = dat$maxkm

plot(NULL,xlim = c(1,max(rowno)),ylim =c(0,10),
ylab = "length range", xlab = "", axes = FALSE)
newcol = ifelse(dat$genelem1==7,1,ifelse(dat$genelem1==4,2,8))
segments(rowno,max,rowno,min, col = newcol, lwd = 3)

axis(rowno, at = rowno, side = 1, labels = dat$element,
pos =.3, las = 2, tick = FALSE, line = NA)
axis(c(0,10), side = 2)

text(7,900, "Gaelic elements", pos = 2)
segments(8,900,13,900, col = 8)
text(7,850, "Scots elements", pos = 2)
segments(8,850,13,850, col = 1)

dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 3.46

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "safe",
host = "localhost", password = "mellon")
res <- dbSendQuery(con, paste("select element, min(nont_score) as minnont,
avg(nont_score) as avgnont, max(nont_score) as maxnont, genelem1
from geogscore join genelem2table
where concat(geogscore.genelem1,geogscore.genelem2) = genelem2table.code
 and altave is not null
 and count > 11
```

```
group by concat(geogscore.genelem1,geogscore.genelem2)
order by avg(nont_score)")) #
dat <- fetch(res, n=7000)

rowno = as.numeric(rownames(dat))
ave = dat$avgnont
min = dat$minnont
max = dat$maxnont

plot(NULL,xlim = c(1,max(rowno)),ylim =c(0,10), ylab = "nont range",
xlab = "", axes = FALSE)
newcol = ifelse(dat$genelem1==7,1,ifelse(dat$genelem1==4,2,8))
segments(rowno,max,rowno,min, col = newcol, lwd = 3)

axis(rowno, at = rowno, side = 1, labels = dat$element,
pos =.3, las = 2, tick = FALSE, line = NA)
axis(c(0,10), side = 2)

dbClearResult(res)
mysqlCloseConnection(con)
```

Figure 4.1

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), username = "root", dbname = "rndb", port = 3306)
res <- dbSendQuery(con, paste("SELECT round(min_semtype_score,2) as min,
 round(non_adj_sem,2) as avg,  round(max_semtype_score,2) as max, semtype, type,
concat_ws(': ',sem,replace(concat(pn, adj),'None','')) as name, amount
FROM semtypescore join semtypetable
where semtypescore.semtype = semtypetable.code
order by round(max_semtype_score,2)-round(min_semtype_score,2)")) #
dat <- fetch(res, n=7000)

rowno = as.numeric(rownames(dat))
ave = dat$avg
min = dat$min
max = dat$max

dbClearResult(res)
mysqlCloseConnection(con)

gray.colors(7, start = 0.3, end = 0.9, gamma = 2.2)

op <- par(mar=c(0,0,0,0), oma=c(18,4,.5,.5))
plot(NULL,xlim = c(1,max(rowno)),ylim =c(1.5,9.5), xlab = "", axes = FALSE)
newcol = ifelse(dat$type=="Adjective",gray.colors(1),ifelse(dat$type=="Natural Feature",
gray.colors(2),ifelse(dat$type=="Topography",gray.colors(3),ifelse(dat$type=="Human",
gray.colors(4),ifelse(dat$type=="Situation",gray.colors(5),gray.colors(6))))))
segments(rowno,max,rowno,min, col = newcol, lwd = 3)
points(rowno, ave, pch = 19, cex = 1.5)
axis(rowno, at = rowno, side = 1, cex.axis = .8, labels = dat$name, pos = 2,
las = 2, tick = FALSE, line = NA)
```

```
axis(c(0,10), side = 2)
par(op)
```

Figure 4.23

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "safe",
 host = "localhost", password = "mellon")
res <- dbSendQuery(con, paste("select semtype.id, type, name, alt_score
from semtype join geogscore
where semtype.id = geogscore.semtype and score is not null")) #
dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

xaxis = c("Colour","Smell / Taste / Feel","Manner","Sound",
"Temperature","Other","Course","Effect / Character","Dimensions","Bed",
"Number","Moistness",
"Age","Elevation","Flora","Fauna","Concavity e.g. Glen*",
"Convexity e.g. Ben*","Land around*",
"Water feature*","Specific natural feature","Non specific agricultural area",
"Specific person / occupation","Specific man made area",
"Agricultural object / structure",
"Non specific settlement / building","Supernatural entity","Event",
"Relation to other features",
"Boundary","Crossing","Other","Material / Object","Weather / Air",
"Visibility")

id = dat$id
score = dat$alt_score

gray.colors(7, start = 0.3, end = 0.9, gamma = 2.2)

op <- par(mar=c(0,0,1.5,0), oma=c(13.5,4,.5,.5))
plot(NULL,xlim = c(1,35),ylim =c(0,10), ylab = "score range",
xlab = "", axes = FALSE)

points(jitter(id-1, factor = 1), jitter(score, factor = 1), pch = 19)

axis( side = 1, labels = xaxis, at = c(1:35), pos =.3, las = 2,
tick = FALSE, line = NA)
axis(c(0,10), side = 2)

par(op)
```

Figure 4.24

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), username = "root",
dbname = "rndb", port = 3306)
```

```
res <- dbSendQuery(con, paste("select semtype.id, type, name,
pos_score from semtype join geogscore
where semtype.id = geogscore.semtype and score is not null")) #
dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

xaxis = c("Colour","Smell / Taste / Feel","Manner","Sound",
"Temperature","Other","Course","Effect / Character","Dimensions",
"Bed","Number","Moistness",
"Age","Elevation","Flora","Fauna","Concavity e.g. Glen*",
"Convexity e.g. Ben*","Land around*",
"Water feature*","Specific natural feature","Non specific agricultural area",
"Specific person / occupation","Specific man made area",
"Agricultural object / structure",
"Non specific settlement / building","Supernatural entity",
"Event","Relation to other features",
"Boundary","Crossing","Other","Material / Object",
"Weather / Air","Visibility")

id = dat$id
score = dat$pos_score

gray.colors(7, start = 0.3, end = 0.9, gamma = 2.2)

op <- par(mar=c(0,0,1.5,0), oma=c(13.5,4,.5,.5))
plot(NULL,xlim = c(1,35),ylim =c(0,10), ylab = "score range",
xlab = "", axes = FALSE)

points(jitter(id-1, factor = 1), jitter(score, factor = 1), pch = 19)

axis( side = 1, labels = xaxis, at = c(1:35), pos =.3,
las = 2, tick = FALSE, line = NA)
axis(c(0,10), side = 2)

par(op)
```

Figure 4.23

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "safe",
 host = "localhost", password = "mellon")
res <- dbSendQuery(con, paste("select semtype.id, type, name, alt_score
from semtype join geogscore
where semtype.id = geogscore.semtype and score is not null")) #
dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

xaxis = c("Colour","Smell / Taste / Feel","Manner","Sound",
```

```
"Temperature","Other","Course","Effect / Character","Dimensions","Bed",
"Number","Moistness",
"Age","Elevation","Flora","Fauna","Concavity e.g. Glen*",
"Convexity e.g. Ben*","Land around*",
"Water feature*","Specific natural feature","Non specific agricultural area",
"Specific person / occupation","Specific man made area",
"Agricultural object / structure",
"Non specific settlement / building","Supernatural entity","Event",
"Relation to other features",
"Boundary","Crossing","Other","Material / Object","Weather / Air",
"Visibility")

id = dat$id
score = dat$alt_score

gray.colors(7, start = 0.3, end = 0.9, gamma = 2.2)

op <- par(mar=c(0,0,1.5,0), oma=c(13.5,4,.5,.5))
plot(NULL,xlim = c(1,35),ylim =c(0,10), ylab = "score range",
xlab = "", axes = FALSE)

points(jitter(id-1, factor = 1), jitter(score, factor = 1), pch = 19)

axis( side = 1, labels = xaxis, at = c(1:35), pos =.3, las = 2,
tick = FALSE, line = NA)
axis(c(0,10), side = 2)

par(op)
```

Figure 4.26

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "safe",
 host = "localhost", password = "mellon")
res <- dbSendQuery(con, paste("select semtype.id, type, name,
km_score from semtype join geogscore
where semtype.id = geogscore.semtype and score is not null")) #
dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

xaxis = c("Colour","Smell / Taste / Feel","Manner","Sound",
"Temperature","Other","Course","Effect / Character","Dimensions",
"Bed","Number","Moistness","Age","Elevation","Flora","Fauna",
"Concavity e.g. Glen*","Convexity e.g. Ben*","Land around*",
"Water feature*","Specific natural feature",
"Non specific agricultural area","Specific person / occupation",
"Specific man made area","Agricultural object / structure",
"Non specific settlement / building","Supernatural entity",
"Event","Relation to other features","Boundary","Crossing",
"Other","Material / Object","Weather / Air","Visibility")
```

```
id = dat$id
score = dat$km_score

gray.colors(7, start = 0.3, end = 0.9, gamma = 2.2)

op <- par(mar=c(0,0,1.5,0), oma=c(13.5,4,.5,.5))
plot(NULL,xlim = c(1,35),ylim =c(0,10), ylab = "score range", xlab = "", axes = FALSE)

points(jitter(id-1, factor = 1), jitter(score, factor = 1), pch = 19)

axis( side = 1, labels = xaxis, at = c(1:35), pos =.3, las = 2, tick = FALSE, line = NA)
axis(c(0,10), side = 2)

par(op)
```

Figure 5.9

```
library(RMySQL)
MySQL(max.con = 99, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), username = "root", dbname = "rndb", port = 3306)

res2 <- dbSendQuery(con, paste("select genelem1, count(*) as amount from list2
where specelem = 2
and genelem1 is not null and genelem1 != 8
group by concat(specelem, genelem1)
order by genelem1", sep = ""))
dat2 <- fetch(res2, n=7000)
dbClearResult(res2)

res3 <- dbSendQuery(con, paste("select genelem1, count(*) as amount from list2
where specelem = 3
and genelem1 is not null and genelem1 != 8
group by concat(specelem, genelem1)
order by genelem1", sep = ""))
dat3 <- fetch(res3, n=7000)
dbClearResult(res3)

res4 <- dbSendQuery(con, paste("select genelem1, count(*) as amount from list2
where specelem = 4
and genelem1 is not null and genelem1 != 8
group by concat(specelem, genelem1)
order by genelem1", sep = ""))
dat4 <- fetch(res4, n=7000)
dbClearResult(res4)

res5 <- dbSendQuery(con, paste("select genelem1, count(*) as amount from list2
where specelem = 5
and genelem1 is not null and genelem1 != 8
group by concat(specelem, genelem1)
order by genelem1", sep = ""))
dat5 <- fetch(res5, n=7000)
dbClearResult(res5)
```

```
res6 <- dbSendQuery(con, paste("select genelem1, count(*) as amount from list2
where specelem = 6
and genelem1 is not null and genelem1 != 8
group by concat(specelem, genelem1)
order by genelem1", sep = ""))
dat6 <- fetch(res6, n=7000)
dbClearResult(res6)

res7 <- dbSendQuery(con, paste("select genelem1, count(*) as amount from list2
where specelem = 7
and genelem1 is not null and genelem1 != 8
group by concat(specelem, genelem1)
order by genelem1", sep = ""))
dat7 <- fetch(res7, n=7000)
dbClearResult(res7)

mysqlCloseConnection(con)

#put 0s in the data

for(i in 2:7) if(length(which(dat2$genelem1 == i))==0)
dat2[max(as.integer(row.names(dat2)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat3$genelem1 == i))==0)
dat3[max(as.integer(row.names(dat3)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat4$genelem1 == i))==0)
dat4[max(as.integer(row.names(dat4)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat5$genelem1 == i))==0)
dat5[max(as.integer(row.names(dat5)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat6$genelem1 == i))==0)
dat6[max(as.integer(row.names(dat6)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat7$genelem1 == i))==0)d
at7[max(as.integer(row.names(dat7)))+1,] = c(i,0)

dat2 = dat2[order(dat2$genelem1), ]
dat3 = dat3[order(dat3$genelem1), ]
dat4 = dat4[order(dat4$genelem1), ]
dat5 = dat5[order(dat5$genelem1), ]
dat6 = dat6[order(dat6$genelem1), ]
dat7 = dat7[order(dat7$genelem1), ]

op <- par(mfrow = c(3, 2), mar=c(0,2,1.5,0), oma=c(5,4,5,.5))

barplot(c(round((dat2[,2]/sum(dat2[,2]))*100,2)),
main = "Old-Celtic", family = "serif")
barplot(c(round((dat3[,2]/sum(dat3[,2]))*100,2)),
main = "Old-Celtic / P-Celtic", family = "serif")
barplot(c(round((dat4[,2]/sum(dat4[,2]))*100,2)),
main = "Old / P-Celtic", family = "serif")
barplot(c(round((dat5[,2]/sum(dat5[,2]))*100,2)),
main = "P-Celtic / Gaelic", family = "serif")
barplot(c(round((dat6[,2]/sum(dat6[,2]))*100,2)),
main = "Gaelic", family = "serif")
```

```
axis(side = 1, at = c(1:5), labels = c("OC / P",
" P", " P/G", "      G", "    Sc"), family = "serif", tck = FALSE, lty = 0)
barplot(c(round((dat7[,2]/sum(dat7[,2]))*100,2)),
main = "Scots", family = "serif")
axis(side = 1, at = c(1:5), labels = c("OC / P", " P",
" P/G", "      G", "    Sc"), family = "serif", tck = FALSE, lty = 0)

par(op)

title(main = paste(
"Comparison of linguistic stratum of a \n generic element and specific element"
), family = "serif")
title(xlab = paste("Stratum of generic element"), family = "serif")
title(ylab = paste("Percentage occurence of for each stratum"),
family = "serif")
```

Figure 5.10

```
#function(cond){
#if(missing(cond)) stop("condition is missing")

library(RMySQL)
MySQL(max.con = 99, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb",
port = 3306, user = "root",
host = "localhost", password = "")

res2 <- dbSendQuery(con, paste("select semtype,
count(*) as amount from list2
where specelem = 2
and semtype is not null and semtype != 8
group by concat(specelem, semtype)
order by semtype", sep = ""))
dat2 <- fetch(res2, n=7000)
dbClearResult(res2)

res3 <- dbSendQuery(con, paste("select semtype,
count(*) as amount from list2
where specelem = 3
and semtype is not null and semtype != 8
group by concat(specelem, semtype)
order by semtype", sep = ""))
dat3 <- fetch(res3, n=7000)
dbClearResult(res3)

res4 <- dbSendQuery(con, paste("select semtype,
count(*) as amount from list2
where specelem = 4
and semtype is not null and semtype != 8
group by concat(specelem, semtype)
order by semtype", sep = ""))
dat4 <- fetch(res4, n=7000)
dbClearResult(res4)
```

```
res5 <- dbSendQuery(con, paste("select semtype,
count(*) as amount from list2
where specelem = 5
and semtype is not null and semtype != 8
group by concat(specelem, semtype)
order by semtype", sep = ""))
dat5 <- fetch(res5, n=7000)
dbClearResult(res5)

res6 <- dbSendQuery(con, paste("select semtype,
count(*) as amount from list2
where specelem = 6
and semtype is not null and semtype != 8
group by concat(specelem, semtype)
order by semtype", sep = ""))
dat6 <- fetch(res6, n=7000)
dbClearResult(res6)

res7 <- dbSendQuery(con, paste("select semtype,
count(*) as amount from list2
where specelem = 7
and semtype is not null and semtype != 8
group by concat(specelem, semtype)
order by semtype", sep = ""))
dat7 <- fetch(res7, n=7000)
dbClearResult(res7)

mysqlCloseConnection(con)

for(i in 2:36) if(length(which(dat2$semtype == i))==0)
dat2[max(as.integer(row.names(dat2)))+1,] = c(i,0)
for(i in 2:36) if(length(which(dat3$semtype == i))==0)
dat3[max(as.integer(row.names(dat3)))+1,] = c(i,0)
for(i in 2:36) if(length(which(dat4$semtype == i))==0)
dat4[max(as.integer(row.names(dat4)))+1,] = c(i,0)
for(i in 2:36) if(length(which(dat5$semtype == i))==0)
dat5[max(as.integer(row.names(dat5)))+1,] = c(i,0)
for(i in 2:36) if(length(which(dat6$semtype == i))==0)
dat6[max(as.integer(row.names(dat6)))+1,] = c(i,0)
for(i in 2:36) if(length(which(dat7$semtype == i))==0)
dat7[max(as.integer(row.names(dat7)))+1,] = c(i,0)

dat2 = dat2[order(dat2$semtype), ]
dat3 = dat3[order(dat3$semtype), ]
dat4 = dat4[order(dat4$semtype), ]
dat5 = dat5[order(dat5$semtype), ]
dat6 = dat6[order(dat6$semtype), ]
dat7 = dat7[order(dat7$semtype), ]

xaxis = c("Colour","Smell / Taste / Feel","Manner","Sound",
"Temperature","Other","Course","Effect / Character",
"Dimensions","Bed","Number","Moistness",
```

```
"Age","Elevation","Flora","Fauna","Concavity e.g. Glen*",
"Convexity e.g. Ben*","Land around*",
"Water feature*","Spec. nat. feature","Non spec. agric. area",
"Non spec. settlement","Spec. man made area","Agricultural object",
"Specific person","Supernatural entity","Event","Relation to features",
"Boundary","Crossing","Other","Material / Object","Weather / Air","Visibility")

rowno = as.numeric(rownames(dat6))

op <- par(mfrow = c(3, 2), mar=c(0,2,1.5,0), oma=c(10,4,5,.5),
xaxp = c(2,36,1))

#barplot(c(round((dat2[,2]/sum(dat2[,2]))*100,2)),
main = "Old-Celtic", family = "serif")

plot(NULL, main = "Old-Celtic", family = "serif",
xlim = c(1,35), ylim = c(1,40), axes= FALSE)
axis(2)
segments(as.numeric(dat2$semtype), 0, as.numeric(dat2$semtype),
c(round((dat2[,2]/sum(dat2[,2]))*100,2)), lwd = 4)
plot(NULL, main = "Old-Celtic / P-Celtic", family = "serif",
xlim = c(1,35), ylim = c(1,40), axes= FALSE)
segments(as.numeric(dat3$semtype), 0, as.numeric(dat3$semtype),
c(round((dat3[,2]/sum(dat3[,2]))*100,2)), lwd = 4)
plot(NULL, main = "P-Celtic", family = "serif", xlim = c(1,35),
ylim = c(1,40), axes= FALSE)
axis(2)
segments(as.numeric(dat4$semtype), 0, as.numeric(dat4$semtype),
c(round((dat4[,2]/sum(dat4[,2]))*100,2)), lwd = 4)
plot(NULL, main = "P-Celtic / Gaelic", family = "serif",
xlim = c(1,35), ylim = c(1,40), axes= FALSE)
segments(as.numeric(dat5$semtype), 0, as.numeric(dat5$semtype),
c(round((dat5[,2]/sum(dat5[,2]))*100,2)), lwd = 4)
plot(NULL, main = "Gaelic", family = "serif", xlim = c(1,35),
ylim = c(1,40), axes= FALSE)
axis(2)
segments(as.numeric(dat6$semtype), 0, as.numeric(dat6$semtype),
c(round((dat6[,2]/sum(dat6[,2]))*100,2)), lwd = 4)
axis(1, at = 2:36, labels = xaxis, las = 2, tick = FALSE, line = 0)
plot(NULL, main = "Scots", family = "serif", xlim = c(1,35),
ylim = c(1,40), axes= FALSE)
segments(as.numeric(dat7$semtype), 0, as.numeric(dat7$semtype),
c(round((dat7[,2]/sum(dat7[,2]))*100,2)), lwd = 4)
axis(1, at = 2:36, labels = xaxis, las = 2, tick = FALSE, line = 0)
par(op)

title(main = paste("Comparison of linguistic stratum and a semantic type"),
family = "serif")
title(ylab = paste("Percentage occurence of for each stratum"),
family = "serif")
```

Figure 5.11

```
#function(cond){
#if(missing(cond)) stop("condition is missing")

library(RMySQL)
MySQL(max.con = 99, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", port = 3306,
user = "root", host = "localhost", password = "")

res2 <- dbSendQuery(con, paste("select semtype.typeno, count(*) as amount
from list2 join semtype
where semtype.id = list2.semtype
and  specelem = 2
and semtype is not null and semtype != 8
group by concat(semtype.typeno)
order by semtype", sep = ""))
dat2 <- fetch(res2, n=7000)
dbClearResult(res2)

res3 <- dbSendQuery(con, paste("select semtype.typeno, count(*) as amount
from list2 join semtype
where semtype.id = list2.semtype
and  specelem = 3
and semtype is not null and semtype != 8
group by concat(semtype.typeno)
order by semtype", sep = ""))
dat3 <- fetch(res3, n=7000)
dbClearResult(res3)

res4 <- dbSendQuery(con, paste("select semtype.typeno, count(*) as amount
from list2 join semtype
where semtype.id = list2.semtype
and specelem = 4
and semtype is not null and semtype != 8
group by concat(semtype.typeno)
order by semtype", sep = ""))
dat4 <- fetch(res4, n=7000)
dbClearResult(res4)

res5 <- dbSendQuery(con, paste("select semtype.typeno, count(*) as amount
from list2 join semtype
where semtype.id = list2.semtype
and  specelem = 5
and semtype is not null and semtype != 8
group by concat(semtype.typeno)
order by semtype", sep = ""))
dat5 <- fetch(res5, n=7000)
dbClearResult(res5)

res6 <- dbSendQuery(con, paste("select semtype.typeno, count(*) as amount
from list2 join semtype
where semtype.id = list2.semtype
and  specelem = 6
```

```
and semtype is not null and semtype != 8
group by concat(semtype.typeno)
order by semtype", sep = ""))
dat6 <- fetch(res6, n=7000)
dbClearResult(res6)

res7 <- dbSendQuery(con, paste("select semtype.typeno, count(*) as amount
from list2 join semtype
where semtype.id = list2.semtype
and  specelem = 7
and semtype is not null and semtype != 8
group by concat(semtype.typeno)
order by semtype", sep = ""))
dat7 <- fetch(res7, n=7000)
dbClearResult(res7)

mysqlCloseConnection(con)

for(i in 2:7) if(length(which(dat2$typeno == i))==0)
dat2[max(as.integer(row.names(dat2)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat3$typeno == i))==0)
dat3[max(as.integer(row.names(dat3)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat4$typeno == i))==0)
dat4[max(as.integer(row.names(dat4)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat5$typeno == i))==0)
dat5[max(as.integer(row.names(dat5)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat6$typeno == i))==0)
dat6[max(as.integer(row.names(dat6)))+1,] = c(i,0)
for(i in 2:7) if(length(which(dat7$typeno == i))==0)
dat7[max(as.integer(row.names(dat7)))+1,] = c(i,0)

dat2 = dat2[order(dat2$typeno), ]
dat3 = dat3[order(dat3$typeno), ]
dat4 = dat4[order(dat4$typeno), ]
dat5 = dat5[order(dat5$typeno), ]
dat6 = dat6[order(dat6$typeno), ]
dat7 = dat7[order(dat7$typeno), ]

xaxis = c("Adjective", "Ecosystem", "Topography", "Human", "Situation", "Other")

rowno = as.numeric(rownames(dat6))

op <- par(mfrow = c(3, 2), mar=c(0,2,2,0), oma=c(2.5,3,4,.5), xaxp = c(2,36,1))
barplot(c(round((dat2[,2]/sum(dat2[,2]))*100,2)), main = "Old-Celtic",
family = "serif", xlim = c(0,6.3), ylim = c(1,90), axes= T)
#axis(2)
#segments(as.numeric(dat2$typeno), 0, as.numeric(dat2$typeno),
c(round((dat2[,2]/sum(dat2[,2]))*100,2)), lwd = 30, col = "grey")
barplot(c(round((dat3[,2]/sum(dat3[,2]))*100,2)), main = "Old-Celtic / P-Celtic",
family = "serif", xlim = c(0,6.3), ylim = c(1,90), axes= T)
#segments(as.numeric(dat3$typeno), 0, as.numeric(dat3$typeno),
c(round((dat3[,2]/sum(dat3[,2]))*100,2)), lwd = 30, col = "grey")
barplot(c(round((dat4[,2]/sum(dat4[,2]))*100,2)), main = "P-Celtic",
```

```
family = "serif", xlim = c(0,6.3), ylim = c(1,90), axes= T)
#axis(2)
#segments(as.numeric(dat4$typeno), 0, as.numeric(dat4$typeno),
c(round((dat4[,2]/sum(dat4[,2]))*100,2)), lwd = 30, col = "grey")
barplot(c(round((dat5[,2]/sum(dat5[,2]))*100,2)), main = "P-Celtic / Gaelic",
family = "serif", xlim = c(0,6.3), ylim = c(1,90), axes= T)
#segments(as.numeric(dat5$typeno), 0, as.numeric(dat5$typeno),
c(round((dat5[,2]/sum(dat5[,2]))*100,2)), lwd = 30, col = "grey")
barplot(c(round((dat6[,2]/sum(dat6[,2]))*100,2)), main = "Gaelic",
family = "serif", xlim = c(0,6.3), ylim = c(1,90), axes= T,
names.arg = xaxis)
#segments(as.numeric(dat6$typeno), 0, as.numeric(dat6$typeno),
c(round((dat6[,2]/sum(dat6[,2]))*100,2)), lwd = 30, col = "grey")
#axis(1, at = 2:7, labels = xaxis, las = 2,line = 1)
#axis(2)
barplot(c(round((dat7[,2]/sum(dat7[,2]))*100,2)), main = "Scots",
family = "serif", xlim = c(0,6.3), ylim = c(1,90), axes= T, axisnames = TRUE,
names.arg = xaxis)
#segments(as.numeric(dat7$typeno), 0, as.numeric(dat7$typeno),
c(round((dat7[,2]/sum(dat7[,2]))*100,2)), lwd = 30, col = "grey")
#axis(1, at = 2:7, labels = xaxis, las = 2, line = 1)
par(op)

title(main = paste("Comparison of linguistic stratum and a semantic type"), family = "serif")
#title(xlab = paste("Stratum of generic element"), family = "serif")
title(ylab = paste("Percentage occurence of for each stratum"), family = "serif")
```

Figure 6.1

```
G = c(1.7, 4.1, 7.0)
S = c(2.3, 4.9, 9.4)
x <- rbind(G, S)
#colnames(x)
barplot(x, beside = TRUE, ylab = "average geogscore",
col = c("black", "grey"))
axis(1, 1:9,  tick = FALSE, labels = c("", "Minimum", "", "",
"Average", "", "", "Maximum",""))
axis(2, 1:10)
legend(2,8, c("Gaelic", "Scots"), col = c("black", "grey"),
text.col = "black", lty = 5, lwd = 5, bty = "n")

#graph showing min, average and max geogscore for RNs with a Gaelic and Scots phonology
# got from the following MySQL code::
#   select stratum, min(score), avg(score), max(score) from geogscore join specelem
#   where source = "OS" and specelem.id = geogscore.phono group by phono
```

D.4 R source Code for Functions

genelemfinder

```
genelemfinder <- function(id,orig = "N", specelem = NULL){
if(missing(id)) stop("id is missing")
```

```
library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res1 <- dbSendQuery(con, paste("SELECT rn, specelem, genelem1,
km_score as km, nont_score as nont, pos_score as pos, alt_score as alt
from geogscore where id = ", id))
dat1 <- fetch(res1, n=7000)

if(orig == "Y") orig = dat1$specelem else orig = 8
if(is.null(specelem) == FALSE) dat1$specelem = paste(" = ",specelem) else dat1$specelem =
paste(" between 0 and 8 ")

dbClearResult(res1)
mysqlCloseConnection(con)

library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("select
`genelem2table`.`element` AS `element`,
min(`geogscore`.`km_score`) AS `minkm`,
avg(`geogscore`.`km_score`) AS `avgkm`,
max(`geogscore`.`km_score`) AS `maxkm`,
min(`geogscore`.`nont_score`) AS `minnont`,
avg(`geogscore`.`nont_score`) AS `avgnont`,
max(`geogscore`.`nont_score`) AS `maxnont`,
min(`geogscore`.`alt_score`) AS `minalt`,
avg(`geogscore`.`alt_score`) AS `avgalt`,
max(`geogscore`.`alt_score`) AS `maxalt`,
min(`geogscore`.`pos_score`) AS `minpos`,
avg(`geogscore`.`pos_score`) AS `avgpos`,
max(`geogscore`.`pos_score`) AS `maxpos`
 from  (`geogscore` join `genelem2table`)
where  (concat(`geogscore`.`genelem1`,
`geogscore`.`genelem2`) = `genelem2table`.`code`)
 and geogscore.specelem ", dat1$specelem,

"and left(`genelem2table`.`code`,1) <= ", orig,

 "group by `genelem2table`.`code`")) #

dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

data1 = subset(dat, dat1$km  <= dat$maxkm)
data2 = subset(data1, dat1$km >= data1$minkm)
data3 = subset(data2, dat1$nont  <= data2$maxnont)
data4 = subset(data3, dat1$nont >= data3$minnont)
data5 = subset(data4, dat1$pos  <= data4$maxpos)
data6 = subset(data5, dat1$pos >= data5$minpos)
```

```
data7 = subset(data6, dat1$alt  <= data6$maxalt)
data8 = subset(data7, dat1$alt >= data7$minalt)
data8
}
```

genelemfinder2

```
genelemfinder2 <- function(id, tol, range){
if(missing(id)) stop("id is missing")
if(missing(tol)) tol = 0
if(missing(range)) range = 90000

library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("SELECT rn, score, LAT, LON from geogscore
where id = ", id))
dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res1 <- dbSendQuery(con, paste("SELECT semtype.name as meaning,
count(*) as amount, concat(round((count(*)/

(SELECT count(*) FROM geogscore join semtype where
 geogscore.score between", dat$score - tol,  " and ",  dat$score + tol, "
 and  geogscore.lat between ( select  lat-", range,  " from  geogscore
where id = ", id, ")
 and  ( select  lat+", range,  " from geogscore where id = ", id, ")
 and  geogscore.lon between ( select   lon-", range,  "  from  geogscore
 where  id = ", id, ")
 and  ( select   lon+", range,  "  from  geogscore where  id = ", id, ")
 and  geogscore.semtype = semtype.id)

)*100,2),'\%') as percentage FROM geogscore join semtype
where
 geogscore.score between", dat$score - tol,  " and ",  dat$score + tol, "
 and  geogscore.lat between ( select  lat-", range,  " from  geogscore
where id = ", id, ")
 and  ( select  lat+", range,  " from geogscore where id = ", id, ")
 and  geogscore.lon between ( select   lon-", range,  "  from  geogscore
where  id = ", id, ")
 and  ( select   lon+", range,  "  from  geogscore where  id = ", id, ")
 and  geogscore.semtype = semtype.id
 group by semtype.id
 order by  count(*) desc")) #

dat1 <- fetch(res1, n=7000)
```

```
dbClearResult(res1)
mysqlCloseConnection(con)

dat1
}
```

specelemfinder

```
specelemfinder <- function(id,out = "p"){
if(missing(id)) stop("id is missing")

library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("SELECT rn, km_score as km,
 nont_score as nont, pos_score as pos, alt_score as alt
from geogscore where id = ", id))
dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res1 <- dbSendQuery(con, paste("SELECT stratum,
round(abs(", ((dat$nont+dat$km+dat$pos+dat$alt)/4),
 "-((avgkm+avgnont+avgpos+avgalt)/4)),2) as prob,
minkm, maxkm, minnont, maxnont, minalt, maxalt, minpos,
 maxpos FROM specelemfinder g
where", dat$km, "between minkm and maxkm and",
dat$nont, "between minnont and maxnont and",
dat$pos, "between minpos and maxpos and",
dat$alt, "between minalt and maxalt
order by prob")) #

dat1 <- fetch(res1, n=7000)

dbClearResult(res1)
mysqlCloseConnection(con)

if(out == "l") writeClipboard((print(xtable(dat))), format = 1) else (dat1)

}
```

specelemfinder2

```
specelemfinder2 = function(id, tol, range){
if(missing(id)) stop("id is missing")
if(missing(tol)) tol = 0
if(missing(range)) range = 90000

library(RMySQL)
```

```
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("SELECT rn, score, LAT, LON from
geogscore where id = ", id))
dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res1 <- dbSendQuery(con, paste("SELECT specelem.stratum as stratum,
count(*) as amount, concat(round((count(*)/

(SELECT count(*) FROM geogscore join specelem where
 geogscore.score between", dat$score - tol,  " and ",  dat$score + tol, "
 and  geogscore.lat between ( select  lat-", range,  " from  geogscore
where id = ", id, ")
 and  ( select  lat+", range,  " from geogscore where id = ", id, ")
 and  geogscore.lon between ( select   lon-", range,  "  from  geogscore
where  id = ", id, ")
 and  ( select   lon+", range,  "  from  geogscore where  id = ", id, ")
 and  geogscore.specelem = specelem.id
and specelem.id != 1)

)*100,2),'\\%') as percentage FROM geogscore join specelem
where
 geogscore.score between", dat$score - tol,  " and ",  dat$score + tol, "
 and  geogscore.lat between ( select  lat-", range,  " from  geogscore
where id = ", id, ")
 and  ( select  lat+", range,  " from geogscore where id = ", id, ")
 and  geogscore.lon between ( select   lon-", range,  "  from  geogscore
where  id = ", id, ")
 and  ( select   lon+", range,  "  from  geogscore where  id = ", id, ")
 and  geogscore.specelem = specelem.id
and specelem.id != 1
 group by specelem.id
 order by  count(*) desc")) #

dat1 <- fetch(res1, n=7000)

dbClearResult(res1)
mysqlCloseConnection(con)

dat1
}

specelemfinder2(3761)
```

specelem1

```
specelem1 = function(cond){
```

```
if(missing(cond)) stop("condition is missing")

library(RMySQL)
MySQL(max.con = 99, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "root",
host = "localhost", password = "")

res2 <- dbSendQuery(con, paste("select ",cond,"_score as ",cond,",
count(*) as amount from geogscore
where specelem = 2
group by ",cond,"_score", sep = ""))
dat2 <- fetch(res2, n=7000)
dbClearResult(res2)

res3 <- dbSendQuery(con, paste("select ",cond,"_score as ",cond,",
 count(*) as amount from geogscore
where specelem = 3
group by ",cond,"_score", sep = ""))
dat3 <- fetch(res3, n=7000)
dbClearResult(res3)

res4 <- dbSendQuery(con, paste("select ",cond,"_score as ",cond,",
count(*) as amount from geogscore
where specelem = 4
group by ",cond,"_score", sep = ""))
dat4 <- fetch(res4, n=7000)
dbClearResult(res4)

res5 <- dbSendQuery(con, paste("select ",cond,"_score as ",cond,",
count(*) as amount from geogscore
where specelem = 5
group by ",cond,"_score", sep = ""))
dat5 <- fetch(res5, n=7000)
dbClearResult(res5)

res6 <- dbSendQuery(con, paste("select ",cond,"_score as ",cond,",
count(*) as amount from geogscore
where specelem = 6
group by ",cond,"_score", sep = ""))
dat6 <- fetch(res6, n=7000)
dbClearResult(res6)

res7 <- dbSendQuery(con, paste("select ",cond,"_score as ",cond,",
count(*) as amount from geogscore
where specelem = 7
group by ",cond,"_score", sep = ""))
dat7 <- fetch(res7, n=7000)
dbClearResult(res7)

mysqlCloseConnection(con)

op <- par(mfrow = c(3, 2), mar=c(0,0,1.5,0), oma=c(5,4,5,.5))
```

```
plot(NULL, xlim =c(1,10), ylim= c(1,50),
tck = FALSE, labels = FALSE,
mgp = c(4,.5,0), main = "Old Celtic", family = "serif")
lines(dat2[,1],  round((dat2[,2]/sum(dat2[,2]))*100,2))
axis(2, labels = TRUE, family = "serif")

plot(NULL, xlim =c(1,10), ylim= c(1,50),
tck = FALSE, labels = FALSE,
mgp = c(4,.5,0), main = "Old Celtic / P-Celtic", family = "serif")
lines(dat3[,1],  round((dat3[,2]/sum(dat3[,2]))*100,2))

plot(NULL, xlim =c(1,10), ylim= c(1,50),
tck = FALSE, labels = FALSE,
mgp = c(4,.5,0), main = "P-Celtic", family = "serif")
lines(dat4[,1],  round((dat4[,2]/sum(dat4[,2]))*100,2))
axis(2, labels = TRUE, family = "serif")

plot(NULL, xlim =c(1,10), ylim= c(1,50),
tck = FALSE, labels = FALSE,
mgp = c(4,.5,0), main = "P-Celtic / Gaelic", family = "serif")
lines(dat5[,1],  round((dat5[,2]/sum(dat5[,2]))*100,2))

plot(NULL, xlim =c(1,10), ylim= c(1,50),
tck = FALSE, labels = FALSE,
mgp = c(4,.5,0), main = "Gaelic", family = "serif")
lines(dat6[,1],  round((dat6[,2]/sum(dat6[,2]))*100,2))
axis(2, labels = TRUE, family = "serif")
axis(1, labels = TRUE, family = "serif")

plot(NULL, xlim =c(1,10), ylim= c(1,50),
tck = FALSE, labels = FALSE,
mgp = c(4,.5,0), main = "Scots", family = "serif")
lines(dat7[,1],  round((dat7[,2]/sum(dat7[,2]))*100,2))
axis(1, labels = TRUE, family = "serif")

par(op)

title(main = paste("Occurence of ", cond, "_score for each stratum",
sep = ""), family = "serif")
title(xlab = paste(cond, "_score", sep = ""), family = "serif")
title(ylab = paste("percentage occurence of for each stratum"),
family = "serif")
}
```

specelem2

```
specelem2 = function(cond){
if(missing(cond)) stop("condition is missing")

#cond = "pos"

library(RMySQL)
MySQL(max.con = 1, fetch.default.rec =7000, force.reload = F)
```

```
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "root",
host = "localhost", password = "")
res <- dbSendQuery(con, paste("select min(", cond,"_score) as min,
avg(", cond,"_score) as avg, max(", cond,"_score) as max, specelem
from geogscore
where specelem !=1
group by specelem", sep = ""))
dat <- fetch(res, n=7000)

plot(dat$specelem, dat$avg, type = "l", ylab = "", xlab = "",
ylim =c(0,10), xlim=c(1,8), axes = FALSE)
lines(dat$specelem, dat$avg)
points(dat$specelem, dat$avg)
lines(dat$specelem, dat$max)
points(dat$specelem, dat$max)
lines(dat$specelem, dat$min)
points(dat$specelem, dat$min)
axis(1, 2:7, labels = c("Old Celtic", "Old /\n P-Celtic",
"P-Celtic", "P-Celtic \n/ Gaelic", "Gaelic", "Scots"),
line = 0, tck = .02, family = "serif")
axis(2, 1:10, labels = 1:10, tck = -.02, line = -3)
#legend(7, 10, "max")
#axis(4, 1:10, labels = c("min", "", "avg","","","","","","max",""),
#tick = FALSE, pos = 6.9, padj = 0)
text(7.5,dat$max[6], "max")
text(7.5,dat$avg[6], "avg")
text(7.5,dat$min[6], "min")
text(0.8, 6, paste(cond, "_score", sep = ""), srt = 90,
family = "serif")
title(main = paste("min, avg and max for ",cond, "_score", sep = ""),
family = "serif")

}
```

semtypefinder

```
semtypefinder <- function(id){
if(missing(id)) stop("id is missing")

library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "root",
host = "localhost", password = "")
res1 <- dbSendQuery(con, paste("SELECT rn, specelem,
 km_score as km, nont_score as nont, pos_score as pos, alt_score as alt
from geogscore where id = ", id))
dat1 <- fetch(res1, n=7000)

dbClearResult(res1)
mysqlCloseConnection(con)

library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
```

```
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "root",
host = "localhost", password = "")
res <- dbSendQuery(con, paste(" select
  `semtype`.`name` AS `name`,
min(`geogscore`.`km_score`) AS `minkm`,
avg(`geogscore`.`km_score`) AS `avgkm`,
max(`geogscore`.`km_score`) AS `maxkm`,
min(`geogscore`.`nont_score`) AS `minnont`,
avg(`geogscore`.`nont_score`) AS `avgnont`,
max(`geogscore`.`nont_score`) AS `maxnont`,
min(`geogscore`.`alt_score`) AS `minalt`,
avg(`geogscore`.`alt_score`) AS `avgalt`,
max(`geogscore`.`alt_score`) AS `maxalt`,
min(`geogscore`.`pos_score`) AS `minpos`,
avg(`geogscore`.`pos_score`) AS `avgpos`,
max(`geogscore`.`pos_score`) AS `maxpos`
 from  (`geogscore` join `semtype`)
where  ((`geogscore`.`semtype` = `semtype`.`id`)
 and  (`geogscore`.`source` <> _utf8'F1')
 and  (`geogscore`.`semtype` <> 1))
 and geogscore.specelem =", dat1$specelem,
" group by `semtype`.`id`"))
dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

data = subset(subset(dat, dat1$km <= dat$maxkm),dat1$km >= dat$minkm)
data = subset(subset(data, dat1$nont <= dat$maxnont),dat1$nont >= dat$minnont)
data = subset(subset(data, dat1$pos <= dat$maxpos),dat1$pos >= dat$minpos)
data = subset(subset(data, dat1$alt <= dat$maxalt),dat1$alt >= dat$minalt)
data = subset(data, as.numeric(minkm) == TRUE) #to remove NAs
data
}
```

semtypefinder2

```
semtypefinder2 <- function(id, tol, range){
if(missing(id)) stop("id is missing")
if(missing(tol)) tol = 0
if(missing(range)) range = 90000

library(RMySQL)
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res <- dbSendQuery(con, paste("SELECT rn, score, LAT, LON from
geogscore where id = ", id))
dat <- fetch(res, n=7000)

dbClearResult(res)
mysqlCloseConnection(con)

library(RMySQL)
```

```
MySQL(max.con = 5, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)
res1 <- dbSendQuery(con, paste("SELECT semtype.name as meaning,
count(*) as amount, concat(round((count(*)/

(SELECT count(*) FROM geogscore join semtype where
 geogscore.score between", dat$score - tol,  " and ",  dat$score + tol, "
 and  geogscore.lat between ( select  lat-", range,  " from
geogscore where id = ", id, ")
 and  ( select  lat+", range,  " from geogscore where id = ", id, ")
 and  geogscore.lon between ( select   lon-", range,  "  from
geogscore where  id = ", id, ")
 and  ( select   lon+", range,  "  from  geogscore where  id = ", id, ")
 and  geogscore.semtype = semtype.id)

)*100,2),'\\%') as percentage FROM geogscore join semtype
where
 geogscore.score between", dat$score - tol,  " and ",  dat$score + tol, "
 and  geogscore.lat between ( select  lat-", range,  " from
geogscore where id = ", id, ")
 and  ( select  lat+", range,  " from geogscore where id = ", id, ")
 and  geogscore.lon between ( select   lon-", range,  "  from
geogscore where  id = ", id, ")
 and  ( select   lon+", range,  "  from  geogscore where  id = ", id, ")
 and  geogscore.semtype = semtype.id
 group by semtype.id
 order by  count(*) desc")) #

dat1 <- fetch(res1, n=7000)

dbClearResult(res1)
mysqlCloseConnection(con)

dat1
}
```

stratum

```
stratum = function(cond){
if(missing(cond)) stop("condition is missing")

library(RMySQL)
MySQL(max.con = 99, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), group = "rs-dbi", port = 3306)

res1 <- dbSendQuery(con, paste("select avg(", cond, ") as cond,
 alt_score as altave from geogscore
where source = 'OS' and score is not null
group by round(alt_score,0)"))
dat1 <- fetch(res1, n=7000)
dbClearResult(res1)

res2 <- dbSendQuery(con, paste("select avg(", cond, ") as cond,
```

```
 km_score as km from geogscore
where source = 'OS' and score is not null
group by geogscore.km_score"))
dat2 <- fetch(res2, n=7000)
dbClearResult(res2)

res3 <- dbSendQuery(con, paste("select avg(", cond, ") as cond,
nont_score as nont from geogscore
where source = 'OS' and score is not null
group by geogscore.nont_score"))
dat3 <- fetch(res3, n=7000)
dbClearResult(res3)

res4 <- dbSendQuery(con, paste("select avg(", cond, ") as cond,
pos_score as pos from geogscore
where source = 'OS' and score is not null
group by geogscore.pos_score"))
dat4 <- fetch(res4, n=7000)
dbClearResult(res4)

mysqlCloseConnection(con)

ylim = c(min(as.numeric(c(dat1$cond, dat2$cond, dat3$cond, dat4$cond))),
max(as.numeric(c(dat1$cond, dat2$cond, dat3$cond, dat4$cond))))

op <- par(mfrow = c(2, 2), mar=c(0,0,1.5,0), oma=c(5,4,3,.5))

#nont_score

plot(NULL, xlim =c(1,10), ylim= ylim,
tck = FALSE, labels = FALSE,
mgp = c(4,.5,0), main = "nont_score", family = "serif")
lines(dat3$nont, dat3$cond)
axis(2, labels = TRUE, family = "serif")

#km_score

plot(NULL, xlim =c(1,10), ylim= ylim, xlab ="lingscore",
tck = FALSE, labels = FALSE,
lwd = .2,   mgp = c(4,.5,0), main = "km_score", family = "serif")
lines(dat2$km, dat2$cond)

#alt_score

plot(NULL, xlim =c(1,10), ylim= ylim, xlab ="lingscore",
tck = FALSE, labels = FALSE,
  mgp = c(4,.5,0),main = "alt_score", family = "serif")
lines(dat1$alt, dat1$cond)
axis(1, labels = TRUE)
axis(2, labels = TRUE)

#pos_score
```

```
plot(NULL, xlim =c(1,10), ylim= ylim, xlab ="lingscore",
 tck = FALSE, labels = FALSE,
  mgp = c(4,.5,0), main = "pos_score", family = "serif")
lines(dat4$pos, dat4$cond)
axis(1, labels = TRUE)

par(op)

title(xlab = "components of geogscore", family = "serif")
title(ylab = paste(cond, "                                              ",
 cond), family = "serif")

all <- par(xpd = TRUE)

legend(x = 2.5, y = 11.4, legend = c(cond),  col = c("black"),
pch = 19, horiz = TRUE, bty = "n")

par(all)

}
```

Onymagic

```
#Jake's Onymagical device

require(tcltk)      # Load the TclTk package
tclRequire("BWidget")
tt <- tktoplevel()  # Create a new toplevel window
tktitle(tt) <- "ONYMAGIC"  # Give the window a title

cond1 <- tclVar(cond1)
cond2 <- tclVar(cond2)

entry.cond1 <-tkentry(tt,width="30",textvariable=cond1)
entry.cond2 <-tkentry(tt,width="30",textvariable=cond2)

#legend
legend1 <- tclVar(legend1)
legend2 <- tclVar(legend2)

entry.legend1 <-tkentry(tt,width="20",textvariable=legend1)
entry.legend2 <-tkentry(tt,width="20",textvariable=legend2)

#database radiobuttons

OS.but <- tkradiobutton(tt)
RNDB.but <- tkradiobutton(tt)

dbValue <- tclVar("2")
tkconfigure(OS.but,variable=dbValue,value=1)
tkconfigure(RNDB.but,variable=dbValue,value=2)

#point colour radiobuttons
```

```
Pointcolblack1.but <- tkradiobutton(tt)
Pointcolgrey1.but <- tkradiobutton(tt)

colValue1 <- tclVar("1")
tkconfigure(Pointcolblack1.but,variable=colValue1,value=1)
tkconfigure(Pointcolgrey1.but,variable=colValue1,value=2)

Pointcolblack2.but <- tkradiobutton(tt)
Pointcolgrey2.but <- tkradiobutton(tt)

colValue2 <- tclVar("2")
tkconfigure(Pointcolblack2.but,variable=colValue2,value=1)
tkconfigure(Pointcolgrey2.but,variable=colValue2,value=2)

#pointtypes

ptypebox1 <- tclVar(ptypebox1)#@
ptypebox2 <- tclVar(ptypebox2)

points <- c("Circle","Square","Triangle","Bullet")
ptypebox1 <- tkwidget(tt,"ComboBox",editable=FALSE,values=points)
ptypebox2 <- tkwidget(tt,"ComboBox",editable=FALSE,values=points)

#contour options
contour <- tclVar(contour)
entry.contour <-tkentry(tt,width="30",textvariable=contour)

#xyrange

ALL.but <- tkradiobutton(tt)
MOST.but <- tkradiobutton(tt)
MINMAX.but <- tkradiobutton(tt)

xyValue <- tclVar("3")
tkconfigure(ALL.but,variable=xyValue,value=1)
tkconfigure(MOST.but,variable=xyValue,value=2)
tkconfigure(MINMAX.but,variable=xyValue,value=3)

#title
title <- tclVar(title)
entry.title <-tkentry(tt,width="40",textvariable=title)

#northarrow
na <- tkcheckbutton(tt)
naValue <- tclVar(naValue)

#Close button
OnOK <- function()
{
    ptypeChoice1 <<-points[as.numeric(tclvalue(tcl(ptypebox1,"getvalue")))+1]
    ptypeChoice2 <<-points[as.numeric(tclvalue(tcl(ptypebox2,"getvalue")))+1]
tkdestroy(tt)
}
```

```
OK.but <-tkbutton(tt,text="   OK   ",command=OnOK)

#Layout

cond1label = tklabel(tt,text="First SQL query:")
cond2label = tklabel(tt,text="Second SQL query:")
tkgrid(cond1label,cond2label)
tkgrid(entry.cond1, entry.cond2)

legend1label = tklabel(tt,text="Legend 1:")
legend2label = tklabel(tt,text="Legend 2:")
tkgrid(legend1label,legend2label)
tkgrid(entry.legend1, entry.legend2)

#point colours
tkgrid(tklabel(tt,text="Point colour 1:"), tklabel(tt,text="Point colour 2:"))
blacklab1 = tklabel(tt,text="Black ")
blacklab2 = tklabel(tt,text="Black ")
greylab1 = tklabel(tt,text="Grey ")
greylab2 = tklabel(tt,text="Grey ")
tkgrid(blacklab1,Pointcolblack1.but, blacklab2,Pointcolblack2.but)
tkgrid(greylab1,Pointcolgrey1.but, greylab2,Pointcolgrey2.but)
tkgrid.configure(Pointcolgrey1.but,Pointcolblack1.but,
Pointcolblack2.but,Pointcolgrey2.but, sticky = "w")
tkgrid.configure(greylab1,blacklab1, blacklab2,greylab2,sticky = "e")

#point types
ptype1label = tklabel(tt,text="Point type 1:")
ptype2label = tklabel(tt,text="Point type 2:")
tkgrid(ptype1label,ptype2label)
tkgrid(ptypebox1, ptypebox2)

tkgrid(tklabel(tt,text="Overall title:"), columnspan=2)
tkgrid(entry.title, columnspan=2)

contourlabel = tklabel(tt,text="Contour Z paramater:")
tkgrid(contourlabel, entry.contour)

tkgrid(tklabel(tt,text="Pick a data source:"))
tkgrid(tklabel(tt,text="OS "),OS.but)
tkgrid(tklabel(tt,text="RNDB "),RNDB.but)
#tkgrid.configure(tklabel(tt,text="RNDB "),tklabel(tt,text="OS "), sticky = "e")
#tkgrid.configure(OS.but,RNDB.but, sticky = "w")

tkgrid(tklabel(tt,text="Pick a range:"))
tkgrid(tklabel(tt,text="ALL "),ALL.but)
tkgrid(tklabel(tt,text="MOST "),MOST.but)
tkgrid(tklabel(tt,text="MINMAX "),MINMAX.but)

tkconfigure(na,variable=naValue)
tkgrid(tklabel(tt,text="Northarrow: "),na)

tkgrid(OK.but)
```

```
tkfocus(tt)
tkwait.window(tt)

# Test the result

dbVal <- as.character(tclvalue(dbValue))

if (dbVal == 1) db = "OS"
if (dbVal == 2) db = "RNDB"

cond1 <- as.character(tclvalue(cond1))
cond2 <- as.character(tclvalue(cond2))

legend1 <- as.character(tclvalue(legend1))
legend2 <- as.character(tclvalue(legend2))

colVal1 <- as.character(tclvalue(colValue1))

if (colVal1 == 1) pointcol1 = "black"
if (colVal1 == 2) pointcol1 = "grey"

colVal2 <- as.character(tclvalue(colValue2))

if (colVal2 == 1) pointcol2 = "black"
if (colVal2 == 2) pointcol2 = "grey"

xyVal <- as.character(tclvalue(xyValue))

if (xyVal == 1) area = "ALL"
if (xyVal == 2) area = "MOST"
if (xyVal == 3) area = "MINMAX"

title <- as.character(tclvalue(title))

naVal <- as.character(tclvalue(naValue))

if (naVal=="1") narrow = 1
if (naVal=="0") narrow = 0

#pointype # if no pointtype is chosen, it seems to fail without these lines:
if(length(ptypeChoice1) == 0) ptypeChoice1 = ""
if(length(ptypeChoice2) == 0) ptypeChoice2 = ""

if(ptypeChoice1 == "Circle")  pointtype1 = 19
if(ptypeChoice1 == "Bullet")  pointtype1 = 20
if(ptypeChoice1 == "Square")  pointtype1 = 22
if(ptypeChoice1 == "Diamond") pointtype1 = 23

if(ptypeChoice2 == "Circle")  pointtype2 = 19
if(ptypeChoice2 == "Bullet")  pointtype2 = 20
if(ptypeChoice2 == "Square")  pointtype2 = 22
if(ptypeChoice2 == "Diamond") pointtype2 = 23
```

```
#contour

contour <- as.character(tclvalue(contour))
if(contour == "amount") contouravg = "count(*)" else contouravg =
paste("avg(",contour,")")
if(contour == "amount") contour = "id"

#start the process

#Get map data.
library(maptools)
data <- read.shape("scotland.shp")
mappolys <- Map2poly(data)

#Run first query.
library(RMySQL)

if(is.null(cond1) == TRUE) cond1 = "id is not null"
if(cond1 == "") cond1 = "id is not null"

if(db == "RNDB") dbname = "rndb" else dbname = "allos"
MySQL(max.con = 5, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb",
user = "root", host = "localhost", password = "")
if(db == "RNDB") query = paste("SELECT LAT, LON from list2 where", cond1,
 "and LAT != 100000") else query = paste("select name, EASTING as LON,
NORTHING as LAT from allos.osscot where", cond1)
if(contour != "") query = paste("SELECT LAT, LON, ", contouravg, "as zaxis from list2
where LAT != 100000 and", contour, "is not null and", cond1, "
group by concat(LAT,LON)")
res1 <- dbSendQuery(con, query) #
dat1 <- fetch(res1, n=7000)

dbClearResult(res1)
mysqlCloseConnection(con)

#Run second query
if(cond2 == "") cond2 = "id is null"

if(db == "RNDB") dbname = "rndb" else dbname = "allos"
MySQL(max.con = 5, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb",
user = "root", host = "localhost", password = "")
if(db == "RNDB") query = paste("SELECT LAT, LON from list2 where", cond2,
"and LAT != 100000") else query = paste("select EASTING as LON, NORTHING as LAT
from allos.osscot where", cond2)
res2 <- dbSendQuery(con, query) #
dat2 <- fetch(res2, n=7000)

dbClearResult(res2)
mysqlCloseConnection(con)

#Put it all together!
```

```
LAT = c(dat1$LAT, dat2$LAT)
LON = c(dat1$LON, dat2$LON)

if(area == "AOS") xlim = c(220000,420000) else if(area == "ALL")
xlim = c(000000,500000) else if(area == "MINMAX")
xlim = c(min(LAT)-15000,max(LAT)+15000) else  xlim =  c(050000,420000)
if(area == "AOS") ylim = c(680000,870000) else if(area == "ALL")
ylim = c(500000,1250000) else if(area == "MINMAX")
ylim = c(min(LON)-15000,max(LON)+15000) else  ylim = c(520000,1000000)

##
if(contour == "") {
par(mar = c(.1,.1,3,.1))
plot(mappolys, tck = FALSE, xlim = xlim, ylim = ylim,labels = TRUE)
title(main = title, family = "serif")

if(max(as.numeric(rownames(dat2)))+max(as.numeric(rownames(dat1))) > 1000)
 cex = .5 else cex = 1

points(dat1$LAT, dat1$LON, col = pointcol1, bg = pointcol1,
pch = pointtype1, cex = cex)
if(cond2 == "") NULL else points(dat2$LAT, dat2$LON, col = pointcol2,
 bg = pointcol2, pch = pointtype2, cex = cex)

x = cbind(dat1$LAT, dat1$LON)
hpts = chull(x)
hpts = c(hpts, hpts[1])
if(outpol == "YES") polygon(x[hpts, ], border = "black")

x = cbind(dat2$LAT, dat2$LON)
hpts = chull(x)
hpts = c(hpts, hpts[1])
if(outpol == "YES") polygon(x[hpts, ], border = "black")
} else {

library(akima)
surface <- interp(dat1$LAT, dat1$LON, dat1$zaxis)
par(oma = c(0,0,0,0))
filled.contour(surface, color = colorRampPalette(c("black", "grey", "white")),
zlim = c(min(dat1$zaxis),max(dat1$zaxis)),
xlim = xlim, ylim = ylim,
nlevels = 20,
plot.axes={plot(mappolys, add = TRUE, xlim = xlim, ylim = ylim)
#key.axes = axis(4, ""axis),
title(main = title, family = "serif")}
)
}

if(contour == "") if(legend2 == "")legend(locator(n = 1, type = "n"),
legend = legend1, pch = pointtype1, pt.bg = pointcol1, col = pointcol1,
bg = "white", par(family = "serif")) else legend(locator(n = 1, type = "n"),
legend = c(legend1, legend2), pch = c(pointtype1, pointtype2),
```

```
 col = c(pointcol1, pointcol2), pt.bg = c(pointcol1, pointcol2), bg = "white")

#northarrow
if(narrow == 1) northarrow() #Can't turn this off! problem at tcl level.
```

lingscorecomparer

```
library(outliers)
require(tcltk)      # Load the TclTk package
tt <- tktoplevel()  # Create a new toplevel window
tktitle(tt) <- "lingscore comparer"  # Give the window a title

cond1 <- tclVar(cond1)
cond2 <- tclVar(cond2)

entry.cond1 <-tkentry(tt,width="50",textvariable=cond1)
entry.cond2 <-tkentry(tt,width="50",textvariable=cond2)

cond1label = tklabel(tt,text="First SQL query:")
cond2label = tklabel(tt,text="Second SQL query:")

text1 = tclVar(text1)
text2 = tclVar(text2)

entry.text1 = tkentry(tt,width="30",textvariable=text1)
entry.text2 = tkentry(tt,width="30",textvariable=text2)

jit <- tkcheckbutton(tt)
jitValue <- tclVar(jitValue)

trim <- tkcheckbutton(tt)
trimValue <- tclVar(trimValue)

adj <- tkcheckbutton(tt)

avg <- tkcheckbutton(tt)

text1label = tklabel(tt,text="First label")
text2label = tklabel(tt,text="Second label")

tkgrid(cond1label,cond2label)
tkgrid(entry.cond1, entry.cond2)

tkgrid(text1label,text2label)
tkgrid(entry.text1, entry.text2)

tkconfigure(jit,variable=jitValue)
tkgrid(tklabel(tt,text="Add jitter: "),jit)

tkconfigure(trim,variable=trimValue)
tkgrid(tklabel(tt,text="Trim outliers: "),trim)

tkconfigure(adj,variable=adjValue)
```

```
tkgrid(tklabel(tt,text="Adjust: "),adj)

tkconfigure(avg,variable=avgValue)
tkgrid(tklabel(tt,text="Plot average: "),avg)

OnOK <- function()
{
tkdestroy(tt)
}
OK.but <-tkbutton(tt,text="   OK   ",command=OnOK)

tkgrid(OK.but)
tkfocus(tt)
tkwait.window(tt)

cond1 <- as.character(tclvalue(cond1))
cond2 <- as.character(tclvalue(cond2))

text1 <- as.character(tclvalue(text1))
text2 <- as.character(tclvalue(text2))

# configure query
adjVal <- as.character(tclvalue(adjValue))
adjVal #@

if (adjVal=="1") adjer = 1
if (adjVal=="0") adjer = 0
adjer

avgVal <- as.character(tclvalue(avgValue))
avgVal

if (avgVal=="1") avger = 1
if (avgVal=="0") avger = 0

lingscore1 = (paste("lingscore"))
lingscore2 = (paste("lingscore"))

if(adjer == 1){
if(any(grep("semtype",cond1))) lingscore1 =
(paste("((avg_specelem_score + avg_genelem_score)/2)"))
if(any(grep("genelem",cond1))) lingscore1 =
(paste("((avg_specelem_score + avg_semtype_score)/2)"))
if(any(grep("specelem",cond1))) lingscore1 =
(paste("((avg_semtype_score + avg_genelem_score)/2)"))

if(any(grep("semtype",cond2))) lingscore2 =
(paste("((avg_specelem_score + avg_genelem_score)/2)"))
if(any(grep("genelem",cond2))) lingscore2 =
(paste("((avg_specelem_score + avg_semtype_score)/2)"))
if(any(grep("specelem",cond2))) lingscore2 =
(paste("((avg_semtype_score + avg_genelem_score)/2)"))
if(any(grep("id ",cond2))) lingscore2 = lingscore1
```

```
}

if(cond2 == "") cond2 = ("id is null")

jitter = 0

jitVal <- as.character(tclvalue(jitValue))

if (jitVal=="1") jitter = 1
if (jitVal=="0") jitter = 0

trimmer = 0

trimVal <- as.character(tclvalue(trimValue))

if (trimVal=="1") trimmer = 1
if (trimVal=="0") trimmer = 0

#get the data

library(RMySQL)
MySQL(max.con = 9, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "root",
host = "localhost", password = "")
res1 <- dbSendQuery(con, paste("SELECT nont_score, km_score, alt_score,
pos_score,", lingscore1, " as lingscore from finalscore
where", cond1, "and concat(nont_score, km_score, alt_score, pos_score,
finalscore.lingscore) is not null")) #
dat1 <- fetch(res1, n=7000)

dbClearResult(res1)
mysqlCloseConnection(con)

q1 = paste("SELECT nont_score, km_score, alt_score, pos_score,",
 lingscore1, " as lingscore from finalscore where", cond1,
"and concat(nont_score, km_score, alt_score, pos_score, finalscore.lingscore) is not null")

library(RMySQL)
MySQL(max.con = 9, fetch.default.rec =7000, force.reload = F)
con <- dbConnect(dbDriver("MySQL"), dbname = "rndb", user = "root",
host = "localhost", password = "")
res2 <- dbSendQuery(con, paste("SELECT nont_score, km_score,
alt_score, pos_score,", lingscore2, " as lingscore from finalscore
where", cond2, "and concat(nont_score, km_score, alt_score,
pos_score, finalscore.lingscore) is not null")) #
dat2 <- fetch(res2, n=7000)

dbClearResult(res2)
mysqlCloseConnection(con)

q2 = paste("SELECT nont_score, km_score, alt_score, pos_score,",
lingscore2, " as lingscore from finalscore where", cond2,
"and concat(nont_score, km_score, alt_score, pos_score,
```

```
finalscore.lingscore) is not null")

#set variables so we can lay about with them
d1lingscore = dat1$lingscore
d1kmscore = dat1$km_score
d1nontscore = dat1$nont_score
d1posscore = dat1$pos_score
d1altscore = dat1$alt_score

d2lingscore = dat2$lingscore
d2kmscore = dat2$km_score
d2nontscore = dat2$nont_score
d2posscore = dat2$pos_score
d2altscore = dat2$alt_score

#add jitter if needed

if(jitter == 1) d1nontscore = jitter(d1nontscore, factor = .6)
if(jitter == 1) d1kmscore = jitter(d1kmscore, factor = .6)
if(jitter == 1) d1altscore = jitter(d1altscore, factor = .6)
if(jitter == 1) d1posscore = jitter(d1posscore, factor = .6)
if(jitter == 1) d1lingscore = jitter(d1lingscore, factor = .6)

if(jitter == 1) d2nontscore = jitter(d2nontscore, factor = .4)
if(jitter == 1) d2kmscore = jitter(d2kmscore, factor = .4)
if(jitter == 1) d2altscore = jitter(d2altscore, factor = .4)
if(jitter == 1) d2posscore = jitter(d2posscore, factor = .4)
if(jitter == 1) d2lingscore = jitter(d2lingscore, factor = .4)

#trim if needed

if(trimmer == 1) d1nontscore = rm.outlier(d1nontscore)
if(trimmer == 1) d1kmscore = rm.outlier(d1kmscore)
if(trimmer == 1) d1altscore = rm.outlier(d1altscore)
if(trimmer == 1) d1posscore = rm.outlier(d1posscore)
if(trimmer == 1) d1lingscore = rm.outlier(d1lingscore)

if(trimmer == 1) d2nontscore = rm.outlier(d2nontscore)
if(trimmer == 1) d2kmscore = rm.outlier(d2kmscore)
if(trimmer == 1) d2altscore = rm.outlier(d2altscore)
if(trimmer == 1) d2posscore = rm.outlier(d2posscore)
if(trimmer == 1) d2lingscore = rm.outlier(d2lingscore)

#start the graph

op <- par(mfrow = c(2, 2), mar=c(0,0,1.5,0), oma=c(5,4,3,.5))

#nont_score

x1= cbind(d1lingscore, d1nontscore)
plot(NULL,xlim =c(1,10), ylim=c(1,10),
tck = FALSE, labels = FALSE,
mgp = c(4,.5,0), main = "nont vs lingscore", family = "serif")
```

```
axis(2, labels = TRUE, family = "serif")
points(x1, col = "red", pch = 19, cex = 0.6, lwd = .2)

x2 = cbind(d2lingscore, d2nontscore)
if(nrow(x2) == 1) points(x2, xlim =c(1,10), ylim=c(1,10),
pch = 19, col = "black", bg = "red", cex = 1.5) else points(x2, xlim =c(1,10),
ylim=c(1,10), col = "black", pch = 19, cex = 0.6, lwd = .2)

hpts = chull(x1)
hpts = c(hpts, hpts[1])
polygon(x1[hpts, ], border = "red")

hpts = chull(x2)
hpts = c(hpts, hpts[1])
polygon(x2[hpts, ], border = "black")

if(avger ==1) points(mean(d1lingscore), mean(d1nontscore),
cex = 1.5, pch = 22, bg = "red", col = "white")
if(avger ==1) points(mean(d2lingscore), mean(d2nontscore),
 cex = 1.5, pch = 22, bg = "black", col = "white")

#km_score

x1 = cbind(d1lingscore, d1kmscore)
plot(NULL,xlim =c(1,10), ylim=c(1,10), xlab ="lingscore",
tck = FALSE, labels = FALSE,
lwd = .2,   mgp = c(4,.5,0), main = "km vs lingscore", family = "serif")
points(x1, col = "red", pch = 19, cex = 0.6)

x2 = cbind(d2lingscore, d2kmscore)
if(nrow(x2) == 1) points(x2, xlim =c(1,10), ylim=c(1,10),
pch = 19, col = "black", bg = "red", cex = 1.5) else points(x2, xlim =c(1,10),
ylim=c(1,10), col = "black", pch = 19, cex = 0.6, lwd = .2)

hpts = chull(x1)
hpts = c(hpts, hpts[1])
polygon(x1[hpts, ], border =
 "red")

hpts = chull(x2)
hpts = c(hpts, hpts[1])
polygon(x2[hpts, ],  border = "black")

if(avger ==1) points(mean(d1lingscore), mean(d1kmscore),
cex = 1.5, pch = 22, bg = "red", col = "white")
if(avger ==1) points(mean(d2lingscore), mean(d2kmscore),
cex = 1.5, pch = 22, bg = "black", col = "white")

#alt_score

x1 = cbind(d1lingscore, d1altscore)
plot(NULL,xlim =c(1,10), ylim=c(1,10), xlab ="lingscore",
```

```
tck = FALSE, labels = FALSE,
  mgp = c(4,.5,0),main = "alt vs lingscore", family = "serif")
points(x1, col = "red", pch = 19, cex = 0.6, lwd = .2)
axis(1, labels = TRUE)
axis(2, labels = TRUE)

x2 = cbind(d2lingscore, d2altscore)
if(nrow(x2) == 1) points(x2, xlim =c(1,10), ylim=c(1,10), pch = 19,
col = "black", bg = "red", cex = 1.5) else points(x2, xlim =c(1,10),
ylim=c(1,10), col = "black", pch = 19, cex = 0.6)


hpts = chull(x1)
hpts = c(hpts, hpts[1])
polygon(x1[hpts, ], border = "red")

hpts = chull(x2)
hpts = c(hpts, hpts[1])
polygon(x2[hpts, ], border = "black")

if(avger ==1) points(mean(d1lingscore), mean(d1altscore),
cex = 1.5, pch = 22, bg = "red", col = "white")
if(avger ==1) points(mean(d2lingscore), mean(d2altscore),
 cex = 1.5, pch = 22, bg = "black", col = "white")

#pos_score

x1 =cbind(d1lingscore, d1posscore)
plot(NULL,xlim =c(1,10), ylim=c(1,10), xlab ="lingscore", tck = FALSE, labels = FALSE,
  mgp = c(4,.5,0), main = "pos vs lingscore", family = "serif")
points(x1, col = "red", pch = 19, cex = 0.6, lwd = .2)
axis(1, labels = TRUE)

x2 = cbind(d2lingscore, d2posscore)
if(nrow(x2) == 1) points(x2, xlim =c(1,10), ylim=c(1,10),
pch = 19, col = "black", bg = "red", cex = 1.5) else points(x2, xlim =c(1,10),
ylim=c(1,10), col = "black", pch = 19, cex = 0.6)

hpts = chull(x1)
hpts = c(hpts, hpts[1])
polygon(x1[hpts, ], border = "red")

hpts = chull(x2)
hpts = c(hpts, hpts[1])
polygon(x2[hpts, ], border = "black")

if(avger ==1) points(mean(d1lingscore), mean(d1posscore),
cex = 1.5, pch = 22, bg = "red", col = "white")
if(avger ==1) points(mean(d2lingscore), mean(d2posscore),
cex = 1.5, pch = 22, bg = "black", col = "white")

par(op)
```

```
title(xlab = "lingscore                                    lingscore", family = "serif")
title(ylab = "score                                            score", family = "serif")

all <- par(xpd = TRUE)

if(cond2 == "id is null") legend(x = 5.5, y = 11.4,
legend = c(text1),  col = c("black"), pch = 19, horiz = TRUE,
 bty = "n")  else legend(x = 2.5, y = 11.4, legend = c(text1,text2),
col = c("red","black"), pch = 19, horiz = TRUE, bty = "n")

par(all)

filename = "scotsvsgaelic"
dev.print(png, filename=paste("/home/user/academic/latexpics/", ".png", filename, sep =""),
width=500, height=500)
```