1 Setting the Scene

One of J. R. R. Tolkien's major academic contributions is his O'Donnell Lecture on the historical relationships between the languages of Britain, published under the title *English and Welsh*. He proclaims:

> The north-west of Europe, in spite of its underlying differences of linguistic heritage – Goidelic, Brittonic, Gallic; its varieties of Germanic; and the powerful intrusion of spoken Latin – is as it were a single philological province, a region so interconnected in race, culture, history, and linguistic fusions that its departmental philologies cannot flourish in isolation. (Tolkien 1963: 33)

Leaving the dated reference to 'race' aside,[1] this statement accurately reflects scholars' long-standing interest in framing the linguistic past of the region as a history of interaction, overlap, and convergence.

I approach this history as a linguist with a primary interest in sound patterns. Phonological phenomena have offered specialists in these languages much food for thought, and I aim to demonstrate that they indeed cast much light on many well-studied linguistic, cultural, and historical issues. This chapter offers a very brief glance at the kinds of shared phonological features observed in the region (section 1.1), an overview of the sociohistorical situation in the region with an emphasis on language contact (section 1.2), and a summary of the book's overall argument (section 1.3).

1.1 Phonological Features of the Northern European Area

The literature pinpoints several non-trivial phonological features whose concentration in the languages of northern Europe requires an account. We have already mentioned the 'polytonicity' of Jakobson (1929, 1931b), which more recent literature describes under the rubric of 'pitch accents' or 'tonal accents'. Some version of this system is found all across northern Europe, from Britain and Ireland to West Germanic vernaculars to the Baltic and Finnic languages in the east. The exact nature of the contrast and the role of pitch vary quite significantly across the languages. In the extreme case, the phenomenon is completely

[1] See Fimi (2009) for in-depth discussion of race in Tolkien's *œuvre*.

divorced from pitch or tone, and is analysed as a different kind of pattern, such as laryngealization/'stød' or ternary quantity.

Preaspiration, usually defined as a period of voiceless frication between the modal voicing of a vowel and the main articulation of a following consonant, is another phonological feature that is especially frequently discussed in a northern European areal context. The North Germanic language Icelandic shows what is probably the best-known (and most-studied) example. It is also attested in the other Insular Nordic language, Faroese. Among the continental North Germanic languages, it had long been assumed that preaspiration was a feature of only a small number of dialects. More recent scholarship has reconsidered these assumptions and provided more reliable data, forcing a re-evaluation of the role of preaspiration in the North Germanic languages.

Outwith North Germanic, the domain of preaspiration includes the Sámi languages, a branch of the Uralic family whose traditional territory covers much of the northern half of the Fennoscandian region, and dialects of Scottish Gaelic, belonging to the Celtic group of Indo-European. Preaspiration has often been considered to be cross-linguistically rare, and its apparent concentration in northern Europe has occasioned scholarly comment since at least Marstrander (1932b). Both Scottish Gaelic and the Sámi languages have experienced prolonged and quite intensive interactions with North Germanic, and this has led many scholars to offer contact explanations for this convergence.

Another relevant phenomenon is 'preocclusion', where a sonorant – usually a nasal but sometimes also a lateral – is realized as a homorganic cluster of a stop, normally of the 'voiced' or 'lenis' series, and the sonorant, as in **nn* > *dn*, or **ll* > *dl*. In our area, this feature is attested amply in (Western) Norse, in some Sámi varieties, in the Insular Celtic language Cornish, and in Manx, another representative of the Gaelic group within Celtic. Although 'preocclusion' by itself is not vanishingly rare cross-linguistically – partially nasal segments of this type are found in regions such as Australia, insular South-East Asia, and South America – it is not especially common in Western Eurasia, including the rest of Europe.

The list can be expanded even further: for example, the so-called 'consonant gradation' in Sámi and Finnic is often compared to other morphologically conditioned alternations, like Celtic initial mutation (Wagner 1964) and root-final alternations in Germanic conditioned by Verner's Law or Kluge's Law (e.g. Schrijver 2013). Initial stress is another feature common in the region (Salmons 1992a). Many languages demonstrate intrusive/excrescent vowels in sonorant-initial clusters such as *lg* or *rm* (Iosad & Maguire in preparation).

I give a more comprehensive overview of the literature on phonological areality in northern Europe in Chapter 2. In this book, the empirical focus in chapters 5 to 8 is on the case of the study of preaspiration, while preocclusion is considered in more detail in Chapter O1 available separately on the publisher's website. I treat the synchronic and diachronic typology of tonal

accents in northern Europe in more detail in Iosad (in preparation[b]), which can be considered a companion volume to this book.

1.2 Historical Background of the Languages and Speakers

In geographical terms, the focus of this book is on northern Europe, particularly on the North Sea region, covering the languages spoken currently and historically in mainland Scandinavia, Britain, Ireland, and other islands of the North Atlantic and the North Sea littoral. A basic orientation map is provided in Figure 1.1.

Figure 1.1 Map of key locations in north-western Europe

Northern Europe has always been characterized by long-term historical and cultural contacts. Here, as is many other places, distinct subgroupings emerged out of dialect continua only for later contact and convergence to blur the genealogical lines, and conquest and settlement left a lasting legacy of both culture and language. This pervasiveness of contact also seems to provide a ready explanation for the unusual phonological convergences between the languages sketched in section 1.1.

I will not discuss the background primarily with reference to genealogical relatedness, but will instead proceed by region. We start on the continent in mainland Fennoscandia, and move in a clockwise direction through the Baltic shores and then westwards to the North Sea littoral and across the sea to the islands of the North Atlantic.

1.2.1 The Fennoscandian Mainland

In the modern era, the two main language groupings in the Scandinavian peninsula and in the islands to the west are the North Germanic and the Sámi languages; to emphasize this dual belonging, I eschew the term 'Scandinavian' in reference to the North Germanic language(s) – the Sámi languages, after all, are just as 'Scandinavian'.

The earliest reliable evidence is Germanic, particularly runic inscriptions dating back to possibly the second century CE. Their language might represent the common North-West Germanic stage ancestral to both North and West Germanic, or a variety not too far removed from it. The evidential base for North Germanic remains scarce until the Viking Age, which begins in the eighth century. At that time the language was uniform enough for it to be referred to as a single entity called *dǫnsk tunga* (Berg 2016). By the start of the Viking Age, we can observe some dialect divisions (Barnes 1997, 2003b, H. Williams 2007), but the language remains essentially unified for a long time afterwards. It would take several more centuries for this Viking Age North Germanic to develop into the language known from the handbooks as 'Old Norse' – a construct based mostly on Old Icelandic manuscripts from the twelfth century onwards (Hagland 2013).

Alongside Germanic, the Fennoscandian region is home to two branches of Uralic: Finnic and Sámi. The Sámi languages are spoken across an area ranging from the shores of the White Sea in the east to the inland areas of Central Scandinavia in Trøndelag and Jämtland: this region is often referred to as Sápmi.[2] Recent scholarship (Aikio 2006, 2012) sees the present-day Sámi languages as an offshoot of the northern reaches of a dialect continuum that can be localized to the north of the Gulf of Finland and northwards into the Finnish/Karelian

[2] This is the Northern Sámi version of the name. While the designation *Lapland* is still occasionally used in English, the name *Lapp* and its derivatives are now generally considered derogatory.

lakeland in the early Iron Age. It began spreading northwards in the early first millennium BCE. The ancestor of the Finnic and Saamic languages remained in close contact with Indo-European languages, with the earliest contacts sometimes dated to a period prior to the emergence of the familiar Indo-European branches (Kallio 2012). A later stratum of borrowings evidences typical Germanic developments such as Grimm's Law and the development of syllabic sonorants into **un*, **ur*. On the other hand, they also participate in (and predate) the extensive reorganization of the Sámi vowel system known as the Great Vowel Shift (Sammallahti 1998, Aikio 2006, Zhivlov 2023). This stratum is followed by numerous borrowings specifically from Proto-North Germanic, or at least a stage not too far removed from it. This stage is both much more extensive in the number of loanwords and is characterized by a different set of correspondences.

The earliest contact with Germanic likely dates to a time when Sámi still formed the northern part of a dialect continuum with Finnic. Aikio (2009, 2012) argues that languages related to Sámi were spoken far to the south of the present-day Sápmi, in the Finnish and Karelian lakeland until these territories became Finnish-speaking in relatively recent times. Under this hypothesis, the Germanic influence on the Finnic-Sámi dialect complex can be connected to the eastward cultural currents observed in the archaeological record of the Bronze Age up to the middle of the first millennium BCE (Carpelan & Parpola 2017).

It is less clear whether the relevant Germanic varieties are ancestral to attested North Germanic. Present-day Germanic vernaculars of the eastern Baltic littoral do not show any traits that would place their eastward spread earlier than the Viking Age, or even the later medieval period (see Ahola, Frog & Schalin [2014] on Åland and Ivars & Huldén [2002] on Finland). It is possible that the some of the earlier borrowings into Finnic and Saamic are rather 'para-North-Germanic', related to but not identical to the ancestor of attested North Germanic (Aikio 2019).

Irrespective of what we can say about the Germanic-Sámi contacts in eastern Fennoscandia, by the Viking Age the Sámi languages are firmly established in much of the Scandinavian peninsula. Traditionally, the presence of the Sámi was seen as restricted to the northern part of the region until the early modern period and the rise of large-scale reindeer herding. Under this view, northern Scandinavia experienced a northward colonization by Germanic-speaking agriculturalists at the expense of hunter-gatherer Sámi-speaking communities. This would potentially be accompanied by an asymmetrical, perhaps antagonistic, cultural relationship in which the Sámi were very much the junior partners.

In recent years, this picture has been re-evaluated, driven by advances in archaeology and cultural history. Archaeologically, new finds have moved the limits of the Sámi presence in Viking-Age Scandinavia ever further southwards, towards Uppland in Sweden and almost to the shores of the Oslofjord further

west (Zachrisson 1997, Hansen & Olsen 2014): the 'Nordic' and 'Sámi' geographical domains in the Iron Age appear to overlap to a much greater degree than previously thought. Culturally, it has been increasingly argued that the traffic between Norse and Sámi societies was essentially two-way (Mundal 2004, Kusmenko 2008, Wang 2023).

A shift away from models of conquest and domination is particularly pronounced in archaeology. As described in Hansen & Olsen (2014), the ethnic dynamics of Fennoscandia are now increasingly seen as involving shifting identities related to the spread of new economic and cultural models. Under these assumptions, both Nordic and Sámi language and culture spreads northwards in the Iron Age by way of being adopted by the local population, which itself remained essentially stationary. The modern Sámi languages are the product of a northward diffusion from the Sámi homeland in south-eastern Fennoscandia into territories previously occupied by speakers of an unknown language, referred to as Palaeo-Laplandic (see Günther et al. [2018], Lamnidis et al. [2018] on the genetic prehistory of the region). This imposition of the Sámi culture in the north likely had far-reaching effects on the Sámi languages, at the very least in the introduction of extensive layers of non-Uralic vocabulary (Aikio 2012, 2019). Further to the south the Sámi also came into direct contact with speakers of North Germanic. The data shows that by the time of contact with North Germanic dialect differentiation in Sámi had begun, but only relatively recently; the two groups have been in close contact ever since.

1.2.2 The Baltic Sea Area

Even in the modern era, the Gulf of Finland and surrounding areas present a three-way interface between Finnic, Baltic, and North Germanic languages. North Germanic varieties, traditionally classified as dialects of Swedish, survive in Åland and along the western and southern coasts of modern Finland; until very recently, they also had a presence along the south coast of the Gulf of Finland, in mainland Estonia and particularly in the western Estonian archipelago. To the south of the Gulf of Finland, we find an interface between the Finnic and Baltic languages. As noted earlier, contact between (at least parts of) Finnic and Baltic is ancient. From a Baltic perspective, this interface area would have been quite peripheral: as discussed, for instance, by Dini (2014), a presence for Indo-European languages of the 'Baltic' type can be postulated, primarily on the basis of river names, over a wide area stretching from the Vistula, if not the Oder, in the west to the Oka and the middle Volga in the east, and from the Gulf of Riga to the north to the southern fringe of the eastern European forest zone in the south. The present-day Finnic-Baltic border was established as a result of northward expansion of the Baltic varieties (e.g. Balode & Holvoet 2001). In historical times, the border lies between Estonian to the

north-east and Livonian either side of the Gulf of Riga and Latvian, including its dialects such as Tamian (with a Livonian substrate) and High Latvian (Latgalian) in the east, near the border with Slavic. There is widespread agreement that these languages were in intense contact, showing non-trivial similarities on many structural levels (Stolz 1991, Koptjevskaja-Tamm & Wälchli 2001), not least phonologically (e.g. Seržant 2010, Daugavet 2013).

Further south, a convergence zone involving Baltic, East Slavic (Belarusian) and West Slavic (Polish) vernaculars has been identified with reference to multiple levels of linguistic structure (e.g. Wiemer 2004), with Ivanov (2005) linking it specifically to the late medieval period and the influence of the Grand Duchy of Lithuania.

On the south shore of the Baltic we find a Slavic/Germanic contact zone, albeit one that is difficult to recover following the population exchanges and redrawing of borders in the postwar period. Further to the west lies the nexus between West and North Germanic. The question of the exact relationship between the different branches of the Germanic family is traditionally vexed, but recent research emphasizes the essential unity of Northwest Germanic (Ringe & Taylor 2014: §2.3). Even though we can identify changes that 'cleanly' separated the West Germanic clade from North Germanic, multiple later innovations spread across these boundaries (Hartmann 2023). This suggests that the North and West Germanic varieties remained sufficiently close geographically and linguistically to form a dialect continuum or *linkage* (François 2015). Much the same picture of internal diversity but essential unity can be painted for West Germanic itself (e.g. Ringe 2012, Salmons 2017).

No discussion of language contact on the shores of the Baltic Sea is complete without a mention of the role of Low German. It remained for a long time the prestige language throughout much of the region and would exert a powerful influence on the local varieties. This was particularly pronounced in the lexicon, and to a certain extent in the morphosyntax. Bilingualism with German was not particularly widespread and remained restricted to the upper classes, which would nevertheless facilitate an impact on the elaboration of standard languages. In the other direction, local varieties of German underwent some influence from the Baltic and Slavic languages (e.g. Lehiste 1965).

The *primus motor* for the spread of Low German influence around the Baltic and North Seas was the Hanseatic League, which reached the height of its power in the fourteenth and fifteenth centuries. Particular attention has been drawn to the Low German influence on the North Germanic languages. It is pervasive in the lexicon, but the presence of Low German speakers also made a structural impact (e.g. Jahr 1995, Nesse 2002).

Low German and North Germanic remain in close contact today, particularly in Schleswig/Slesvig. The vernaculars on both sides of the Danish-German border experienced rather spectacular convergence across multiple levels of linguistic structure (Höder 2011, 2016).

The area at the border between North and West Germanic, stretching from Jutland to the mouth of the Rhine, is, of course, also the ultimate homeland of the language that would become English. Before we turn to the tangled questions of its origins, we will briefly consider what the linguistic situation was in Britain and Ireland and their adjacent islands before the *aduentus Saxōnum*.

1.2.3 The British and Irish Isles: The Celtic Period

The eastern periphery of the Atlantic Ocean contains several islands and island groups. Although Britain and Ireland, the largest of these, are separated by the Irish Sea and the North Channel, maritime traffic between them is relatively straightforward: the Mull of Kintyre in Scotland and Torr Head in County Antrim are but 13 miles apart, whilst in the Irish Sea itself the large islands of Anglesey and Man further shorten the distances. It is not surprising, therefore, that the water did not have to be a sharp linguistic barrier.

We do not know the full extent of linguistic diversity in Britain and Ireland before the *aduentus Romanōrum*. Crumbs of information can be gathered from works such as Ptolemy's *Geographia* (Rivet & Smith 1979, G. Rhys 2015). The received view is that most of the population of the islands at the time of the Roman Conquest spoke some Celtic language. Specialists still disagree on the relative importance of different isoglosses within the family (for recent discussions, see Schrijver 2015, Jørgensen 2022), but it is clear that the Insular languages – the two relatively well-defined groups Brythonic and Goidelic – share non-trivial similarities setting them apart against their continental relatives.

It is generally assumed that Britain and Ireland had been Celticized some time before their appearance on the Roman horizon: for recent discussions, see Sims-Williams (2020), Patterson et al. (2021), Mallory (2023). It is less clear whether they coexisted with pre-Indo-European populations in historical times. Certainly, many place names present difficult etymological problems. The names of rivers such as Humber, Farrar or Ness, or islands such as Lewis, Uist, Achill, Iona, Wight remain, to various degrees, contentious. Related to this issue is the 'problem of the Picts' in present-day Scotland. It had long been supposed that the Picts were the remains of a pre-Indo-European population, which has been reinforced by long-standing ideas about their 'difference' with regard to the Celtic culture and population of the rest of the islands: their supposed matrilineal inheritance, the symbol stones that still defy convincing interpretations, and the numerous Ogham inscriptions that have resisted attempts at reading them in an Indo-European language. These facts had to be accommodated with the considerable evidence for a Brythonic language being spoken north of the Forth.

Jackson (1955) posited a Brythonic 'Pictish' language coexisting with a pre-Indo-European one. More recently, this view has been subject to significant

scrutiny (Woolf 2017, Noble & Evans 2022), especially as the Picts' exoticism somewhat dissipated in the light of sustained enquiry. Forsyth (1997) argued that the inscriptional evidence for non-Indo-European Pictish cannot bear the weight placed on it (but see Rodway 2020). The onomastic and lexical evidence (James 2011, 2013, G. Rhys 2020) suggests a Brythonic dialect continuum straddling the Roman walls of Hadrian and Antoninus rather than a sharp break. Thus, it is plausible, although not fully assured, that Britain before the Romans was solidly Brittonic Celtic in language (almost) throughout.

As for Ireland, in historical times it is the province of the Goidelic, or Gaelic, branch of the Celtic family.[3] The question of when it came to be so is fiercely contested. The proposals enumerated by Koch (1991) run the gamut from 4,500 BCE (assuming the arrival of Celtic with Neolithic farming) to as late as the mid first millennium CE. It is even less clear whether and how long the pre-Goidelic language of Ireland survived. A pre-Indo-European or at least pre-Goidelic population can be surmised on the basis of problematic place names such as Rathlin and Achill and apparently non-Indo-European items in the Irish lexicon, not all of them shared with Brythonic (cf. OIr *sinnach* 'fox', *lacha* 'duck', *ness* 'weasel' and others).

Schrijver (2009) has argued that Goidelic Celtic was only brought to Ireland thanks to the upheavals in sub-Roman Britain, as Celtic speakers from Britain's Highland Zone fled the disruption and took over the social structures of Ireland. This would explain both the linguistic closeness of Gaelic and Brythonic and the persistence of pre-Goidelic in Ireland almost until historical times, but the question remains far from settled (Stifter & White 2023).

The problems raised by this relative lack of linguistic distance between Goidelic and Brythonic also rear their head further north. By the time written history begins in present-day Scotland, much of the western seaboard is firmly within the Gaelic cultural and political sphere – even perhaps at its centre (Ó Muircheartaigh 2014). Traditionally, the presence of Gaelic in the west is explained by the expansion of the kingdom of Dál Riata from its historical base in today's County Antrim into the Rhinns of Galloway and up the west coast into Argyll. However, this relies on outdated interpretations of both historical and place-name evidence (Woolf 2007: 322–340, Clancy 2011). A sharp linguistic boundary between Gaels and Brythonic-speaking Picts in Highland Scotland is now clearly more problematized.

Schrijver (e.g. 2007) has also proposed that in Lowland Britain Latin was not only a living language in the Roman period but in fact *the* main language of

[3] I will generally use 'Gaelic languages' to refer to the branch consisting of Irish, Scottish Gaelic, and Manx, to emphasize both its essential unity across the islands and the continuity from the early stage to the present day. Given the weight of tradition, I will continue to use the terms 'Old Irish' and 'Early Irish' to refer to the Gaelic language as recorded from the mid first millennium CE, even though 'Old Gaelic' would have been preferable to emphasize the fact that its domain covered both Ireland and much of northern Britain (Ó Muircheartaigh 2014).

the population. For him, the surviving Brythonic languages are the descendants of a 'Highland' British Celtic that was only superficially Latinized during the Roman period, but then underwent rapid upheaval once the incoming Anglo-Saxons caused an influx of Latin speakers from the Lowlands who shifted to Highland British and thus effected phonological and grammatical influence on the language. This issue remains live, but in any case language contact looms large in the linguistic history of Britain and Ireland before Germanic varieties enter the scene.

1.2.4 Britain and Ireland: The Germanic Era

The linguistic landscape of Britain changed dramatically between the years 200 and 700 CE. With Latin all but gone and the Celtic domain shrunk towards the western seaboard, the Germanic languages made their presence felt from Kent and Devon in the south to the Firth of Forth and Cumberland in the north. Varieties of West Germanic, which we now call Old English, became dominant in what is now England and parts of Scotland. The Celtic languages remained in Wales and in the south-western peninsula. In the north, Brythonic remained for a while alive in Cumbria and Strathclyde.

The question of how long Celtic survived in today's England remains unresolved. It is uncontroversial that some Germanic speakers were present in Britain already in fifth century, and likely earlier, but this need not imply continuity from this fact to the later cultural and linguistic changes. What happened to the Celtic- or Latin-speaking population of Lowland Britain is less clear (e.g. Higham 2004, Halsall 2007, Parsons 2023). The traditional picture of conquest, expulsion, and genocide, supported among other arguments by the relative lack of pre-English place names especially in the east of Britain and the near-total absence of Celtic loanwords in general English, had come under increasing scrutiny, especially from frameworks that problematized 'ethnic' models of identity (e.g. Harland 2021), while at the same time advances in archaeogenetics have reaffirmed that a degree of population movement is nevertheless something to be reckoned with (e.g. Schiffels et al. 2016, Gretzinger et al. 2022).

The debate around whether English experienced contact influence from the pre-Germanic population remains open (for an up-to-date overview, see Walkden, Klemola & Rainsford 2023). It is clear, contrary to some recent claims, especially in the popular literature (see the discussion in Sims-Williams 2022), that the Germanic era in the linguistic history of Britain and Ireland does not begin until the fifth century at the earliest. However, no sooner had the linguistic shock waves of the *aduentus Saxōnum* more or less settled down that a new intrusion of Germanic speech occurred across the archipelago. The advent of the Viking Age initiated a much longer-lasting North Germanic linguistic presence over an area stretching from the east of England to Ireland, the Outer Hebrides, and Shetland.