

Mind Your Language: Explaining the Retreat of the Irish Language Frontier*

Alan Fernihough[†] Christopher L. Colvin[‡]
Eoin McLaughlin[§] Philip R. Neary[¶]

May 2026

Abstract

Why do individuals choose one language over another? We study the shift from Irish to English in nineteenth-century Ireland, combining a model of household language choice with individual-level census data. The model shows how local interaction patterns and intergenerational transmission can produce a gradual, spatially uneven retreat of a language frontier. In the data, Irish-language outcomes depend strongly on both local and wider economic geography and on the education of earlier generations. Higher literacy among elders—shaped by a decentralised network of voluntary primary schooling—is associated with lower Irish use among later cohorts, consistent with rising returns to English.

Keywords: language loss, bilingualism, education policy, census data, Ireland.

JEL Classification: I24, I28, N33, N93, R12, Z13.

Current version available at: <https://go.qub.ac.uk/irishlanguage>

*We thank audiences at the ASREC Europe Conference (London, October 2022), Maynooth University (March 2023), the EHS Annual Conference (Warwick, April 2023), Queen’s University Belfast (May 2023), the Education and Development in History Workshop (Valencia, July 2023), Martin Luther University Halle-Wittenberg (July 2023), the World Congress of Cliometrics (Dublin, July 2023), the EHES Conference (Vienna, September 2023), the University of Manchester (October 2023), Heriot-Watt University (November 2023), the University of Galway (January 2024), the University of Antwerp (May 2024), the Royal Irish Academy (January 2025), University College Cork (October 2025) and the University of Bergamo (January 2026) for comments. We thank Luís Guimarães, Máire Ní Chiosáin, Niall Ó Ciosáin, Cormac Ó Gráda, Kevin O’Rourke, Rowena Pecchenino, Roland Rathelot, Kyle Richmond and John Turner for comments on earlier drafts. We thank Mark McManus for excellent research assistance. Any and all errors remain ours alone. Funding was provided by the Royal Irish Academy and the North South Research Programme of the Government of Ireland’s Shared Island Fund.

[†]Queen’s University Belfast, CEPH and CEPR. Email: a.fernihough@qub.ac.uk.

[‡]Queen’s University Belfast and CEPH. Email: chris.colvin@qub.ac.uk.

[§]Heriot-Watt University and CEPH. Email: e.mclaughlin@hw.ac.uk.

[¶]Royal Holloway, University of London. Email: philip.neary@rhul.ac.uk.

I'm ashamed I don't speak the language myself. I'm told it's a grand language by them that knows.

Old Mother Grogan says to Haines (James Joyce, *Ulysses*, 1904).

1 Introduction

What drives shifts in language frontiers? Max Weber, writing at a time when new ethnolinguistic nation states were being founded all across central Europe, argued that language is artificial—a political tool used by elites to dominate their society (Weber, 1921). For Weber, language communities are defined in opposition to outsiders, and changes in language use are the outcome of policies of linguistic standardisation and national education (Tada, 2018); language frontiers are determined exogenously through government intervention. Pierre Bourdieu, meanwhile, conceptualised a linguistic market, in which languages compete and language shifts occur as individuals and communities respond to incentives linked to social and economic power (Bourdieu, 1977). For Bourdieu—who grew up speaking Béarnese, a Gascon dialect of south-west France—language frontiers are endogenous to social, cultural and economic forces rather than products of deliberate state design (Grenfell, 2011).

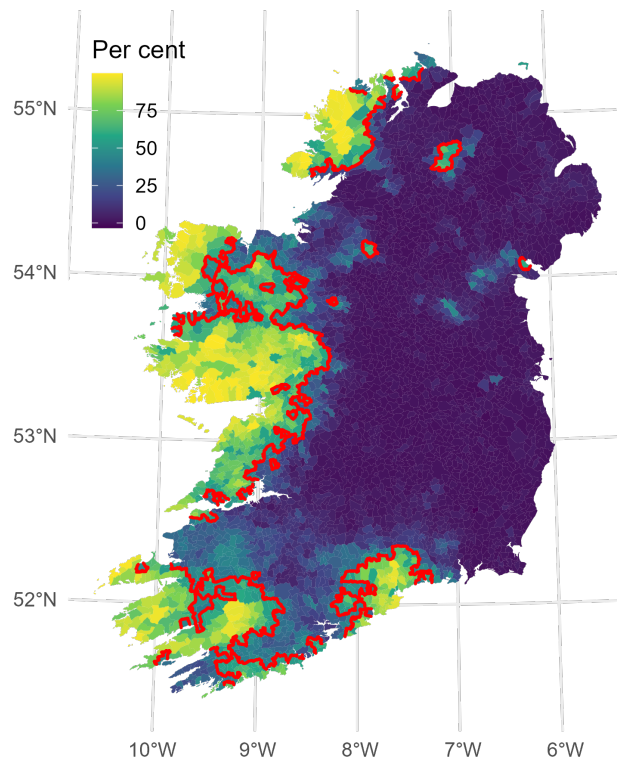
Within economics, this discussion has been approached in two main ways. The first is empirical and treats language frontiers as exogenous. Such studies evaluate language-shifting policies, typically enacted by sub-national or post-colonial governments to promote indigenous or minority languages through education systems and labour market institutions.¹ The second approach is more closely aligned with Bourdieu in that it explores endogenous reasons for linguistic diversity and language acquisition. These contributions are usually more theoretical in nature, relying on mathematical modelling to generate stylised facts about changes in language use across time and place.²

We add to this literature an empirical investigation of a marked change in language use in Ireland that we argue was not the consequence of a particular government policy intervention, but was rather the confluence of various socioeconomic forces. Figure 1 maps the Irish–English language frontier in 1901, showing both a sharp spatial boundary and the extent to which Irish had become a minority language. Drawing on a new conceptual

¹Ginsburgh and Weber (2020) argue that most studies find negative economic consequences of language policies. For example, Angrist and Lavy (1997) documents adverse effects of a 1983 language-of-instruction reform in Morocco on educational attainment, while Chakraborty and Bakshi (2016) find negative labour-market effects of abolishing English teaching in Indian primary schools. Related work examines identity formation: Aspachs-Bracons et al. (2008) show that only compulsory instruction policies affect identity in Catalonia and the Basque Country, and Blanc and Kubo (2025) argue that nineteenth-century centralised schooling promoted national assimilation in France.

²For example, Clingensmith (2015) uses network theory to explain the emergence of global languages, and John and Özgür (2021) adapt endogenous growth models to capture long-run trends in linguistic diversity. See John (2016, 2026) for a discussion of the various strands in this literature.

Figure 1: Irish Language Proficiency of Population Aged 10+ in 1901



Source: 1901 Census of Ireland Microdata.

framework that models household language choice as a coordination game shaped by local linguistic interactions, we track and explain the near-disappearance of Irish and the consolidation of English as the dominant language of communication across late-nineteenth-century Ireland. We show how exogenous changes—such as shifts in education provision, demographic structure, and state capacity—altered local incentives, but that their effects operated through endogenous household responses. Our contribution is therefore a model-guided empirical analysis of language frontier change arising from the interaction between external shocks and decentralised behavioural adjustment, a combination largely absent from the economics literature.

Linguistically, Irish—a Celtic language—is very distant from English, implying substantially higher acquisition costs than for closely related language pairs such as English and French, and making transitions between the two more akin to Basque–Castilian than to Catalan–Castilian.³ Following a century of policies aimed at reviving it, Irish today remains a minority language classified as “definitely endangered” by UNESCO (UNESCO, 2003; Moseley, 2010). It was used on a daily basis by just 115,525 people

³Ginsburgh and Weber (2020) discuss classifications of linguistic distance and their economic implications. Irish differs from English across all major dimensions: cladistic, lexicostatistic, phonetic, grammatical, and syntactic. There is also substantial dialectal variation within Irish itself (Kessler, 1995). See McMahon and McMahon (2005) for quantitative evidence, and Fabb (2016) and Chen and Fabb (2026) for an introduction to the relevant linguistic theory for economists.

across both Irish jurisdictions at the time of the last censuses in 2022, or 1.6 per cent of the population—all of whom also spoke English.⁴ Despite being a mandatory core subject across all Republic of Ireland government-funded primary and secondary schools, Irish is not even the most commonly spoken second language in Ireland today.⁵ However, the language has not fallen into complete disuse; rather, it continues to be sustained within small self-reinforcing speech communities, consistent with the idea that language use change is gradual and that stable outcomes can exist in which a minority language persists alongside a dominant one.⁶

Irish speaking was not always so rare. On the eve of the Great Irish Famine in the mid-nineteenth century, over two million people may have used the language (Hindley, 1990).⁷ The first official estimates of the number of Irish speakers is post-Famine, from 1851, and suggest a stark decline: 1.5 million Irish speakers, or 23.3 per cent of the population (British Parliamentary Papers, 1855, p. xvii). This decline continued apace; by 1901 the share of Irish speakers had halved (FitzGerald, 1984, 2003). Slomanson (2012) frames this development as ‘probably the most dramatic case of a major western European language, being spatially and demographically reduced to an utterly peripheralised minority language in the space of one or two generations’.

We analyse the Irish language’s decline in use across the space of these one or two generations with the help of individual-level data from the 1901 census. This high-quality source allows us to track the spatiotemporal change in Irish usage across birth cohorts from the 1820s to the 1890s. This approach is superior to using the aggregated data sources employed in previous quantitative analyses of Irish language decline by FitzGerald (1984, 2003) because it enables us to: (1) control for demographic and other individual-level characteristics; and (2) explore different levels of spatial aggregation to ascertain which is the most relevant unit of analysis for the study of language transmission.

Our empirical strategy evaluates a set of non-exclusive hypotheses about the decline of Irish within a unified regression framework. These hypotheses are motivated by a theoretical model in which a language frontier emerges endogenously from local interactions

⁴While Irish has been a compulsory subject in the Republic of Ireland since 1934, only 10% of speakers report high proficiency in 2022, and just 71,968 use it daily (Central Statistics Office, 2023a). In Northern Ireland, 12% of the population reports some ability, but only 43,557 report that they use Irish daily (Northern Ireland Statistics and Research Agency, 2022, 2023). Previous research finds that modern Irish speakers earn a wage premium (Borooah et al., 2009), though its magnitude and underlying mechanisms remain contested (Watson and Nic Ghiolla Phádraig, 2011).

⁵There are more people who use Polish than Irish at home: 123,968 use Polish daily in the Republic of Ireland (Central Statistics Office, 2023b).

⁶Ireland’s modern-day Gaeltacht districts—officially designated Irish-speaking areas—form a small subset of the Irish-speaking regions shown in Figure 1. Even within these areas, however, Irish has continued to decline, with reduced intergenerational transmission and growing dominance of English in daily use (Ó Giollagáin and Charlton, 2015).

⁷This is a conservative estimate. Williams and Ford (1992) estimate that 80% spoke Irish at the turn of the nineteenth century. Other estimates range from 45% to 80% (Doyle, 2015). Wolf (2014) reports a pre-Famine Irish-speaking headcount of 4.1 million.

between households embedded in a social network. In the model, the retreat of the Irish-speaking frontier is gradual and spatially uneven, reflecting changes in neighbourhood composition and intergenerational transmission rather than a single, economy-wide shock. This perspective directs attention to endogenous forces—such as evolving social connectivity, household educational choices, second-order economic geography and demographic turnover—that alter the local costs and benefits of second-language acquisition. Our model also allows for the possibility that exogenous influences, including institutional change and broader policy environments, shape these local incentives.

This approach responds to a long-standing gap in the literature. As Slomanson (2012), following de Fréine (1960), argues, there has been a ‘great silence’ both about explaining Ireland’s language shift itself and about investigating its wider social consequences. In its place, a simplified popular narrative has emerged in which the decline of Irish is attributed to the coercive imposition of English as a colonial language.⁸ A less extreme rendition of this view would emphasise that government policy deprived Irish of a space in the public sphere—within politics, the courts, the army, and schools—thereby reducing its prestige and discouraging intergenerational transmission.

Alongside such accounts are alternative explanations that are more endogenous in nature, focusing on what economists would recognise as shifts in the demand for and supply of language skills. This literature is reviewed in Appendix A. Taken together, we conclude that existing scholarship offers limited guidance on the *relative* importance of these competing mechanisms, partly because explanations are typically framed in monocausal terms.⁹ Additionally, Corrigan (2003) cautions that many of these explanations are shaped by the language ideologies of their authors. Our contribution is to bring these explanations together within a single quantitative empirical framework, allowing us to assess their relative importance and to distinguish between endogenous adjustment processes and exogenous shocks in shaping the nineteenth-century decline of Irish.

It is important to reiterate that we are not evaluating the impact of a specific government policy designed to influence language use. Rather, we are exploring the interactions between different cultural, geographic and socioeconomic factors which can influence language acquisition, none of which come anywhere close to resembling an official language policy intervention. We are therefore much closer to scholars who use economic theory to explore the dynamics of language frontiers and who focus on endogenous reasons for spatiotemporal change, like Clingingsmith (2015) and John and Özgür (2021). And so we are closer to Bourdieu than to Weber in viewing language communities as significantly endogenous, responding to economic incentives and constraints.

We focus our analysis on areas of the island that contained *any* Irish speakers before

⁸This narrative has also influenced economists; e.g., Hynes (2014) characterises the rise of English in Ireland as an instance of ‘linguistic imperialism’.

⁹Hindley (1990) is a rare exception, but his analysis is qualitative and largely confined to the twentieth century, after the language shift had already taken place.

the Great Irish Famine, a mid-century cataclysm that led to approximately one million excess deaths and a further one million migrants (Blum et al., 2026). The areas we target are in the provinces of Connacht, Munster and the western fringes of Ulster, the location of a bilingual society at the island’s language frontier. These are precisely the areas where Irish was “lost” during the subsequent language shift; parts of the island where no Irish was spoken pre-Famine, even as a second language, had no Irish to lose.

We ascertain that the intergenerational transmission of language occurs across multiple geographical contexts. Probit regressions on our target population suggest, unsurprisingly, that Irish-language ability among household elders predicts language use among the next generation. More surprisingly perhaps, we show that this intergenerational transmission mechanism is just as strong if we expand our definition of the elder generation geographically; Irish prevalence in neighbourhood (townland) elders has as much influence as Irish speaking within the household. This relationship disappears when even larger geographic units are used—a result that helps us identify the optimal geographic unit for studying factors that affect a language’s communicative value.

Having established the importance of endogenous intergenerational transmission, we turn to explanatory factors. Two results emerge. First, schooling, proxied by literacy, matters. However, marginal effects indicate that Irish language ability depends more on historical district-level literacy than on contemporary literacy or an individual’s own ability to read and write. This is consistent with the channels identified in our intergenerational transmission model: community-level conditions from earlier generations exerts a stronger influence on language outcomes than contemporaneous individual factors. In other words, the educational environment in which a grandparent grows up does not affect their own Irish-speaking ability; instead, it shapes that of their grandchildren by influencing the language they choose to transmit.

Second, language is shaped by both first- and second-nature economic geography. Consistent with our conceptual framework, language retreats systematically along a geographical frontier: individuals closer to this frontier are less likely to maintain Irish than those surrounded by Irish-speaking neighbours. Second-nature geography, captured by market access, also plays a substantial role. Language loss is greater where pre-rail transport made markets more accessible, and the arrival of railways accelerated this process.

Our study continues by setting out a conceptual framework in Section 2, building a game-theoretic model with households connected in a social network to examine how Ireland’s language shift could have occurred. Section 3 describes our dataset, while Section 4 outlines our empirical strategy and key findings. This section ends by putting all our explanatory factors in one parsimonious model, to discern their relative importance. Section 5 relates our empirical model back to our theoretical framework, and discusses conclusions, limitations and policy implications. Appendices review the historical litera-

ture, develop our model, describe our data sources and address selection issues.

2 Conceptual Framework

Our goal in this section is to develop a conceptual framework through which to understand the process and pattern of language shifts in the Irish context. The decision-making unit in our model is always the household, some of which are native Irish speaking and the remaining native English speaking. Each household has a strategic choice to make: should we invest in learning the language of the other group? In our model, the returns to learning the other group’s language depend on the frequency with which it will be used, which in turn depends on societal structure—on how well connected each household is to its own group and the other group.

We make two assumptions on societal structure, both based on the map in Figure 1. These are that the native English speaking households are connected in a contiguous way, and that each native English speaking household has enough other native English speaking households as neighbours that the returns to learning Irish do not compensate for the cost. From this, our model has a unique equilibrium. Moreover, when we consider dynamics—language portfolio choices over time—the remoteness of some native Irish speaking households rules out the near-complete disappearance of Irish in one fell swoop. Rather, the Irish-speaking frontier will retreat in discrete jumps, where each new jump is a consequence of returns to English increasing. When combined with the literature on the disappearance of the Irish language reviewed in Appendix A, our modelling approach generates a set of testable hypotheses about the relative impact of factors which affect the returns to English language acquisition.

Section 2.1 introduces the environment. In Section 2.2 we demonstrate how optimal behaviour is classified by a *threshold rule* that can be described informally as: it is optimal to learn the other group’s language if and only if at least a *threshold fraction* of neighbouring households from the other group are monolingual. We then show how our assumptions on societal structure guarantee a unique equilibrium where English-speaking and Irish-speaking households are separated by a linguistic frontier. Section 2.3 illustrates how the Irish-speaking frontier can only retreat in discrete steps, where each step must be due to changes in the underlying environment: the returns to speaking English must have increased in some noticeable way. Throughout, we present our model using a diagrammatic and intuitive exposition.

2.1 The Environment

Our model builds on the two-type *language game* of Neary (2011, 2012). The novelty is that each group’s less preferred strategy is replaced with a bilingual option (Goyal and

Figure 2: The three pairwise interactions, G^{EE} , G^{II} , and G^{EI} .

G^{EE}				G^{II}			
		$\mathcal{E}$	$\mathcal{B}$			$\mathcal{I}$	$\mathcal{B}$
$\mathcal{E}$		1, 1	1, 1 - c	$\mathcal{I}$		1, 1	1, 1 - c
$\mathcal{B}$		1 - c, 1	1 - c, 1 - c	$\mathcal{B}$		1 - c, 1	1 - c, 1 - c

G^{EI}			
		$\mathcal{I}$	$\mathcal{B}$
E	$\mathcal{E}$	0, 0	1, 1 - c
	$\mathcal{B}$	1 - c, 1	1 - c, 1 - c

Janssen, 1997; Galesloot and Goyal, 1997). Let $\mathcal{N}$ denote a set of n (≥ 2) *households* that are connected via a social network, whose structure is given by a graph $G = (\mathcal{N}, E)$, where the set of households, $\mathcal{N}$, is the set of vertices and E is the set of *edges* that captures what households are connected. For a given household i , we write N_i for the set of households that i is connected to and refer to these households as *neighbours* of i , and we write $d_i = |N_i|$ for the size of household i 's neighbourhood.

Doyle (2015) documents three distinct historical language communities in Ireland over time, broadly as monolingual Irish and English speakers and bilingual groups. We simplify this in that every household in our model belongs to one of two groups: Native English, Group E , or Native Irish, Group I . The strategic decision faced by each household is what *language portfolio* to adopt, where we assume that each household is already endowed with a native language: $\mathcal{E}$ for households in Group E and $\mathcal{I}$ for those in Group I . Given this, each household must decide whether or not to invest and learn the language of the other group, behaviour that we denote by $\mathcal{B}$.¹⁰

Three kinds of pairwise interaction are possible: the two within-group interactions E with E and I with I , and the across-group interaction E with I . We assume that meetings between neighbours are fruitful, yielding both a payoff that we normalise to 1, if and only if both households share a common language; otherwise both receive zero. With the cost of learning a second language given by $c \in (0, 1)$, the three pairwise interactions are described by G^{EE} , G^{II} , and G^{EI} , as in Figure 2.

We write s_i for household i 's language portfolio, and write $\mathbf{s}_{-i}$ for the language portfolio of the $n-1$ households other than i . We also write $N_i(E)$ and $N_i(I)$ to denote the set of neighbouring households of i that belong to each group, so that, for every household

¹⁰While we use $\mathcal{B}$ for the bilingual option of both groups, we note that the meaning for each group is different. For those in E , it means “learn Irish” while for those in I it means “learn English”.

i , we have that $N_i = N_i(E) \cup N_i(I)$. With this, payoffs are as follows: assuming that i is a household in E and that j is a household in I , the utility to each household from the choice of either language portfolio are given by the following:

$$U_i(\mathcal{E}, \mathbf{s}_{-i}) := \frac{1}{d_i} \left(\#\{k : k \in N_i(E)\} + \#\{k : k \in N_i(I) \text{ and } s_k = \mathcal{B}\} \right) \quad (1)$$

$$U_i(\mathcal{B}, \mathbf{s}_{-i}) := \frac{\#\{k : k \in N_i\}}{d_i} - c = 1 - c \quad (2)$$

$$U_j(\mathcal{I}, \mathbf{s}_{-j}) := \frac{1}{d_j} \left(\#\{k : k \in N_j(I)\} + \#\{k : k \in N_j(E) \text{ and } s_k = \mathcal{B}\} \right) \quad (3)$$

$$U_j(\mathcal{B}, \mathbf{s}_{-j}) := \frac{\#\{k : k \in N_j\}}{d_j} - c = 1 - c \quad (4)$$

Note that, for each household, the sum of payoffs earned in all pairwise interactions is scaled by the size of its neighbourhood. Let us now explain equations (3) and (4) that capture the situation faced by a Native Irish household.¹¹ The first term in (3), $\#\{k : k \in N_j(I)\}$, is the number of neighbouring households of j that are Native Irish. The language portfolio of these households is irrelevant because successful communication can and will occur through Irish. The second term, $\#\{k : k \in N_j(E) \text{ and } s_k = \mathcal{B}\}$, counts the number of neighbouring Native English households that are bilingual. Observe that the returns to monolingualism are strictly increasing in the number of neighbours from the other group who are bilingual.

Equation (4) confirms that bilingualism guarantees a utility of $1 - c$. This holds because an Irish household that is bilingual is guaranteed to communicate successfully with each neighbouring household, regardless of their group affiliation or their language portfolio. Note, however, that being bilingual is solely a cost when interacting with those from the same group.¹²

2.2 Threshold rule and equilibrium

2.2.1 Optimal behaviour

We begin by determining what constitutes optimal behaviour. To fix ideas, let us consider the environment from the perspective of a Native Irish-speaking household.¹³ The household already has Irish in its language portfolio, so the fraction of Native Irish speaking neighbouring households – whether they are bilingual or not – does not influence the household's decision since successful communication with those neighbours is already guaranteed. An immediate consequence of this is that a Native Irish-speaking household without a Native English-speaking neighbour will never choose to become bilingual.

¹¹Equations (1) and (2) for native English speaking households have mirror-image interpretations.

¹²This can also be seen from the pairwise interaction G^{II} in which the strategy $\mathcal{B}$ is strictly dominated.

¹³As before, a mirror image analysis holds for optimal behaviour of the Native English-speaking households.

What matters is the language portfolios of Native English neighbours. In particular, if “enough” Native English neighbours are monolingual so that the benefit from communicating with these extra neighbours justifies the cost, then bilingualism is optimal. By comparing equations (3) and (4), we can compute the fraction that is enough. The expression in (5) below makes this formal.

$$\begin{aligned}
U_i(\mathcal{B}) &\geq U_i(\mathcal{I}) \\
\iff 1 - c &\geq \frac{\#\{j : j \in N_i(I)\}}{d_i} + \frac{\#\{j : j \in N_i(E) \text{ and } s_j = \mathcal{B}\}}{d_i} \\
\iff \frac{\#\{j : j \in N_i(E) \text{ and } s_j = \mathcal{E}\}}{d_i} &\geq c
\end{aligned} \tag{5}$$

where the second inequality follows by definition from equations (3) and (4) and the third follows by rearranging and some algebra.

From equation (5), we conclude that it is optimal for a Native Irish speaking household to learn English only if the fraction of monolingual Native English households exceeds a *threshold*. And, due to the normalisation on payoffs of successful and failed communication, this threshold fraction is precisely equal to the cost of learning English given by $c \in (0, 1)$.

In summary, optimal behaviour for a household, be it Native Irish-speaking or Native English-speaking, is given by the following *threshold rule*. Following Morris (2000), for simplicity we assume that a household indifferent over language portfolios chooses the bilingual option.

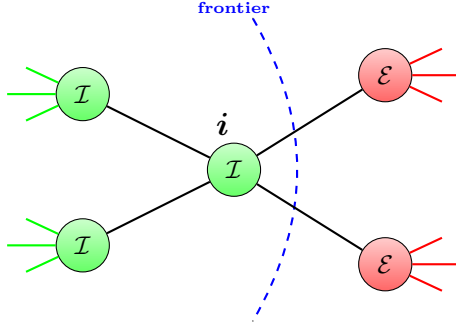
Threshold Rule. *It is a best-response for a household to become bilingual if the fraction of monolingual neighbours from the other group exceeds the cost of learning, c*

In words, the threshold rule says the following: given that communication is already feasible with same-group neighbours and bilingual households from the other group, learning the other group’s language raises utility only if the marginal benefit from newly accessible interactions exceeds the associated cost.

We now illustrate the threshold rule with an example. Consider an Irish-speaking household, i , which has four neighbouring households, two of which are Native Irish and two are monolingual Native English. The local neighbourhood of household i is depicted below in Figure 3, with the Native English speaking households represented by **red** nodes and the Native Irish speaking households by **green** ones. Edges between pairs of nodes represent connections. We have drawn three edges emanating from each of the other households to indicate that each has three other neighbouring households (not including household i). For the Native English households, these edges are colour-coded **red** to indicate that these neighbouring households are Native English.

Suppose that the cost of bilingualism is $c = 0.55$ (this value is not innocuous: it is

Figure 3: The language frontier



carefully chosen for illustrative purposes). Since two of household i 's four neighbours are monolingual English, equation (5) confirms that the benefit to learning English for household i is not great enough to justify its cost (because $\frac{1}{2} < 0.55$) and so household i remains monolingual Irish.¹⁴ This means that household i lies on the Irish-speaking side of the frontier. The part of the frontier relevant to this quadrant of the social network is depicted by the blue dashed line in the figure.

2.2.2 Equilibrium

As it stands, the model can admit many pure strategy equilibria. The final part of the set up involves placing assumptions on network architecture, i.e., on how society is structured. These assumptions are motivated by our data (recall Figure 1) and have the effect of removing the potential equilibrium multiplicity.

Informally, our two assumptions are that (i) the Native English households are connected in a *contiguous* way,¹⁵ and (ii) every Native English household has at least some fraction of Native English households as neighbours. To make this formal, we need some additional terminology.

A *path*, P , in a graph $G = (V, E)$ is a sequence of distinct vertices $v_1, v_2, \dots, v_p$ such that there is an edge from each vertex in the sequence to its successor. A graph is said to be *path connected* if there is a path between every pair of distinct vertices in V . Path connectedness formalises the notion of *contiguous* that we mentioned above. Given a vertex $v \in V$, the *deletion of v from G* results in the graph $G - v = (V \setminus \{v\}, E \setminus E_v)$, where E_v is the set of edges in E incident to v . A graph G' is called a *subgraph* of G if G' is obtained from G by deleting a subset of vertices $V' \subseteq V$.

We denote the subnetwork of Native English speakers and the subnetwork of Native Irish speakers by $G[E]$ and $G[I]$ respectively. That is, $G[E] = G - I$ and $G[I] = G - E$. Our first assumption on network structure relates to the subpopulation of native English

¹⁴Recall that the language portfolio of neighbouring Irish-speaking households is not relevant to i 's strategic decision since successful communication with these households is guaranteed.

¹⁵That is, for any pair of Native English households, it is possible to reach one from the other in the network by passing only through Native English households.

speaking households.

Assumption 1. The subnetwork of Native English speaking households, $G[E]$, is path connected (i.e., contiguous).

Our next assumption says that, in addition to being path connected, the Native English speaking households are also well connected.

Assumption 2. For every household in E , the fraction of neighbouring households that are also in E is strictly greater than $1 - c$. That is, for every $i \in E$, $\frac{|N_i(E)|}{d_i} > 1 - c$.

Note that Assumptions 1 and 2 do not place any constraints on the structure of subnetwork of the Native Irish speaking households, $G[I]$. Indeed, Figure 6 illustrates that the subnetwork of Native Irish households is sparsely connected. The reason for Assumption 2 will appear later: it ensures that those in Group E will never learn Irish.

We now state our result. We then sketch the proof, illustrating how it hinges on households following the threshold rule described before. The formal details are found in Appendix B.

Proposition 1. Given Assumptions 1 and 2, there is a unique pure strategy Nash equilibrium in which

1. all Native English speaking households remain monolingual, and
2. all Native Irish speaking households that have a fraction greater than c of their neighbours in E become bilingual while remaining households in I remain monolingual.

To understand Proposition 1, we begin by considering the subnetwork of Native English speakers: since each Native English-speaking household has fraction strictly greater than $1 - c$ of Native English-speaking households, no such household can have fraction c of Native Irish-speaking neighbours. As such, regardless of the language portfolio of the Native Irish-speaking households, it is always optimal for a Native English speaking household to remain monolingual.

Given that all Native English-speaking households remain monolingual, classifying optimal behaviour for a Native Irish-speaking household becomes a straightforward task. By appealing again to the threshold rule, we have that a Native Irish-speaking household chooses to become bilingual if and only if the fraction of Native English-speaking households (all of whom are monolingual English) exceeds the threshold c .

2.3 The Stepwise Retreat of the Irish-Speaking Frontier

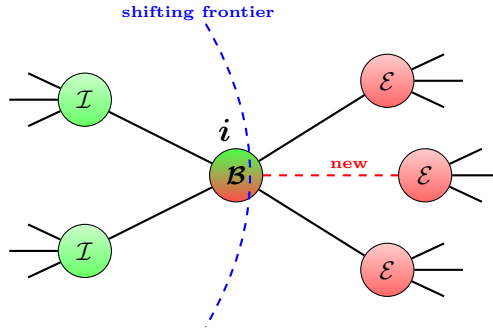
The unique equilibrium determined in Proposition 1 provides an immediately testable hypothesis.

Hypothesis 1. *A Native Irish household that is completely isolated from (i.e., not connected to any) Native English household will never learn English.*

Then how is it that the Irish language retreated so quickly across the second half of the nineteenth century? For bilingualism to become more appealing to native Irish speakers, the returns to speaking English must increase. As mentioned earlier, there are several ways in which this can occur. With Figure 3 as our starting point, we focus on one mechanism by which the language frontier can retreat in a stepwise fashion: societal structure changes in such a way that Irish speakers become better connected to the English-speaking population, coupled with demographic shifts.

With the parameters chosen, the returns to speaking English are not enough for household i to incur the cost of bilingualism. Therefore, in order for household i to adopt English, something about its environment has to change. In Figure 4 we depict a change to household i 's local environment whereby it has a new English-speaking neighbour. This new connection is depicted as a dashed red connection and labelled as “new”.

Figure 4: The language frontier shifts



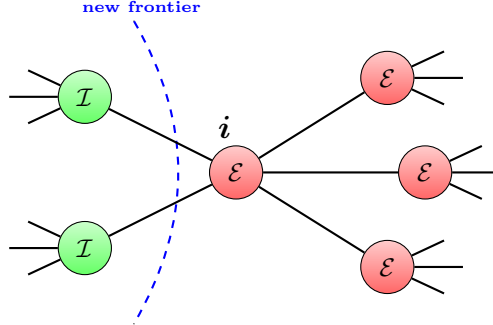
Importantly, learning English is now worthwhile for household i because the benefit from doing so has increased to $\frac{3}{5}$, which exceeds the cost of 0.55. We show this in two ways in the figure: the language portfolio of household i is changed from $\mathcal{I}$ to $\mathcal{B}$, and we change the colour of household i from green to a hybrid of green and red. Since this additional connection induces household i to become bilingual, we note that the frontier has shifted, with household i now residing on it.¹⁶

Now, let us show how the frontier can shift further due to generational attributes of households. Suppose that while household i is a multi-generational bilingual household, not all individuals in the household are bilingual. In particular, let us suppose that Irish is the language of the oldest generation (the grandparents), that English is that of the youngest one (the grandchildren), and that the middle generation are bilingual (the parents). If and when the oldest generation passes away, household i will lose its monolingual Irish speakers. And as one more generation passes, the household will become a

¹⁶We focused on the addition of a new connection. Note that the same resulting shift would also have occurred if, for example, household i had lost a connection to one of its monolingual Irish neighbours (for example through emigration or death).

monolingual English-speaking household. This scenario is represented in Figure 5 where household i is now coloured red.

Figure 5: The language frontier shifts further



Finally, note that once household i becomes an English-only household, the environment of i 's Irish neighbours has now changed since their fraction of neighbours that are monolingual English has increased. This change could be enough to induce one or both of these households to pursue bilingualism.

We conclude by contrasting our model with that of Morris (2000), who studies a coordination game played on a social network in which all individuals have identical preferences and therefore a common threshold.¹⁷ In Morris's model, full contagion can occur because once a new and better action starts to gain ground, a chain reaction of adoption can ensue. However, the takeover of a new spoken language is unlikely to occur in one fell swoop unless the network is extremely well connected. If not, it must be that the environment itself changes on occasion. This is precisely what we observe in the data for the Irish case.

Hypothesis 2. *The Irish-speaking frontier will retreat incrementally. Moreover, the environment has to change in order for Irish to retreat further.*

The historical literature (summarised in Appendix A) outlines several hypotheses that are consistent with the mechanisms underlying our model. A central argument in this literature is that the shift was driven by a changing demand for English. In the context of the model, this corresponds to a shift in the frontier (Figure 4). The national education system is also frequently cited: it was state-run and operated through the medium of English rather than Irish, reinforcing this shift. This connects to the role of the broader expansion of the English-speaking state, which lowered the effective cost of adopting English. Other factors highlighted in the literature include the ambivalence of both the Catholic Church and the nationalist movement towards the language. We test each of these channels in our empirical specifications. However, our benchmark model emphasises

¹⁷Neary and Newton (2017) extend Morris's setup to a (finite population) model where individuals may have different thresholds.

family-based transmission within a local interaction network as the mechanism shaping network structure and, in turn, the dynamics of language shift.

3 Data Construction and Sample

We use census data to retrospectively track spatiotemporal decline of the Irish language. While FitzGerald (1984, 2003) relied on the published aggregated census returns of 1851, 1861, 1881, and 1911 in his analysis, our primary data source is the full 1901 census returns, recorded at household level (held at the National Archives of Ireland). We believe this census source is superior for two reasons. Firstly, we use the full individual returns, meaning we have distinct demographic data for over four million observations. The full returns allow us to explore various levels of spatial aggregation, unlike FitzGerald (1984), who was restricted to the geographically aggregated Baronial system ($N = 336$). Secondly, while FitzGerald (2003) used more geographically disaggregated district electoral division ($N > 3,400$) spatial units from 1911, this source restricts the pre-Famine birth cohorts to all individuals over 60 years of age. Our 1901 census data are less likely to suffer from age-related sample selection complications.¹⁸

The 1901 returns required household heads to specify whether each individual household member spoke Irish, distinguishing bilingual from monolingual speakers. Notably monolingual English speakers did not have to complete this section, highlighting the English language’s dominance. This prevalence of English was reflected in the returns, as 20 per cent of the population 20 years of age and older indicated some proficiency in Irish, but only 1 per cent were returned as “Irish Only” speakers. Given the extremely low number of monolingual Irish speakers, we focus our analysis on both bilingual and monolingual Irish.

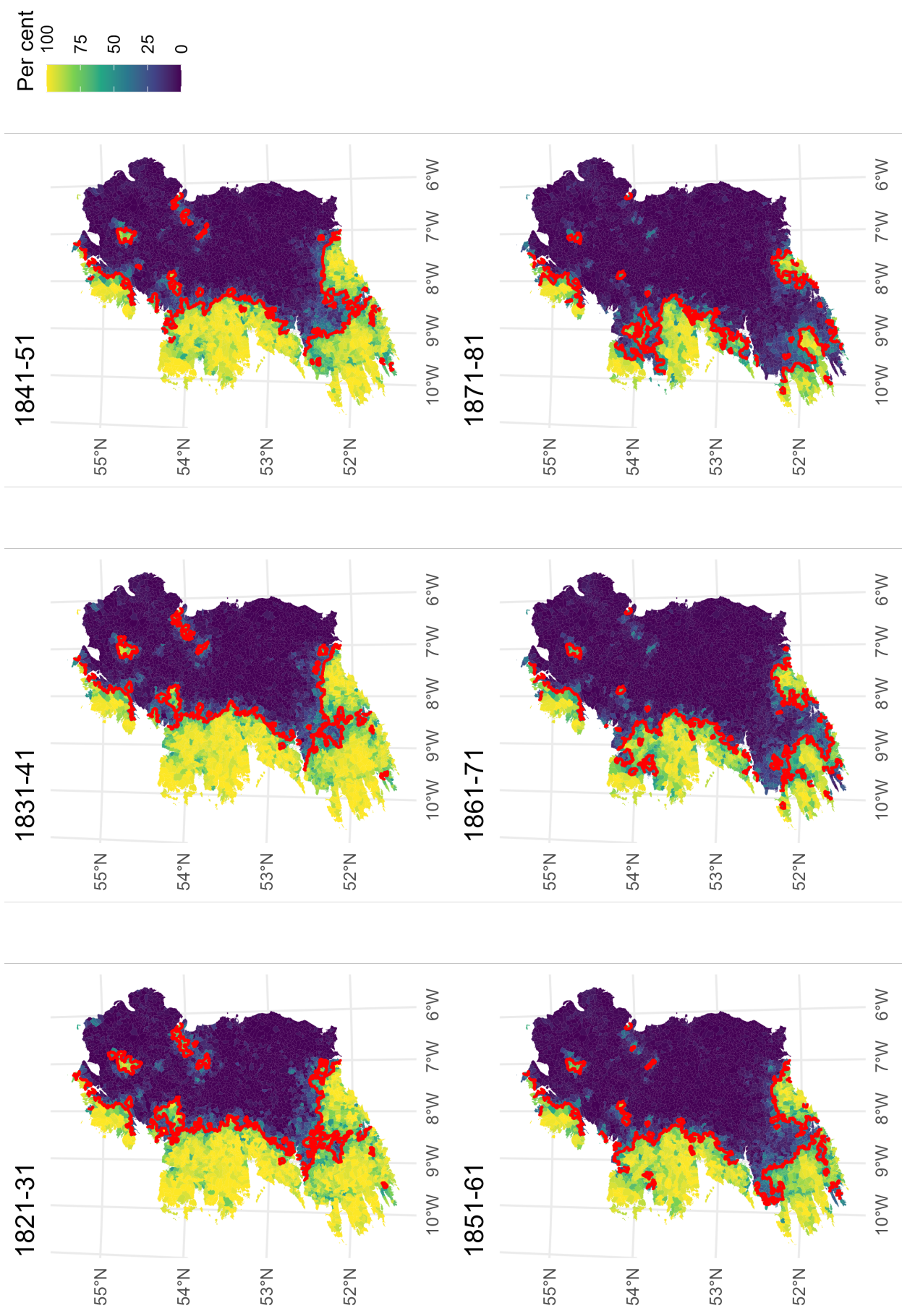
Figure 6 displays the Irish language’s retreat over the nineteenth century across pre- and post-Famine birth cohorts, aggregated to the District Electoral Division (henceforth district) level.¹⁹ Before the Famine, Irish-speaking majorities scarcely existed outside the most Western and Southern parts of the island. This frontier weakened in the second half of the nineteenth century, with all of the decline—unsurprisingly given that Irish was essentially extinct elsewhere—occurring in formerly Irish-speaking strongholds.

Also evident in Figure 6 is the bimodal distribution of the Irish-speaking share across the island. Districts either had full Irish coverage (over 90 per cent of the cohort

¹⁸The 1911 census also suffers from a well-known age-exaggeration issue caused by the introduction of the state pension in 1909 (de Bromhead et al., 2021; Heaney, 2026).

¹⁹Districts were Ireland’s basic territorial unit used for land value rating and census purposes, and to organise poor relief. They were amalgamated into larger areas for defining polling districts and political constituencies. Devised in the mid-nineteenth century to create equally populated units, they were small (on average 25 km²) but varied in size; they reflected local interests and historical and geographical considerations.

Figure 6: Irish Language Proficiency by District across Decennial Birth Cohorts, 1821–1881



Notes: The linguistic frontier, marking the boundary between predominantly Irish-speaking and predominantly non-Irish-speaking areas, is highlighted in red.

population with Irish-speaking ability) or a complete absence (fewer than 10 per cent). For pre-Famine birth cohorts, 20 per cent of districts could be categorised as fully Irish speaking, and Irish was completely absent in nearly half of these areas. Post-Famine, the share of districts where over 90 per cent of the population was Irish speaking fell from 20 per cent to 5 per cent, while the equivalent monolingual English share rose from half to two-thirds.

Three scenarios describe how the prevalence of Irish language ability fared over the century. Firstly, Irish remained popular in areas with high initial concentrations of Irish speakers. Secondly, Irish speaking declined in areas where Irish had been popular in past generations. Finally, Irish speaking remained rare, without exception, in areas where Irish had already disappeared. Informed by the information in Figure 6, we compose our analysis sample by creating a data subset consisting of all individuals aged between 20 and 39 (born between 1861 and 1881) resident in pre-Famine Irish-speaking majority districts. Segregating our analysis sample means that we target individuals located in areas where a majority of the previous generation spoke Irish—precisely the group where Irish was lost in the subsequent generation because individuals from Irish minority districts did not take up Irish and therefore had no language ability to lose.

Table 1 reports the summary statistics of the variables used in our analysis. Appendix C provides a complete outline of the data sources and data generation procedures. We discard all observations where the return indicated that they were neither the household head nor a relative. This removes people residing in institutions and more transient groups such as boarders and lodgers. The majority of these variables account for either individual-level differences or district group-level differences.

Overall, around half of the observations indicated they possessed Irish-speaking ability in our data. As expected, this share is far greater than the 14.5 per cent in these age cohorts returned as Irish-speaking in the complete census returns because our sample is comprised of individuals resident in formerly Irish-speaking majority districts. The sex ratio is balanced; 49 per cent of the sample is female. The age of the sample skews slightly younger, with three-fifths belonging to the 1871-81 birth cohort compared to the two-fifths in the 1861-71 group. On average, the observations were residents in districts that lost around one-third of their population (0.29 log points) in the Famine through excess mortality and emigration (between the 1841 and 1851 censuses) and nearly 20 per cent (0.17 log points) in the post-Famine period through emigration (between the 1851 and 1871 censuses).

We measure distance from the linguistic frontier by tracing a line around districts where Irish-speaking fell below 50 per cent among pre-Famine birth cohorts (i.e., those aged over 60), and then measuring the shortest distance from each district to this line. An average of 27 km indicates that most districts were close to the frontier, although the furthest was still 116 km away. Geographic location is also captured using market

Table 1: Descriptive Statistics

Variable	N	Mean	St. Dev.	Min	Max
Irish Language Ability	290,440	0.496	0.500	0	1
Female	290,440	0.490	0.500	0	1
Born 1871–81	290,440	0.605	0.489	0	1
Population Decline 1841–51	290,440	0.295	0.231	−0.988	1.059
Population Decline 1851–71	290,440	0.167	0.243	−0.650	2.196
Linguistic Frontier Proximity (km)	290,440	26.921	21.647	1.923	116.356
Market Access 1841 (Log Points)	290,440	16.094	0.167	15.700	16.464
Market Access Change 1841–81 (Log Points)	290,440	0.334	0.102	0.171	0.555
Irish Speakers (Household, Previous Generation)	229,882	0.819	0.367	0.000	1.000
Irish Speakers (Townland, Previous Generation)	289,578	0.789	0.248	0.000	1.000
Irish Speakers (District, Previous Generation)	290,440	0.716	0.230	0.004	1.000
Irish Speakers (County, Previous Generation)	290,440	0.653	0.217	0.034	0.895
Literate (Read and Write)	289,320	0.864	0.342	0	1
District Literacy (Pre-National Schools)	290,440	17.620	7.107	0.000	37.234
District Literacy (National School Expansion)	290,440	23.092	11.713	−11.352	68.481
District Literacy (Post-Famine Expansion)	289,320	45.399	10.856	5.439	81.963
Roman Catholic	290,427	0.958	0.200	0	1
Roman Catholic Name Index	275,921	70.568	24.517	0.000	100.000
Traditional Name Index	275,921	26.313	24.098	0.000	69.610
Teachers per Capita (District)	290,440	0.006	0.005	0.000	0.043
Clergy per Capita (District)	290,440	0.003	0.006	0.000	0.131
Police per Capita (District)	290,440	0.003	0.004	0.000	0.030
Resident in Birth County	289,815	0.945	0.228	0	1
District Share Born in County	289,815	93.934	7.257	14.188	100.000
Pre-Famine Land Value per Acre	290,440	0.421	0.460	0.008	3.190
Pre-National Schools per 1841 Population	290,291	0.001	0.001	0.000	0.017
Pre-National School Pupils per 1841 Population	290,291	0.057	0.066	0.000	1.121
National Schools per 1841 Population	290,291	0.0003	0.0005	0.000	0.006
National School Pupils per 1841 Population	290,291	0.036	0.055	0.000	0.643
Occupation (Irish-Speaking)	175,622	0.242	0.092	0.000	1.000

Notes: Authors' calculations using data described in Appendix C.

access—the cumulative population of Ireland and Britain weighted by ease of access (i.e., the time and cost of travel via land, waterway, road, or rail). Following Fernihough and Lyons (2022), we measure market access both before and after the advent of rail. We therefore include pre-rail (circa 1841) market access and the growth in market access attributable to rail as explanatory variables. Rail increased market access by around 0.33 log points.

Intergenerational transmission is measured across several geographic strata. All individuals in our data reside in formerly Irish-speaking districts; the average Irish-speaking at the household level among elders (40 years or older and a relative) was around 82 per cent. This means over four-fifths of the sample had an Irish-speaking parent. The sample size drops from around 290k to 229k for this variable, as a minority share of the population did not live with an older relative. The share of the previous generation with Irish ability declines gradually as we increase the geographic scope from the household

(82 per cent) to the county (65 per cent) via townlands (79 per cent) and districts (72 per cent).

It is worth noting the scale of these geographical units. In 1901 there were nearly 70k townlands with an average townland population of just 64 people. districts were larger, with more than 3.5k distinct areas recorded and an average population of 772. Counties are far larger still; Ireland’s 32 counties had an average population of 138k in 1901. Aggregating neighbourhoods at different resolutions allows us to investigate the relevant geographic level through which intergenerational transmission occurs.

The establishment and expansion of national schools from the 1830s on was an important driver of literacy in nineteenth-century Ireland. This is evident in the statistics on individual literacy in our data, with 86 per cent indicating they could both read and write. However, school attendance was not uncommon, albeit irregular, before the emergence of national schools (FitzGerald, 2013), and we can track changes over the century more broadly. The 1841 census report (British Parliamentary Papers, 1843) provides data on a similar scale to districts and allows us to create a measure of literacy that pre-dates national schools. Since this measure of community-level literacy pre-dates the birth of the sample observations, it serves as a proxy for historical levels of schooling. Aggregate pre-national school literacy stood at a mere 18 per cent. Furthermore, we measure the influence of pre-National Schools using data from the *1826 Education Inquiry*, capturing schooling provision along both the extensive (schools per capita) and intensive (pupils per capita) margins (Fernihough, 2025).

Using the 1901 data, we can discern the immediate impact of national schooling by looking at the literacy difference between those aged 60 and over (born before 1841) and the previous literacy total based on the 1841 official report. Table 1 suggests the introduction of national schools led to a 23-percentage point increase in literacy in the years that immediately followed. This trend continued and the difference between the pre- and post-Famine birth cohorts (the literacy gap between our district cohorts: 20–39-year-olds and the group aged 60 and above) was 45 per cent, as schooling became almost universal across the island.

We explore the influence of religion on both the extensive and intensive margin. The extensive margin, one’s religious affiliation, is not particularly illuminating because the sample is overwhelmingly (96 per cent) Catholic. However, like Connor (2021) we can use first names to measure religious adherence at the intensive margin. We create a name index that measures names considered distinctly more “Catholic”. A score of zero indicates no Catholics held this name; a score of one is the opposite extreme. We also replicate Connor by constructing a traditional name index that measures historically popular names. The presence of the state and church is potentially important, and we create variables measuring the relative frequency of teachers, policemen, and all religious-based occupations. The latter category was mainly Catholic nuns and priests (94 per cent

of religious-based occupations in our sample districts).

We define all people resident in counties outside their declared county of birth as internal migrants. This variable provides insight into the influence of migration on Irish speaking. 94 per cent of our observations resided in their county of birth at the time of the census. Migration at the broad district level is also measured with an almost identical level of community migration evident (94 per cent). The low share of internal migrants in our data reflects the nature of our sample, as few migrants were moving into districts with pre-Famine Irish-speaking majorities.

Land values, reported in the 1851 census but surveyed before the Famine, are also included. Note that the value per acre, essentially a fair annual rent, varies considerably. If a substantial number of people shifted away from Irish to improve their labour market prospects, then occupational advancement may have contributed to the language's decline. To capture this channel, we construct an occupation-based measure of linguistic environment: for each individual, we take their stated occupation and assign the average level of Irish speaking among individuals in that occupation. This allows us to proxy the extent to which different occupations are associated with Irish or English use, and hence the potential incentives to adopt English. Not all census returns include a legible occupation, and we restrict attention to occupational categories with more than 200 observations; however, coverage remains high, with 97 per cent of individuals assigned a value for this variable.

4 Results

4.1 Empirical Strategy

Our empirical strategy is designed to take our theoretical model of decentralised language choice to the data. In the model, households are embedded in a social network and belong to one of two native-language groups, English or Irish, with initial fluency in their native language only. Households may invest in bilingualism at a cost, gaining the ability to communicate with all neighbours, while communication between neighbouring households is productive only when a common language is shared. Because bilingualism yields benefits solely through interactions with neighbours from the other language group, and provides no additional returns within one's own linguistic community, the incentive to adopt a second language depends on local social exposure. Language adoption is therefore optimal only when exposure to monolingual neighbours from the other group is sufficiently high to justify the cost. Changes in social connectivity or in household composition across generations can shift these local incentives, inducing language adoption at the boundary between communities. The theoretical model thus implies a gradual and spatially uneven retreat of the Irish-speaking frontier, driven by endogenous local interactions rather than

by uniform, economy-wide forces.

Our empirical specifications reflect this theoretical mechanism by relating individual Irish-language ability to covariates that capture both endogenous transmission and exposure channels, as well as exogenous shocks—such as demographic upheaval or institutional change—that alter local environments in ways orthogonal to individual choice. We estimate a series of probit regressions in which an indicator for Irish-language ability is modelled as a function of multiple explanatory factors motivated by the historiography of Irish language change presented in the appendix. These covariates proxy for variation in local social environments and intergenerational dynamics, and for shocks that plausibly shifted neighbourhood composition or altered the relative returns to English.

In practical terms, we run a series of multivariate probit regression models with different specifications accounting for different sources of language loss. To ease interpretation, we report all coefficients expressed as marginal effects and rescale all continuous explanatory variables by normalising them to have a zero mean and 0.5 standard deviation. This means that all marginal effects represent the predicted probability change based on a two-standard deviation (2SD) increase in the associated explanatory variable. The advantage of this approach is that it makes all effects comparable—the largest estimated marginal effect thus offers the most explanatory power throughout.

Whilst these data are at the individual level, long-range spatial autocorrelation, or spatial unit roots, remain a potential concern. We assess the extent of this issue using the spatial unit root tests proposed by Müller and Watson (2024).²⁰ We construct an aggregated dataset from our district subsample and measure the average level of Irish speaking. The $I(0)$ test for a spatial unit root fails to reject (p -value of 0.317), whereas the $I(1)$ test is rejected (p -value of 0.084). On this basis, we rule out Irish-speaking as a spatial unit root process subject to low-frequency noise. In addition, we use county-clustered standard errors throughout to account for shorter-range spatial correlation.

4.2 Spatio-Linguistic Geography

How does language transfer from one generation to the next? Table 2 reports probit regression marginal effects from several model estimates that link an individual’s Irish-speaking ability to the ability of their household and neighbours. We measure intergenerational transmission by restricting the Irish-speaking environment, whatever the level of aggregation, to represent individuals aged 40 and above. This removes the potential for reverse causality, a common problem in the peer effects literature (Angrist, 2014).

Column (1) of Table 2 measures the intra-household intergenerational transmission of Irish ability. The baseline regressors indicate that those born in 1871–81 were 19 percentage points less likely to speak Irish relative to their 1861–71 counterparts. The

²⁰Tests implemented using Becker et al. (2025).

Table 2: Spatio-Linguistic Geography, Probit Regression Marginal Effects

	Irish Language Ability			
	(1)	(2)	(3)	(4)
Female	0.002 (0.008)	-0.029*** (0.010)	-0.002 (0.006)	-0.034*** (0.007)
Born 1871-81	-0.186*** (0.012)	-0.119*** (0.011)	-0.220*** (0.010)	-0.146*** (0.008)
Irish-Speaking: Household	0.689*** (0.045)		0.439*** (0.021)	
Irish-Speaking: Father		0.372*** (0.046)		0.212*** (0.021)
Irish-Speaking: Mother		0.473*** (0.054)		0.377*** (0.030)
Irish-Speaking: Townland			0.424*** (0.075)	0.440*** (0.094)
Irish-Speaking: District			0.234*** (0.049)	0.232*** (0.047)
Irish-Speaking: County			0.046 (0.081)	0.016 (0.091)
Average Irish Speaking	0.495	0.469	0.495	0.469
McFadden's R-Squared	0.195	0.185	0.320	0.303
Observations	229,882	95,059	229,882	95,056

Notes: Columns (1) to (4) report marginal effects from individual-level probit regressions where the dependent variable is an indicator for Irish language ability. Marginal effects are reported for a two-standard-deviation change in continuous explanatory variables, and for a discrete change from zero to one for binary indicators. Column (1) focuses on intergenerational transmission within the household, relating individual Irish-speaking outcomes to household-level Irish use. Column (2) replaces the household measure with separate indicators for the Irish-speaking ability of the mother and father. Column (3) extends the analysis to incorporate wider linguistic environment effects. Column (4) replicates this specification for the son/daughter subsample. Standard errors are clustered at the county level. Statistical significance is indicated by $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

estimated female marginal effect indicates no sex difference in Irish-speaking ability. The household transmission marginal effect is 0.7. Given that roughly half of the sample were Irish speakers, this means Irish speaking was absent among residents in households where the previous generation did not speak Irish. But if the previous generation were Irish speaking, this boosts the likelihood of Irish speaking by 70 percentage points. If the next generation wholly adopted, without deviation, the previous generation's language ability, we would expect to find a marginal effect of one. This suggests that while households are important, they are not the only influence on language acquisition.

Column (2) takes a focused view of intergenerational transmission by restricting the estimation sample to “sons” and “daughters” currently living with both their mother and

father.²¹ We find that both mothers and fathers act as influential conduits. Interestingly, mothers exert around 27 per cent more influence than fathers. This result lends support to Mulholland (2025, p. 216–217), who posits that female life-cycle migration—the growth of agricultural in-service, with women moving from the Irish-speaking west to the east and then returning to start families in the nineteenth century—helped to break the intergenerational transmission of language.

In columns (3) and (4), we expand the geographic scope over which intergenerational transmission occurs. These broader measures incorporate Irish speaking, for those aged 40 and older, at the townland, district, and county levels. Column (3) replicates (1) and shows that both the townland and district are independently associated with Irish speaking. Transmission at the townland level appears as important as that within the household. There is a decay in influence as the geographic scope increases. The county-level estimates are small and not statistically significant. Thus, the generational link appears to cease at geographic units larger than 5 km.²² The results in column (4) are similar: they emphasise the disproportionate influence of mothers and again underline the local nature of intergenerational language transmission. One caveat is that we cannot disentangle the intercorrelations between the different spatial units. As a result, the district- and county-level effects are likely attenuated, since their influence on Irish speaking may be partially mediated through the household or townland.

4.3 Economic Geography

Our conceptual framework highlights the relevance of geography and the shifting language frontier. Table 3 examines this from both first- and second-nature perspectives. Column (1) uses proximity to the frontier (the inverse of distance) to capture first-nature effects. The result is clear: the closer an individual is to the language frontier (as defined in Section 3), the greater the likelihood of language loss.

However, linguistic and economic space is not defined solely by geographic coordinates. Instead, place and space are embedded in a transport network, which in turn shapes the concentration of economic activity (proxied by population size). This captures second-nature influences. Column (2) shows that these are more influential than their first-nature counterparts, both statically and dynamically. Market access derived from Ireland’s pre-rail transport network yields a marginal effect of -0.7 . Moreover, the substantial increase in market access following the introduction of railways reduced Irish speaking in a similar manner, with an effect of -0.5 . Column (3) shows that all three geographic variables act as independent sources of language decline.

²¹Despite the reduction in sample size, replicating the specification in column (1) yields qualitatively similar results.

²²The average district is 25 km².

Table 3: First- and Second-Nature Economic Geography, Probit Marginal Effects

	Irish Language Ability		
	(1)	(2)	(3)
Female	-0.030*** (0.009)	-0.026*** (0.008)	-0.028*** (0.008)
Born 1871-81	-0.142*** (0.013)	-0.143*** (0.013)	-0.151*** (0.013)
Linguistic Frontier Proximity	-0.407*** (0.064)		-0.310*** (0.067)
Market Access: 1841		-0.699*** (0.097)	-0.536*** (0.127)
Market Access Rise: 1841-81		-0.463*** (0.138)	-0.413*** (0.146)
Average Irish Speaking	0.496	0.496	0.496
McFadden's R-Squared	0.105	0.102	0.145
Observations	290,440	290,440	290,440

Notes: This table reports marginal effects from probit regressions where the dependent variable is an indicator for Irish language ability. Column (1) includes first-nature economic geography only, captured by proximity to the linguistic frontier. Column (2) replaces the frontier measure with second-nature economic geography, measured by baseline market access in 1841 and the change in market access between 1841 and 1881. Column (3) includes both first- and second-nature economic geography jointly. All specifications include controls for sex and birth cohort. Standard errors are clustered at the county level. Statistical significance is indicated by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Changes in market access provide a mechanism for the observed shift and help reconcile our findings with Lee (1989). While Lee emphasises the role of commerce in driving the adoption of English prior to mass migration, he argues that this alone cannot explain the abandonment of Irish. Our results suggest that improvements in market access, through the expansion of railways and access to overseas markets, link the growth of commercial integration to the decline of Irish.

4.4 Education

The introduction of publicly funded primary schooling ('national schools') is often cited as a key factor in the language shift (see Appendix A). Education focused on the 3 R's, with English as the medium of instruction (Corcoran, 1925; Williams and Ford, 1992; Coolahan, 2001). However, this narrative downplays the role of agency, particularly within the network of voluntary private schools that pre-dated the national system. If state schooling were the primary channel through which bilingualism was discouraged, the literacy gains associated with its expansion should exhibit the largest marginal effect. Instead, pre-existing variation in literacy is the dominant factor, while gains linked to national schools are secondary in magnitude. This pattern suggests that

Table 4: Education and Irish Language Ability, Probit Regression Marginal Effects

	Irish Language Ability					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.022*** (0.008)	-0.032*** (0.006)	-0.027** (0.011)	-0.042*** (0.009)	-0.027*** (0.009)	-0.027*** (0.009)
Born 1871–81	-0.105*** (0.014)	-0.142*** (0.014)	-0.098*** (0.011)	-0.128*** (0.012)	-0.125*** (0.011)	-0.127*** (0.011)
Fully Literate	-0.226*** (0.027)	-0.075*** (0.016)	-0.123*** (0.021)	-0.008 (0.013)		
District Literacy: Pre-National School		-0.417*** (0.049)		-0.427*** (0.052)		
District Literacy: National School Rise		-0.237*** (0.049)		-0.259*** (0.052)		
District Literacy: Post-Famine Rise		-0.084* (0.049)		-0.123** (0.062)		
Father Fully Literate			-0.087*** (0.016)	-0.026** (0.012)		
Mother Fully Literate			-0.159*** (0.025)	-0.051*** (0.016)		
Pre-National Schools per Capita					-0.096*** (0.032)	
National Schools per Capita					-0.030 (0.042)	
Pupils in Pre-National Schools per Capita						-0.112*** (0.034)
Pupils in National Schools per Capita						-0.062* (0.036)
Sample	Full	Full	Sons/Daughters	Sons/Daughters	Full	Full
Average Irish Speaking	0.496	0.496	0.469	0.469	0.496	0.496
McFadden's R-Squared	0.028	0.171	0.052	0.181	0.018	0.024
Observations	289,320	289,320	93,750	93,750	290,291	290,291

Notes: Columns (1) to (6) report marginal effects from individual-level probit regressions where the dependent variable is an indicator for Irish language ability. Marginal effects are reported for a two-standard-deviation change in continuous explanatory variables and for a discrete change from zero to one for binary indicators. Column (1) introduces individual literacy as a control, while column (2) adds measures of local literacy and schooling intensity at the district level. Columns (3) and (4) replicate these specifications for the son/daughter subsample, distinguishing between paternal and maternal literacy effects. Columns (5) and (6) use alternative sources of district-level schooling. Standard errors are clustered at the county level. Statistical significance is indicated by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

decentralised voluntary schooling played a more significant role than the emerging state system, implying that (grand)parents were willing—rather than compelled—to send their children to English-medium schools.

Table 4 explores the role of education empirically, both at the individual level and aggregate district level over time. We again report probit regression marginal effects, where the explanatory variables are either dichotomous dummy variables or have been rescaled so that a one-unit change reflects a 2SD shift.

Column (1) examines the basic literacy–Irish link and suggests that literacy is associated with a 23 percentage point reduction in Irish-speaking ability. In column (2), we expand the analysis to incorporate environmental variation in literacy. Community-wide differences offer greater explanatory power, with historical variation in literacy that pre-dates national schooling emerging as the primary driver. A 2 SD shift (a 14 pp

change relative to the variable’s mean of 18 pp) in pre-national school literacy reduces the probability of Irish speaking in our sample by 0.42. The impact of national schools is important, but the estimated marginal effect of -0.24 is approximately half that of pre-national school literacy. Literacy among contemporaries—the post-Famine rise—is the least important contextual factor, although its marginal effect is still comparable to that of individual literacy (both around -0.08).

The results in column (2) underline the importance of community-level intergenerational transmission, although they omit within-household parent-to-child transmission. Columns (3) and (4) address this by focusing on our subsample of sons and daughters. Column (3) shows that parental literacy reduces Irish-speaking ability by 16 and 9 percentage points for mothers and fathers, respectively. Taken together, parental literacy matters more than individual literacy. Thus, even within households, the influence of education operates on an intergenerational basis. In column (4), we again highlight the importance of geographic context, as the marginal effects of parental literacy diminish once the three district-level literacy variables are included. All three district literacy measures yield similar marginal effects to those in column (2), with the largest effect associated with pre-national school literacy. Column (4) effectively replicates the specification in column (2) while incorporating parental literacy. Notably, an individual’s own literacy status no longer has explanatory power in this specification.

Columns (5) and (6) explore this further by incorporating data from the *1826 Education Inquiry* (Fernihough, 2025), alongside information on the establishment of National Schools between 1833 and 1843 (Lougheed, 2018). These sources capture variation in school provision through both the number of schools in each district and the number of pupils attending.²³ These measures are admittedly imperfect. For example, we cannot observe school quality, and restricting school density to within-district counts ignores cross-district attendance. However, both the 1826 (12,013 schools serving 572,791 pupils) and 1833–43 (3,375 schools with 363,162 pupils) snapshots are subject to similar limitations, and because we standardise the variables, their magnitudes are directly comparable. In both columns (5) and (6), the density of pre-national schools provides greater explanatory power, despite capturing the educational landscape at an earlier point in time. The national schools per capita variable is not statistically significant, whereas the pre-national schools variable is, and yields a substantially larger marginal effect. This pattern also holds when using pupil numbers instead of school counts.

Table 4 highlights two key features of language decline. First, education—and thus education policy—operates at the collective rather than the individual level. Second, this collective effect unfolds gradually over time. Changes do not appear to manifest within an individual’s life course; instead, they operate through the community, which gradually internalises them, generating a cumulative pattern of language retreat across successive

²³Both measured on a per capita basis to adjust for population size in 1841.

generations.

4.5 Church and State Capacity

Our results thus far point to the importance of education operating on a multidecade intergenerational basis. Next, we look at the influence of the Catholic church and the state more broadly. The probit marginal effect results are in Table 5. Column (1) shows that Catholicism is a strong predictor of Irish language ability. The estimated marginal effect of 0.47 relative to the Irish-speaking sample mean of 0.5 effectively suggests that the few (around 4 per cent) non-Catholics in our sample were not Irish speakers. This result is unsurprising, but given the complex nature of religious faith and national identity/ethnicity, it would be absurd to argue that these results demonstrate that the church promoted Irish speaking during this period.

Column (2) addresses this issue by using first names as a marker of Catholicism on the intensive margin. Following Connor (2021) and Fernihough and Henderson (2025), we argue that a person’s first name conveys information about their religious background and upbringing. Those with more distinctly Catholic first names were more likely to adhere to the Church’s position on various issues. Connor shows that this measure of Catholicism explains fertility, while Fernihough and Henderson find that individuals with more distinctly Catholic names were more likely to be illiterate. We also construct a traditional-name index to capture the persistence of popular names that may endure for non-religious reasons.

Interestingly, the marginal-effect estimate on the Roman Catholic name index is 0.07, meaning that a 2SD increase in this measure increases Irish-speaking propensity by 7 percentage points. While the Catholic Church’s attitude in nineteenth-century Ireland was, at best, ambivalent toward the preservation of Irish speaking, the evidence here suggests that this attitude did not transmit to the Church’s followers. Traditional names, which arguably serve as a proxy for more conservative social values (or simply a lack of openness to new ideas), also explain the upkeep of Irish speaking (marginal effect of 0.05). The relatively small magnitude of both name marginal effects cautions against overemphasis in our interpretation—although, given that both variables only serve as rough proxies for religious adherence and traditional values, this implies a degree of measurement error that could downplay their statistical influence.

Column (4) reveals that despite the positive Catholic–Irish language relationship, the presence of Catholic clergy in a person’s district predicts lower Irish speaking. However, this effect diminishes and becomes statistically insignificant once per capita measures for Police and Teachers are included in column (4). Of the church and state agents, teachers appear to provide most of the sway. Again, these results do not support the hypothesis that the Catholic Church undermined the Irish language, at least in a manner

Table 5: Church and State Capacity, Probit Regression Marginal Effects

	Irish Language Ability				
	(1)	(2)	(3)	(4)	(5)
Female	-0.023** (0.009)	-0.015** (0.007)	-0.014** (0.007)	-0.014** (0.007)	-0.015** (0.007)
Born 1871–81	-0.135*** (0.012)	-0.133*** (0.012)	-0.134*** (0.012)	-0.135*** (0.012)	-0.134*** (0.013)
Roman Catholic	0.473*** (0.052)	0.457*** (0.050)	0.457*** (0.050)	0.457*** (0.050)	
Roman Catholic Name Index		0.068*** (0.018)	0.067*** (0.018)	0.066*** (0.017)	0.065*** (0.018)
Traditional Name Index		0.046*** (0.011)	0.046*** (0.011)	0.045*** (0.011)	0.046*** (0.012)
Clergy per Capita			-0.062** (0.031)	-0.038 (0.026)	
Police per Capita				-0.045 (0.032)	
Teachers per Capita				-0.057*** (0.021)	
Average Irish Speaking	0.496	0.497	0.497	0.497	0.515
McFadden’s R-Squared	0.043	0.045	0.048	0.052	0.017
Observations	290,427	275,912	275,912	275,912	264,919
Catholic Only Subsample	N	N	N	N	Y

Notes: Columns (1) to (5) report marginal effects from individual-level probit regressions where the dependent variable is an indicator for Irish language ability. Marginal effects are reported for a two-standard-deviation change in continuous explanatory variables and for a discrete change from zero to one for binary indicators. Column (1) introduces individual religious affiliation, while column (2) augments this with name-based measures of religious and cultural identity. Column (3) adds a measure that captures the presence of Catholic clergy, and column (4) further incorporates measures of state capacity, police and teachers. Column (5) restricts the sample to Catholics. Standard errors are clustered at the county level. Statistical significance is indicated by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

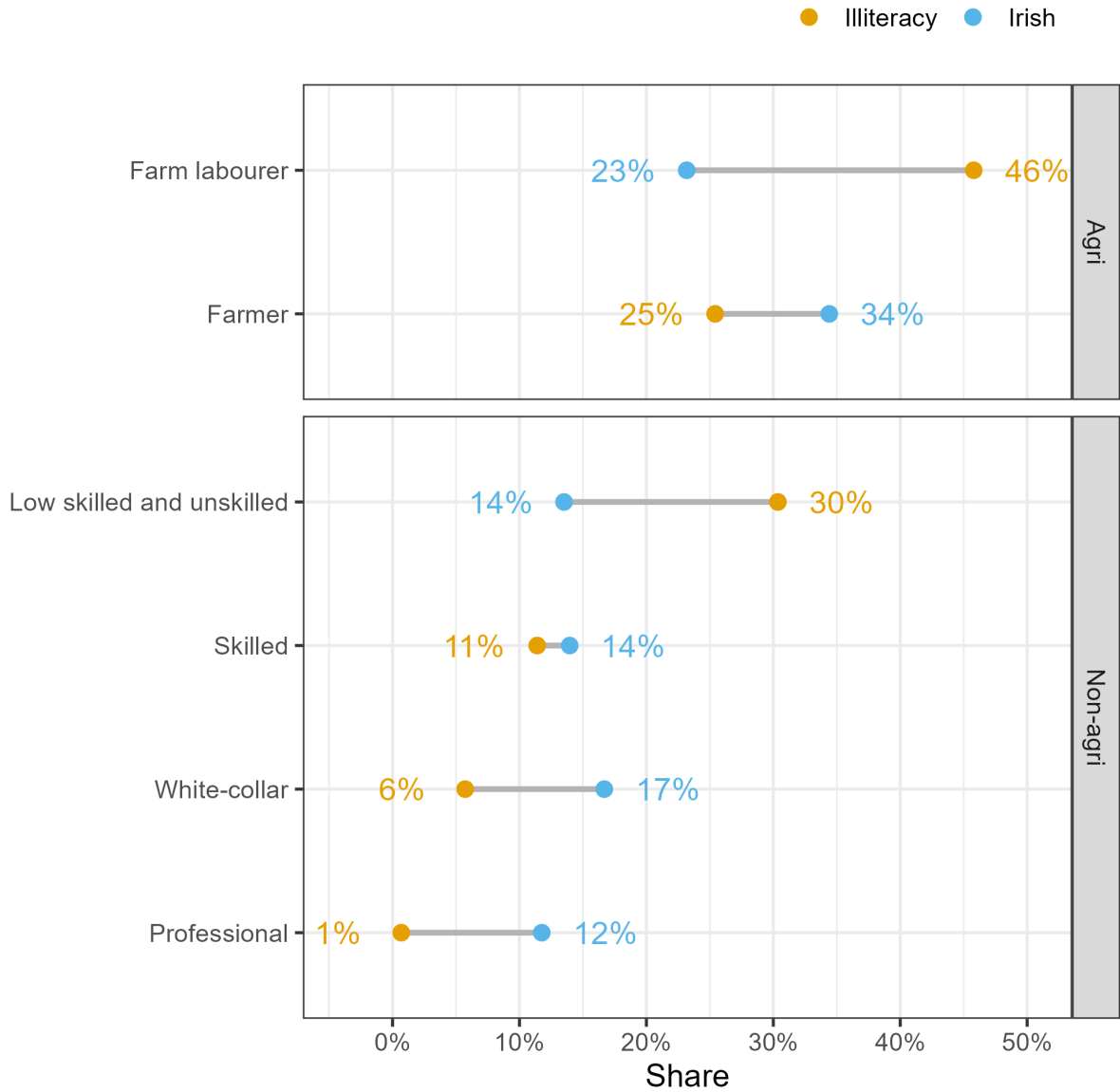
that manifests in our data. Column (5) reruns the first column’s specification on a Catholic-only subsample. The estimated marginal effects for both name-index variables are the same, demonstrating that the number of non-Catholics does not influence results.

4.6 Economics and Sociodemographics

The simultaneous decline of the Irish language and the demographic upheaval of the mid-to-late nineteenth century is suggestive. In this section, we investigate these changes formally. Figure 7 connects Irish speaking and occupational-level literacy for six categories in the complete 1901 census returns.²⁴ The socioeconomic status–literacy gradient

²⁴Non-occupations that appeared in this field, such as “scholar” or “retired”, were discarded. Those returned as “Read Only” are deemed illiterate in this application.

Figure 7: Literacy-Irish Occupational Connection



Notes: Data from full census returns of all males aged 30 and older ($N = 792,594$). Occupational categories follow Fernihough and Ó Gráda (2025).

is clear. Within the agricultural sector, farm labourers were almost twice as likely to be illiterate as the broad “farmers” category (46 vs. 25 per cent). Outside agriculture, the gradient is also evident: nearly 1 in 3 “low skilled and unskilled” workers were unable to fully read and write, whereas illiteracy was essentially nonexistent among the professional class.

Irish speaking does not follow the same trajectory. Within farming, farmers (1 in 3) were more likely to speak Irish than agricultural labourers (1 in 4). Moreover, we find a flat gradient in non-agricultural occupations. Around 1 in 7 low-skilled and unskilled workers were Irish speakers, compared with 1 in 8 among the professional classes. Indeed,

the more socioeconomically advanced skilled (1 in 7) and white-collar (1 in 6) workers were either equally or more likely to be Irish speaking than their low-skilled and unskilled counterparts. The Irish-speaking white-collar group—composed mostly of workers with public-facing roles, such as priests, teachers, and shopkeepers—had high rates of both literacy and Irish speaking. The high rate of bilingualism among priests is consistent with Wolf’s (2014) assertion that the Catholic Church facilitated bilingualism. Moreover, it demonstrates that Irish speaking is not correlated with socioeconomic disadvantage. If socioeconomic advancement was a reason to abandon bilingualism, we would expect to find a closer relationship between occupational class and Irish speaking.

Table 4.6 takes a closer look at this issue. In the first column, we add variables that measure the share of Irish speakers in each individual’s occupation and a dummy variable indicating whether the observation resides in their county of birth. The positive sign on the occupation’s Irish-speaking level is as expected, but the size of this marginal effect (0.15) is small relative to comparable explanatory forces. The relatively small effect here reflects the heterogeneous nature of Irish speaking. Consistent with our conceptual framework, language choice and socioeconomic position are more contingent on geographical position and the frontier than on raw occupational standing.

Another facet of the literature is the role of emigration in driving the shift to English. Lee (1989) argues that only the Irish adopted English before emigrating, linking language shift directly to migration incentives. Consistent with this, we find a strong relationship between internal migration and Irish speaking: those who remained in their county of birth were 20 percentage points more likely to be Irish speakers. Despite only measuring internal migration, this finding is consistent with the hypothesis that Irish speaking declined, in part, because of migration’s appeal, as parents abandoned the native language to prepare their children for life in Ireland’s English-speaking urban centres and overseas elsewhere in the Anglosphere. It also highlights our concerns about sample selection. If non-Irish speakers were more likely to leave, our sample overestimates the number of Irish speakers who originated from these communities. We address this concern more formally in Appendices E and F.

Column (2) of Table 6 introduces several district-level contextual variables. Famine (1841–51) and post-Famine (1851–71) population declines are negatively associated with language loss. The intergenerational context is again to the fore, as Famine population loss yields an estimated marginal effect of -0.11 compared with the post-Famine equivalent of -0.06 , although neither effect suggests that the Famine was a decisive factor.

The share of residents born outside their county of residence in the district shows a weak positive, but statistically insignificant, relationship with Irish speaking. That migration is associated with fewer Irish speakers matches the individual-level result, albeit with substantially less explanatory power. Irish was less likely to fall in areas with a larger pre-Famine population density and far more likely to fall in areas with more

Table 6: Economics and Sociodemographics, Probit Regression Marginal Effects

	Irish Language Ability		
	(1)	(2)	(3)
Female	-0.003 (0.019)	-0.005 (0.017)	0.016 (0.013)
Born 1871–81	-0.113*** (0.014)	-0.124*** (0.015)	-0.142*** (0.015)
Population Decline: 1841–51		-0.110** (0.046)	-0.114** (0.047)
Population Decline: 1851–71		-0.064* (0.038)	-0.067* (0.039)
Occupation: Irish Speaking	0.154*** (0.013)	0.127*** (0.012)	
Non-Migrant: Resides in Birth County	0.198*** (0.045)	0.150*** (0.041)	0.152*** (0.040)
District Share Born in County		0.081 (0.065)	0.083 (0.066)
Population Density: 1841		0.193 (0.166)	0.194 (0.169)
Pre-Famine Land Value per Acre		-0.320** (0.147)	-0.325** (0.152)
Occupation: Professional			-0.103*** (0.031)
Occupation: White-Collar			0.007 (0.017)
Occupation: Low Skilled and Unskilled			0.031* (0.016)
Occupation: Farmer			0.110*** (0.032)
Occupation: Farm Labourer			0.069*** (0.024)
Average Irish Speaking	0.506	0.506	0.506
McFadden’s R-Squared	0.036	0.075	0.069
Observations	175,312	175,312	175,312

Notes: Columns (1) to (3) report marginal effects from individual-level probit regressions where the dependent variable is an indicator for Irish language ability. Marginal effects are reported for a two-standard-deviation change in continuous explanatory variables and for a discrete change from zero to one for binary indicators. The estimation sample excludes all observations without a defined occupation. Column (1) introduces individual occupational characteristics and migration status. Column (2) augments this specification with local economic and demographic conditions, including population density and land values. Column (3) replaces the occupational Irish-speaking indicator with a full set of occupation group dummies. “Skilled” workers are the reference group. Standard errors are clustered at the county level. Statistical significance is indicated by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

valuable land—although we will see later how controlling for additional geographical features causes this effect to vanish. The final column (3) repeats the specification from (2) but measures occupation as a categorical explanatory variable. The results here broadly match those in Figure 7.

4.7 Including All Explanatory Factors

The preceding subsections provided a thematic analysis of the Irish language’s decline. In this subsection, we examine a fully specified model that unifies these themes. Our results, displayed in Figure 8, show the marginal effects stratified by whether the variable is recorded at the individual or place level, which we further divide into contemporary and historical measures. All explanatory variables are again scaled to facilitate comparisons.

Most of the marginal effects shown in Figure 8 are similar to their counterparts in Tables 2 to 6. Aside from the expected Roman Catholic effect (since we know the small number of other religions were almost entirely non-Irish speaking), the collection of literacy and economic geography variables provides the most explanatory power, with variation in pre-national school literacy being the most prominent. A 2SD increase in pre-national school literacy is associated with a 0.37 decrease in the probability that residents several decades later are Irish speaking. The increase attributable to the establishment of national schools is smaller (-0.23), but still of substantial importance. The post-Famine increase in literacy, among contemporaries, also remains relevant, with an estimated marginal effect of -0.12 .

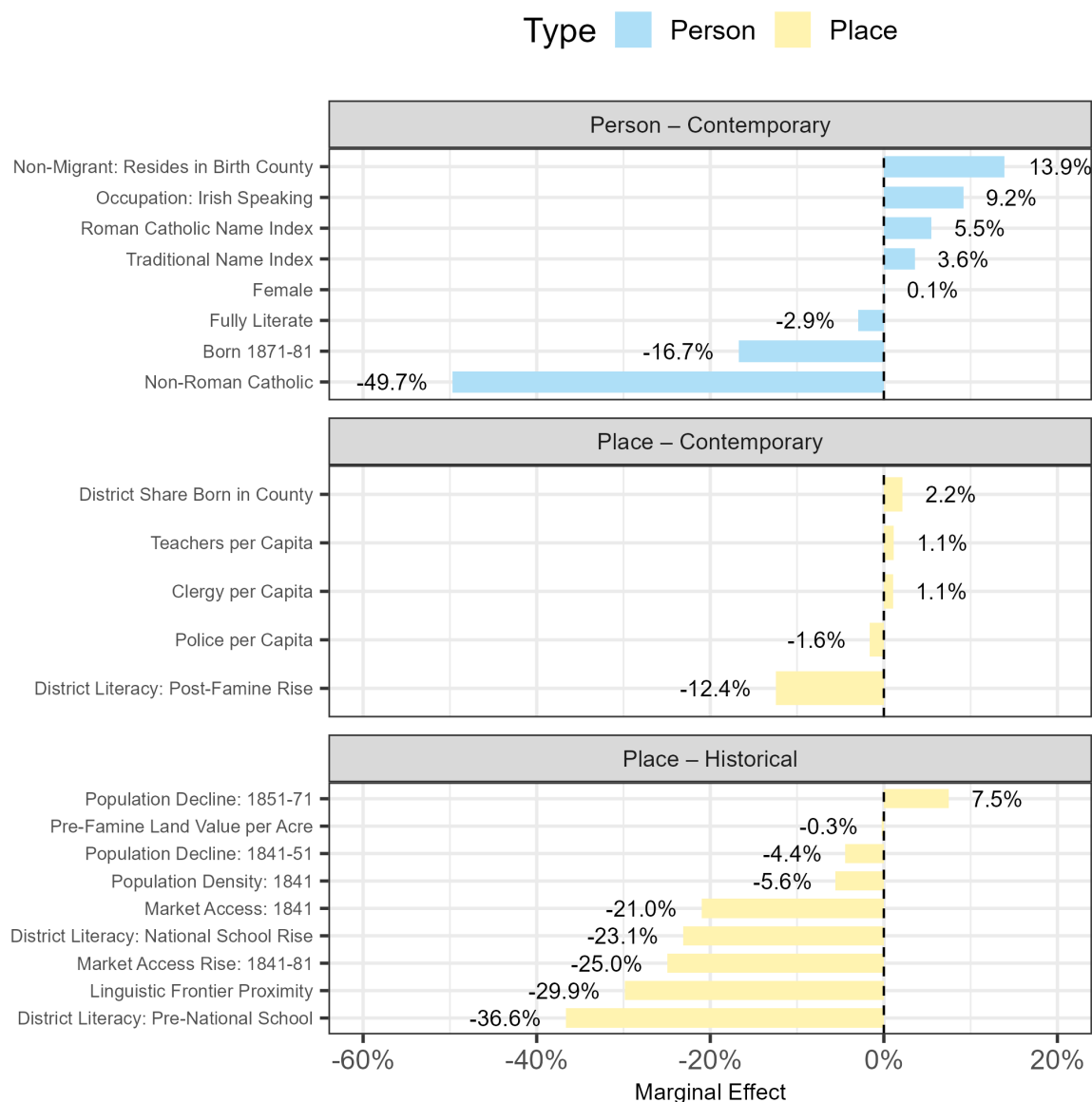
The proximity to the linguistic frontier is associated with a marginal effect of -0.30 , while rail-induced market access improvements contribute a further -0.25 , alongside a pre-rail market access effect of -0.21 . Overall, these different dimensions of economic geography offer considerable explanatory power in our model. The contributions of the additional explanatory variables are more limited. Interestingly, pre-Famine land values drop in importance, as does the population decline linked to the Famine. The post-

Table 7: Variable category importance decomposition

Category	Explanatory power (per cent)
Baseline	5
Spatio-linguistic geography	37
Education	39
Church and state	14
Economics and sociodemographics	5

Notes: The sum of the squared standardised probit coefficients represents the total explained variation.

Figure 8: Fully Specified Probit Model Marginal Effects



Notes: Depicted bars are marginal effects from a 2SD increase.

Famine population loss marginal effect changes sign: Figure 8 lists a marginal effect of 0.08, meaning a 2SD increase in population loss between 1851 and 1871 maps onto an 8 percentage point increase in Irish-speaking propensity.

Table 7 decomposes the explanatory power of each set of regressors by theme, taking the squared standardised probit coefficients and assuming that their sum represents the total explained variation. Thus, the values sum to 100. Baseline characteristics—year-of-birth cohort and sex—account for 5 per cent of explained variation. In line with the results in Figure 8, we find that variables relating to education offer substantial explanatory power, contributing at least 39 per cent of the explained variation. This underscores the importance of the literacy and education channel. The spatio-linguistic

geography—first- and second-nature geography—has an almost equal independent effect, absorbing 37 per cent of the variation. The contribution of the other variable sets is less important.

5 Discussion and Conclusion

Our analysis traces the contours of Irish language loss across the nineteenth century. Ireland’s language shift is striking in its apparent speed and permanency. Bilingualism proved to be a transitional stage towards a new English-language monolingualism. Subsequent policy interventions designed to revive the language have barely halted further loss; areas that the Irish government today recognises as places where the Irish language is a predominant vernacular represent just a small subset of the bilingual districts we analyse in our study. A major report into the modern use of Irish in these regions by Ó Giollagáin and Charlton (2015) warned that the language was on the verge of becoming dysfunctional even there; Irish has a social function only among the older generation, while younger speakers have just a passive or weak knowledge of Irish.

Bringing back a language that was so emphatically lost was always going to be a tall order. But part of the problem, we argue, is a failure by Ireland’s policymakers to understand why the island’s indigenous language fell out of favour in the first place. We contend that the decline of Irish was not due to direct exogenous forces alone. Rather, we find that the language frontier moved further and further west because of the interplay between endogenous and exogenous socioeconomic factors. The decline was the outcome of individuals and families making calculated trade-offs between the costs and benefits of language acquisition and retention, conditioned by the structure and linguistic composition of their social networks. Bourdieu, not Weber, explains the dynamics of the Irish case; the shrinking size of Ireland’s bilingual community was endogenous to social, cultural and economic forces rather than the consequence of state intervention.

Our findings point to a coherent account of Irish language decline as the outcome of interacting economic and social forces operating through decentralised decision-making. At the core was a rising differential in the communicative and economic returns to English relative to Irish, driven by migration opportunities and the growing global reach of the Anglosphere. This shift in relative returns altered the incentives facing households, who increasingly chose to acquire and transmit English through existing educational institutions. These decisions were embedded in local social networks, such that as the share of English speakers increased within a community, the value of Irish further diminished. The result was a self-reinforcing dynamic: educational choices and horizontal social interactions weakened the intergenerational transmission of Irish, producing a tipping process from stable bilingualism to English monolingualism.

In particular, our econometric analysis highlights the role of intergenerational trans-

mission through educational institutions as a key mechanism within this broader process. This was not primarily an exogenous shift imposed through state-funded, centrally planned education policy. Rather, it was the pre-existing education institutions that most influenced the Irish language border. Parents enrolled their children in these hedge schools to acquire a language—English—that would give them access to economic opportunity. These institutions’ decentralised and voluntary nature empirically supports the hypothesis that language boundaries during the modern period—an era marked by the emergence of the homogenised nation-state—were at least partially shaped by endogenous forces.

A further implication of this framework concerns the spatial resolution of language transmission. Our individual-level census data from a complete enumeration allow us to discern that vertical transmission within households was important, but it was not the only influence on language acquisition. Consistent with our game-theoretic model of Ireland’s language shift, we find that horizontal transmission between households within a geographic locale is probably of greater, but certainly of equal, importance for language transmission and retention. Finally, we also highlight the importance of time in the analysis of language change. That the effects of pre-national schooling could manifest themselves in a census conducted over 50 years later underlines that the economic analysis of language necessitates a long-run, historical, perspective.

Our analysis entails several limitations. Methodologically, we offer a unified quantitative approach that traces the impact of contributory sources on Irish-speaking ability. This approach does not, however, fully delineate these explanatory variables. They are not mutually exclusive, and we cannot provide a framework that neatly compartmentalises these various elements. While our results offer strong evidence on the mechanisms of language loss, they cannot fully account for why the decline was so severe and—as argued by Slomanson (2012)—historically unprecedented from a comparative perspective. We believe this is a strength rather than a weakness of our approach; the multicausal nature of our findings is truer to the historical process, in which language shift emerges from the interaction of exogenous shocks and endogenous responses, rather than from any single, isolable cause.

Following Ferguson (1959), the diglossia framework has been used to characterise stable bilingual societies in which a language with a high social status co-exists with another with a lower, more informal status.²⁵ Our analysis, in contrast, focuses exactly on the dynamics: an unstable society with a marked change in language use over time. Our theoretical model frames the shift in Ireland’s language frontier as the outcome of decentralised decisions shaped by local social networks, where the costs and benefits of using each language depend on the linguistic composition of those networks. As these networks evolved, the returns to the low-status language declined relative to the high-

²⁵For a historical application to Irish, see Ó Murchú (1988).

status language, leading to a tipping process in which the modal outcome became not to use Irish. Our evidence points to the modal decision becoming not to pass on the low-status language to the next generation. An alternative interpretation is that Irish lost its status even as the island's low-status language and that English supplanted Irish for all the social, cultural and economic uses historically associated with the informal indigenous language.

Europe today has several prominent language frontiers that do not correspond with national boundaries. Additionally, various actively spoken minority languages are used alongside national languages within many regions of Europe's constituent polities. A number of these minority languages even straddle political borders. We speculate that the Irish case differs from these others because of the nature of the language with which it competed, and hence the relative returns described above. The communicative value of English, a global language already by the nineteenth century when Irish saw its biggest decline in use, was enormous. In essence, Irish simply could not compete; it did not offer its users the full range of functions offered by English.

This comparative perspective reinforces the central role of relative economic returns in shaping language outcomes. Where Finnish competed in Finland's bilingual society with a Swedish language originating from a relatively small neighbouring economy (Sweden), Irish competed with the language used not only by its industrial superpower neighbour (Great Britain) but also in the principal countries that hosted its migrant diaspora (United States, Canada and Australia). The comparative fate of the Welsh language is apt. Unlike Ireland, nineteenth-century Wales industrialised, offering rural Welsh speakers economic opportunities within their home nation (Thomas, 1959; Baines, 1986). Conversely, Ireland's lack of industrialisation and economic opportunities meant the Irish had to migrate to where the industrial jobs were located—either elsewhere in the United Kingdom or overseas in one of its English-speaking offshoots (Ó Gráda, 1995), thereby increasing the relative returns to English.

What makes the fate of the Irish language unique is the interaction of economic conditions, mass migration, and the rise of the Anglosphere as the dominant cultural and political force of the nineteenth century. Consistent with our broader framework, education rather than demographic decline holds the most explanatory leverage in our econometrics. This is not paradoxical, as our models rely on statistical variation at the intensive, not extensive, margin. Thus, the uniform macro-demographic landscape does not translate into population change at the micro level. Here, education is crucial, with communities self-selecting into both education and language loss. In terms of linguistic diglossia, English, the high-status language of law and commerce, now supplanted Irish to additionally function as the low-status language. Schools acted as incubators where English was not only the medium of instruction but also that of daily interaction in the wider community.

Our interpretation is necessarily speculative. Unfortunately, the *relative* communicative value of the Irish language’s competing vernacular is not something that can be easily captured in a study of just two languages. An empirical approach involving multiple countries and diverse linguistic experiences, or the application of our methodology to other contexts, could be interesting and a fruitful area for future research.

References

- Adams, W. F. (1932). *Ireland and Irish Emigration to the New World 1815 to the Famine*. New Haven: Yale University Press.
- Angrist, J. D. (2014). The perils of peer effects. *Labour Economics* 30, 98–108.
- Angrist, J. D. and V. Lavy (1997). The effect of a change in language of instruction on the returns to schooling in Morocco. *Journal of Labour Economics* 15(1), S48–S76.
- Aspachs-Bracons, O., J. Costa-Font, I. Clots-Figueras, and P. Masella (2008). Compulsory language educational policies and identity formation. *Journal of the European Economic Association* 6(2/3), 434–444.
- Baines, D. (1986). *Migration in a mature economy: Emigration and internal migration in England and Wales 1861–1900*. Cambridge: Cambridge University Press.
- Becker, S. O., P. D. Boll, and H.-J. Voth (2025). Testing and correcting for spatial unit roots in regression analysis. *The Stata Journal*.
- Blanc, G. and M. Kubo (2025). The making of France. Mimeo, Simon Fraser University. URL: https://www.guillaumeblanc.com/files/theme/BlancKubo_French.pdf.
- Blum, M., C. L. Colvin, and E. McLaughlin (2026). Scarring and selection in the Great Irish Famine. *The Economic History Review* 79(1), 189–220.
- Borooah, V. K., D. A. Dineen, and N. Lynch (2009). Language and occupational status: Linguistic elitism in the Irish labour market. *The Economic and Social Review* 40(4), 435–460.
- Bourdieu, P. (1977). The economics of linguistic exchanges. *Social Science Information* 16(6), 645–668.
- British Parliamentary Papers (1825). *First Report of the Commissioners on Education in Ireland*, Volume (400). London: HMSO: House of Commons.
- British Parliamentary Papers (1843). *Report of the Commissioners Appointed to Take the Census of Ireland for the Year 1841*. London: HMSO.
- British Parliamentary Papers (1855). *The Census of Ireland for the Year 1851: Part IV: Report on Ages and Education*. London: HMSO: House of Commons.
- Central Statistics Office (2023a). Census 2022 Profile 8 – The Irish Language and Education. CSO Statistical Publication, 19 December 2023, Dublin. URL: <https://www.cso.ie/en/releasesandpublications/ep/p-cpp8/census2022profile8-theirishlanguageandeducation/irishlanguageandthegaeltacht/>.
- Central Statistics Office (2023b). Census of Population 2022 – Summary Results. CSO Statistical Publication, 30 May 2023, Dublin. URL: <https://www.cso.ie/en/releasesandpublications/ep/p-cpsr/censusofpopulation2022-summaryresults/migrationanddiversity/>.

- Chakraborty, T. and S. K. Bakshi (2016). English language premium: Evidence from a policy experiment in India. *Economics of Education Review* 50, 1–16.
- Chen, K. and N. Fabb (2026). Whorfian economics. In S. Weber and V. Ginsburgh (Eds.), *The Palgrave Handbook of Economics and Language* (2 ed.), pp. 3–29. London: Palgrave Macmillan.
- Clingingsmith, D. (2015). Are the world’s languages consolidating? The dynamics and distribution of language population. *Economic Journal* 127(599), 143–176.
- Connor, D. S. (2021). In the name of the father? Fertility, religion, and child naming in the demographic transition. *Demography* 58(5), 1793–1815.
- Coolahan, J. (2001). Education as cultural imperialism: The denial of the Irish language to Irish speakers, 1831–1922. *Paedagogica Historica* 37(1), 17–33.
- Corcoran, T. (1925). The Irish language in the Irish schools. *Studies: An Irish Quarterly Review* 14(55), 377–388.
- Corrigan, K. P. (2003). The ideology of nationalism and its impact on accounts of language shift in nineteenth century Ireland. *AAA: Arbeiten aus Anglistik und Amerikanistik* 28(2), 201–229.
- de Bromhead, A., A. Fernihough, and C. Ó Gráda (2021). 175 years of mismeasuring Ireland?: Some traps set by historical data. *Journal of the Statistical Social Inquiry Society of Ireland* LI, 128–155.
- de Fréine, S. (1965/1960). *The Great Silence*. Mercier Press. [1965 translation of the original 1960 Irish edition].
- Doyle, A. (2015). *A History of the Irish Language: From the Norman Invasion to Independence*. Oxford: Oxford University Press.
- Fabb, N. (2016). Linguistic theory, linguistic diversity and Whorfian economics. In V. Ginsburgh and S. Weber (Eds.), *The Palgrave Handbook of Economics and Language*, pp. 17–60. London: Palgrave Macmillan.
- Ferguson, C. A. (1959). Diglossia. *Word* 15(2), 325–340.
- Fernihough, A. (2025). Ireland’s educational foundations: Digitising and mapping the 1826 Education Inquiry. Digital Repository of Ireland. Collection. Accessed: 2026-03-20.
- Fernihough, A. and S. Henderson (2025). Protestantism and human capital: Evidence from early 20th century Ireland. *Explorations in Economic History* 95, 101647.
- Fernihough, A. and R. C. Lyons (2022). Agglomeration and emigration: The economic impact of railways in post-Famine Ireland. QUCEH Working Paper Series, Paper No. 22-11. URL: <https://www.quceh.org.uk/uploads/1/0/5/5/10558478/wp22-11.pdf>.
- Fernihough, A. and C. Ó Gráda (2025). Across the sea to Ireland: Return Atlantic migration before the First World War. *European Review of Economic History* 29(1), 49–67.
- FitzGerald, G. (1984). Estimates for baronies of minimum level of Irish-speaking amongst successive decennial cohorts: 1771–1781 to 1861–1871. *Proceedings of the Royal Irish Academy. Section C: Archaeology, Celtic Studies, History, Linguistics, Literature* 84C, 117–155.
- FitzGerald, G. (2003). Irish-speaking in the pre-Famine period: A study based on the 1911 census data for people born before 1851 and still alive in 1911. *Proceedings of the Royal Irish Academy: Archaeology, Culture, History, Literature* 103(5), 191–283.
- FitzGerald, G. (2013). Irish primary education in 1824. In G. FitzGerald (Ed.), *Irish Primary Education in the Early Nineteenth Century: An Analysis of the First and Second Reports of the Commissioners*

- of *Irish Education Inquiry, 1825–6*. Dublin: Royal Irish Academy.
- Galesloot, B. M. and S. Goyal (1997). Costs of flexibility and equilibrium selection. *Journal of Mathematical Economics* 28(3), 249 – 264.
- Ginsburgh, V. and S. Weber (2020). The economics of language. *Journal of Economic Literature* 58(2), 348–404.
- Goyal, S. and M. C. Janssen (1997). Non-exclusive conventions and social coordination. *Journal of Economic Theory* 77(1), 34 – 57.
- Grenfell, M. J. (2011). Bordieu, language and linguistics. In M. J. Grenfell (Ed.), *Bordieu, language and linguistics*, pp. 35–63. London: Continuum.
- Heaney, T. (2026). *Essays in Irish Economic Development*. PhD thesis, Queen’s University Belfast, Belfast.
- Hindley, R. (1990). *The Death of the Irish Language: A Qualified Obituary*. London: Routledge.
- Hynes, W. (2014). To what extent were economic factors important in the separation of the south of Ireland from the United Kingdom and what was the economic impact? *Cambridge Journal of Economics* 38(2), 369–397.
- John, A. (2016). Dynamic models of language evolution: The economic perspective. In V. Ginsburgh and S. Weber (Eds.), *The Palgrave Handbook of Economics and Language*, pp. 101–120. London: Palgrave Macmillan.
- John, A. (2026). Dynamic models of language survival and language death: The economic perspective. In S. Weber and V. Ginsburgh (Eds.), *The Palgrave Handbook of Economics and Language* (2 ed.), pp. 73–97. London: Palgrave Macmillan.
- John, A. and O. Özgür (2021). Linguistic diversity in the very long run. *The Economic Journal* 131(635), 1186–1214.
- Kessler, B. (1995). Computational dialectology in Irish Gaelic. In S. P. Abney and E. W. Hinrichs (Eds.), *Seventh Conference of the European Chapter of the Association for Computational Linguistics*, Dublin, Ireland. Association for Computational Linguistics. URL: <https://aclanthology.org/E95-1009>.
- Lee, J. J. (1989). *Ireland 1912–1985: Politics and Society*. Cambridge: Cambridge University Press.
- Lougheed, K. (2018). After the manner of the Irish schools: The influence of Irish national education in the British Empire. *Journal of Historical Geography* 60, 1–10.
- McMahon, A. and R. McMahon (2005). *Language Classification by Numbers*. Oxford: Oxford University Press.
- Morris, S. (2000, January). Contagion. *Review of Economic Studies* 67(1), 57–78.
- Moseley, C. (Ed.) (2010). *Atlas of the World’s Languages in Danger* (3rd ed.). Paris: United Nations Educational, Scientific and Cultural Organization. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000187026>.
- Mulholland, M. (2025). *At the Rising of the Moon: The Peasantry and Ireland from the Tudor Conquest to the Fall of Landlordism*. Dublin: Eastwood Books.
- Müller, U. K. and M. W. Watson (2024). Spatial unit roots and spurious regression. *Econometrica* 92(5), 1661–1695.
- Neary, P. R. (2011). *Multiple-Group Games*. Ph. D. thesis, University of California, San Diego.

- Neary, P. R. (2012). Competing conventions. *Games and Economic Behavior* 76(1), 301 – 328.
- Neary, P. R. and J. Newton (2017). Heterogeneity in preferences and behavior in threshold models. *The Journal of Mechanism and Institution Design* 2(1), 141–159.
- Northern Ireland Statistics and Research Agency (2022). Main Statistics for Northern Ireland: Language. Statistical Bulletin, 22 September 2022, Belfast. URL: <https://www.nisra.gov.uk/system/files/statistics/census-2021-main-statistics-for-northern-ireland-phase-1-statistical-bulletin-language.pdf>.
- Northern Ireland Statistics and Research Agency (2023). Census 2021 Main Statistics Language Tables. 30 November 2023, Belfast. URL: <https://www.nisra.gov.uk/publications/census-2021-main-statistics-language-tables>.
- Ó Giollagáin, C. and M. Charlton (2015). Nuashonrú ar an Staidéar Cuimsitheach Teangeolaíoch ar Úsáid na Gaeilge sa Ghaeltacht: 2006—2011. Technical report. URL: <https://www.rte.ie/documents/news/udaras-na-gaeltachta.pdf>.
- Ó Gráda, C. (1995). *Ireland: A New Economic History 1780–1939*. Oxford: Oxford University Press.
- Ó Murchú, M. (1988). Diglossia and interlanguage contact in Ireland. *Language, Culture and Curriculum* 1(3), 243–249.
- Slomanson, P. (2012). On the great silence: A gap in Irish historiography and consequences for language education in Ireland. *Nordic Irish Studies* 11(2), 95–114.
- Tada, M. (2018). Language, ethnicity, and the nation-state: On Max Weber’s conception of imagined linguistic community. *Theory and Society* 47, 437–466.
- Thomas, B. (1959). Wales and the Atlantic economy. *Scottish Journal of Political Economy* 6(3), 169–192.
- UNESCO (2003). Language vitality and endangerment. International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages, Paris, 10–12 March 2003. URL: <https://ich.unesco.org/doc/src/00120-EN.pdf>.
- Watson, I. and M. Nic Ghiolla Phádraig (2011). Linguistic elitism: The advantage of speaking Irish rather than the Irish-speaker advantage. *The Economic and Social Review* 42(4), 437–454.
- Weber, M. (1968/1921). *Economy and Society: An Outline of Interpretive Sociology*. Berkeley: University of California Press. [1968 translation of original 1921 German edition].
- Williams, J. E. C. and P. K. Ford (1992). *The Irish Literary Tradition*. Cardiff: University of Wales Press.
- Wolf, N. M. (2014). *An Irish-Speaking Island: State, Religion, Community, and Linguistic Landscape in Ireland, 1770–1870*. Madison, Wisconsin: University of Wisconsin Press.

Appendices

A Literature Review on Irish Language Decline

The literature on the decline of spoken Irish advances a range of explanations, most of which are framed as monocausal. Some operate primarily as exogenous shocks—such as demographic upheaval, institutional expansion, or state intervention—while others act through household responses to changing incentives. This literature—largely written by social and cultural historians, with important contributions from linguists—can be organised around four broad categories corresponding to the structure adopted below: (1) changing demand for English, (2) voluntary education, (3) the expansion of the English-speaking state, and (4) religion and nationalism.

Rather than competing explanations, we treat these categories as context-specific influences on the costs and benefits of language acquisition and the dynamics of intergenerational transmission. We view them as non-mutually exclusive and potentially interacting. In our quantitative analysis in Section 4, we decompose their effects into more endogenous and more exogenous components. Exogenous forces reshape local environments, while endogenous responses reflect household-level adjustments to these changes through language choice and intergenerational transmission.

This approach is consistent with our theoretical model sketched in Section 2, in which the endogenous decision faced by households is the choice of language portfolio adopted by younger members. This decentralised decision is strategic, but its optimality depends on factors that are exogenous from the perspective of individual households. Whether bilingualism or language switching is optimal is determined by the balance of costs and benefits shaped by local social environments, demographic composition, institutional context, and, importantly, the language portfolios of neighbouring households. The literature reviewed in this current appendix therefore provides historical counterparts to these elements of the model, identifying the social, economic, and institutional mechanisms emphasised by Irish language scholars through which the perceived costs and benefits of language acquisition were reshaped over time and across space.

A.1 Changing Demand for English

The nineteenth-century decline of Irish is closely associated with a gradual increase in the perceived returns to English, which shaped household language choices through labour-market incentives, migration prospects, and access to key institutions. Historical work points to a gradual increase in the economic and social demand for English rather than a sudden linguistic rupture (Doyle, 2015). English facilitated access to labour markets, migration opportunities, commerce, and socially valued professions such as the army and

the church (O’Brien, 1989). As older Irish-speaking generations passed away, parental preferences increasingly favoured English, allowing language shift to appear rapid at the aggregate level while remaining gradual at the household level.

Several contributors, most notably Nic Craith (2002), argue that language shift reflected growing commercial integration with Britain, with English functioning as the language of commerce and Irish increasingly associated with poverty. However, Lee (1989) contends that this explanation accounts for the adoption of English but not for the abandonment of Irish. Other small economies competing in similar export markets—such as Denmark—did not experience comparable language loss. Wolf (2014) further argues that language itself did not constitute a binding barrier to market access, as Irish speakers could rely on bilingual intermediaries.

Demographic change reinforced these incentives. Nineteenth-century Ireland experienced mass depopulation, driven largely by emigration. Emigration was already rising before the Great Famine, particularly to English-speaking destinations (Adams, 1932).²⁶ The Famine intensified existing trends and produced an ageing population (Colvin et al., 2025). Its proximate cause was an exogenous ecological shock—*Phytophthora infestans*—and migration was the primary non-governmental response (Ó Gráda, 1995; Ó Murchadha, 2011).

Although Ravenstein (1879) raised emigration as a potential driver of language decline, he discounted its importance on the grounds that migrants tended to come from English-speaking areas. Corrigan (2003) likewise argues that Famine-induced migration from Irish-speaking areas was quantitatively insufficient to account for the scale of language shift. Nevertheless, Irish-speaking regions were disproportionately poor before the Famine (Fernihough and Ó Gráda, 2022) and experienced the largest absolute scope for change in language use.

Rather than migration per se, Fitzpatrick (1986) emphasises emigration’s role in accelerating literacy—necessarily literacy in English. This argument focuses primarily on female emigration and understates male literacy trends. Lee (1989) observes that, unlike other European migrants, the Irish often acquired English proficiency prior to migration rather than at the destination.

A complementary mechanism operated through changes in labour-market structure, particularly the post-Famine expansion of domestic service. Following the collapse of domestic weaving after the Famine, a growing share of rural women entered domestic service, often at a young age and in middle-class, English-speaking households (Mulholland, 2025). This form of employment separated young women from Irish-speaking networks and placed them in settings where English was required for both communication and literacy. As a result, Irish-language use was constrained during critical years of linguistic

²⁶Ireland had one of the highest emigration rates per 1,000 population in Europe (Hatton and Williamson, 1994).

formation, weakening the intergenerational transmission of Irish when these women later became mothers (Mulholland, 2025).

A.2 Voluntary Education

Primary education in early nineteenth-century Ireland was provided by institutions known as ‘hedge schools’. About 9,000 in number by the 1820s, this was a highly decentralised system with little meaningful state involvement (FitzGerald, 2013). Detailed contemporary evidence on the location, management, curriculum, and language of instruction in these schools is provided by the *1826 Education Inquiry*, now digitised and georeferenced in the Digital Repository of Ireland, which documents the scale and heterogeneity of voluntary primary education across the island (Fernihough, 2025). As Corrigan (2003) emphasises, it is important to distinguish the effects of this locally-provided voluntary system of education from those of later state-funded education, since they operate through different mechanisms and have distinct implications for language choice.

There was substantial variation in hedge school registration and attendance by region and birth cohort, but overall participation rates were high by European standards. At the time of the 1821 census, around 44 per cent of boys and 26 per cent of girls attended school between the ages of six and thirteen (Blum et al., 2017). Children in Irish-speaking areas were marginally less likely to attend, though this difference is small in descriptive regressions of school attendance (Ó Gráda, 2013). Drawing on contemporary reports and administrative evidence, Heaney (2026) suggests that hedge schools did not constitute a binding supply constraint on education. Instead, he argues that schooling decisions reflected household strategies and opportunity costs, with irregular attendance and early withdrawal common as children were required for agricultural or domestic labour.

The curricula of hedge schools provide insight into the preferences of fee-paying parents. Instruction was overwhelmingly conducted in English, even in Irish-speaking districts (Wall, 1969), and parents sent their children to school precisely to acquire the language of economic opportunity (Mac Giolla Chríost, 2005). Besides scripture, arithmetic was the principal subject taught (McManus, 2002). Poor households demanded this practical curriculum to enhance their children’s prospects of social advancement. In this sense, hedge schools functioned less as a mechanism for cultural reproduction than as an endogenous response to rising returns to English, reducing the cost of second-language acquisition well before the introduction of national schooling.

Importantly, the persistence of hedge schools after the introduction of national schooling in 1831 does not appear to reflect resistance to education. Rather, they often coexisted with national schools, with their continued operation shaped by parental preferences, local institutional inertia, and irregular attendance patterns (FitzGerald, 2013). From the perspective of our model, hedge schools therefore facilitated language shift by lowering the

effective cost of English acquisition at the household level, smoothing the later transition induced by the expansion of state-funded education rather than acting as a barrier to it.

Attempts to teach through the medium of Irish were unsuccessful. Giving evidence in 1825, Henry Monck Mason, the secretary of the Irish Society for promoting the Education of the Native Irish through the medium of their own language stated under oath ‘that the schools for teaching in the Irish language exclusively did not thrive, because the parents prefer in general having their children taught English.’ (British Parliamentary Papers, 1825, p. 742)

A.3 The English-Speaking State

The expansion of the English-speaking state influenced language outcomes not only through coercion or settlement, but through a broad set of policies and institutions that altered the relative costs and benefits of language use. While there is debate over the extent to which explicitly coercive measures—such as the penal laws—shaped linguistic behaviour, what matters analytically here is that broader state policies systematically shifted the incentives surrounding language use. From the perspective of our model, the state constitutes a largely exogenous force that reshaped local environments by determining the language of administration, law, education, and public authority. As Corrigan (2003) emphasises, state-sponsored institutions affect language incentives in ways that differ fundamentally from decentralised schooling or market interaction, operating instead through standardisation and scale.

Within this broader context, Doyle (2015) identifies three historical language communities in Ireland: Gaeil (Irish), Sean-Ghaill (Old English), and Nua-Ghaill (New English). The English-speaking frontier in 1500 was confined to the English Pale (Doyle, 2015, Figure 3.1), but expanded gradually with the growth of Tudor and Stuart state power. This early phase of state expansion established English as the language of governance and law, even while Irish remained dominant among the population at large.

The Cromwellian conquest of the 1640s constituted a major political and demographic upheaval, precipitating the wider dissemination of English through the arrival of English and Scottish settlers in parts of Ireland beyond the Pale. Despite this expansion, English speakers remained in the absolute minority for much of the period (Doyle, 2015, Figure 4.1). The timing of the transition from an Irish-speaking to an English-speaking majority remains debated, with Doyle (2015) placing it in the mid-eighteenth century, while Wall (1969) and others date it closer to the mid-nineteenth century.

Some authors associate the subsequent retreat of Irish with the collapse of Gaelic Ireland, linking it to the penal laws imposed on the non-Anglican majority (Williams and Ford, 1992). Others argue that the decline set in only after these legal restrictions were largely repealed (Wall, 1969). For Lee (1989), it was instead the post-1800 *growth*

of the English-speaking state itself that proved decisive, as bureaucracy, administration, and legal authority operated exclusively in English.²⁷

A central institutional mechanism through which this expanding state influenced language outcomes was the introduction of state-funded primary education. Ireland was the first polity of the United Kingdom to receive a national system of schooling, beginning in 1831.²⁸ The National Education Board set a uniform curriculum in which English was the exclusive medium of instruction, with no accommodation for Irish until the end of the nineteenth century (e.g., Corcoran, 1925). This system has long been interpreted as an assimilationist policy intended to promote social and cultural integration (Akenson, 1970; Coolahan, 1981; Nic Craith, 2002). Williams and Ford (1992) argues that there is ‘no doubt at all that the education policies of the Board turned an Irish-speaking nation into an English-speaking one’ (p. 256), while Coolahan (2001) views national schooling as a supportive force that hastened, rather than caused, language decline.

The scale of the system expanded rapidly: the number of national schools rose from around 1,000 in its first decade to nearly 9,000 by the turn of the century (British Parliamentary Papers, 1851, 1900). The state paid teachers’ salaries and financed a substantial share of school construction, and although the system was intended to be non-denominational, a de facto denominational structure emerged (Coolahan, 1981). The core curriculum focused on reading, writing, and arithmetic, with English always the language of instruction. As Comerford (2003) observes, ‘as the Irish population acquired literacy it did so in English’. Subsequent revivalist efforts to restore Irish through compulsory instruction met limited success, in part because secondary and university education also operated exclusively in English (Wall, 1969), and because of the weak Irish-language print culture that developed after the mid-nineteenth century (Ó Ciosáin, 2024).

A.4 Religion and Nationalism

Historically, Ireland’s religious divide also coincided with its linguistic divide: Catholicism was associated with Irish and Protestantism with English (Wolf, 2014). Roman Catholicism was a central pillar of Irish nationalist identity and, in the context of an Anglican state, also functioned as an alternative source of social control (O’Brien, 1989). The fact that this dimension of Irish nationalism was arguably indifferent, if not unsupportive, of the Irish language is therefore highly significant for understanding language dynamics.

The post-Famine period witnessed a shift from what Jackson (1999) characterises as ‘traditional lax and heterodox practice’ towards a more formal, centralised, and rigorous Catholicism. This transformation increased mass attendance and enhanced the political

²⁷One indicator of the state’s expanding reach is the sharp increase in policing across the island over the nineteenth century (McLaughlin et al., 2021).

²⁸National schools did not immediately replace hedge schools; some Catholic dioceses continued to support voluntary provision (Barr, 2009).

and social influence of the clergy. Catholic religious services were not conducted through Irish (or, indeed, English); mass itself was typically delivered in Latin (Cullen, 1990). There was also little institutional support for the Irish language in clerical training.²⁹

While popular narratives frequently apportion blame to the Catholic clergy for the demise of Irish, Wall (1969) argues that this explanation lacks nuance, as it overlooks the much earlier erosion of the language's social status. Wolf (2014) likewise disputes the claim that the Catholic Church was antipathetic towards Irish, arguing instead that clerics often encouraged and facilitated bilingualism rather than monolingual English adoption.³⁰ This interpretation also struggles to account for variation at the relevant margin: although Protestants rarely spoke Irish, the language frontier lay almost entirely within Catholic populations, implying that confessional differences alone cannot explain patterns of language shift.

A related strand of the literature emphasises nationalist politics. As the central figure in the mass nationalist movement of the early and mid-nineteenth century, Daniel O'Connell (1775–1848) is frequently identified as a key culprit in the decline of the Irish language for failing to support it (Wall, 1969). Known as 'The Liberator' for his role in Catholic emancipation, O'Connell's views are widely quoted:

I am sufficiently utilitarian not to regret its gradual abandonment. A diversity of tongues is no benefit; it was first imposed on mankind as a curse, at the building of Babel. It would be of vast advantage to mankind if all the inhabitants of the earth spoke the same language. Therefore, although the Irish language is connected with many recollections that twine round the hearts of Irishmen, yet the superior utility of the English tongue, as the medium of all modern communication, is so great that I can witness without a sigh the gradual disuse of Irish.³¹

The principal charge against O'Connell and his supporters is that they did little to encourage a Gaelic cultural revival alongside political mobilisation (Murphy, 1948). While Catholic emancipation was a mass movement, its medium was English.³² Ironically, the subsequent generation of nationalists, the so-called 'Young Irelanders', promoted Irish under the influence of continental romanticism linking language and national identity, yet none of its leaders could speak Irish, and its main publication, *The Nation*, was published in English (Doyle, 2015).

Prominent figures in the Irish language revival movement at the turn of the twentieth century included Douglas Hyde, D.P. Moran, Horace Plunkett, and Ernest Blythe. In

²⁹Maynooth College, a state-funded Catholic seminary, appointed a cleric to its chair in Irish who had not met the advertised standard and did not subsequently develop the language there (McMahon, 2008).

³⁰For example, John Kiely, a Catholic clergyman from Co. Cork, stated that clergy were 'obliged to catechize and instruct in both languages' (British Parliamentary Papers, 1825, p. 320).

³¹Quoted in Murphy (1948); Ó Cuív (1969); Ó Tuathaigh (1974)

³²Doyle (2015) speculates that O'Connell would have been incapable of encouraging Irish, suggesting his Irish never developed beyond what he learnt as a child. Wolf (2014) disputes this, arguing that O'Connell was willing to use Irish when it aided communication of his political message.

addition to differing in religious affiliation, these individuals also differed substantially in their political views on Irish independence. However, Walsh (2011) argues that they were united by a shared belief that the Irish language provided a foundation for cultural renewal, social cohesion, and economic development. They viewed language promotion not merely as a symbolic project, but as a practical instrument for national development—an idea that subsequently informed twentieth-century language policy, including efforts to promote Irish usage through schooling and to foster economic development in the Gaeltacht regions (Walsh, 2022).

B Proof of Proposition 1

First, we restate Proposition 1.

Proposition 1. Given Assumptions 1 and 2, there is a unique pure strategy equilibrium in which

1. all Native English speaking households remain monolingual, and
2. all native Irish-speaking households with a fraction greater than c of neighbours in E become bilingual; the remainder in I remain monolingual.

Proof. Consider first those in E . By Assumption 2, each household in E has fraction at least $1 - c$ of neighbours in E . As such, regardless of the language portfolio a household in I chooses, the payoff to a household in E from remaining monolingual strictly exceeds $1 - c$, which is the guaranteed payoff from becoming bilingual. Therefore, the unique best-response for every household in E is to choose $\mathcal{E}$.

For the Native Irish households, I , begin by partitioning them into $I' \cup I'' \cup I'''$ where I' are those households with no English speaking neighbours, I'' are the households with fraction less than c Native English speaking neighbours, and I''' are the households with a fraction at least c English speaking neighbours. We will show that the unique pure strategy equilibrium involves all players in I' choosing $\mathcal{I}$, all players in I'' choosing $\mathcal{I}$, all players in I''' choosing $\mathcal{B}$, and all players in E choosing $\mathcal{E}$.

Now consider a household j in I''' . Above we showed that all Native English speaking neighbours remain monolingual, and so, by (3) the utility from $\mathcal{I}$, $U_j(\mathcal{I}, \mathbf{s}_{-j})$ is equal to $\frac{1}{d_j} \left(\#\{k : k \in N_j(I)\} \right)$ (since the set $\{k : k \in N_j(E) \text{ and } s_k = \mathcal{B}\}$ is empty). Moreover, by the definition of those in I''' , they have at least c neighbours in E and so strictly less than $1 - c$ neighbours in I and so $U_j(\mathcal{I}, \mathbf{s}_{-j}) < 1 - c$. Given that $U_j(\mathcal{B}, \mathbf{s}_{-j}) = 1 - c$, the unique best response for all those in I''' is to become bilingual.

Given the behaviour of those in E and those in I''' , an almost identical argument to the above shows that $\mathcal{I}$ is the best response for those in I'' .

Lastly, it is immediate that those in I' remain monolingual, and so we are done. $\square$

C Data Sources and Variable Generation Procedures

The 1901 census of Ireland provides strong grounds for confidence that systematic self-reporting bias in the Irish-language question was limited. The census was taken on Sunday 31 March by the Royal Irish Constabulary, with the assistance of enumerators. A total of 3,941 enumerators assisted with the 1901 census, equating to an average of 1,097 people per enumerator, with no marked difference in predominantly Irish-speaking counties (1,040 on average).

Each enumerator was required to visit every house in their assigned district and was legally obliged to certify, under statutory declaration, that the information collected had been ‘truly and faithfully taken’ and was correct to the best of their knowledge. Failure by an enumerator to perform these duties incurred a fine of £5, while household members could also be fined £5 for refusing to answer or for willfully providing false information (with the sole exemption applying to religion). These penalties were explicitly stated in census advertisements.

The Irish-language question appeared directly on Form A, alongside demographic information such as age, occupation, birthplace, and religion. Respondents were given clear instructions to record either ‘Irish’ or ‘Irish and English’ where applicable, leaving the column blank otherwise. Both the head of household and the enumerator were required to sign the completed return, and enumerators were specifically instructed to assist those unable to complete the form themselves. Moreover, enumerators were empowered to question household members directly where clarification was needed.

Taken together, these procedures suggest that while some misreporting is possible, there is little reason to expect systematic under- or over-reporting of Irish-speaking ability, particularly one that would differ consistently across regions (British Parliamentary Papers, 1902, 589-640).

The General Register Office (GRO) undertook the administration of the census and the Irish GRO was held in high regard by the General Register Office of England & Wales (Higgs, 2004). The Irish census included several unique aspects compared to its English counterpart, including the recording of religion, the Irish language, and occupations.

Table C1: Data Sources and Variable Generation Procedures

Statistic	Description	Source
Irish Language Ability	A binary indicator assigned a value of one when the observation is either a bilingual or monolingual Irish speaker.	National Archives of Ireland (2019).
Female	A binary indicator assigned a value of one when the observation is female.	Same as above.
Born 1871-81	A binary indicator assigned a value of one if the observation's birth year falls between 1871 and 1881, corresponding to a census age of 20-29. The omitted group in our analysis are those born between 1861-71 or aged 30-39.	Same as above.
Population Decline 1841-51	Represents the population decline associated with the Famine, with aggregated values varying at the district electoral division level.	Kelly and Fotheringham (2011).
Population Decline 1851-71	Represents the population decline in post-Famine period, with aggregated values varying at the district electoral division level.	Same as above.
Linguistic Frontier Proximity	The inverse distance to districts where Irish-speaking fell below 50 per cent among pre-Famine birth cohorts (i.e., those aged over 60).	National Archives of Ireland (2019)
Market Access 1841	The sum of other locations' population, weighted by bilateral trade costs.	See Appendix D.
Market Access Rise 1841-81	Market access growth caused by the introduction of railways.	Same as above.
Irish Speakers (Household, Previous Generation)	The share of relatives (mostly parents) in the observation's household who are 40 years old and above who are Irish speaking (bilingual and monolingual).	National Archives of Ireland (2019).
Irish Speakers (Townland, Previous Generation)	The share of residents in the observation's townland who are 40 years old and above who are Irish speaking (bilingual and monolingual).	Same as above.
Irish Speakers (District, Previous Generation)	The share of residents in the observation's district electoral division who are 40 years old and above who are Irish speaking (bilingual and monolingual).	Same as above.
Irish Speakers (County, Previous Generation)	The share of residents in the observation's county who are 40 years old and above who are Irish speaking (bilingual and monolingual).	Same as above.
<i>Continued on next page</i>		

Table C1 continued from previous page

Statistic	Description	Source
Literate (Read and Write)	A binary indicator assigned a value of one when the observation can both read and write.	Same as above.
District Literacy (Pre-National Schools)	The share of residents with the ability to both read and write in 1841. These data were areally interpolated (Prenner and Revord, 2019) by superimposing civil parish polygons (the spatial units reporting literacy in 1841) over the district electoral division boundaries used in this application.	British Parliamentary Papers (1843).
District Literacy (National School Expansion)	The difference between the share of fully literate aged 60 and above (born in 1841 or before) and the pre-national school literacy variable at the district electoral division level. The vast majority of these individuals would have been taught in the national school system and therefore this difference represents the literacy boost stimulated by national schools.	British Parliamentary Papers (1843) and National Archives of Ireland (2019).
District Literacy (Post-Famine Increase)	The difference between the share of fully literate aged 20-39 (born 1861-1881) and the level of literacy amongst those 60 and above. Again, this variable is measured at the district electoral division level. This variable measures the contemporary level of literacy amongst each observation's peers. We omit each observation from each district's calculation. Thus, someone who is illiterate will live in a more literate district compared to a person in the same district who is literate.	National Archives of Ireland (2019).
Roman Catholic	Binary indicator that takes a unit value when the observation is a Roman Catholic.	Same as above.
Roman Catholic Name Index	An index capturing the likelihood of someone's religious affiliation based on their first name.	National Archives of Ireland (2019) and see Connor (2021).
Traditional Name Index	A complementary index to that outlined above that accounts for the historic prevalence and popularity of certain names.	Same as above.
Teachers per Capita (District)	The district population of teachers aged 40 and above.	National Archives of Ireland (2019).
<i>Continued on next page</i>		

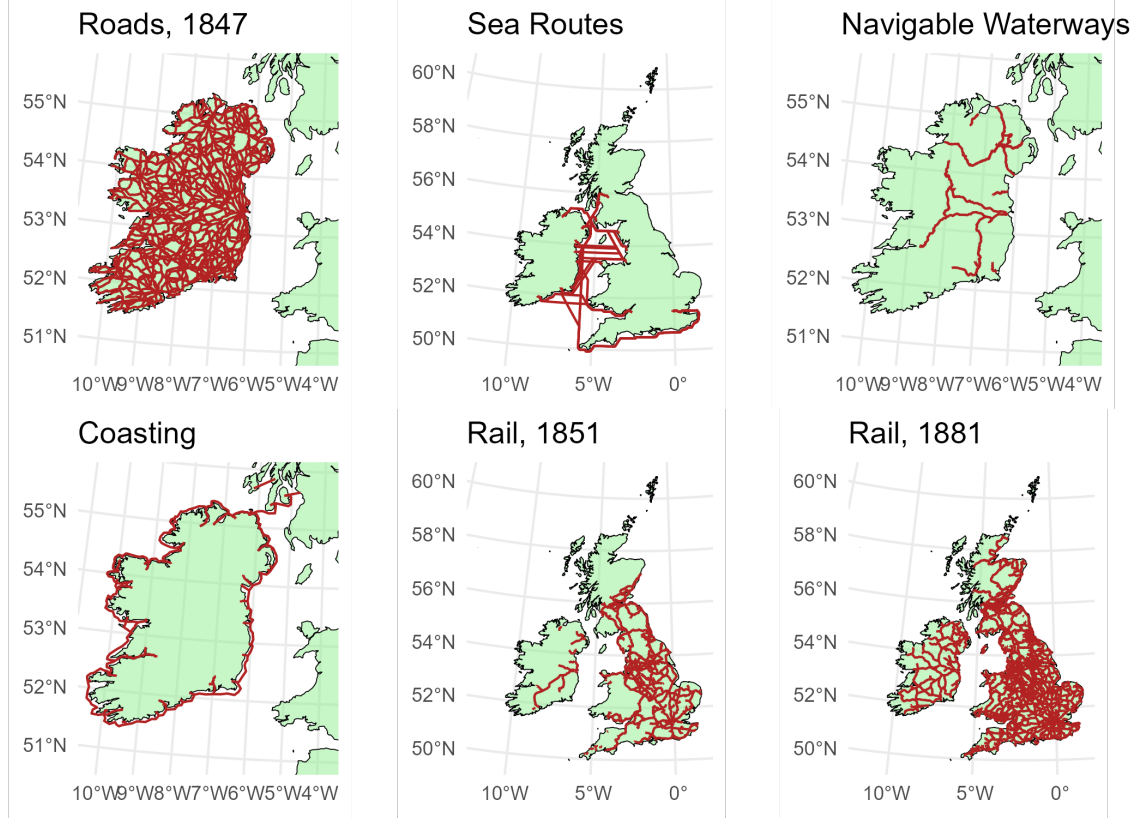
Table C1 continued from previous page

Statistic	Description	Source
Clergy per Capita (District)	The district population (both men and women, for all religions) aged 40 and above employed in religious occupations.	Same as above.
Police per Capita (District)	The district population aged 40 and above employed in the police force (RIC).	Same as above.
Resident in Birth County	Binary indicator that takes a unit value when the observation is resident in their county of birth. A measure of non-migrant status.	National Archives of Ireland (2019).
District Share Born in County	Share of district population who reside outside their county of birth.	Same as above.
Pre-Famine Land Value per Acre	Poor Law land values, a rough measure of annual land rent.	Fernihough and Ó Gráda (2022) and Fernihough and Lyons (2022).
Pre-National Schools per 1841 Population	Pre-National schools per district, adjusted for 1841 population.	Fernihough (2025)
Pre-National School Pupils per 1841 Population	Pre-National school pupils per district, adjusted for 1841 population.	Same as above.
National Schools per 1841 Population	National schools per district, adjusted for 1841 population.	Lougheed (2018).
National School Pupils per 1841 Population	National school pupils per district, adjusted for 1841 population.	Same as above.
Occupation (Irish Speaking)	The Irish-speaking share of the individual's profession.	National Archives of Ireland (2019).

D Market Access Construction

This paper follows a market access approach, requiring the calculation of bilateral transport costs between all district pairs. We construct a multimodal transport network, shown in Figure D1, comprised of roads, railways, inland waterways, and coastal and cross-channel shipping routes. Prior to the introduction of rail, travel occurs via road and water. The network incorporates Irish and British rail systems in 1851 and 1881, together with sea routes linking the two islands. Road data are taken from Lewis (1847), navigable rivers and railways from Great Britain (1907), coasting routes from Great Britain (1838a), steamship routes from Bradshaw (1850), and the British rail network from Satchell et al. (2018); Marti-Henneberg et al. (2018). Irish railway line openings follow Johnson (1998).

Figure D1: Transport Network in 19th Century Ireland



We define market access as in Fernihough and Lyons (2022), as the weighted cost of accessing other locations' population in 1851, integrating both Irish and British markets. Market access for district i at time t is given by:

$$MA_{it} = \sum_{j \neq i} \tau_{ijt}^{-\theta} N_j, \quad (6)$$

where N_j denotes population and τ_{ijt} bilateral transport costs. Trade costs follow an

Table D1: Transport speeds and costs by mode

Speeds:		
Mode	Speed (MPH)	Sources
Rail	25	Great Britain (1838b); Lee (1976).
Inland Waterway & Coasting	2.5	Great Britain (1838b).
Sea	12	Bradshaw (1850).
Road	7	Great Britain (1838b).
Costs:		
Mode	Cost (d per ton mile)	Sources
Rail	2	Great Britain (1838b); Lee (1976).
Inland Waterway & Coasting	2.5	Great Britain (1838b).
Sea	0.168	Bogart et al. (2022); Great Britain (1836).
Road	6	Great Britain (1838b).

iceberg formulation:

$$\tau_{ijt} = 1 + \frac{t_{ijt}}{\bar{P}}, \quad (7)$$

where t_{ijt} is the cost of moving one ton of freight between i and j , and $\bar{P}$ is the average value of a ton of freight Great Britain (1838a).

To construct bilateral costs t_{ijt} , we combine information on mode-specific freight prices and travel speeds. Costs are benchmarked to rail ($2d$ per ton-mile), with waterways and coastal shipping at $2.5d$, roads at $6d$, and cross-channel steamship transport at $0.168d$, plus a fixed port-handling charge of $22.9d$. Speeds differ substantially across modes, and we incorporate these differences through a composite cost measure:

$$C_m = 0.9 \frac{p_m}{p_{rail}} + 0.1 \frac{s_{rail}}{s_m}, \quad (8)$$

where p_m and s_m denote price and speed. All costs are expressed in rail-equivalent pence per ton-mile.

Bilateral costs are computed as least-cost paths over the network. District centroids are matched to the nearest network nodes, and paths are computed across all modes, including inland and cross-channel segments.

We compute market access using two network configurations. First, we construct a pre-rail measure based on the transport network as it stood before railways were introduced:

$$MA_i^{1841} = \sum_{j \neq i} \tau_{ij,1841}^{-\theta} N_j. \quad (9)$$

Second, we compute market access using the expanded network in 1881, incorporating

the Irish and British railway systems:

$$MA_i^{1881} = \sum_{j \neq i} \tau_{ij,1881}^{-\theta} N_j. \quad (10)$$

We set $\theta = 4$ throughout. Changes in market access reflect reductions in transport costs arising from railway expansion. We also hold population values (N_j) constant at 1851 levels, removing the possibility for market access to influence population.

E Comparison with the 1881 Census

The role of sample selection in our previous results is of concern. If those who left previously Irish-speaking districts were most likely to be Irish speakers, this means we are overestimating the proportion of Irish speakers. Of greater worry is that emigrants were systematically selected in a way that is correlated with our regressors of interest, creating bias in our marginal effect estimates.

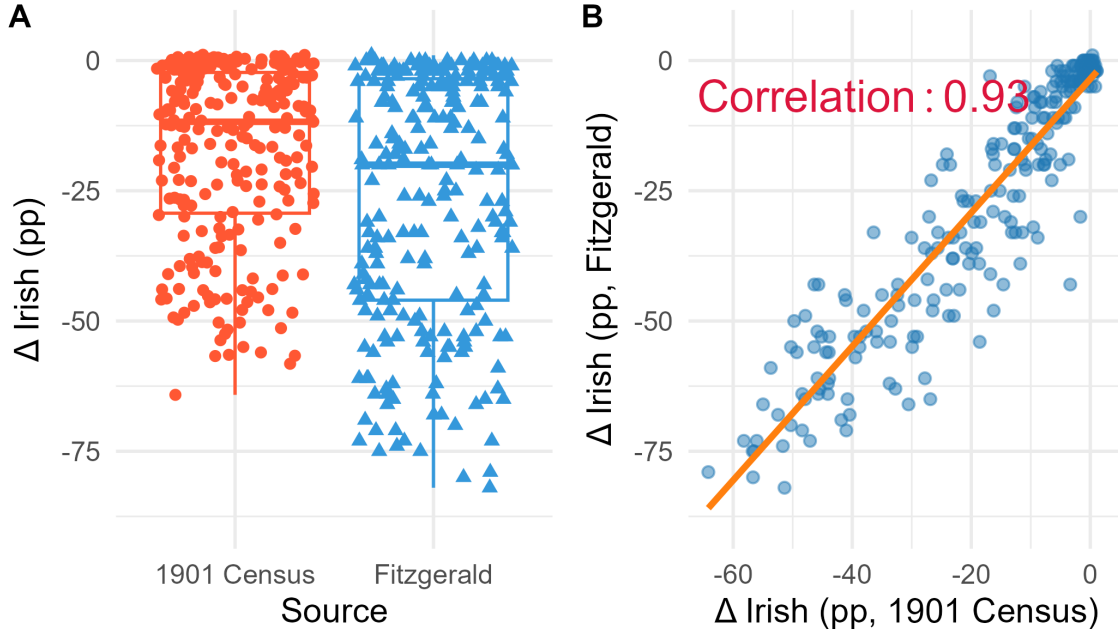
Garret FitzGerald used baronial-level data to estimate Irish speaking for decennial birth cohorts up to 1861-71 for 247 of Ireland’s 336 (74 per cent) baronies (FitzGerald, 1984). The omitted baronies were mostly in the east, with non-existent levels of Irish speaking in the nineteenth century. Whilst not directly comparable, we can reshape our source, the individual 1901 census returns, to match Fitzgerald’s data. By superimposing a baronial GIS shapefile, we can accurately approximate baronial-level data. Fitzgerald’s primary source was the 1881 census report and this means we do not have data for the 1871-81 cohort, used in our analysis.

To compare the sources, we create a differenced variable measuring the change in Irish speaking between birth cohorts from 1821-31 and 1861-71. The Fitzgerald source for measuring the decline in Irish speaking is less likely to suffer from migration sample selection bias because the latter birth cohort was only 10-19 years of age in his source and therefore less likely to have emigrated relative to those aged 30-39 in our 1901 census source.

Figure E1 provides a comparative synopsis of the 247 baronies for both sources. The box plot in panel A shows that our data underestimate the scale of the language shift. On average, Irish fell by 26 percentage points according to Fitzgerald, where we only see a 17.5 percentage point drop. This discrepancy is consistent with our data overestimating the degree of Irish speaking in post-Famine birth cohorts.

Panel B depicts the association between the two sources. This image is more reassuring as a very strong relationship connects the two sources. The estimated correlation coefficient of 0.93 points to an almost one-for-one relationship that implies that the 1901 data are a scaled-down version of the earlier 1881 data. While emigration reduces our estimates of the Irish language decline, these reductions are proportionate thus validating

Figure E1: Comparison with FitzGerald (1984)



our use of 1901 in regression analysis.

The consequences of using the 1901 source data can be further probed through regression. Table E1 contains the regression results for two model specifications that run parallel with both measures of the Irish language featuring as the dependent variable. To match our earlier models, we trim the sample and focus on the 95 baronies where the majority share has Irish ability in the pre-Famine period. Again, we have rescaled all of the explanatory variables to represent 2SD changes.

Table E1 presents a comparison between the 1901-based estimates (columns 1–2) and those using Fitzgerald’s earlier source (columns 3–4). The key result is that the coefficients are broadly consistent across data sources, which weakens concerns that sample selection drives the main findings.

Across specifications, the patterns are stable. With all regressors scaled to two standard deviations, coefficient magnitudes are directly comparable. Frontier proximity and education emerge as the dominant forces, while market access and demographic change play a more limited role. Higher pre-national school literacy is consistently associated with lower subsequent Irish-speaking, and the literacy gains linked to national schools also have a sizeable negative effect. In this table, the national school effect appears somewhat larger than the pre-national literacy effect, in contrast to the main results; however, given the smaller sample and larger standard errors, this is not conclusive.

Contemporaneous (post-Famine) literacy growth produces mixed results. It is negative but statistically insignificant in the Fitzgerald data, and positive in the 1901-based estimates. This remains consistent with an intergenerational interpretation: earlier literacy improvements shape later language outcomes, while contemporaneous changes

Table E1: Baronial OLS Regressions Explaining Irish Language Ability in 1871 (pp)

	Source: 1901 Census		Source: Fitzgerald	
	(1)	(2)	(3)	(4)
Linguistic Frontier Proximity	−3.595*** (1.323)	−4.078*** (1.364)	−3.009** (1.252)	−2.977** (1.359)
Literacy: Pre-National School	−1.527*** (0.368)	−1.957*** (0.385)	−2.005*** (0.389)	−2.261*** (0.446)
Literacy: National School Rise	−4.404*** (1.314)	−3.236** (1.312)	−5.184*** (1.412)	−4.942*** (1.351)
Literacy: Post-Famine Rise	1.465** (0.677)	1.501* (0.912)	−0.778 (0.794)	−0.963 (0.996)
Population Decline: 1841-51		−0.786 (0.774)		−0.889 (0.822)
Population Decline, 1851-71		0.209 (0.961)		0.464 (1.046)
Market Access: 1841		−1.936** (0.853)		−0.496 (0.907)
Market Access Rise: 1841-81		2.552*** (0.867)		2.502** (1.031)
Average Irish Speaking, 1871	53.269	53.269	36.926	36.926
Observations	95	95	95	95

Notes: Columns (1)–(2) and (3)–(4) report regressions of baronial shares of Irish speakers within the 1861–71 birth cohort, using our dataset and Fitzgerald’s dataset, respectively. The dependent variable is expressed in percentage points. All explanatory variables are standardised; a one-unit change corresponds to a shift of two standard deviations. Robust (heteroskedasticity-consistent) standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

have limited immediate impact within these cohorts.

Proximity to the linguistic frontier is negative and statistically significant throughout, with coefficients in the range of -3 to -4 , making it one of the largest effects in all specifications. The market access results are less uniform. Higher initial (1841) market access is associated with greater Irish language loss, in line with the main results. However, changes in market access—driven by railway expansion—enter positively, implying a slower relative decline in Irish-speaking in areas experiencing improved access.

One explanation for this sign reversal is the coarser spatial resolution of baronial data. Aggregation compresses within-area variation in railway exposure and blends treated and untreated locations, which can attenuate estimates and alter signs when effects differ across space. In particular, large access improvements often occurred in more peripheral, initially Irish-speaking areas. At the baronial level, this can mechanically associate access gains with higher Irish-speaking shares, even if finer-grained data would reveal that improved access facilitated language shift.

F Selective Migration and Synthetic Cohorts

In Appendix E, we found that selective migration overestimates the prevalence of Irish speaking in our analysis. Migration was a key feature of nineteenth-century Irish demography, and it appears that those who left were less likely to speak Irish. What would our results look like if we included those who left? Given that our analysis emphasises explanatory factors that differ at a spatially aggregated level (by district), and that we have accurate measures of cohort depletion in these areas, we can generate synthetic cohorts reflecting the profiles of emigrants. Appending these observations to our data allows us to probe the robustness of our main results.

We produce these synthetic cohorts by the following method. We aggregate our data up to the district level. Armed with population information from the decennial censuses, we construct two estimates of cohort depletion. The first assumes that the year-of-birth cohorts for each district can be approximated by taking the current cohort present in 1901 and scaling it up so that the true population is the current population plus a mark-up equal to the size of the cohort depletion occurring between late childhood and early adulthood. For example, the Graigabbey district in Co. Galway had 23 observations in our dataset born between 1861 and 1871. The population in the district dropped from 383 to 348 between the 1881 and 1891 censuses, a fall of around 9 per cent. Extrapolating this loss to our sample size implies two missing data points ($-23 \times -0.09 \approx 2$).

The second estimate, less conservatively, boosts the population to counteract all post-1841 population loss. In the small number of districts where the population increases (5.8% of the pre-Famine majority Irish-speaking districts with the first method and 2.6% with the second), no synthetic observations are generated. After ascertaining the size of these cohorts, we simulate Irish-speaking ability. We account for the possibility that Irish-speaking was lower among those who left by using the average Irish-speaking ability of those with the same county of birth as the districts, but who resided elsewhere. For example, 74 per cent of Galway’s residents born in the county were Irish speakers, compared to 28 per cent of those born in Galway but resident elsewhere. Once we attach a likely “migrant speaking” probability to each county of birth, we use this statistic to simulate draws that match this likelihood. So in the case of Graigabbey, we generate observations where Irish-speaking ability is a 0-1 indicator, where $Pr(Irish = 1) = 0.28$.

The results of this exercise are reported in Table F1. Our primary focus is on education, particularly variables that map directly onto district-level information. Since these district-level factors are fixed by construction, they do not require simulation. Column (1) reproduces the baseline specification, indicating that a two standard deviation increase in pre-national school literacy reduces the probability of Irish-speaking by 34 percentage points. The estimated marginal effects for national and post-national schooling are consistent with those reported in the main text. We also recover the first- and second-

Table F1: Probit regression marginal effects with synthetic cohorts

	Irish Language Ability		
	(1)	(2)	(3)
Year of Birth: 1871-81	-0.159*** (0.013)	-0.137*** (0.014)	-0.113*** (0.011)
Linguistic Frontier Proximity	-0.261*** (0.040)	-0.240*** (0.037)	-0.197*** (0.027)
District Literacy: Pre-National School	-0.339*** (0.069)	-0.294*** (0.059)	-0.224*** (0.039)
District Literacy: National School Rise	-0.214*** (0.049)	-0.185*** (0.040)	-0.162*** (0.030)
District Literacy: Post-Famine Rise	-0.093** (0.038)	-0.075** (0.030)	-0.080*** (0.021)
Market Access: 1841	-0.236* (0.127)	-0.209* (0.114)	-0.174** (0.080)
Market Access Rise: 1841-81	-0.252** (0.120)	-0.208** (0.104)	-0.148** (0.069)
Average Irish Speaking	0.496	0.467	0.402
McFadden's R-Squared	0.213	0.182	0.135
Observations	290,440	321,713	429,472
Synthetic Observations	0	31,273	139,032

Notes: Columns (1) to (3) report marginal effects from probit regressions where the dependent variable is an indicator for Irish language ability. Marginal effects are reported for a two-standard-deviation change in continuous explanatory variables. Column (1) uses observed individuals only. Columns (2) and (3) progressively augment the sample with synthetic observations constructed to represent earlier birth cohorts, increasing effective cohort coverage while preserving the underlying covariate distributions. The row “Synthetic Observations” reports the number of such constructed observations included in each specification. Standard errors are cluster-robust and clustered at the county level. Statistical significance is indicated by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

nature geography results for the linguistic frontier and market access variables.

Columns (2) and (3) incorporate the synthetic cohorts. Emigrants were less likely to speak Irish, implying that selective attrition biases upward the observed mean level of Irish-speaking ability. This is reflected in both supplementary samples: average Irish-speaking declines from 50 to 47, and then to 40 per cent. The magnitude of the marginal effects attenuates, although only modestly. Under the counterfactual assumption of no population loss following the Famine, the effect of pre-national schooling declines from 34 to 22 percentage points. By contrast, the more plausible birth-cohort depletion scenario reported in column (3) yields a 30-percentage point effect. These results indicate that even under extreme assumptions, the estimated relationships remain stable, and the inclusion of synthetic cohorts leaves the main findings substantively unchanged.

Appendix References

- Akenson, D. H. (1970). *The Irish Education Experiment: The National System of Education in the Nineteenth Century*. London: Routledge & K. Paul.
- Barr, C. (2009). MacHale, John. In *Dictionary of Irish Biography*. Dublin: Royal Irish Academy. URL: <https://doi.org/10.3318/dib.005220.v1>.
- Blum, M., C. L. Colvin, M. Laura, and E. McLaughlin (2017). Women of an uncertain age: Quantifying human capital accumulation in rural Ireland in the nineteenth century. *The Economic History Review* 70(1), 187–223.
- Bogart, D., E. Alvarez-Palau, M. Satchell, and L. Shaw-Taylor (2022). Multi-modal models for pre-steam English and Welsh freight transportation systems. PDF file.
- Bradshaw, G. (1850). *Bradshaw's Monthly Railway and Steam Navigation Guide for Great Britain and Ireland, 1850: March – No. 200*. Reprinted by Phoenix Print (Devon) Ltd., Devon.
- British Parliamentary Papers (1825). *The Select Committee of the House of Lords appointed to inquire into the state of Ireland, more particularly with reference to the circumstances which may have led to disturbances in that part of the United Kingdom*, Volume (521). London: HMSO: House of Lords.
- British Parliamentary Papers (1843). *Report of the Commissioners Appointed to Take the Census of Ireland for the Year 1841*. London: HMSO.
- British Parliamentary Papers (1851). *The Seventeenth Report of the Commissioners of National Education in Ireland (For the Year 1850.) Vol. I*, Volume 2304. London: HMSO: House of Commons. [1405] [1405-II], H.C., pp 101, 641.
- British Parliamentary Papers (1900). *The Sixty-Sixth Report of the Commissioners of National Education in Ireland, Year 1899-1900*. London: HMSO. [Cd.285] [Cd.286] [Cd.287] [Cd. 288] [Cd.289], H.C., pp xxiii, 189, 251, 455, 585, 733.
- British Parliamentary Papers (1902). *Census of Ireland, 1901: General Report*. London: HMSO: House of Commons. [Cd. 1190].
- Colvin, C. L., S. Henderson, and E. McLaughlin (2025). Age structure and age heaping: Solving Ireland's post-Famine digit preference puzzle. *European Review of Economic History* 29(1), 28–47.
- Comerford, R. V. (2003). *Ireland: Inventing the Nation*. London: Bloomsbury.
- Connor, D. S. (2021). In the name of the father? Fertility, religion, and child naming in the demographic transition. *Demography* 58(5), 1793–1815.
- Coolahan, J. (1981). *Irish Education: Its History and Structure*. Dublin: Institute of Public Administration.
- Coolahan, J. (2001). Education as cultural imperialism: The denial of the Irish language to Irish speakers, 1831-1922. *Paedagogica Historica* 37(1), 17–33.
- Corcoran, T. (1925). The Irish language in the Irish schools. *Studies: An Irish Quarterly Review* 14(55), 377–388.
- Corrigan, K. P. (2003). The ideology of nationalism and its impact on accounts of language shift in nineteenth century Ireland. *AAA: Arbeiten aus Anglistik und Amerikanistik* 28(2), 201–229.
- Cullen, L. M. (1990). Patrons, teachers and literacy in Irish: 1700–1850. In M. Daly and D. Dickson (Eds.), *The Origins of Popular Literacy in Ireland: Language Change and Educational Development*

- 1700–1920, pp. 15–44. Dublin.
- Doyle, A. (2015). *A History of the Irish Language: From the Norman Invasion to Independence*. Oxford: Oxford University Press.
- Fernihough, A. (2025). Ireland’s educational foundations: Digitising and mapping the 1826 Education Inquiry. Digital Repository of Ireland. Collection. Accessed: 2026-03-20.
- Fernihough, A. and R. C. Lyons (2022). Agglomeration and emigration: The economic impact of railways in post-Famine Ireland. QUCEH Working Paper Series, Paper No. 22-11. URL: <https://www.quceh.org.uk/uploads/1/0/5/5/10558478/wp22-11.pdf>.
- Fernihough, A. and C. Ó Gráda (2022). Population and poverty in Ireland on the eve of the Great Famine. *Demography* 59(5), 1607–1630.
- FitzGerald, G. (1984). Estimates for baronies of minimum level of Irish-speaking amongst successive decennial cohorts: 1771–1781 to 1861–1871. *Proceedings of the Royal Irish Academy. Section C: Archaeology, Celtic Studies, History, Linguistics, Literature* 84C, 117–155.
- FitzGerald, G. (2013). Irish primary education in 1824. In G. FitzGerald (Ed.), *Irish Primary Education in the Early Nineteenth Century: An Analysis of the First and Second Reports of the Commissioners of Irish Education Inquiry, 1825–6*. Dublin: Royal Irish Academy.
- Fitzpatrick, D. (1986). ‘A share of the honeycomb’: Education, emigration and Irishwomen. *Continuity and Change* 1(2), 217–234.
- Great Britain (1836). Select Committee on State of Coal Trade: Report, Minutes of Evidence, Appendix. Parliamentary papers, vol. 11, London, UK: HMSO.
- Great Britain (1838a). General System of Railways for Ireland: Second Report, Appendices and Index. Parliamentary Papers, Vol. 35 35, London, UK: HMSO.
- Great Britain (1838b). General System of Railways for Ireland: Second Report, Appendices and Index. Parliamentary Papers, Vol. 35 35, London, UK: HMSO.
- Great Britain (1907). Vice-Regal Commission on Irish Railways, including Light Railways. Appendix to the First Report. Parliamentary Papers, Vol. 37 35, London, UK: HMSO.
- Hatton, T. J. and J. G. Williamson (1994). International migration 1850–1939: An economic survey. In T. J. Hatton and J. G. Williamson (Eds.), *Migration and the International Labour Market 1850–1939*. London: Taylor & Francis.
- Heaney, T. (2026). *Essays in Irish Economic Development*. PhD thesis, Queen’s University Belfast, Belfast.
- Higgs, E. H. (2004). *Life, Death and Statistics: Civil Registrtaion, censuses and the work of the General Register Office, 1836-1952*. Amsterdam: Halstan Co.
- Jackson, A. (1999). *Ireland 1798–1998: Politics and War*. Oxford: Blackwell Publishing.
- Johnson, S. (1998). *Johnson’s Atlas and Gazetteer of the Railways of Ireland*. Midland Publishing.
- Kelly, M. and A. S. Fotheringham (2011). The online atlas of Irish population change 1841–2002: A new resource for analysing national trends and local variations in Irish population dynamics. *Irish Geography* 44(2-3), 215–244.
- Lee, J. (1976). The Railways in the Irish Economy. In L. M. Cullen (Ed.), *The Formation of the Irish Economy*, Chapter 6, pp. 77–87. Cork, IR: Mercier Press.

- Lee, J. J. (1989). *Ireland 1912–1985: Politics and Society*. Cambridge: Cambridge University Press.
- Lewis, S. (1847). *A Topographical Dictionary of Ireland : Comprising the Several Counties; Cities; Boroughs; Corporate, Market, and Post Towns; Parishes and Principal Villages; With Historical and Statistical Descriptions: Embellished With Engravings of the Arms of the Cities, Bishoprics, Corporate Towns, and Boroughs; And of the Seals of the Several Municipal Corporations*, Volume 2. S. Lewis and Company.
- Lougheed, K. (2018). After the manner of the Irish schools: The influence of Irish national education in the British Empire. *Journal of Historical Geography* 60, 1–10.
- Mac Giolla Chríost, D. (2005). *The Irish Language in Ireland: From Góidel to Globalisation*. Abingdon: Routledge.
- Marti-Henneberg, J., M. Satchell, X. You, L. Shaw-Taylor, and E. Wrigley (2018). 1881 England, Wales and Scotland Rail Lines. Data set.
- McLaughlin, E., C. L. Colvin, and M. Blum (2021, 12). Anthropometric history: Revisiting what’s in it for Ireland. *Irish Economic and Social History* 48(1), 3–26.
- McMahon, T. M. (2008). *Grand Opportunity: The Gaelic Revival and Irish Society, 1893–1910*. Syracuse, New York: Syracuse University Press.
- McManus, A. (2002). *The Irish Hedge School and its Books, 1695–1831*. Dublin: Four Courts.
- Mulholland, M. (2025). *At the Rising of the Moon: The Peasantry and Ireland from the Tudor Conquest to the Fall of Landlordism*. Dublin: Eastwood Books.
- Murphy, G. (1948). The Gaelic background. In M. Tierney (Ed.), *Daniel O’Connell: Nine Centenary Essays*. Dublin: Browne and Nolan Limited.
- National Archives of Ireland (2019). Census of Ireland 1901/1911 and census fragments and substitutes, 1821–51. [Data set]. Census years: 1901. URL: <http://census.nationalarchives.ie/pages/1901/>.
- Nic Craith, M. (2002). ‘The tiger no longer speaks Celtic’: Economic conditioning and the Irish language. In U. Kockel (Ed.), *Culture and Economy: Contemporary Perspectives*, pp. 175–95. Aldershot: Ashgate.
- Ó Ciosáin, N. (2024). *Print and the Celtic Languages: Publishing and Reading in Irish, Welsh, Gaelic and Breton, 1700–1900*. Abingdon: Routledge.
- Ó Cuív, B. (1969). Irish in the modern world. In Ó Cuív, B. (Ed.), *A View on the Irish Language*. Dublin: The Stationary Office.
- Ó Gráda, C. (2013). School attendance and literacy in Ireland before the Great Famine: A simple baronial analysis. In G. FitzGerald (Ed.), *Irish Primary Education in the Early Nineteenth Century: An Analysis of the First and Second Reports of the Commissioners of Irish Education Inquiry, 1825–6*. Dublin: Royal Irish Academy.
- Ó Gráda, C. (1995). *Ireland: A New Economic History 1780–1939*. Oxford: Oxford University Press.
- Ó Murchadha, C. (2011). *The Great Famine: Ireland’s Agony 1845–1852*. London: Bloomsbury.
- Ó Tuathaigh, G. (1974). Gaelic Ireland, popular politics and Daniel O’Connell. *Journal of the Galway Archaeological and Historical Society* 34, 21–34.
- O’Brien, G. (1989). The strange death of the Irish language, 1780–1800. In G. O’Brien (Ed.), *Parliament, Politics and People: Essays in Eighteenth-Century Irish History*, pp. 149–170. Dublin: Irish Academic Press.

- Prener, C. and C. Revord (2019). areal: An R package for areal weighted interpolation. *Journal of Open Source Software* 4(37).
- Ravenstein, E. G. (1879). On the Celtic languages in the British Isles: A statistical survey. *Journal of the Statistical Society of London* 42(3), 579–636.
- Satchell, M., E. Wrigley, L. Shaw-Taylor, X. You, and J. Henneberg (2018). 1851 England, Wales and Scotland Rail Lines. Data set.
- Wall, M. (1969). The decline of the Irish language. In Ó Cuív, B. (Ed.), *A View of the Irish Language*. Dublin: The Stationary Office.
- Walsh, J. (2011). *Contests and Contexts: The Irish Language and Ireland’s Socio-Economic Development*. Bern: Peter Lang.
- Walsh, J. (2022). *One Hundred Years of Irish Language Policy, 1922–2022*. Oxford: Peter Lang.
- Williams, J. E. C. and P. K. Ford (1992). *The Irish Literary Tradition*. Cardiff: University of Wales Press.
- Wolf, N. M. (2014). *An Irish-Speaking Island: State, Religion, Community, and Linguistic Landscape in Ireland, 1770–1870*. Maddison, Wisconsin: University of Wisconsin Press.