Preposition drop and ditransitives in British English:

a corpus and survey approach

JONATHAN STEVENSON

PHD

University of York

Language and Linguistic Science

July 2024

Abstract

The present thesis maps geographical variation in the relative frequency of variants of two syntactic variables: ditransitives [DATALT] and preposition dropping [P/D-DROP]. Several data sources are used: large scale geolocated Twitter corpora, combined with traditional corpora.

These data are analysed in the context of models of diffusion, and patterns of historical migration. Comparing the present-day distribution to historical linguistic data, migration data and social history, allows for the estimation of the likely time periods of variant innovation and diffusion. Similarly, understanding the present-day geographical variation of linguistic features also provides a window into historical migration patterns.

Using mapped data, it is possible to find sharp boundaries, where the relative rates of variants change abruptly over the space of a few miles, indicating multiple transition zones. These boundaries are shown to correspond to other linguistic features. Most importantly, the isoglosses of various manifestations of *definite article reduction* [DAR], found in the Survey of English Dialects, are shown to correspond tightly to northern variants of [P/D-DROP], suggesting a relationship that has hitherto not been made. The data additionally provide quantitative evidence to support the diagnostics used in contemporary syntactic studies of the variables in question, and is able to locate, the likely place where a particular grammar should exist.

The linguistic boundaries also provide numerous opportunities for further investigation, using additional methodologies. In the present work, a further study is carried out in the North West of England, using a syntactic judgement survey, delivered through sixth form colleges in the region. The survey data are used to triangulate the Twitter data and to test for the existence of the grammar predicted from the mapped usage data.

Contents

List of Figures

8

9

14

15

16

List of Tables

Acknowledgements

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

"It is the task of sociolinguistics to describe and explain the patterns of variation that occur within a linguistic community given the theoretical limits of this variation uncovered by generative linguistics" (Barbiers, 2005, p.235)

CHAPTER **1**

Introduction

1.1 Introduction

The present thesis takes a combined approach to the study of dialect varia-
tion in British English, with an ultimate focus on the North West of England.
The analysis centres on the distribution of two syntactic variables: the *dative
alternation* [DATALT] and *preposition-determiner dropping* [P/D-DROP]; and one
phonological variable: *definite article reduction* (DAR). These variables are in-
troduced in §1.2.

Several data sources are employed: for the UK and Ireland, a geolocated
Twitter corpus (where social media posts are located based on users' phone lo-
cation) is used alongside traditional corpora. In the North West, a second Twit-
ter corpus—which uses an alternative method for assigning user-location—
is used in combination with a grammaticality judgement survey, delivered
through sixth-form colleges across the region.[1]

The geographic distribution and structural properties of the two syntac-
tic variables have been the subject of recent inquiry. The present work sup-
plements this inquiry by providing greater geographical resolution, historical
depth, and structural detail. Regions and their boundaries are identified based

---

[1] Sixth form colleges are equivalent to K12 in the US system.

not simply on the distribution of individual variants but by the clustering of the relative frequencies of each variant in each location.

Geographic variability in the distribution of [P/D-DROP] is set alongside the established literature on concurrent processes of DAR. A connection between [P/D-DROP] and DAR has, until now, not been explicitly made. The present thesis argues that at least some variants of DAR are directly linked to variants of [P/D-DROP], while other DAR variants additionally provided the conduit into which one particular [P/D-DROP] variant diffused.[2]

The linguistic variables that will be the subject of analysis are outlined in §1.2 and discussed in detail in Chapter 2. Following this, I provide a geographic overview of the North West region, which is the ultimate focus of the project, as well as the South East, which also plays an important part in the story. Finally, I provide an orientation to the present work and where it sits theoretically and methodologically between distinct linguistic subfields.

1.2 Background to variables

The dative alternation [DATALT]

The dative alternation [DATALT] has been much discussed in the literature. It refers to the availability of two word orders that may occur with a range of ditransitive verbs—such as *give, send* and *lend*—and which are characterised principally as involving *transfer of possession* (Levin, 1993).[3] Ditransitive verbs are so-called because, along with the subject, they select two objects, usually referred to as the THEME and the GOAL.[4] The GOAL is the object to which the THEME object is being *given* or *sent*.

(1)　　[DATALT]

　　　　a.　Susan gave the book to Mary　　　　*prepositional dative* (PDAT)
　　　　b.　Susan gave Mary the book　　　　*GOAL-THEME ditransitive* (GTD)

In the above examples, (1a) is referred to as the prepositional dative (PDAT).

---

[2]The broad nature of the work set out in this document means that it touches upon many theoretical questions that emerge from the data presented. The goal of this thesis is to shed light on some of these questions and set the conditions for future work. A full theoretical account of the phenomena is beyond the scope of the present work.

[3]Alternating verbs may also include additional characteristics, such as *motion towards a GOAL*, as in *send* or *bring*, or variations on *possession transfer*, such as *denial of possession*, in *refuse*. Variability in the extent to which such characteristics play into the ability of a given verb to alternate are important considerations, as we will see.

[4]THEME and GOAL are also referred to as *indirect* and *direct* objects, respectively.

Here, the THEME is *the book* and the GOAL is *Mary*. With prepositional datives, the THEME precedes the GOAL which is introduced by a preposition. In the second example (1b), referred to in this thesis as a GOAL-THEME ditransitive (GTD), the order of the objects is reversed. Here the GOAL occurs without a preposition and precedes the THEME.[5]

Where both objects are pronouns, an additional third option (2c) is available and used preferentially by a majority of speakers in the Midlands and North West of England (Gerwin, 2014; Haddican, 2010; Siewierska & Hollmann, 2007; Stevenson, 2016; Yáñez-Bouza & Denison, 2015).

(2)   [DatAlt]-pronominal

    a.   Susan gave it to her   *pronominal prepositional dative* ($_p$PDAT)

    b.   Susan gave her it   *pronominal GOAL-THEME ditransitive* ($_p$GTD)

    c.   Susan gave it her   *pronominal THEME-GOAL ditransitive* ($_p$TGD)

This third option has come under some scrutiny in recent literature (Biggs, 2018; Haddican, 2010; Myler, 2013). The question on the table asks whether $_p$TGD (_*it me*) is underlyingly $_p$PDAT (_*it to me*) (2a) with a dropped preposition or whether it is akin to $_p$GTD (_*me it*) with a reversed object order.

(3)   a.   Susan gave it [TO] her

    b.   Susan gave it ←→ her

The answer hinges, in part, on the availability, or not, of the TGD with full-DP complements (4).

(4)   TGDs with DP complements

    a.   Susan gave it the cat   *pronoun-determiner phrase TGD* ($_{PDP}$TGD)

    b.   Susan gave the letter the bank   *2-determiner phrase TGD* ($_{DPDP}$TGD)

For the majority of those speakers who accept pronominal TGDs (_*it me*), the sentences in (4) range from OK to fully ungrammatical. This cline in acceptability, along with a range of other diagnostics, leads Haddican (2010) to conclude that, for most speakers in his Manchester sample, $_p$TGDs (_*it me*) are underlyingly akin to $_p$GTDs (_*me it*), rather than $_p$PDAT (_*it to me*) with an

---

[5]What is referred to as GOAL-THEME ditransitives (GTD) in this thesis is also commonly referred to as the *double object construction* in other work.

elided preposition.[6] The rationale, briefly, is that if it were an elided preposition, there should be no obvious reason why the preposition could not be elided with two full DPs.

However, a subset of speakers does accept both of the sentences in (4). Haddican thus reports that there is some degree of inter-speaker variation in Manchester: for those speakers, THEME-GOAL ditransitives with two full DP complements (and by extension those with only pronominal complements) likely do involve a mechanism that allows for preposition drop.

Further, the availability of THEME-GOAL ditransitives with full DPs appears to correlate with the availability of another instance of preposition dropping evident in the North West: (in)transitive and ditransitive verbs (*come, go, throw, send*) which involve *motion towards a directional GOAL* (5) (Biggs, 2018; Haddican, 2010; Myler, 2013).

(5)   Preposition drop with directional verbs (available in the North West)

       a.   Susan went [TO] the pub   *preposition-drop* (P-DROP)

       b.   John took his bag [TO] the park   ditransitive P-DROP

       c.   John brought the boys [TO] the game   ditransitive P-DROP

       d.   Jay sent the cat the vet   ditransitive P-DROP

Biggs (2018) extends this observation, finding that a majority of speakers in Liverpool also accept P-DROP with ditransitive verbs that involve *transfer of possession*, as in (4b) above.[7] This, she observes, coincides with widespread speaker acceptance of P-DROP with *directional verbs* shown in (5).

Where both complements are full DPs, THEME-GOAL ditransitives (4b) involving *transfer of possession* (e.g. *give*) and *ditransitive* P-DROP (5b)-(5c) involving *motion→GOAL* appear similar. However, the distinction between *transfer of possession* and *motion towards a directional* has structural consequences (Myler, 2013).[8] Namely, *motion→GOAL* ditransitives are standardly acceptable only with a preposition present and the GOAL is limited to places or destinations. Additionally, while an inanimate THEME is possible with *ditransitive* P-DROP

---

[6]The full range of diagnostics are presented in the next chapter. Here, for the purpose of brevity, I present the most salient one.

[7]It is not clear from Biggs (2018) the extent to which there may be inter-speaker variation in Liverpool as this is not made explicit, though older respondents are reported as patterning more like Haddican's Manchester set, indicating that this may be a recent innovation.

[8]Motion here usually means the act of physically moving towards a location, but the present data reveals it is also possible abstractly, as expressed in examples such as *I have come the conclusion* or *I went straight the character death scene*.

(5b), it is more frequently animate (5c).

The boundary between ditransitive P-DROP and TGD is not always so clear. The verb *send* carries with it both a *transfer of possession* and a *motion to GOAL* reading. The animacy of the GOAL is also crucial for a *transfer of possession reading,* which makes sense, given that we do not generally think of inanimate objects as being able to possess anything.[9] These facts are apparent in examples such as:

(6)   *Send*

    a.   I sent the kids [TO] the shop (*motion to GOAL*)

    b.   I sent the contract [TO] the landlord (*transfer of possession*)

Whilst the primary implication of (6b) is that of *transfer of possession*, there is an additional implied sense of *motion* imparted on the THEME (the contract) towards the GOAL (the landlord). (6a), however, implies a sense of *motion* of the THEME (the kids) towards the GOAL (the shop), with an additional sense of a command issued by the subject, and crucially no transfer of possession. Thus, *send* presents ambiguities that allow it to occur in either structural grouping. We can immediately see this structural distinction between (6a) and (6b) when we attempt to invert the THEME and GOAL objects.

(7)   a.   *I sent the shop [TO] the kids

    b.   *I sent the landlord [TO] the contract

    c.   I sent the landlord the contract

These distinctions have been much discussed in the literature (cf. Bresnan & Nikitina, 2009). The goal of the present work is not so much to further the formal theoretical discussion (though this is intended in separate work), the goal is, rather, to show how these constraints play into, and may explain, some of the geographic variability evident in the data.

Crucially, as the data presented will show, the existence of *ditransitive* P-DROP does not, on its own, predict the availability of $_{DPDP}$TGD: high frequency use of the pronominal form of the THEME-GOAL ditransitive ($_p$TGD) is also required. In the North West, we find that it is this *motion*, or *spatial*, reading of THEME-GOAL ditransitives with full DPs that characterises P-DROP phenomena.

---

[9]Some exceptions to this might be *giving a car a lick of new paint.* There is some sense here that the car is in possession of its paint. The non-categoricity of how verbs are used in natural language production is explicated in Bresnan and Nikitina (2009)

The non-spatial, *transfer of possession* reading—as with the verb *give*—is only available for $_{\text{DPDP}}$TGDs where speakers also have access to $_{\text{P}}$TGDs.

Finally, TGDs are available to passivise for some speakers. Haddican and Holmberg (2012) show that THEME-passives (8) are available to speakers in their Manchester data only if they also accept TGDs in active contexts, which they argue should hold more generally.

(8)　　a.　It was given him
　　　　b.　The letter was given the bank (by Susan)

While Haddican and Holmberg (2012) show that there is a correlation in acceptance between TPASS (8b) and TGDs, indicating the availability of a grammar in the North West that possesses *passive symmetry*, it is specifically the availability of TPASS with derived DP subjects that Biggs (2018) claims implicates P-DROP.[10]

Preposition and determiner dropping [P/D-DROP]

The phenomenon of P-DROP can be viewed as part of a four-way variable [P/D-DROP] with three other logical variants in British English, all of which are possible.

(9)　　[P/D-DROP]

　　　　a.　Susan went to the pub (NO-DROP)
　　　　b.　Susan went the pub (P-DROP)
　　　　c.　Susan went pub (PD-DROP)
　　　　d.　Susan went to pub (D-DROP)

The variants of [P/D-DROP] (9), have been reported as showing substantial geographic variation across England. While intransitive P-DROP (*go the N*) is documented in the North West (Biggs, 2018; Myler, 2013), PD-DROP (*go N*) is reported in the South-East (Bailey, 2018b; Hall, 2019), with additional reports of its availability in Manchester (Bailey, 2018b). Between Bailey (2018b) and Hall (2019), there is some disagreement as to whether the absent preposition and determiner remain structurally present, or whether they are 'radically ab-

---

[10]Passive symmetry means that both THEME and GOAL objects in a ditransitive are able to passivise, a feature of a number of other languages, such as Norwegian, Kinyarwanda and Kinande. This contrasts to most varieties of English which are classified as asymmetrically passive in that only GOAL arguments may passivise (Haddican & Holmberg, 2012).

sent', with the preposition incorporating into the verb. More recently, Gopal et al. (2021) and Stevenson (2021) have independently shown using Twitter corpora, that PD-DROP is in fact widespread in England, with high rates found across the Midlands and North West. Gopal et al. (2021) additionally show, by comparing the present-day distribution to simulated models, that PD-DROP appears, quite convincingly, to have undergone a recent diffusion from London towards the North West. This diffusion, Gopal et al. (2021) show, follows a *gravity model* (cf. Trudgill, 1974), with the variant occurring earlier and with greater relative frequency in large settlements than in less populated intermediate areas.

However, the picture is likely more nuanced and complex than a relatively simple diffusion model would suggest. As will be discussed (§2.5), [P/D-DROP] is reported at least as far back as the 1870s in Cheshire, North Staffordshire (Ellis, 1889; Wright, 1898), and Leicestershire (Evans, 1848, 1881). In the mid-20th century, it is also reported in South Zeal, Devon (Harris, 1967), and the East Midlands is again reported (Braber & Robinson, 2018) as hosting [P/D-DROP] along with Cambridgeshire (Ojanen, 1985). Further, the electronic atlas eWAVE shows that [P/D-DROP] is typologically common across global varieties of English.

1.3 Historical overview of geography, migration and communications

The main region of interest is ultimately the North West, which is the area chosen for closer inspection in the survey (Chapter 5). However, all regions are discussed to some extent in their relation to the overall distribution of variables. Here, the South East and Midlands are particularly relevant to the story. I therefore additionally provide a brief background to these regions here as well. As we will see, the South-East is an important player in the discussion of [P/D-DROP]. It was here that the PD-DROP variant is suggested to have been innovated (Bailey, 2018b; Gopal et al., 2021), subsequently diffusing to other places.

The United Kingdom and Ireland

Patterns of internal migration across the British Isles during the 19th and 20th centuries played a crucial role in shaping regional dialects and in the devel-

Figure 1.1: Map of the United Kingdom and Ireland

opment of urban vernaculars. In the context of this thesis, having a sense of these demographic movements is important to interpreting the distribution of the variables under investigation.

During the 19th century, the Industrial Revolution introduced unprecedented levels of rural-to-urban migration. Large numbers of people left agricultural regions across England, Wales, Scotland, and Ireland to seek employment in rapidly expanding industrial cities (such as Liverpool, Manchester, Birmingham and London), bringing regional dialect features into urban centres (Cham-

pion, 2005; Pooley & Turnbull, 2005). Importantly, these migrations tended to be highly local, with cities drawing from nearby counties and only London experiencing in-migration from across the country during this period (op. cit.).

Transport infrastructure and communications are critical considerations here. As Beal (2004, p.7) reports, journeys in the 18th century were by carriage and times were measured in days rather than hours. The building of better roads reduced journey times; from York to London, for instance, the time was reduced from three days in the 1730s to one day in the 1780s. Such journeys were, however, reserved for those who could afford it. The development of the railways in the mid 19th century dramatically reduced journey times and cost but were still out of reach for most people. It was not until the end of the 19th century that travel by train became viable to a larger population (Pooley & Turnbull, 2005). The motor car, invented in 1885, was not available until 1908, with the introduction of the Model T Ford, however, mass automobile use was not seen until the "last quarter of the 20th century" (Beal, 2004, p.7), made possible by cheaper cars and the national road network: A-roads from the 1930s and motorways from 1959.

Systems of mass communication were similarly not widely available in Britain in the first half of the 19th century. The introduction of the Penny Post in 1840 saw a rise in the number of letters sent in Britain from 82 million to 917 million by the 1870s (Beal, 2004). The wider availability of personal written communication is particularly relevant here. Where communications in the 18th century tended to subject to greater normative pressure, influenced by the rise of 18th century grammar guides (Yáñez-Bouza, 2016a), this was loosened in the latter part of the 19th century. As Beal (2004) reports:

> "[T]he relatively informal style of personal letters allowed new syntactic structures to be introduced into the written medium, even as grammarians were railing against them." Beal (2004, p.9)

As we will see, the process of liberation from standardising norms in written communications was accelerated in the late 20th century with introduction of computer-mediated communication which precipitated what Ferrara, Brunner, and Whittemore (1991) termed 'interactive written discourse'. Today, personal written messages are routine, informal, mobile and measured in billions of exchanges.

New technology and infrastructure in transport and communications, then,

were key drivers of changes in mobility and linguistic interaction. Alongside this, the post-war period saw a partial reversal of the trend towards urbanisation, with counter-urbanisation seeing populations moving into rural and semi-rural suburbs following the decline of industry in urban centres, the rise of commuting, and new forms of economic activity (Britain, 2010). This brought newly forged linguistic varieties, which had formed in the context of substantial dialect mixing in urban centres, into the surrounding counties. Meanwhile inner-city areas were often re-populated via international migration, which brought with it new waves of dialect formation, as documented for London (Cheshire, Kerswill, Fox, & Torgersen, 2011; Fox, 2010) and Birmingham (Fox, Khan, & Torgersen, 2011; Khan, 2006).[11]

1.3.1 A socio-demographic and historical overview of the North West

The North West—which includes Liverpool, Manchester, Wigan, and intermediate towns—was particularly affected by these demographic shifts, owing first, in the 19th century, to the rapid rise of textile, shipping, and manufacturing industries.[12]

Within Lancashire, population movement was primarily driven by intra-regional migration from smaller rural settlements to nearby urban centres (Pooley & Turnbull, 2005). As is well-documented, Liverpool, in particular, saw a significant influx of Irish migrants, especially in the wake of the Great Famine (1845—1852), resulting in an Irish-born population of over 20% by the mid-19th century (Cardoso, 2015). In addition, migrants from North Wales contributed substantially to the population growth of Liverpool and Birkenhead (op.cit.).

These patterns of migration created the conditions for extensive dialect contact. This, Honeybone (2007) argues, led the development of Liverpool English as a distinct variety via *New Dialect Formation* in the context of a highly diverse population during the 19th century.[13] As a result, Liverpool English "stands

---

[11]The emergence of new multiethnic varieties has been documented, under similar contexts, such as in Denmark (Quist, 2008), Sweden (Stroud, 2004), Germany (Paul, Wittenberg, & Wiese, 2008) and across Europe (Wiese, 2009)

[12]The area referred to as the *North West*, in this thesis, largely points to the southern, more urbanised end of the greater North-West region shown in Figure 1.2, chiefly comprising the neighbouring city regions of Merseyside and Greater Manchester, as well as parts of Cheshire and Lancashire.

[13]*New Dialect Formation* refers to the well-documented mechanism by which new dialects emerge rapidly under conditions of extensive contact between speakers of different varieties (cf. Trudgill,

Figure 1.2: The North-West region

outside of dialect continuum, as a relatively new variety, which has been decisively affected by linguistic contact" (Honeybone, 2007, p.110).

The combination of Irish, Welsh, and local Lancashire inputs created a linguistic environment distinct from the surrounding region. As Knowles (1973, p.15) puts it: "[i]n language, as in history and tradition, Liverpool and Merseyside are in the North of England but not of it". By contrast, Manchester's growth depended on more localised population flows from surrounding towns and villages (Knowles, 1973). At the beginning of the 19th century Liverpool's population was 83,000 and Manchester's was 89,000, by 1891 both cities exceeded half a million (Pooley & Turnbull, 2005, p.85).

The 20th century saw continued forms of internal migration, though its nature was distinct from the initial population boom of the 19th century. The first half of the century witnessed further urbanisation, alongside the early stages of suburbanisation. After the Second World War, government housing policy and urban planning led to a deliberate redistribution of populations from overcrowded inner cities to newly constructed towns. The New Towns Act (1946) facilitated the development of overspill settlements such as Skelmersdale and Runcorn, designed to accommodate excess populations from Liverpool. Similarly, Warrington expanded as a key resettlement area

1986), such as in the *New Town* of Milton Keynes (Kerswill & Williams, 2000).

33

for both Merseyside and Greater Manchester. These policies again brought together speakers from a range of linguistic backgrounds, generating transitional dialect zones characterised by contact, mixing, and potential levelling (Britain 2002; Kerswill 2003). Warrington is of particular interest here, as it lies almost exactly equidistant between Manchester and Liverpool, and still serves as an intermediary location between the two main cities.

By the latter half of the 20th century, the decline of traditional industries led to increased unemployment and a shift in migration patterns with significant populations moving out from the centre of Liverpool to suburbs or neighbouring towns. The initial mixing of populations during industrial boom, and the later dispersal to neighbouring suburbs, in part via state-managed relocation, led to the spread of Liverpool English to the wider region. Knowles (1973, p.14) describes the influence of Liverpool English as spreading outwards, "beyond its former boundaries", to Southport, Maghull, Lydiate, Ormskirk, St.Helens, and to Runcorn and Widnes, as well as across the Mersey river. Interestingly, Knowles (1973, p.14) also describes two competing 'urban fields', that of Liverpool and Manchester, and the existence of a "narrow band of overlap between the two urban fields, running through Preston, Wigan and Warrington".
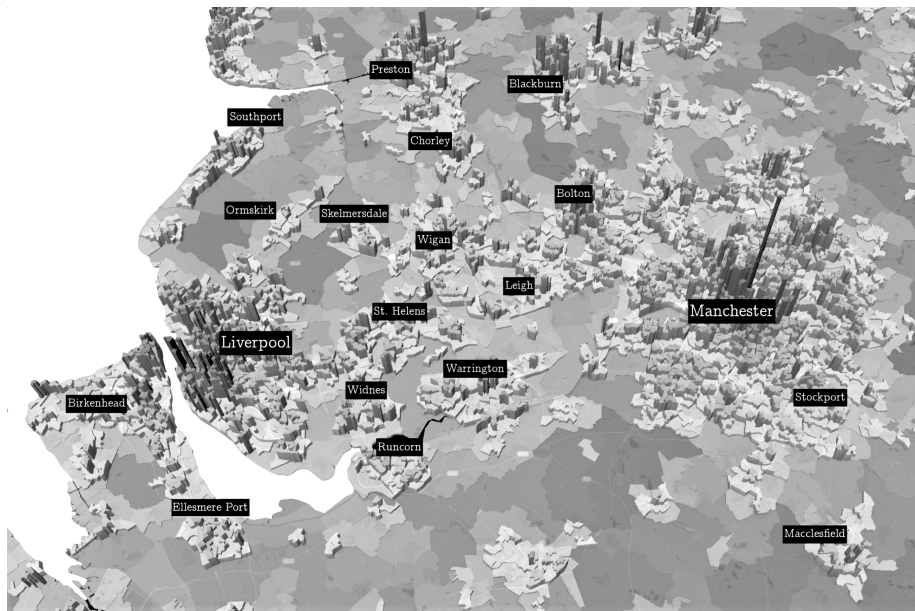
Figure 1.3: 3D representation of the North West region with column height corresponding to population. Source: `https://parallel.co.uk`

The map, presented in Figure 1.3, shows the relative population distribution

in the North West. We can identify the two main conurbations: Liverpool in
the west, and Manchester in the East. In between the two cities, there lies a
semi-urban corridor stretching, almost uninterrupted. The schematic presented
in Figure 1.4 provides an interpretation of this concept of competing *urban
fields* between the two major conurbations (Liverpool and Manchester), with
Warrington essentially caught between the two.
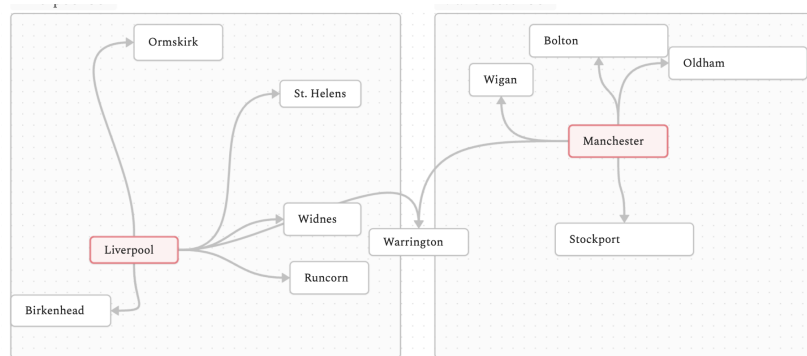
Figure 1.4: Schematic representation of competing *spheres of influence* between Manchester and
Liverpool with Warrington sitting in between the two.

It might be tempting to ascribe the relative distribution of linguistic features
to a *gravity-based diffusion*. The gravity model, borrowed conceptually from the
physical sciences, predicts that interaction will be increased in more populous
locations, and that such interaction will be reduced as a function of distance
from populous centres (Haynes & Fotheringham, 1984). However, the spread
of Liverpool English is more complex than the mechanistic model of *gravity-
based diffusion* might suggest. As West (2015) argues, such an "equation does
not account for the potential effect of social factors, such as the varying contact
that these localities may have with Liverpool" (p.26).

Given late-20th century outmigration from Liverpool, the presence of Liv-
erpool features in the wider region is unsurprising. In addition, both Liverpool
and Manchester exert social influence on the surrounding smaller towns. The
result is that interlying towns each tend to gravitate, in social, cultural and
linguistic terms, towards their closest city. At some point, towards the centre
of these areas, lies a linguistic (and social) boundary or 'isogloss'. People living
in these places are often acutely aware of the presence of this boundary, as can
be seen in the tweet in (10).

(10)   Haha! I'm from Warrington so depending which half of the family I'm

around I either go plastic scouse or proper manc! (Tweet from NWatlas-corp)

Warrington draws particularly mixed accounts. Some report it as 'Manc', and others 'Plastic Scouse'. The Liverpool term *plastic scouse* is similar in meaning to the term *wool* or *woolyback*, denoting people who live outside the boundaries of the city proper (the exact nature and location of which there is ongoing dispute) but who retain Liverpool dialect features. Differentiating between 'plastic scouse' and 'wool' is fine-grained, but 'wool' has connotations of non-urban or rural. The example in (10) is particularly interesting as it suggests not only an awareness of the two distinct varieties, but an ability and willingness to adapt to a given variety on demand.

A similar situation is reported in (11) for the boundary between Wigan and St. Helens, which is also approximately half way between the two main cities, but 20 miles north of Warrington.

(11)   bellaMac73 Noooooooooooo Wigan are the enemy…..😂😂😂 I'm from St Helens 😂😂😂 (Tweet from NWatlas-corp)

There is also evidence of an acute sensitivity to linguistic variation in the region, and the existence of a sharp linguistic boundary (Figure 1.5).

Here is the Greater Manchester / Lancashire / Liverpool City Region border. Less than 15sq miles where you can and can't go the gym. Where can and can't go to your mum's. Where you can and can't go the pub. Where you can and can't have a car boot sale.
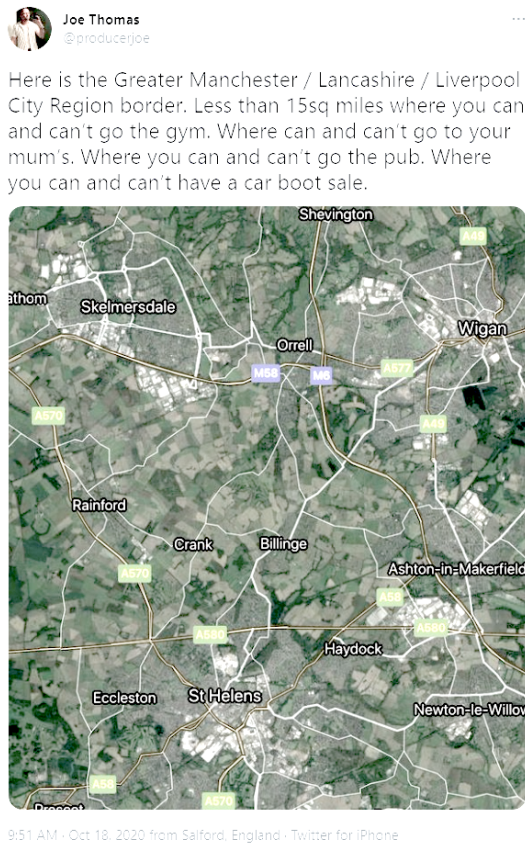
Figure 1.5: Tweet referring to the North West region and use of P-DROP ("or not").

Identity, then, and its role in shaping linguistic production should be taken into account in the analysis of variability in the wider region. This may be particularly relevant in explaining a possible shift from DAR-ð/θ (th') to P-DROP, and its maintenance as a local marker of non-rural identity. Meanwhile, speakers who reside outside of Liverpool *proper*, but who use Liverpool features in this way, are perceived as *woolybacks*, *wools* or *plastic scousers*. There is, then, considerable social awareness of the (contested) boundary to what constitutes legitimate Liverpool-hood, which is captured in a map of the region created by a *Reddit* user, presented in Figure 1.6.
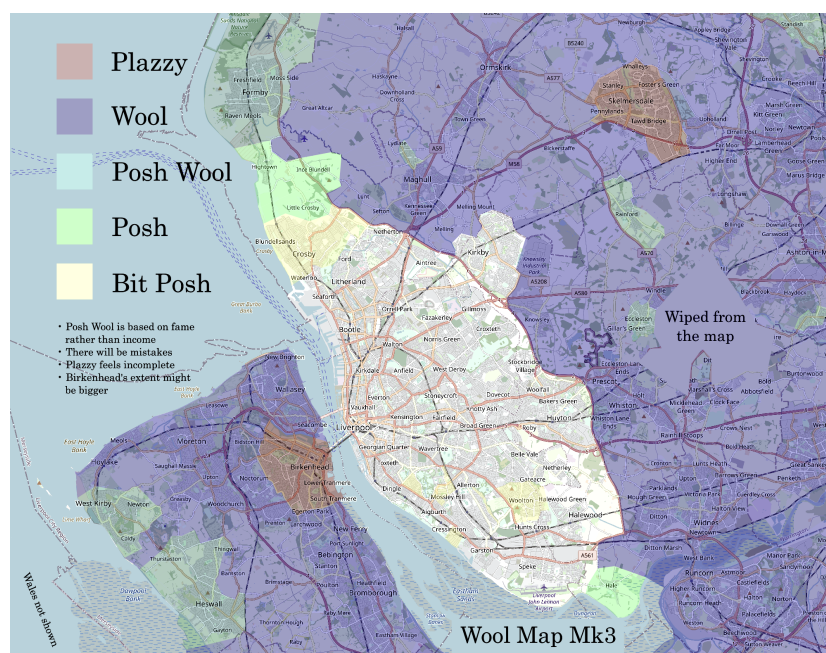
Figure 1.6: "Final Wool Map". Updated user-created map of areas around Liverpool now including 'Wool', 'Posh', 'Bit Posh', 'Posh Wool' and 'Plazzy' (Plastic Scouser). Updated after kickback in the comments for a previous map. From: `https://www.reddit.com/r/Liverpool/comments/whznh3/its_just_the_m57_final_wool_map/`

Shown here is the 'final version' that was resubmitted to the Liverpool Reddit forum, following lengthy discussion and debate over the first iteration. This serves to illustrate the sensitivity of speakers to what constitutes 'authentic' and 'inauthentic' Liverpool-hood. In the context of P-DROP and GTD, it is interesting, as we will see, that P-DROP use goes well into the area considered to be *wool*, where GTD is quite tightly constrained to the Liverpool (and Birkenhead) areas. The difference in distribution may be partly explained by the relative salience of each structure: speakers seem to have some awareness of P-DROP, and its association with Liverpool, but [DATALT] (whether a speaker uses "give me it" or "give it me" is more likely to go unnoticed. Labov's (1972) foundational concept of *markers, indicators and stereotypes*, whereby features are ranked on their relative salience to speakers, and thus the extent to which they may carry social meaning, is likely relevant here.

1.3.2 London, the Midlands, and the South East

Whilst the ultimate focus of this thesis is the North West, the South East and Anglia regions will also prove important to the discussion, particularly in re-

lation to the distribution of [P/D-DROP]. As such, a brief overview of these regions is given here.

Figure 1.7: The South East and Anglia

Patterns of internal migration in London, the Midlands, and the South East across the 19th and 20th centuries were transformative. Marked initially, as with other urban areas, by large-scale rural-to-urban migration, dialect contact, and subsequent linguistic innovation and levelling, the South East experienced a similar pattern of counter-urbanisation in the post-war period which saw the spread of varieties outward to neighbouring counties such as Essex and Kent.

During the 19th century London saw massive population growth rising from around one million at the beginning of the century to over six million by 1900. This growth was fuelled both by natural increase and migration from across the British Isles, including from rural England, Wales, and Ireland (Upton, 2006). The expanding docklands and manufacturing sectors brought in workers, while internal mobility within the South East brought in speakers from neighbouring counties such as Essex, Kent, and Surrey. Pooley and Turnbull (2005) report that during the 19th century, London was the exception in attracting migrants from all around the country; and that, outside of London, "for most people movement within a local region was the most common experience" (p.88). London's unique status as a destination for migrants from further afield has

important consequences for the interpretation of the distribution of PD-DROP, as we will see.

In the post-war period, internal migration in the South East took on a different character. The London County Council and later the Greater London Council implemented housing and planning policies aimed at redistributing London's growing population to surrounding areas. The development of *New Towns* such as Stevenage, Basildon, and Milton Keynes, as well as overspill estates in towns like Harlow, Luton, Crawley, and Slough, aimed to relieve pressure on the capital while maintaining employment connectivity through improved transport infrastructure. These policies generated substantial population shifts from inner London boroughs to suburban and exurban settlements, resulting in high levels of dialect contact and levelling (Kerswill 2003; Britain 2002).

One of the most notable sociolinguistic outcomes of the more recent period is the development of Multicultural London English (MLE), a contact variety that emerged in the late 20th century in multiethnic working-class communities in London, particularly in the East End, where international migration re-populated areas that were once predominantly 'Cockney' (Cheshire et al., 2011; Fox, 2010). While MLE is commonly associated with more recent international migration, it also reflects a much longer history of internal dialect contact, innovation, and levelling. Kerswill, Cheshire, Fox, and Torgersen (2012) argue that MLE is part of a wider process of levelling and innovation in the South East, which has diminished traditional dialect distinctions and produced new urban varieties with wide regional influence.

As we will see, these successive waves of in-migration, and subsequent out-migration are important in the context of the development and spread of PD-DROP. PD-DROP has been associated with London English (Gopal et al., 2021; Hall, 2019), which is supported by the finding, as we will see, that the feature is attested earliest here. More recently, as Bailey (2018b) notes, the emergence of MLE—which is composed of several input varieties which themselves independently feature PD-DROP, such as those from the Caribbean —likely provided an additional driving force behind the establishment of PD-DROP in London.

However, it also seems likely that both PD-DROP and P-DROP developed independently in other parts of the country, such as Leicestershire, Cambridgeshire, Shropshire and Devon. The point here is the timing. Whilst London PD-DROP is attested in 1737, there did not exist the mechanism by which its widespread

diffusion could take place until after the time when the feature was apparently established as far as the North West.

1.4   The wider geographic distribution of variants

Regarding the wider geographic distribution of [DATALT] and [P/D-DROP] variants, the thesis provides substantial new detail.

[P/D-DROP]

For [P/D-DROP], the data reveal that PD-DROP use extends to the South West and the Republic of Ireland. In the North West, parts of the Midlands and Dublin, PD-DROP gives way to P-DROP. In large parts of Yorkshire, the fourth logical variant, D-DROP (9d), is dominant. Close analysis of the data also shows that there is geographic variability within PD-DROP: the type found in the North West follows distinct constraints to that reported for the South East. These structural distinctions, along with the broader geographical distribution, complicate the picture regarding Gopal et al.'s (2021) diffusion model, which appears to show PD-DROP spreading North West from London. While it is likely that some degree PD-DROP diffusion has taken place from the South East, the data suggest that [P/D-DROP] variants were, in addition, independently innovated in other parts of the country, via localised internal processes.

There is currently no real consensus on the chronology of P-DROP and PD-DROP innovation, but both are generally speculated as being relative newcomers in recent literature (Bailey, 2018b; Gopal et al., 2021). Here, I show that both P-DROP and PD-DROP are attested in several locations across the country as far back as the 1850s (Evans, 1848). Meanwhile, I show that the geographic distribution of variants, in the context of what is known about migration patterns, sheds light on their historical origins. Here, an additional connection is made to the well-established phenomenon of *definite article reduction* (DAR), found in Yorkshire, Lancashire and the Midlands, that has hitherto not been discussed.

(12)   DAR

    a.   John went t' ʔ'pub (DAR-t)

    b.   Sally is going t' θ'shops (DAR-θ)

    c.   Mark went to Ø beach (DAR-Ø)

D-DROP (*go to N*) and P-DROP (*go the N*) are shown to exhibit a close spatial alignment with DAR-t (12a) and DAR-θ (12b), respectively, as reported in the Survey of English Dialects (Orton & Dieth, 1962).

Further, an explanation for the present-day distribution of P-DROP is offered, following contact phenomena which would likely have taken place at the time when Liverpool was emerging as an industrial port, with the particular distribution of P-DROP in the North West following 19th century transport and trade routes (§4.5.2).

Meanwhile, the structural distinction between North West PD-DROP (*go N*) and South East PD-DROP, is shown to have likely emerged out of a development— or, at least, influence from— DAR. This is supported by a textual analysis of tweets in comparison with texts from traditional corpora, which reinforces a probable diachronic connection between definite article reduction (DAR) and the various manifestations of [P/D-DROP].

Thus, the thesis argues that there were most likely multiple sites of [P/D-DROP] innovation rather than a single source. Regarding the proposed diffusion from the South East, I suggest that a recent surge in PD-DROP (*go N*) usage has likely occurred, starting in East London, fuelled in part by the rise of MLE, and social media. I suggest that it is this surge which has undergone diffusion and has been interacting with preexisting [P/D-DROP] phenomena in the process of its spread North West. In the South West, based on traditional corpus data, a separate instantiation of PD-DROP is suggested to have taken place. Further, the existence of high rates of PD-DROP in the Republic of Ireland is presented, a fact that has hitherto gone unnoticed and which again complicates a diffusion model.

[DATALT]

Regarding pronominal ditransitives (2), the geographic distribution is found to align with *Survey of English Dialects* (SED) (Orton & Dieth, 1962) data (as mapped by Kirk et al. (2014)), and with (deep-)historical migration patterns, with present data revealing sharp linguistic boundaries that likely have considerable time-depth.[14] The distribution is also found to align with more recent survey data, such as that provided by Britain et al. (2018) and MacKenzie et al.

---

[14]Throughout this these I refer to the Survey of English Dialects (SED), specifically the distribution of ditransitive and DAR variants. These distributions are based on those found through the mapping of these data by secondary sources, largely Kirk et al. (2014), in the case of [DATALT] and Barry (1972), in the case of DAR.

(2022), which further legitimises the validity of the Twitter data. High rates of ₚGTD (_me it_) in the East, North East and Scotland challenge the notion that it is a relative newcomer and lend support to Gast's (2007) observation that ₚGTD may have arisen as a result of contact with Old Norse. In addition, the question of ₚTGD (_it me_) as akin to ₚGTD (_me it_) is addressed via a quantitative analysis of the distribution by verb type. Here, data strongly support Haddican's (2010) findings, but also point to some degree of interspeaker, and inter-regional variation.

1.5   The present thesis

The current thesis presents substantial new data which shed light on these syntactic conundra. The primary data are Twitter corpora spanning ten years from 2011-2021, supplemented with traditional dialect corpora and, in the North West, a large-scale grammaticality judgement survey, conducted with the participation of Sixth Form Colleges.[15] The large scale of the Twitter corpus, with an effective size of approximately 9.8B words, allows for a geographic scope and level of detail that have hitherto been unreported. Meanwhile the traditional corpora and the Sixth-Form survey reinforce the Twitter results, provide an indication of diachronic development as well as an indication of the underlying structure of variants.

Previous observations regarding a connection between [DATALT] and [P/D-DROP], for some speakers in the North West, are broadly supported (with some caveats). Additionally, the Twitter data make it possible to predict where, geographically, this grammar is most likely to exist; a prediction that the survey data confirm is borne out. Where Biggs (2018) focuses on Liverpool for her study, she notes that the claim of her paper "is not that certain linguistic forms will only occur in precise geographical locations rather the availability of such a construction will correspond to systematic and productive variation in other aspects of that individual's grammar" (Biggs, 2018, p.4, footnote 5). In light of this, I show that such a grammar is, in fact, probabilistically located in the region _between_ Liverpool and Manchester; an area that lies at the intersection of the distribution of P-DROP and ₚTGD.

This analysis is underwritten by the corpus and survey data which show that, in contrast to Biggs's (2018) finding that the ₚTGD (_it me_) is well accepted

---

[15]High School students, age 16-18

in Liverpool, it is in fact relatively disfavoured here and in locations in its immediate vicinity, which instead favour ₚGTD (_*me it*). Rather, it is within this intersecting area, where both P-DROP and ₚTGD are dominant, that we find the conditions which result in widespread speaker acceptance of THEME-GOAL ditransitives with two full DP complements (_DPDP_TGD (_*the letter the bank*)). Thus, where previous work has reported inter-speaker variation in the underlying grammar in the North West, the present work reveals the likely geographical source of this grammar.

1.6 Methodological overview

1.6.1 Orientation of the research within linguistic subfields

The work presented in this thesis lies at the intersection of a number of different subfields of linguistics. It is primarily concerned with dialectology, or what has more recently been termed *geolinguistics*, that is, the study of the geographical distribution of linguistic features across a population. Given the close relationship between dialectology and sociolinguistics, the current work also engages with topics that are inherent to sociolinguistics, such as the role of identity (especially as it pertains to place), speaker age and social class.

In terms of data, the heavy lifting is carried out using a range of large-scale linguistic corpora. In this regard, the work is aligned with the field of *corpus linguistics*. Here, the research can be seen as a direct extension of previous work conducted in this area, such as Siewierska and Hollmann's (2007) corpus-based study of ditransitives in Lancashire, or Gerwin's (2014) extensive corpus investigation into *Ditransitives in British English Dialects*. However, in addition, the present work introduces a methodology usually associated with formal syntactic theory, that of acceptability judgements.

There were several initial motivations underlying the choice to include an acceptance judgement survey. First, was continuity with previous research. Previous studies which have investigated the phenomena under investigation were rooted in the field of formal syntactic theory and used speaker judgments as their primary evidence. In order to compare like-for-like, it was therefore deemed appropriate to employ an equivalent methodology for the present work.

Second, was the requirement for negative evidence, that is, where corpus data show us what is possible, they do not show us what is impossible. This is

particularly relevant when looking at syntactic micro-variation between small geographical locations, where a lack of attestation in the corpus may simply be due to insufficient data. This second reason was, in part, a response to Siewierska and Hollmann's (2007) proposed follow-up survey to their corpus study that would "follow a recent trend in dialect grammar research towards triangulation" (Siewierska & Hollmann, 2007, p.100). This proposal followed in the wake of a series of projects which took the approach of combining corpus and judgment data (cf. Cornips & Corrigan, 2005). For example, Cornips and Poletto (2005) point out that a purely observational (corpus-based) approach runs the risk of remaining blind to certain aspects of language use, especially uncommon patterns, while a solely experimental method (written question- naires and oral tasks) also has shortcomings, e.g. the well-known tendency for subjects who speak a nonstandard variety to be influenced by prescriptive norms. Taking a combined approach, then, aims to capitalise on the benefits particular to each method, while minimising the drawbacks.

In terms of respondents being influenced by prescriptive norms, survey re- sponses should be interpreted with this in mind. It is nevertheless perfectly possible to measure the relative acceptability of non-standard features in rela- tion to each other. For instance, both PD-DROP and P-DROP are non-standard, though each will likely be subject to different levels of degradation. How- ever, each will likely be further degraded if they are not present in the local vernacular, or are associated with a competing vernacular. The effects of pre- scriptive norms can further be mitigated through the instructions provided to participants, by explicitly informing them that the survey is interested in local features of English, and is not concerned with standard, or 'correct' language.

Third, the use of survey data was seen as a useful counterpoint to the Twit- ter data. Given that Twitter data still constitute a relatively novel data source, it is reassuring to be able to compare results with those gathered via a funda- mentally different means.

Combining such different approaches is, of course, not without issue. Labov's (1996) article, *When intuitions fail*, for example, casts serious doubt as to whether intuition data can reliably be used as a measure of sociolinguistic variance. In addition, it is not by any means clearcut the extent to which relative accep- tance rates should correlate to relative rates of use. Bermel and Knittl (2012a) attempted such a comparison, for example, in a study of variability in Czech, with mixed results. This said, what is important is the particular theoretical

questions that are being asked, and, crucially, how acceptance judgement questionnaires are framed to study participants. More recently, Jamieson, Smith, Adger, Heycock, and Thoms's (2024) paper entitled *When intuitions (don't) fail*, makes a strong case for the careful use of judgements, alongside sociolinguistic interview data. The position taken in the present thesis aligns with that of Jamieson et al. (2024): different research questions demand different kinds of evidence. In the context of the current work, for example, seeking evidence for the presence of $_{\text{DPDP}}$TGD in the specific location identified as the most likely host for such a structure, required both detailed corpus data to locate the area, and acceptance judgements to test for its presence.

In the present work, a key question revolves around the extent to which the manifestations of P-DROP and PD-DROP that are found in the North West of England are structurally distinct to those reported in the South East. Here, the diagnostics employed by generative syntacticians are used to make predictions about the kinds of contexts in which we might expect each structure to occur. These predictions can then be tested using the various corpora (Twitter and traditional corpora), i.e., do we see evidence of strings in the North West in contexts that would be disallowed in the South East. The same diagnostics are also tested using acceptance judgements in the survey. Here we see how sociolinguistic questions regarding the diffusion (or not) of a given feature may be addressed using diagnostics that were originally developed to answer formal questions about the structure of a single variety.

Finally, the survey for the North West of England, which worked with Sixth Form Colleges, demonstrates a scalable working model for how usage data may be corroborated and extended. It shows, more broadly, a way that universities and Further Education centres may collaborate in a way that is mutually beneficial. For English Language A-Level students, it was an opportunity to participate in university research. For the project, it provided valuable data which could be used to test the predictions made from usage patterns found on Twitter.

1.6.2 The reliability of Twitter data

In terms of the use of Twitter data for linguistic research, the thesis also answers questions of reliability through comparison with traditional corpora, survey data, and two methods for gathering Twitter location metadata. For the UK

and Ireland, location is based on that which is provided by Twitter (based on GPS and other user-profile information). In the North West, an additional corpus, based on where a user once reported having 'grown up' or 'being from', supplements the GPS-based corpus. This second method mirrors that used by Willis (2020) and Gopal et al. (2021) who make the strong case that knowing where a user is from is more useful than knowing where they happened to be when they sent a tweet. Having the two corpora, then, allows for a useful comparison between the two methods for assigning location. It is shown that the two methods in fact show a strong overall correlation.

It is also shown that the nature of Twitter data has changed over the ten-year timespan from which the corpora were collected (§4.3). This shift, which sees a change towards the use of more standard registers, likely reflects an increase in the average age of the people sending Twitter messages. The well-established phenomenon of *age-grading*, where speakers trend to more standard language use as they move into more advanced life stages, is likely at play here. This trend aligns with a similar shift to standard forms observed in Grondelaers et al.'s (2023) analysis of Dutch Twitter messages. The present work proposes that an additional factor underlying this shift is a change in the communication style conducted on the platform, away from conversational messages between individuals towards more public-facing messages, which themselves tend to engender the use of more standard registers.

This is important information for any study that uses Twitter data from a particular time window within the ten-year span. A change towards standard registers additionally sheds light on the register status of individual variants: while PD-DROP use drops dramatically after 2018, P-DROP (the variant found in the North West) is relatively stable. This is taken as an indication of its relative status in the register hierarchy, or its relative invisibility to the speakers who use it. This observation is further supported by the linguistic contexts in which it is found to be used.

1.7 Summary of the present state of knowledge

Here, I present a brief summary of the present state of knowledge regarding the two main variables as well as the use of Twitter data for linguistic research. These points are expanded in the following chapter, but are presented here in order to frame the research questions in the next section.

[P/D-DROP]

1. PD-DROP in London is restricted to a narrow range of verbs and conditions (Hall, 2019). In Kent, it follows similar restriction, though is hypothesised as possibly distinct due to a different contact trajectory (Bailey, 2018b).

2. Regarding P-DROP, conditions underlying the structure have been explored in Liverpool (Biggs, 2018) and Ormskirk (Myler, 2013) and are found to be distinct between the two places. P-DROP in Liverpool appears to be more permissive across wider range of contexts than in Ormskirk.

3. There is a clear structural distinction between P-DROP (as found in Liverpool and Ormskirk) and PD-DROP (in London and the South East). However, where the determiner is absent in the full form, such as in: "going to France", "going to football practice", the two structures are identical at the surface level. Meanwhile, the broader geographic extent of PD-DROP *as distinct from* P-DROP, and their frequencies relative to other variants and the specific extent of P-DROP distribution in the NW (and beyond) are unknown.

4. Bailey (2018b, p.52, fn.6) does report, in passing, an example of PD-DROP in Manchester, which "appears to be the same as the Southeast variety", but cites this as evidence that it is "not geographical spread of the feature". It is not clear, at this stage, whether PD-DROP in Manchester is subject to the same conditions as those reported for the South East; a clearer sense of such conditions might help to shed light on whether Manchester PD-DROP arrived via spread from the South East. Bailey presents a strong argument for independent innovation cross-linguistically, however, more recent research using Twitter data (Gopal et al., 2021), has shown that PD-DROP is used with high frequency across England and patterns geographically as though it has, indeed, undergone a relatively recent spread from the South East towards the North West, via the Midlands.

5. However, we will also see that PD-DROP is reported as far back as the 1850s in the North West, which complicates the picture regarding a proposed SE-NW diffusion model.

6. Finally, a connection between variants of [P/D-DROP] and DAR in the North and Midlands has not been made.

[DATALT]

1. Recent national survey data (Britain et al., 2018; MacKenzie et al., 2022) have provided good detail on the present-day distribution of variants of the pronominal ditransitive. Meanwhile, the distribution of the relative usage rates of variants across the country has been covered (Gopal et al., 2021; Stevenson, 2016) using Twitter data, but is still lacking a degree of fine-grained detail between locations. Specifically, studies to-date have not reported on the relative rates of variants in each location, or analysed the relative distribution of [DATALT] variants by verb and verb type.

2. Subsetting the distribution by verb-type requires substantially more data, but may provide geographically sensitive, quantitative evidence, for Haddican's diagnostic, based on Levin (1993), that different verb-types are used preferentially with different ditransitive structures, for example, that latinate verbs tend to prefer PDAT, while *refuse*-type verbs tend to prefer GTD.

3. Biggs (2018) finds syntactic evidence for a connection between TGD and P-DROP in Liverpool. Biggs links Liverpool acceptance of THEME-passives with full-DP objects, where subject is a DP, to P-DROP. However, there are limited quantitative data on the actual usage rates of TGD in Liverpool. Previous corpus studies, such as Gerwin (2014) necessarily (due to the corpus size) group Liverpool together with the region as a whole. Early Twitter data from Stevenson (2016) does show that Liverpool seems to prefer $_p$GTD rather than $_p$TGD and this pattern is reinforced by recent survey data from MacKenzie et al. (2022).

4. Haddican (2010) finds that TGD is, for most of his Manchester survey respondents, underlyingly more akin to GTD. However, the diagnostics are not always clearcut, and there is a significant amount of inter-speaker variation: some respondents appear to align more with the pattern found by Biggs (2018) in Liverpool, suggesting that for those speakers TGD is underlyingly PDAT with a dropped preposition.

1.8 Research questions

Here I spell out of overall goals of project, and research questions that will be addressed. These questions will be returned to individually at later stages of

the document.

Twitter

(1)   How reliable are geocoded Twitter data? To what extent do they correlate with Twitter data located based on user-reported place of origin? How have Twitter data changed over time?

(2)   What is the comparative geographical spread of variants of (a) the pronominal ditransitive and (b) preposition/determiner dropping?

(3)   What historical factors may explain the geographic distribution uncovered in (2)?

(4)   What quantitative evidence can be brought to bear on the TGD-as-GTD question?

(5)   To what extent do [P/D-DROP] variants align with DAR distribution?

(6)   Was PD-DROP independently innovated in multiple locations, or did it spread from one point of innovation in the South East of England?

(7)   Is there a likely location for the grammar described by Biggs (2018), given the distribution of the Twitter data?

Survey (North West)

(8)   To what extent do speaker judgements correlate with corpus frequencies for a given location?

(9)   To what extent is there evidence for $_{\text{DPDP}}$TGD in the location predicted from the Twitter data? Is there a grammar in the North West for which TGDs are underlyingly PDAT with a dropped preposition and is this grammar geographically restricted?

(10)   Is PD-DROP in the North West akin to that found in the South East?

1.9 Thesis structure

The core thesis is presented in chapters 3, 4 and 5. Chapter 2 introduces the overall lay-of-the-land, motivation for the thesis as a whole, and can be seen as a prelude to chapter 6 which brings together the two core parts.

Chapters 3 and 4 document the mapping of syntactic alternations in the UK using Twitter data, demonstrate the validity of these kinds of data for dialectal research and shed new light on the distribution of several syntactic phenomena. They show how such an atlas may be used to find particular areas of interest that warrant further study.

Such a further study is then fleshed out in the North West of England (NWE) in chapter 5, using a mass participation survey — delivered through sixth form colleges, which focuses on syntactic acceptability judgements. The survey is in part an experiment in methodological approach: testing the viability of gathering linguistic data in collaboration with institutions of further education. It is designed to collect a range of linguistic data, with scope for further analysis in the future. The current dissertation considers a subset of this data most relevant to the present narrative.

Background

2.1 Introduction

The current chapter provides a more in-depth background to the variables
in their social, linguistic and geo-historical contexts. The start of chapter 3
(on building a Twitter corpus) and chapter 5 (on the survey development and
results) provide further background on the literature specific to the distinct
methodological approaches employed.

2.2 The dative alternation in English

The dative alternation—which refers to the structures licensed by 'alternating'
dative verbs such as *give, send, buy, sell etc.*—has received much attention in the
literature. Alternating ditransitive verbs are so-called because they are able to
occur as part of either the prepositional ditransitive (PDAT) where the THEME
precedes the GOAL which is introduced by a preposition (1a), or a double object
construction where the GOAL argument precedes the THEME (GTD) (1b).[1]

(1)   Dative alternation (full-DP)

      a.   John sent the book to Sam (PDAT)

---

[1] The double object construction is often labelled DOC, but here GTD is chosen for continuity
with the THEME-GOAL variant (TGD).

b.   John sent Sam the book (GTD)

c.   ?John sent the book Sam (TGD) (unacceptable for most speakers)

A third option, where the THEME precedes a preposition-less GOAL is unacceptable, for most speakers, where both objects are full noun-phrases (1c). For ditransitives where both objects are pronouns (pronominal ditransitives, pDit), however, the structure is available to a significant portion of speakers in the UK (2c) (pTGD). This is the same surface order as (1a) but with an apparently elided preposition.

(2)   Pronominal ditransitives (pDit)

a.   gave it to him (pPDAT)

b.   gave him it (pGTD)

c.   gave it him (pTGD)

Whether the structure in (2c) is derived from (2a) via a process of preposition elision or dropping, or whether it is (2b) with a reversed object order (via short-object movement), is, it turns out, not a trivial question. As we will see, it appears to depend on which part of the North West of England the question is being asked, and hinges on the localised acceptability of TGDs with full DP objects (DPDPTGD) (1c). Broader geohistorical variation in the relative rates of pDit variants has been well documented, and is detailed in §2.3.

An underlying distinction between Manchester and Liverpool: Haddican 2010; Biggs 2014,2016

Haddican's (2010) study into TGD use in Manchester aims to establish whether the pTGD is underlyingly akin to the pGTD or whether it is better described as a pPDAT with a dropped preposition. The study shows convincingly that, under a range of conditions and for the majority of speakers who took part, the pTGD behaves like a 'true double object construction' (DOC). Meanwhile, Biggs (2018) provides a detailed analysis of the ditransitive in Liverpool, amassing some evidence, using the same criteria as Haddican (2010), that the TGD in Liverpool is perhaps fact best described as a PDAT with elided preposition.[2]

The argument either way is primarily driven by the interaction with two

---

[2]It is not clear the extent to which there is also cross-speaker variation in Liverpool, though it is indicated that there is some here too.

key diagnostics: object type and verb class.[3]

Object type

The first piece of evidence that the $_p$TGD is underlyingly a $_p$GTD in Manchester stems from the unacceptability of the preposition-less THEME-GOAL order with full noun-phrase complements, as in the earlier example (1c) and the example below (3).[4]

(3)   ?? She gave the ball the boy

If the TGD with pronominal objects (2c), repeated below in (4), were underlyingly a prepositional dative, then we would expect (3) to also be grammatical, as there is no obvious reason, a priori, why a hypothetical preposition drop would favour pronominal objects over full noun phrase objects.[5]

(4)   She gave it me ($_p$TGD)

Haddican then shows that, on average, participants in his Manchester survey indeed report degraded acceptability of the TGD with DP objects (Figure 2.1). Interestingly, however, there is considerable cross-speaker variation; for at least some participants, by the same diagnostic, the TGD does in fact seem to behave like the PDAT, and thus suggests that for this subset of speakers, the TGD seems to involve P-DROP.[6] The important fact however, is that overall, the pattern seems to show that TGD is more akin to GTD.

Meanwhile, in Liverpool, the majority of Biggs's (2018) respondents apparently do accept TGDs with full DPs (3) (Biggs, 2014, 2016) resulting, it follows, in the conclusion that, for Liverpool speakers, TGDs are underlyingly $_p$PDAT. The argument that Liverpool speakers might arrive at THEME-GOAL surface orders via preposition drop is bolstered by the fact, as we will see, that

---

[3]Additional diagnostics are used, such as the animacy of the recipient, but the focus here is on the two clearest diagnostics.

[4]It is worth noting that this structure is exceptionally rare (though not unattested) in both contemporary (Gerwin, 2014) and historical (Yáñez-Bouza & Denison, 2015) corpora.

[5]Though as we will see, there may be reason to suspect the reverse to be true: in grammars where preposition drop is indicated for full noun phrase objects, pronominal TGDs may still be derived via object shift, or scrambling, rather than preposition-drop. A parallel can be seen in modern German (gib es mir etc)

[6]How we should interpret this inter-speaker variation is a central issue still to be properly addressed, though as Biggs (2018) points out, fn 6, what we are identifying is the systematic availability of structures in a given location (based on some component of the grammar in that area).

the dropping of prepositions is widespread in Liverpool English with verbs of motion, such as *going, coming, bringing* (discussed in §2.4).

An intermediary structure, where the THEME is pronominal and GOAL is a full DP is less accepted, on average, than where both objects are pronominal, but nevertheless appears to be quite widespread across the region, in both Liverpool, Manchester and further afield. There is again, however, considerable inter-speaker with this structure as well.

(5)   She gave it the boy

Interestingly, the least accepted TGD is found where the THEME is full DP and the GOAL is pronominal. This may be an effect of the *principle of end weight*, a general rule of information structure in English (and other languages), which favours placing complex constituents after lighter ones.

(6)   *she gave the ball him

Figure 2.1: Cline in acceptability ratings with the introduction of full-DP objects.
Source: Haddican (2010, p.6)

Verb class

The second piece of evidence emerges through comparing the acceptance ratings of the three pDit types with different verb 'classes'. So far, we have considered examples with 'alternating' verbs such as *give, send, sell* etc. Haddican, following Levin (1993), groups these verbs into a set labelled the 'give class'. They are known as alternating because they can occur with both PDAT and

GTD. However, there are other classes of verbs that tend to only occur with either PDAT or GTD, and are thus reported as ungrammatical, or at least degraded, when they are placed in the structure with which they do not usually occur.

The verb classes which tend to be accepted with PDAT, but degraded with GTD, are reported as follows: 'verbs of continuous imparting of force' (*carry, pull, push);* 'Manner of communication' verbs (*whisper, yell, bark*) and 'latinate' verbs (*contribute, donate, distribute etc.*) (Levin, 1993). If the TGD is underlyingly GTD, we predict it to be degraded in the same way as the GTD is in these contexts. This prediction is borne out in Haddican's Manchester study, but again, not in Biggs' Liverpool sample where the verbs that are standardly only acceptable with PDAT are also reported as reported as licit with TGD. This can be seen in the following example adapted from Biggs (2016, p.14):

(7)   a.   She pushed/hauled/lifted it to me. (Manchester OK) (Liverpool OK)

       b.   She pushed/hauled/lifted me it. (Manchester*) (Liverpool*)

       c.   She pushed/hauled/lifted it me. (Manchester*) (Liverpool OK)

Verb classes which tend to standardly be accepted with GTD but not PDAT are known as 'prevention of possession verbs' (refuse, cost, deny). Here, again, the acceptance patterns for Liverpool and Manchester are as expected given a PDAT-derived TGD in Liverpool and a GTD-derived TGD in Manchester, with the following examples again adapted from (Biggs, 2018, p.16):

(8)   a.   She refused it to him (Manchester*) (Liverpool*)

       b.   She refused him it (Manchester OK) (Liverpool OK)

       c.   She refused it him (Manchester OK) (Liverpool*)

This result is notable, as it appears to disallow the TGD as GTD-derived in Liverpool, but permit it in Manchester. Whilst the examples in (7) support the Liverpool TGD with DP objects as being permitted via preposition drop, the examples in (8) suggest that this is the only way that the TGD is derived in Liverpool. It should be noted here, however, that the examples in (8b) and (8c) were, in Manchester as well, both quite severely degraded, but critically—for Haddican's argument—not as degraded as (8a). This, Haddican takes as evidence that (8b) and (8c) were related in Manchester English (Figure 2.2).
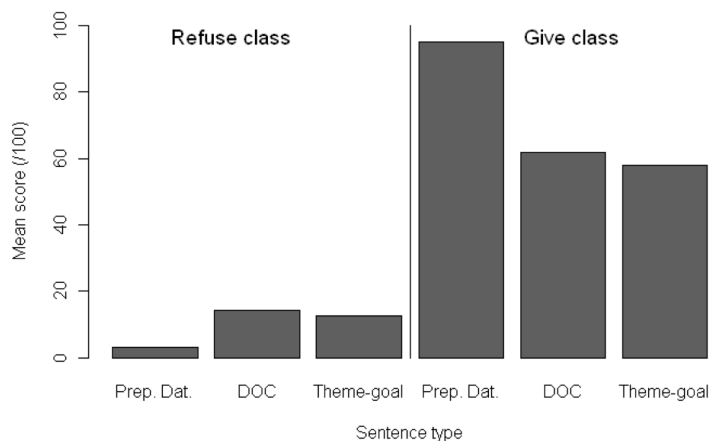
Figure 2.2: Mean acceptance scores, Refuse vs Give class
Source: Haddican (2010, p.8).

Whilst Biggs (2018) reports TGD-*refuse* (8c) as grammatical in Manchester, the chart in Haddican (2010) (figure 2.2) shows that its grammaticality there is by no means clear-cut. Given that Biggs does not supply gradient acceptance results in the way that Haddican does, it is not clear where she draws the line in terms of grammaticality. It may be that *gradient* acceptance ratings for TGD-*refuse* in Liverpool would actually pattern as with Manchester: heavily degraded, relative to TGD-*give*, but less degraded than PDAT-*refuse*. Given this, it is taken to be an open question whether Liverpool English $_p$TGD may also be derived from $_p$GTD (via object shift or similar mechanism). If this were the case, it might better fit the usage patterns that we will see in the Twitter data, where the frequency of $_p$GTD is inflated at the expense of $_p$TGD rather than $_p$PDAT. Whether the Liverpool $_p$TGD may also be derived—as with Manchester—from the $_p$GTD, will be addressed through the gradient acceptance judgements in the survey, presented in Chapter 5.

THEME-passives

A link between the TGD and THEME-passives (TPASS) (9a) has been put forward, initially evidenced by the apparent co-occurrence of the two structures in geographical space: where you find TPASS, you invariably find TGD (Haddican & Holmberg, 2012).

The passivisation of the pronominal TGD produces the structure in (9a),

57

while (9b) represents the equivalent GTD-derived passive.

(9)   a.   It was sent me (TPASS)
      b.   I was sent it (PASS)

Haddican (2010) finds both structures to be accepted by speakers in Manchester. The availability of both (9a) and (9b) leads to the possible classification of the Manchester dialect as a variety that allows 'symmetric' passives, that is, language systems that permit either object to passivise (Woolford, 1993). In this respect, the Manchester dialect system is akin to *symmetric* passive languages such as Kinyarwanda, Norwegian and Swedish. Meanwhile, Standard English belongs to the set of *asymmetric* language systems that permits only "the accusative object with the highest thematic role [to passivise] and no intransitive impersonal passives are possible" (Woolford, 1993, p.1).

Haddican and Holmberg (2012) further characterise the 'inventory of grammars' available in Manchester, and the Northwest, as follows:

Grammar	THEME-GOAL orders in Active Contexts	THEME-GOAL orders in Passive Contexts
1	*	*
2	OK	OK
3	OK	*
4 (unattested)	*	OK

Table 2.1: Inventory of grammars in Manchester.
Reproduced from Haddican and Holmberg (2012)

Crucially for the current investigation, TPASS with definite NP derived subjects as in (10) is only available, according to Biggs, in the Liverpool area. However, it is reported that TPASS is accepted 'in a limited set of environments' in Chester - there, (10a) was accepted but not (10b) (Biggs, 2016, p.3, fn.3). It seems likely, given this, that THEME passivisation, at least for a subset of speakers, exists outside of Liverpool.

(10)   a.   The compass was given the boy ($_{DPDP}$TPASS)
       b.   The tape was sent the music studio ($_{DPDP}$TPASS)

It is unclear what is driving the acceptance of $_{DPDP}$TPASS and $_{DPDP}$TGD. Biggs (2018) argues that there is a silent functional element x which possesses some of properties of an overt preposition (specifically *to* or *at*), a point I will return to in the coming sections.

2.3 Geohistory of the ditransitive

2.3.1 Ditransitivity from Middle English to Modern English

The dative alternation is attested throughout the history of English, as far back as the Old English period (OE, ca. 5th-11th century AD), when word order was considerably freer (De Cuypere, 2015; Zehentner, 2016). With ditransitives, the order of GOAL and THEME arguments was in alternation with both full noun phrases and pronominals (De Cuypere, 2015). It is generally accepted that the loss of case marking between the OE and ME periods gave rise to a greater competition from prepositional constructions (Zehentner, 2019).

In the modern English period, whilst accounts in the literature have tended to label the pTGD as a minority dialectal variant, this may be due, as Yáñez-Bouza and Denison (2015) suggest, to its being limited to situations where both objects are pronouns, a structure that tends to be found in speech rather than in writing (Biber, Johansson, Leech, Conrad, & Finegan, 1999) (Figure 2.3).

Distribution of pronoun sequences as direct and indirect object across registers; occurrences per million words

each ▪ represents 10   ▫ represents less than 5

	CONV	FICT	NEWS	ACAD
indirect object + direct object	▪▪▪▪	▫	▫	▫
direct object + *to*-phrase	▪▪▪▪▪▪▪▪	▪▪▪▪▪▪	▪	▫
direct object + indirect object	▪▪	▪	▫	▫

Figure 2.3: Chart showing distribution of ditransitive types, and the double-object's tendency to occur in speech rather than in writing.

As Biber, Gray, and Staples (2016) point out:

> "Conversational participants share time and place, and they normally also share extensive personal background knowledge. As a result, colloquial features like pronouns and vague expressions are common." (Biber et al., 2016, p.1).

Being predominantly limited to spoken contexts has meant that the quantitative distribution of pronominal ditransitives (pDit) has become clear only recently with the advent of large-scale corpora transcribed from audio recordings. Such corpora made possible Siewierska and Hollmann's (2007) analysis of the pTGD in present-day Lancashire which found that it does in fact occur

with considerable frequency there.[7] This picture is again supported by Gerwin (2013), who finds high $_\text{p}$TGD use predominantly in the North West of England and the Midlands across a range of corpora; the International Corpus of English (ICE), The Freiburg Corpus of English Dialects (FRED) and the British National Corpus (BNC). The level of use of $_\text{p}$TGD is such that, for Gerwin, the label given to the $_\text{p}$GTD as the 'canonical' order is called into question: "it appears that the alternative double object construction was, in fact, the 'canonical' pattern until far into the 20th century" (Gerwin, 2014, p.183).

The tendency of pronominal ditransitives to occur primarily in speech rather than in writing has implications for historical corpus study. The rarity of pDits in written English means that finding sufficient examples may be challenging (Siewierska & Hollmann, 2007). This is compounded by fact that syntactic features are already infrequent (compared to phonological features), and yet more so when focusing on smaller geographical areas. These facts should be considered when looking at the relative historical rates of pronominal ditransitives presented in Yáñez-Bouza and Denison (2015).
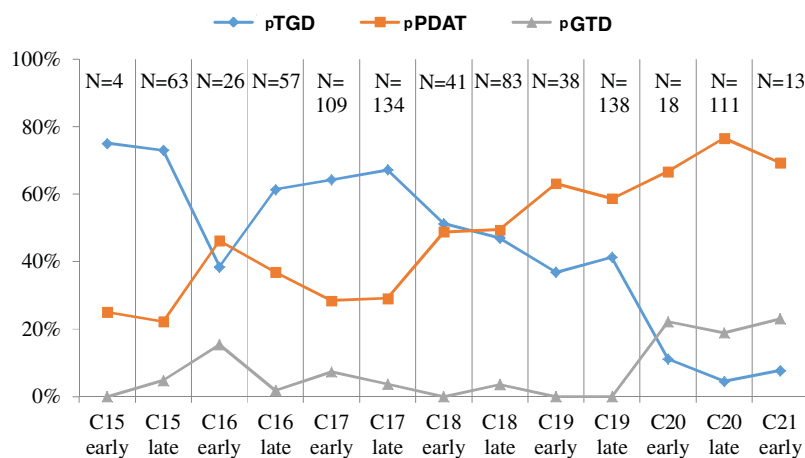
Figure 2.4: Historical distribution of pronominal ditransitives
source: Yáñez-Bouza and Denison (2015)

The plot shown in Figure 2.4 appears to show that $_\text{p}$GTD only recently came to prominence in British English. However, this is likely misleading as the historical data on which the plot is based are not (and cannot be) geographically representative. Historical studies of language change such as this necessar-

---

[7]Siewierska and Hollmann (2007) used FRED, BNC, SED incidental recordings, Helsinki Corpus of BrE Dialects.

ily draw on the available data and achieving geographical representativity is challenging.

I draw attention here to the notable spike in $_p$GTD usage that registers in the plot in the early 16th century. Recent studies have highlighted that $_p$GTD was, in fact prevalent in Middle English (Gast, 2007), while De Cuypere (2015) reports that the distribution of TGD to GTD was relatively even in Middle English. Zehentner (2018) makes the point that the apparent preference towards $_p$TGD reported by Yáñez-Bouza and Denison (2015) may be explained by the exclusive focus on the THEME pronoun *it*, which may be biased towards THEME-GOAL orders due to a general preference for phonologically 'heavier' elements to precede those that are 'weak'.

Gerwin (2014) looked specifically at the distribution of pronominal ditransitives in the Freiburg English Dialect Corpus (FRED) and the British National Corpus (BNC).[8] The results are reproduced in Tables 2.2 and 2.3:

FRED	$_p$GTD	$_p$PDAT	$_p$TGD	Total
Southeast	5 (11%)	41 (87%)	1 (2%)	47 (100%)
Southwest	1 (3%)	30 (91%)	2 (6%)	33 (100%)
Midlands	1 (7%)	10 (67%)	4 (27%)	15 (100%)
North	10 (36%)	13 (46%)	5 (18%)	28 (100%)
Total	**17 (14%)**	**94 (76%)**	**12 (10%)**	**123 (100%)**

Table 2.2: Combinations of two pronominal objects in the FRED corpus (4 main regions), adapted from Gerwin (2014, p.178)

BNCreg	$_p$GTD	$_p$PDAT	$_p$TGD	Total
Southeast	22 (19%)	90 (78%)	4 (3%)	116 (100%)
Southwest	15 (31%)	26 (54%)	7 (15%)	48 (100%)
Midlands	12 (13%)	57 (63%)	21 (23%)	90 (100%)
North	60 (54%)	32 (29%)	19 (17%)	111 (100%)
Total	**109 (30%)**	**205 (56%)**	**51 (14%)**	**365 (100%)**

Table 2.3: Combinations of two pronominal objects in BNCreg (4 main regions), adapted from Gerwin (2014, p.178)

Despite overall low counts, Gerwin's (2014) data nevertheless agree in showing an increase in the rate of $_p$TGD in the North and Midlands regions. Interestingly, Yáñez-Bouza (2016a) reports that there was considerable attention towards which order should be used with ditransitives in early 18th Century

---

[8]I discuss FRED and BNC in §2.6.2 and §2.6.1.

grammar guides. Here, opinions differ; Maittaire 1712:183) indicates that the word order is relatively free:

> "When both cases are Pronouns, 'tis no great matter, which is first; as thou gavest them me, or thou gavest me them: Yet in some, the Dative rather precedes; in others the Accusative: as he gave me this, he gave it me." (Maittaire 1712:183) cited in Yáñez-Bouza (2016b, p.150)

Meanwhile, other texts declare that the ₚGTD order should be avoided in the English of England, deriding it as 'vulgar Scots'.

> Give me it, show me it; Sc.—Give it me, show it me. The former is Scotch, the latter English. (Mitchell, 1799, p.54), cited by Yáñez-Bouza (2016b, p.156).

The fact that the ordering of pronominal ditransitives drew such attention in grammar guides is itself telling. It is an indication of the salience of variance in the order of the structure. Additionally, the link to Scots is notable, as Stevenson (2016) showed and as we will further see, in the current data, ₚGTD is the preferred order in the Scottish Twitter data.

2.3.2   The Survey of English Dialects (SED)

The Survey of English Dialects (SED) (Orton & Dieth, 1962) was a large-scale and comprehensive study of the dialects spoken in England, conducted over about ten years from the early 1950s. The primary aim was to document and analyse regional variation in English speech before these dialects were significantly altered or disappeared due to increasing mobility, urbanisation, and the spread of standardised English.

The survey covered 313 localities across England, chosen such that a wide range of dialects would be included, and focused on *Non-mobile, older rural males* (NORMs). The idea for choosing these subjects was that they would be most representative of local speech practice; less likely to have been influenced by processes of standardisation.

Phonological, lexical and grammatical data were collected. Most relevant to the present study are the data concerning [DATALT] and *definite article reduction* (DAR). The distribution of DAR and its various manifestations are discussed in 4.6, for [DATALT], subjects were asked:

(11)   Jack wants to have Tommy's ball and says to him, not Keep it! but . . .
        (SED, Questionnaire, IX. 8. 2)

Subjects' responses were recorded as either:

(12)   a.   give it me
        b.   give me it
        c.   give it to me

The geographic distribution of these responses is plotted in Figure 2.5. Here, the preference for ₚGTD (_me it) across the East and North East, is revealing. The widespread nature of ₚGTD in the SED is further indication that its preference has considerable time-depth. Further, the distribution appears to align quite strikingly with the known extent of the Danelaw; the area administered by Scandinavian settlers around the 9th Century. This distribution led Gast (2007) to speculate that the ₚGTD may have its roots in contact with Old Norse. He cites that ₚGTD is the preferred option in Icelandic, modern-day Norwegian and that there was a clear tendency towards this order in Old Norse.

The influence of Old Norse on the development of English has been much-discussed, and is at times controversial. The radical claim that modern English is actually directly descended from Old Norse (Emonds & Faarlund, 2014) has found some support (Holmberg, 2016), but is roundly rejected elsewhere. For example, Bech and Walkden (2016) conclude that Emonds and Faarlund's (2014) "manifesto fails to convince on methodological, empirical, and theoretical grounds" (p.93). In any case, it is clear that the impact of Old Norse on the development of English was substantial. In the light of this, the question of the present distribution of ₚGTD may be brought to bear. This I discuss in the Twitter results.

Regarding the distribution of ₚTGD (_it me), the pattern is clear in the SED data mapped by Kirk et al. (2014): an area spreading from the southern Midlands across South and West Yorkshire and up to Lancashire and the North West with additional pockets along the south coast. The prepositional variant ₚPDAT (_it to me) is most dominant in the South West, London and pockets along the Norfolk coast. The distribution of [DATALT] variants, reported in the SED leads Upton (2006) to assert that:
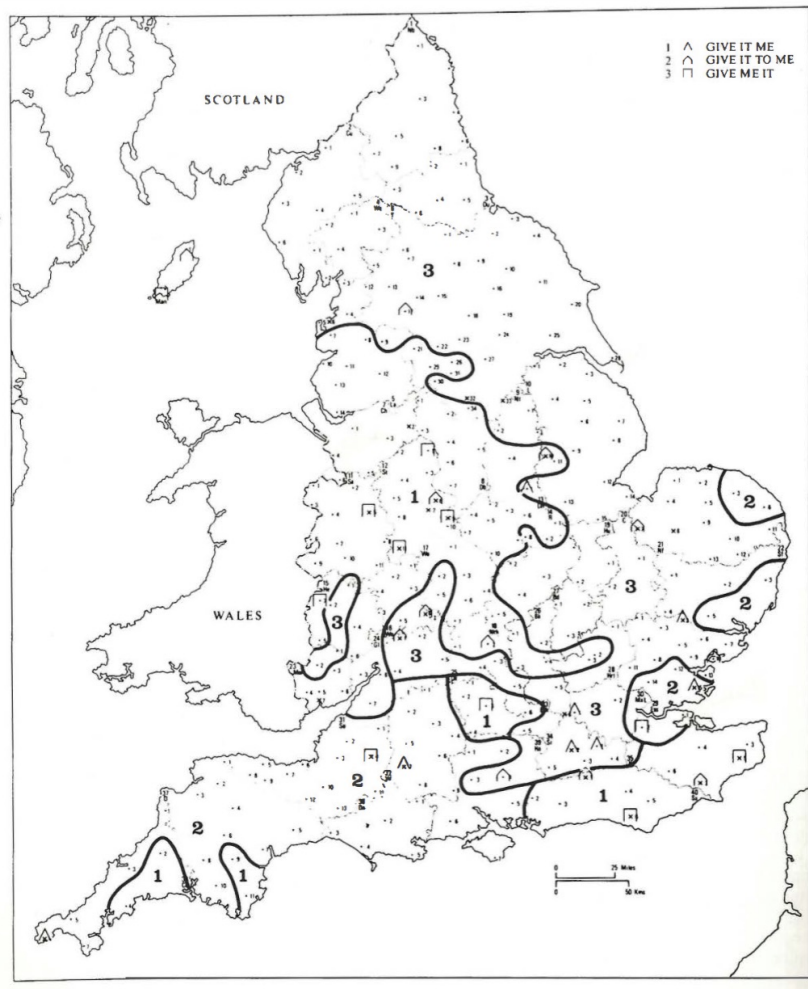
Figure 2.5: Pronominal ditransitives in the Survey of English Dialects
1 = TGD, 2 = PDAT, 3 = GTD
Source: Kirk et al. (2014, p.132)

"No better example exists of a syntactic puzzle than the quite definite regional preferences for the standard *give me it* in northern and eastern England, a non-standard *give it me* in the West Midlands, and an expanded *give it to me* in the south-west, as recorded by SED" Upton (2006, p.329)

It is upon this *syntactic puzzle* that the present Twitter data aims to shed new light.

2.3.3 Recent accounts of ditransitive variation in Britain

The English Dialects App

The *English Dialects App* (Britain et al., 2018; Leemann, Kolly, & Britain, 2018), was developed as a smartphone application and marketed to a broad audience in the UK. The application was essentially a vehicle to deliver a language survey and a method to gather user-recorded segments by having them read a passage of text. This approach is an exciting development in the field of dialectology, offering a means to gather very large quantities of data without the need to conduct extensive and costly fieldwork.

A component of the application gamifies the process of eliciting data by offering the user a prediction of their location based on their submitted data. In this respect, there is considerable scope for its further development in the light of recent advances in machine learning, which could offer powerful ways to process and use the data collected by the application.

One aim of the survey questions was to assess the extent to which they agreed with data from the SED. Of interest to the present work, one question sought speaker judgment on the $_{\text{p}}$TGD, the results of which are presented in Figure 2.6, alongside a map created from SED data.
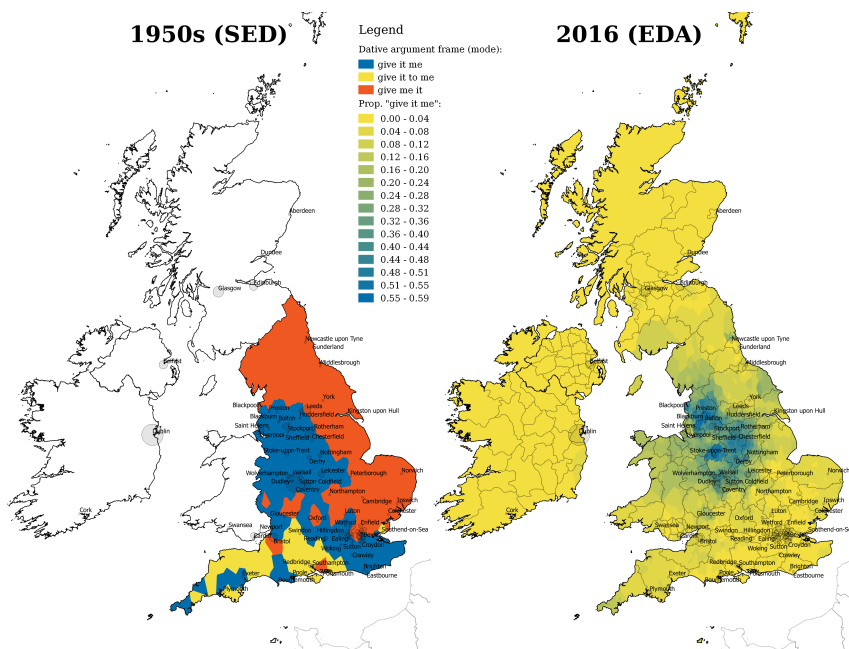
Figure 2.6: Pronominal ditransitives in the Survey of English Dialects
From Britain et al. (2018, p.88)

In the EDA, $_p$TGD is recorded in a concentrated area of the Midlands and North West. There is a clear reduction in the area that it covers compared to the SED map, which shows use extending through to the South East. There is a notable absence of use in a small pocket around Liverpool, which matches Stevenson's (2016) Twitter results, which instead show Liverpool as $_p$GTD rather than $_p$TGD dominant, shown in Figure 2.7.

RELATIVE REGIONAL USE

□pTGD ▨pGTD ▧pPDAT
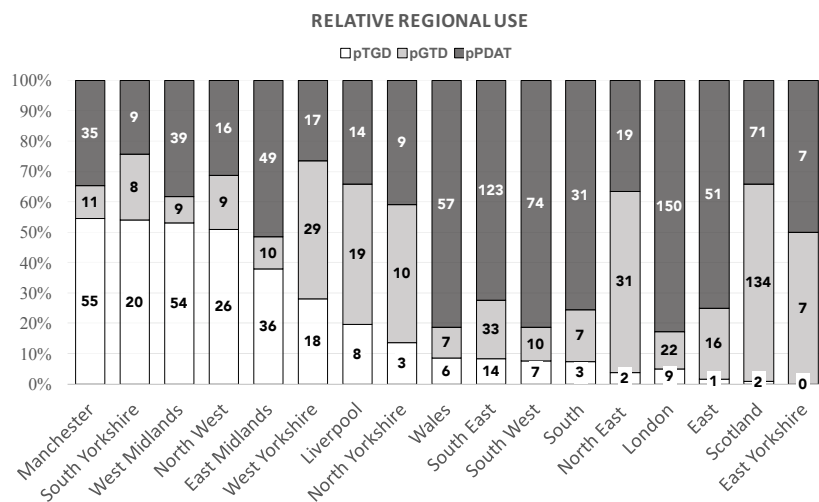
Figure 2.7: Relative rates of pronominal ditransitives, with the verbs *gave* and *sent*
Source: Stevenson (2016)

Our Dialects Project

The *Our Dialects Project* (ODP) (MacKenzie et al., 2022), based at the University of Manchester, tasked successive cohorts of undergraduate students to administer a language survey as part of their introductory linguistics course, originally by paper and subsequently online. Students recruited respondents via their own networks and via social media. This method was effective in gathering a lot of data, quickly becoming the largest of its kind with over fourteen thousand respondents.

It can, however, be problematic to consult university students on dialect use, as they frequently have moved away from the place they grew up and are at a particularly fluid time of linguistic adaptation as they interact with peers in a new and aspirational environment. This may not always be a disadvantage and issues may be mitigated through the manner in which a survey is presented to participants; the instructions and background information can also play a

crucial role. For instance, the rating scale for a particular judgement asked the participant whether they *would use a structure*, or if they would *hear others in their area use a structure*, which is quite different from asking only *how good a sentence sounds to them*.

Like the EDA, survey questions were geared towards providing an updated account of the results of the SED. For the pronominal ditransitive, the survey again focused on the ₚTGD, with respondents asked to evaluate the string *give it me*. Respondents were tasked with rating the string on a scale from (a) to (e) as follows:

(a) I'd say this myself

(b) I wouldn't use it, but some people from my area do

(c) I've heard some people use this form

(d) A speaker of English might say this, but I haven't really heard it

(e) No native speaker of English would say this

The results for the ₚTGD in the ODP are presented in Figure 2.8. The yellow-shaded areas represent respondents who answered either (a), or (b), while the isogloss lines are reproduced from the SED map of the same feature (Orton et al., 1978), which is based on whether SED respondents themselves used the structure (as presented in §2.3.2).

Grouping responses (a) and (b) together, in this way, is more lenient than the SED method, and may partly explain why the yellow-shaded area exhibits a considerably broader spread to the North East than the SED isogloss. This may also explain the discrepancy between the ODP map and the EDA map, presented in Figure 2.5. As MacKenzie et al. (2022, p.59) point out, it is possible that, if SED informants had been given the opportunity to rate ₚTGD outside of it being their "primary variant", that the SED would have shown a wider distribution. It is interesting to see this wider spread of ₚTGD (secondary) acceptance as it potentially provides a more faithful record of its broader use. It would, however, be interesting to see an equivalent map for response (a) in isolation, which would be more directly comparable to the SED data.

In the map, it is again possible (as with the EDA map above) to make out a lighter shaded area around Liverpool, where ₚTGD acceptance is less strong. This distinction between Liverpool, Warrington and Manchester is supported
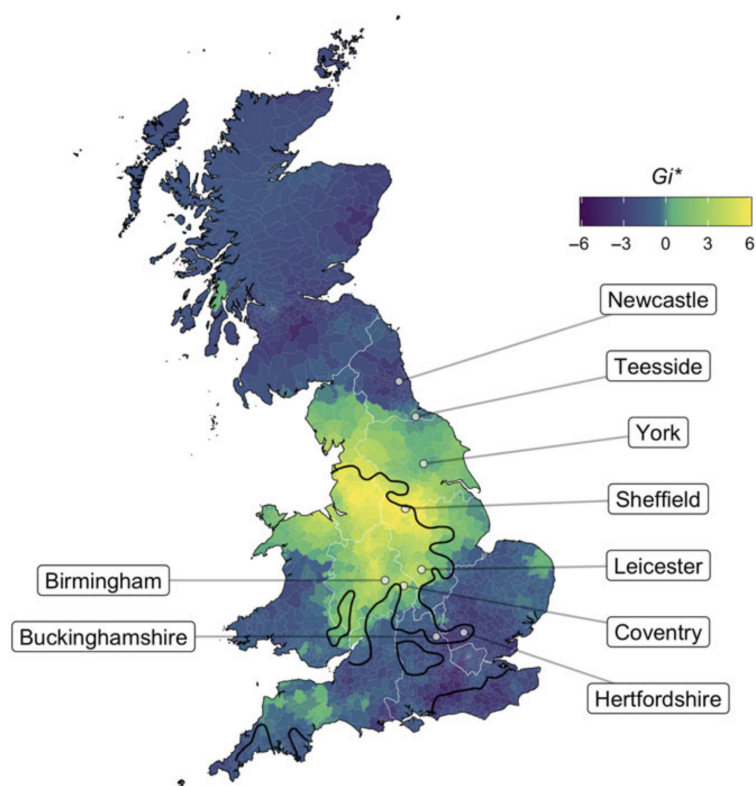
Figure 2.8: Proportion of respondents who reported that they, or those in their area use ₚTGD, in *Our Dialects Project* (MacKenzie et al., 2014), with isoglosses from the Linguistic Atlas of England (Orton et al., 1978)
Source: MacKenzie et al. (2022, p.59)

in the numerical data presented in Table 2.4 below, which measures response (a) against all responses.

Postcode area	ₚTGD (n)	Total (n)	% ₚTGD
Liverpool (L)	130	482	27%
Manchester (M)	665	1113	60%
Warrington (WA)	247	504	49%

Table 2.4: Proportion of people who report using ₚTGD themselves from the *Our Dialects Survey* (MacKenzie et al., 2014)
Data supplied by personal correspondence from George Bailey.

Here we can see that 27% of Liverpool respondents reported using ₚTGD themselves compared to 60% of Manchester respondents. These data are strikingly close to the usage data that were reported in Stevenson (2016), shown in Figure 2.7 above. This is reassuring, given that we are comparing speaker

self-reporting to usage on Twitter.

The data also collectively confirm that Liverpool does not follow the rest of the North West in the high acceptance of the $_P$TGD. This has consequences for Biggs's (2018) analysis of TGDs and P-DROP in Liverpool, which relies on speakers accepting $_P$TGDs. Of course, at least a quarter of Liverpool participants in the ODP do accept the structure. Additionally, as mentioned earlier, Biggs's (2018) claim was not that this grammar necessarily exists in a particular location:

> The claim [...] is not that certain linguistic forms will only occur in precise geographical regions. Rather, the claim is that the availability of such a construction will correspond to systematic and productive variation in other aspects of that individual's grammar. (Biggs, 2018, p.4, fn.5)

The construction to which Biggs refers is the TGD with full DP objects ($_{DPDP}$TGD). This point is crucial, as it predicts that, if there is a location where there is a concentration of aspects individuals' grammars which are predictive of the $_{DPDP}$TGD, then we should find greater acceptance of it there. As we will see, such conditions are met in the intermediate region between Liverpool and Manchester.

Variation by verb

So far, it is unknown the extent to which pronominal ditransitive use varies regionally by verb, though we have some predictions based on Levin (1993), who categorises ditransitive alternations by verb type, as discussed on page 55. Knowing the relative rates of pronominal ditransitives by verb type is important because it would provide a quantitative measure for a crucial part of the analysis conducted by Haddican (2010) and Biggs (2018). Accordingly, variation by verb type is analysed in the results of the present work.

2.4 Preposition and determiner dropping

Myler (2013), 'come the pub' and preposition-dropping in the North West of England

Preposition dropping has been found in a number of places in the North West of England. Structures like those in (13) are reported to be readily available

to speakers in Ormskirk, located roughly 13 miles north of Liverpool (Myler, 2013).

(13) a.   I want to go Chessington.
      b.   John came the pub with me

Myler (2013) investigated the alternation between preposition-dropped and non-preposition-dropped alternatives, finding that the alternation is available with motion verbs GO, RUN, DRIVE, JOG, POP, and NIP, as well as 'at least' the ditransitives TAKE and SEND. In terms of potential sociolinguistic investigation, P-DROP appears to happen below the level of conscious awareness. This is interesting as, outside of the immediate Liverpool sphere of influence, P-DROP is quite noticeable for those speakers who do not use it, as we will see.

A crucial aspect of Myler's (2013) analysis is that the GOAL argument ('the pub' in 'come the pub') carries a combination of properties usually associated with both *direct objects* and *PP objects*. In light of this, Myler argues for the existence of a silent element 'TO' that introduces the GOAL argument as the silent head of a PP.

(14) Myler's (2013) TO is licit with most *motion* verbs

      a.   I nipped [TO] the shops
      b.   Sally drove [TO] the cinema

The GOAL argument then moves to direct-object position where it receives accusative case. Critically, the hypothesised null element TO is counterpart to the overt preposition 'to', in that it cannot take the place of other prepositions, but is only available in constructions with directional readings.

(15) TO is licit only in replacing *to,* with a directional *motion* reading in Ormskirk, according to Myler (2013).

      a.   I'm going the library
      b.   *The path goes the front door
      c.   *I'm working the library
      d.   *He put it the bin

For Myler (2013), this disqualifies a potential reading of this case of preposition drop as happening post-syntax, in phonology.

Ditransitives are also found to be licit with P-DROP (a), (b), which may also be passivised (c) and may occur with full DPs (d).

(16)   a.   Me nan sent me the shops
       b.   John took me the zoo
       c.   I was sent the shops by my dad
       d.   I took the boys the zoo

The fact that we see ditransitives available within the P-DROP paradigm is informative. Speakers who use such structures, it would seem, may be more likely to also accept TGDs with two full DP objects, as Biggs (2018) reports for speakers in Liverpool.

(17)   I sent the letter the bank

The difference here revolves essentially on the animacy of the THEME argument. We might expect, therefore, to find a correlation between speakers who have P-DROP, and speakers who accept $_{DPDP}$TGD, indeed, this correlation is reported by Haddican (2010). According to Haddican (2010), acceptance of $_{DPDP}$TGD depends on the acceptance of preposition dropping in other contexts such as with verbs of motion (*come the pub*) in his Manchester sample.

Microvariation in the North West

Whilst Myler (2013) reports that it is "not known whether or to what extent constraints on the *come the pub* construction vary across the North West", Biggs (2018) does find a different set of constraints in her Liverpool sample.[9] The question of the variability of underlying constraints is, then, an intriguing one.

Innovative null case marker 'ϰ'

Biggs elaborates on Myler's findings in Ormskirk, arguing for the existence of a recently innovated null element in Liverpool English—that she labels 'ϰ'—as distinct from Myler's 'TO'. ϰ permits preposition dropping across a greater range of contexts than that found in neighbouring localities, using Haddican's (2010) study in Manchester and Myler's (2013) study in Ormskirk as principal points of comparison.

---

[9]As we will see, it is clear from the Twitter results, presented in chapter 3, that the construction is in fact widespread in the North West and parts of Staffordshire and the Midlands.

Biggs then goes into some depth developing an account of the properties of ϰ. Crucially for the analysis, whilst ϰ occupies the same position as overt prepositions 'to' and 'at', it is not morphosyntactically equivalent. As with Myler (above) this discounts a potential reading of preposition drop as being phonological in nature. However, whilst Myler argues that the absent preposition incorporates to the verb, Biggs instead argues that ϰ (in Liverpool) is an independent lexical entity (Biggs 2016, p.46) which remains structurally present and licenses the GOAL. This analysis is in part driven by the reported acceptance in Liverpool of structures involving *straight* modification, a classic diagnostic for the structural presence of a prepositional phrase (Emonds, 1972).

(18)   They went straight the pub (*Ormskirk, OK Liverpool)

According to Biggs, the functional range of ϰ is demonstrably broader and more complex than TO. Furthermore, there appears to be a generational divide in the acceptance of the criteria used to validate the existence of ϰ in Liverpool. Whilst participant numbers are small, this is taken as an indication that ϰ may in fact be a recently innovated feature of Liverpool English. Whilst assessing change over apparent time was not the principal goal of Biggs' paper, it is an intriguing suggestion. Indeed, the validation of the finding that ϰ/TO is in fact recently innovated and the subsequent investigation of the scope and range of the diffusion of ϰ across social and geographical space is a focus of the current investigation (see research question 2).

These facts about Liverpool English contra Manchester English again supply an additional diagnostic when investigating whether a given intermediate variety is patterning more like Manchester or Liverpool, and potentially the extent to which ϰ may have spread.

Pseudo-incorporation versus separate functional element

Biggs assigns ϰ as a functional lexical element, as opposed to Myler's TO, which he argues incorporates to a narrow range of compatible verbs in order to be licensed. ϰ is argued to operate independently of the verb, and is able to license the GOAL.

Whilst Biggs suggests that the availability of P-DROP in the context of TGD-fullDP, and passives with fullDPs in derived subject position plausibly extends

to pronominal TGDs, it may also be that $_\text{P}$TGDs are a special case, akin to $_\text{P}$GTDs even where TGD-fullDP is licit. TGD-fullDP may become available to speakers by analogy to P-DROP-fullDP whether or not they have $_\text{P}$TGD in their grammar.

Biggs (2018) finds that $_\text{P}$TGD is accepted in Liverpool, and so, that it apparently follows the rest of the North-West in being $_\text{P}$TGD-dominant. However, there is some discrepancy between Biggs' findings and those reported elsewhere. Stevenson's (2016) data show that, in terms of usage data, Liverpool and its immediate environs are not, in fact, $_\text{P}$TGD-dominant, a finding that is corroborated by MacKenzie et al.'s (2022) survey data (as we will see, in the Twitter and survey data, Liverpool speakers show preference for $_\text{P}$GTD).

This said, it is clear that whilst $_\text{P}$TGD may not be dominant in Liverpool, it remains quite acceptable to a proportion of the population there, albeit, as Biggs (2018) argues, potentially with a distinct underlying structure to that found for $_\text{P}$TGD in the neighbouring region.[10] Biggs (2018, p.3) concludes, referring to $_\text{P}$TGD that "a single surface string [...] can correspond to very different underlying syntactic structures in regional varieties". This conclusion plays directly into MacKenzie's (2019) concept of 'covert representational variability', the idea that the underlying representation of the same surface string may be distinct between speakers, between and within speech communities. Whether an alternative underlying structure for TGDs exists in a specific geographic location (such as Liverpool), elsewhere, or is characteristic of the grammars of specific speakers across multiple locations feeds RQ(9).[11]

RQ (9) To what extent is there evidence for $_\text{DPDP}$TGD in the location predicted from the Twitter data? Is there a grammar in the North West for which TGDs are underlyingly PDAT-P-DROP and is this grammar geographically restricted?

Constraints on PD-DROP in London and the South East

Whilst the North West exhibits P-DROP, recent work has highlighted the existence, in London and Kent of preposition-determiner drop (PD-DROP), where the determiner is absent as well as the preposition (19).

(19)   I'm going shops

---

[10]Biggs (2018) reports $_\text{P}$TGD as accepted in Liverpool and about 25% of Liverpool respondents from MacKenzie et al.'s (2022) survey do, too.

[11]I return to this point in §4.8.1 and on page 278.

The structural properties of PD-DROP are covered extensively in Bailey (2018b); Hall (2019). The two accounts report a broad agreement on the structural properties of PD-DROP, though differ in their syntactic analysis. Structurally, PD-DROP is tightly constrained to the verbs *come* and *go*.[12] With respect to the GOAL noun phrase, Hall lists the following main features:

1. The definite article is obligatorily dropped

2. It cannot be plural

3. It cannot be modified by a PP, a relative clause, or an adjective

4. It must be a familiar/institutional place name

Hall (2019) argues that these facts about London PD-DROP align with a *pseudo-incorporation* (PI) analysis, following that proposed for PD-DROP phenomena in Greek (Gehrke & Lekakou, 2013). PI dictates that not only are the preposition and determiner absent at the surface level, but that there is a 'radical absence of PP structure' in the derivation (Gehrke & Lekakou, 2013; Hall, 2019).

As with P-DROP in the North West, the strongest evidence for the absence of the PP in PD-DROP, is its failure to pass the robust *straight*-modification diagnostic. If a PP is present, it should be possible for it to be modified by *straight/right*.

(20) a. he went straight to the pub
      b. *he went straight pub

As shown in the previous section and repeated below, this contrasts with Biggs's (2018) finding in Liverpool P-DROP, which does permit SM, but, interestingly, not with Myler's (2013) account of Ormskirk P-DROP.

(21) he went straight the pub (OK Liverpool, *Ormskirk)

Bailey, however, argues that in Kent PD-DROP, there is, in fact, full DP and PP structure present. Whilst Bailey and Hall follow alternative explanations for the respective varieties of PD-DROP, the characteristics of both nevertheless

---

[12]For Bailey (2018b), the restriction on the verb is not as strict; whilst the majority are *come* and *go*, she finds that it is also possible with 'semantically weak' verbs, giving the example: "This train calls Sittingbourne, Rainham..."

set South Eastern PD-DROP in sharp distinction to that described for P-DROP in the North West.

Interestingly, Bailey reports (footnote 6, p.52) that the PD-DROP found in Manchester appears to be of the same type as that found in the South East, which she suggests "adds further weight to the claim that this is not geographical spread of a feature". The observation is made as a side-note, based on one example ("You've never been Benidorm"), however. If it is the case that the Manchester variety of PD-DROP is of the same type as the South East, it is not clear why this should be evidence, in itself, against geographical spread. However, the argument that PD-DROP is widespread across multiple, global, varieties of English, as well as other languages, is convincing, in terms of the feature's capacity to emerge independently.

While Bailey finds the variety of PD-DROP in Kent to be structurally similar to that described in London, she suggests that the two owe their presence to different factors. She also acknowledges the potential role for contact in the development of PD-DROP, citing the influence of Multicultural London English (MLE) on London PD-DROP: MLE has amongst its input varieties Caribbean and Subsaharan African Englishes (Cheshire et al., 2011; Fox et al., 2011), varieties which, themselves, are established as exhibiting PD-DROP (Bailey, 2018b).

2.5 The geohistory of [P/D-DROP]

There are multiple accounts of preposition/determiner dropping in the dialect literature. Edwards and Weltens (1985, p.114) observe that P-DROP, specifically with the prepositions *on* and *to*, is a prominent feature of British dialects. Meanwhile, Ramisch (1997, p.224) points out that, in responses to the SED question *What do good people do on Sunday?*, the preposition *to* may be dropped, resulting in *they go church*. This feature he ascribes predominantly to the South West of England. Meanwhile, Upton, Parry, and Widdowson (1994) report that the expression *go church* is attested in the SED in regions across England: Suffolk, Kent, Cheshire, Derbyshire and Staffordshire, Cornwall, Devon, Dorset and Somerset. Similarly, *go school* (SED VIII.6.1) is attested in Cheshire, Staffordshire, Sussex, Somerset, Devon and Cornwall Upton et al. (1994, p.504).

Early accounts in the dialect literature

Going further back, Ellis (1889, p.410) reports "the singular omission of the preposition ['to'] in [Cheshire], which extends also to [North Staffordshire]", as shown in the examples in (22).

(22) Examples of preposition dropping from Ellis (1889, p.410) in the North West Cheshire and North Staffordshire:

    a. *gu, be'd* (go to bed)
    b. *l kam àr áis* (he'll come to our house)
    c. *ev a̯gu :kræ'u* (if I go to Crewe)
    d. *s gu skjiu* (let's go to school)

This albeit brief account of [P/D-DROP] in the North West from the 1880s adds a useful time dimension to the story. The absence of the prepositions *on* and *to* is also reported in Evans's (1881) account of the dialect of 19th century Leicestershire: "With regard to other parts of speech, the most noticeable peculiarities are, perhaps, the position often assigned to the adverb in a sentence, and the frequent omission of the prepositions 'on' and 'to.'" (Evans, 1881, p.44).

(23) From Evans (1881, p.44) (Leicestershire):

    a. I hope to soon get church
    b. He goes Bos'o'th Wednesdays

An earlier report of PD-DROP is provided by Evans (1848), in *The Salamanca Corpus: Leicestershire Words, Phrases and Proverbs*. In the entry for *going*, we find:

(24) GOING, p. For *going* to. 'Are you going Leicester?'.

Ellis (1889, p.476) records the following example in the South Midlands (Market Drayton) in 1882:

(25) (an je dù n) have you done ?-(E) s gu skjiu) let's go [to] school.- (dhe iimer wii) the eamer [shorter] way.

There is no equivalent record of PD-DROP in Ellis (1889) in the descriptions for

London or South East varieties. Similarly, there is no mention of the feature in Parish, Shaw, and Masters's (1888) *A dictionary of the Kentish dialect.*

However, a search of the *Old Bailey Corpus 2.0* (Huber, Nissel, & Puga, 2016), which is a rich record of transcripts of court proceedings in London from 1720-1913, compiled into a searchable corpus of 24.4 million spoken words, revealed:

(26)   a.   and falling asleep I asked him <u>to go Bed</u>, and he did so (1737)

        b.   on <u>my going Dean-street</u> I saw several people (1812)

        c.   I <u>went trample stall</u> to get some apples (1871)

        d.   he should have <u>to go prison</u> or Paris (1880)

        e.   I was released on the Tuesday and <u>went bank</u> on the Friday (1907)

Whilst the trials took place at the Old Bailey, in London, the original birthplace of the speakers it is not clear, however. This said, we do have an earliest attestation of PD-DROP uttered in London in 1737.

More recent accounts

Braber and Robinson (2018) report preposition deletion in present-day East Midlands: "omission of the preposition to, especially in common collocations such as with place names or habitual destinations like school, toilet or shops." (Braber & Robinson, 2018, p.99).

(27)   From Braber and Robinson (2018, p.99):

        a.   I went York Road school (Swadlincote)

        b.   there was a couple as went school with us (Swadlincote)

        c.   there's a lot more women who go football (Leicester)

        d.   my eldest daughter went university (Coalville)

Harris (1967) reports widespread preposition dropping in the dialect of South Zeal, Devon. Specifically, they note (p.133), that "[t]he phrases indicating relationship of place which can occur without a preposition are many and varied". Examples include:

(28)   from Harris (1967, p.132-133), examples of [P/D-DROP] in South Zeal, Devon:

        a.   /jyd katj fas trejn "barn— sto—pi/

You'd catch first train [to] Barnstaple.

b. /jy to. gow "wot mar— kot jy "majnd ty/
You can go [to] what market you mind to

c. /ðej got ty dri: gwen tek am ow— '"kam— m
They got two or three going take them [to] Okehampton

Vasko (2010, §10.3) gives examples of preposition dropping in the Cambridgeshire dialect from her corpus collected mainly during the 1970s.

(29) MG: You didn't have to go to Cambridge yourself very often?
      SC: I used to go Cambridge always. (Lt. Eversden)

Here, Vasko notes that the respondent replies with a non-expressed preposition, despite the fact that the interviewer uses one. This is interesting as it demonstrates a lack of style-shifting, which implies that the variant is operating outside of sociolinguistic monitoring (Labov et al., 2011; Smith & Holmes-Elliott, 2022).

Watts (2005) describes the variable omission of 'to' in Wilmslow, Cheshire (now part of Greater Manchester). The data show that, rather than being a feature of older speech, it was the younger informants who showed the most frequent dropping of the preposition. Her data appear to show that preposition dropping is restricted to the verb *to go*. However, as a study of language use, the absence of use with other verbs does not mean that it was not possible.

(30) Examples of P-Drop in Watts (2005, p.323-324) (Wilmslow, 13 miles south of Manchester city centre):

      a. My dad needs to go the opticians (female, aged 9)
      b. We sometimes go the Bollin

(31) Examples of PD-Drop in Watts (2005, p.323-324) (Wilmslow, 13 miles south of Manchester city centre):

      a. I'll probably go Wilmslow (male, aged 10)
      b. The people that have gone high school (female, aged 11)
      c. I think we're just going Wales this year (female, aged 12)
      d. I can go Manchester whenever I want (male, aged 13)
      e. I prefer going Manchester because there's more shops (male, aged 21)

f.   We used to <u>go George and Dragon</u> but it's shut down now (female, aged 30)

In the examples provided by Watts (2005), there are only two (30) that are definitively P-DROP (with the determiner still present). This distribution would fit the expected pattern for Greater Manchester, with P-DROP associated more with the Merseyside region, following Biggs (2018) and Myler (2013). The rest either would not ordinarily have a determiner present, or are examples of PD-DROP, as in (31f).

In addition to being well attested in British dialects spanning back to at least the 1880s, [P/D-DROP] is quite frequent and widespread across global varieties of English, as catalogued in the *Electronic World Atlas of Varieties of English* (eWAVE) atlas (Kortmann et al., 2020).

Figure 2.9: Typological distribution of preposition dropping, adapted from eWAVE.
Square = 'Traditional L1 variety', Diamond = 'High-contact L1 variety', Triangle = 'English-based creole'
Source: Kortmann et al. (2020), `https://ewave-atlas.org/`
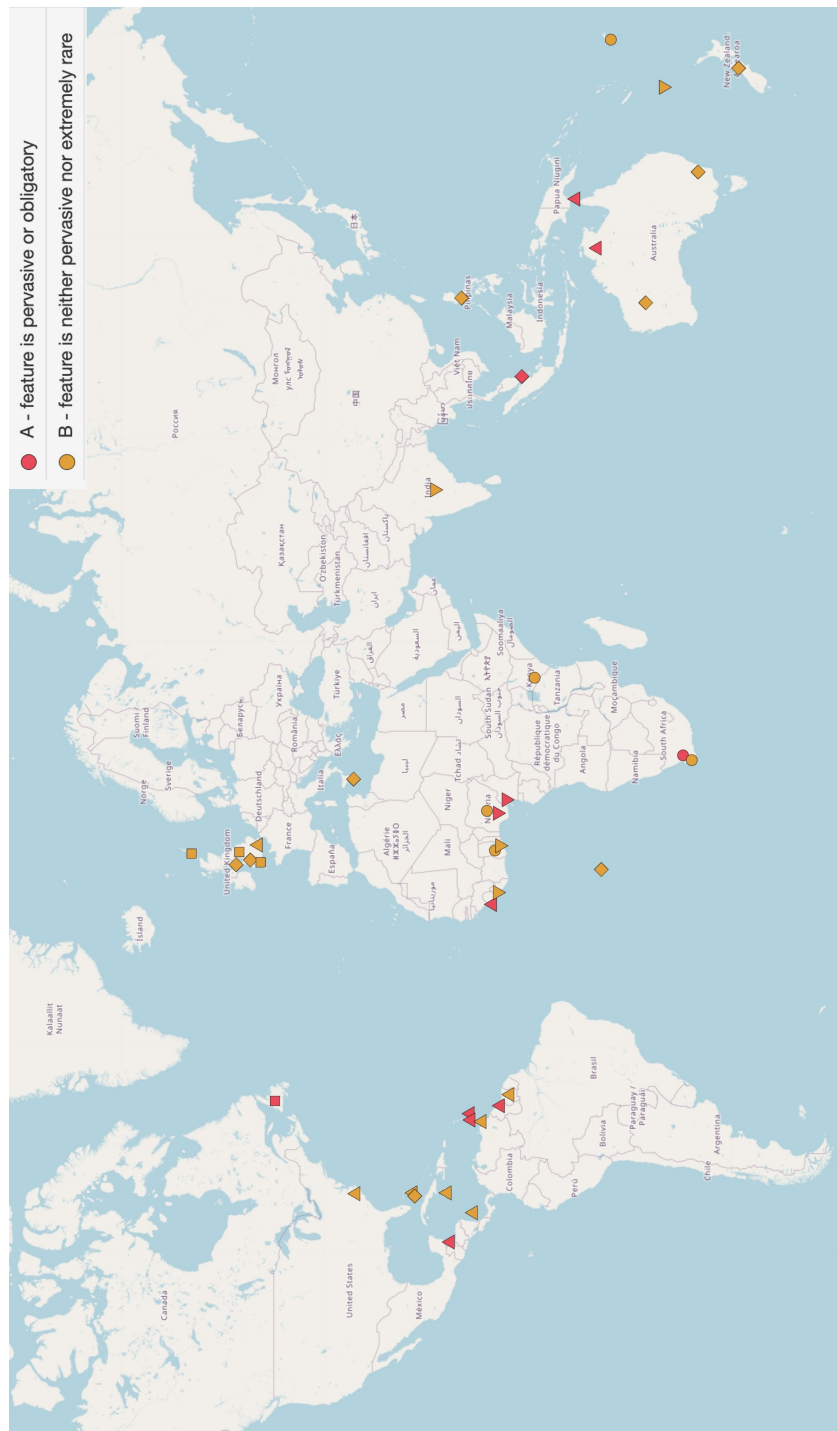
Figure 2.10: The global distribution of PD-Drop.

Square = 'Traditional L1 variety', Diamond = 'High-contact L1 variety', Triangle = 'English-based creole', Inverted triangle = 'English-based pidgin'

Source: eWave

The global distribution of PD-DROP-phenomena sets the stage for the distribution found in the UK. It appears to be a phenomenon that readily occurs in English(es). However, the typological classification for preposition omission used on eWAVE is quite broad, and extends beyond the scope of the classification for P-DROP that is the focus of the current study. The example provided as preposition omission for the North of England, for instance, is:

(32) eWAVE example of P-DROP for the North of England (which is not characteristic of the kind of P-DROP on which this investigation is focused)

    a. I'll see you Saturday

The broader scope of the classification that is used for eWAVE should be considered when interpreting the eWAVE maps shown on the previous pages (Figures 2.9 and 2.10).

The examples provided for Newfoundland English (NFE) are more interesting. NFE was one of the earliest British-settled areas (from the beginning of the 17th century), with a population formed predominantly from Southern Ireland and the South West of England (Kortmann et al., 2020).

(33) eWAVE examples of PD-DROP for Newfoundland

    a. She was born [at] home.
    b. I was born [in] St. John's
    c. taking them [on] tours down [to] the cape on the trail

Given that PD-DROP appears to be well-rooted in south-western British English, it is plausible that the Newfoundland examples have a shared lineage. If this were the case, it would suggest that the feature was present in the South West during the initial settlements of Newfoundland, from as early as the beginning of the 17th century. PD-DROP is also attested on eWAVE for the South West, adding to the previously discussed data from Harris (1967):

(34) eWAVE example of P-DROP for the South West of England

    a. I went Denny's house

Kent is recorded as exhibiting an 'attested absence' of preposition omission. It is not clear why this is the case, considering the feature's now well attested and widespread use in the South East (Bailey, 2018b; Gopal et al., 2021; Hall,

2019). London is shown as having PD-DROP, but interestingly, this is attributed to an 'English-based creole', presumably refering to the presence of the structure in MLE, a point I return to shortly.

In any case, it is in the light of both widespread historical accounts (reported above), and a global propensity for PD-DROP in other varieties of English, that we should consider the proposal by Gopal et al. (2021), that PD-DROP in the UK is the result of a relatively recent diffusion from the South East.

Gopal et al. (2021)

Gopal et al. (2021) investigate the geographical distribution of both [DATALT] and [P/D-DROP] in the UK and Ireland using a large Twitter dataset covering 20 months (October 2017 to May 2019). They confirm that PD-DROP is not geographically restricted to the South-East and Manchester, but well established across the Midlands and North West. Given what we have seen in the historical record—that [P/D-DROP] is attested as far back as the 1880s in Staffordshire and the East Midlands—the presence of the feature in these areas is not surprising. However, the focus of the paper is whether this distribution correlates with a simulated model of *gravity-based* diffusion from London along the 'urban corridor' towards the North-West, favouring large urban centres before interlying, less-populated areas. This is set in contrast to [DATALT] distribution which has been shown to be relatively stable over time.[13]

Twitter users were assigned location using an algorithm which targeted where a user was likely from, rather than using phone GPS data which tells us only where they were when they sent the tweet.[14] This approach is potentially a substantial improvement over previous methods where the principal goal of dialectology is to measure language use endemic to a particular location.[15] They show that the present-day [P/D-DROP] distribution does indeed align with the simulated *gravity model of diffusion* (cf. Trudgill, 1972) as opposed to a wave model (Figure 2.11).

Interestingly Gopal et al.'s (2021) data show [P/D-DROP] to be largely ab-

---

[13]Whilst [DATALT] does appear to be geographically stable, this is in contrast to the historical trajectory plotted in Yáñez-Bouza and Denison (2015), which appears to show a dramatic shift in [DATALT] use from the late 18th Century (see Figure 2.4).

[14]Although Twitter *geolocation data* does employ its own 'geo-enrichment' algorithm which uses a combination of GPS and user-entered location data, it prioritises GPS location. As GPS data availability declined, from around 2015, the algorithm will presumably rely increasingly on user-entered location.

[15]It is this method that was the inspiration for that used in the NWatlas-corp, detailed in §3.3.5. The two methods for gathering location data are compared in §4.2.2.
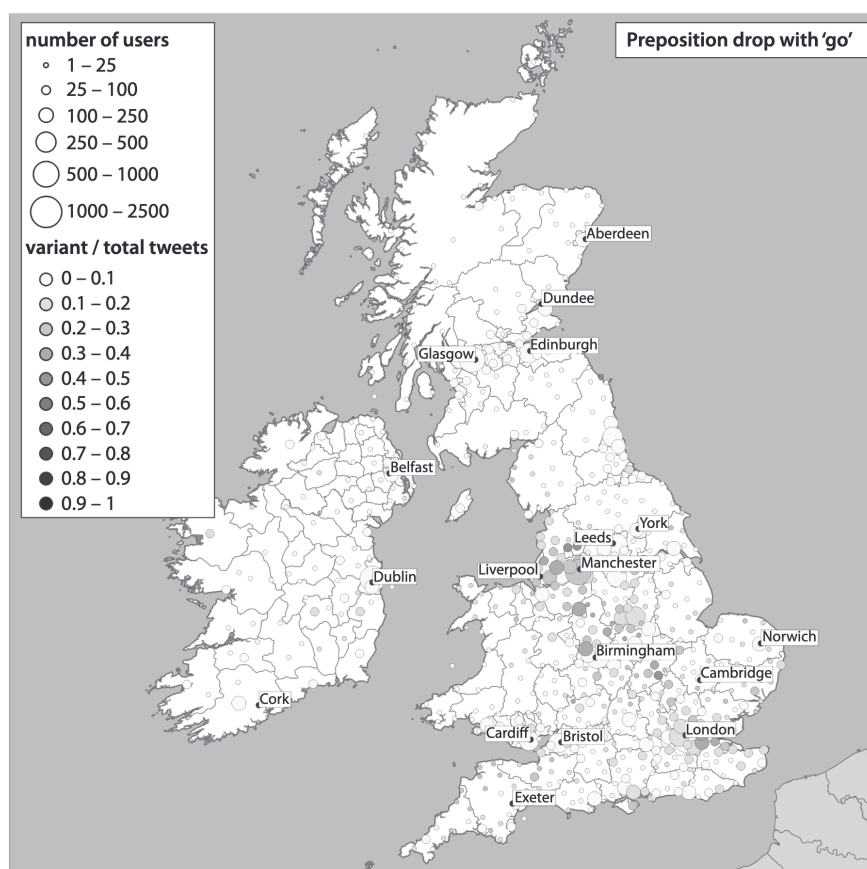
Figure 2.11: Gopal et al.'s (2021) map of preposition drop with the verb *go*. Data from November 2017 to May 2019.
Source: Gopal et al. (2021, p.266)

sent in Ireland, but do show a number of darker points (corresponding to high relative rates) in the South West of England, both of which align with the eWAVE data, and the historical record, presented in the previous section. The presence in the South West is however not commented on in the paper, perhaps due to lower overall rates.

In order to provide a general overview, the authors combine both P-DROP and PD-DROP into one set. The result of this is that there is no visible distinction in the North West, where P-DROP, rather than PD-DROP is reported to be widespread (Biggs, 2018; Myler, 2013). The third logical non-standard variant, where only the determiner is dropped (D-DROP), was not included, perhaps because it is historically distinct and peripheral to the point of their study. However, in terms of the present-day situation, D-DROP constitutes another option in the envelope of variation. The objective of Gopal et al.'s (2021)

research was to test whether, in a generalised sense, the current distribution of P-DROP and PD-DROP patterns as if it had recently diffused from London. The results of this analysis are quite striking. However, it necessarily simplifies the picture and there are nuances to the distribution and evidence from the historical record which do not fit neatly with this conclusion.

2.5.1 Innovation diffusion or regional levelling

The proposal that PD-DROP/P-DROP are new features, innovated in London, that have subsequently traveled northwards is, I suggest, an oversimplification the picture. This is most clearly the case when we consider P-DROP in isolation, which, as we will see, likely has an independent genesis in the North West. It is also problematic with PD-DROP, which is evidently not new, and seems to occur readily across dialects. As Edwards and Weltens (1985, p.114) notes, preposition deletion, "most notably with 'on' or 'to'", is a general tendency of dialects in British English.

If we consider the currently available historical data; attestations of P-DROP and PD-DROP predate the 20th century when the kind of rapid diffusion of a feature would be more viable. During the 19th century, London was exceptional in attracting economic migrants from across the country, where most other industrialising cities drew from more local populations (Pooley & Turnbull, 2005). As (Ellis, 1889, p.110) describes it, London was "an area of continual conflict and mixture of the S., W. , M., and E. populations", and thus dialectally mixed, during the late 19th century. Given this, it seems plausible that, in fact, PD-DROP entered into London from the dialects of the surrounding counties during this period.[16]

The map shown in Figure 2.12 shows the dates, in red, of historical attestations of PD-DROP. Whilst the earliest attestation of PD-DROP is in London, in 1737, it nevertheless seems unlikely that diffusion from London is responsible for the presence of the feature in the more northerly locations during the 19th century, considering the migration patterns, and extant transport infrastructure discussed above.[17].

Whilst, as Pooley and Turnbull (2005) describe, turnpiking reduced journey times between London and Manchester to 30 hours in the 1820s, and

---

[16]As noted earlier, the London attestations are from the Old Bailey, which documents trials that were conducted in London, though the original birthplace of speakers is not apparent.

[17]It is difficult to say for certain, however, without more data, given that there are only a handful of attestations for each place
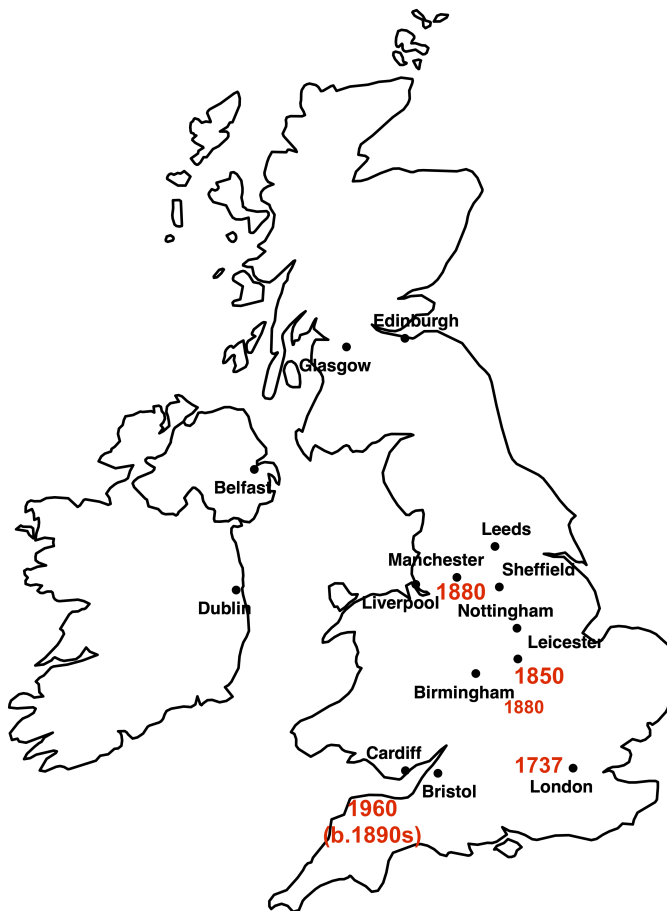
Figure 2.12: Map showing the dates, in red, of the earliest attestations of PD-DROP in the historical record.

the construction of the rail network from the 1840s further reduced this to 7h45m, fares "remained relatively high and for most ordinary working people rail travel was a major and only occasional expense until at least the 1890s, by which period travel time from London to Manchester had further reduced to around 4 hours 15 minutes" (Pooley & Turnbull, 2005, p.56).

It seems more likely that Edwards, Trudgill, and Weltens's (1984) assertion (above) is true; that PD-DROP is a tendency across dialects, and that, in informal speech, there is a propensity in English to drop the preposition. Additionally, if we consider a possible connection to DAR, as I discuss in coming sections, it seems more likely that PD-DROP has multiple sites of innovation.

The attestations in the South West, from Harris (1967), as mentioned previ-

85

ously, were from elderly, non-mobile, male speakers who had lived their whole lives, almost continuously, in the small settlement of South Zeal.

A useful comparison of diffusion from London is that of TH-fronting, the merger of /θ/ and /f/ as [f], as in *thing→fing*.[18] As Kerswill (2003) reports, TH-fronting, likely innovated in London in the early 19th century, diffused rapidly across the country in the post-war period, almost simultaneously reaching parts of the north in the 1960s and 1970s, as shown in Figure 2.13. Kerswill (2003) speculatively attributes the earlier attestation in Bristol, in 1880, to the construction of the railway between Bristol and London, in 1841. Aside from this, during the 19th century, transport infrastructure and migration patterns generally disfavour such a broad spread and rapid spread.
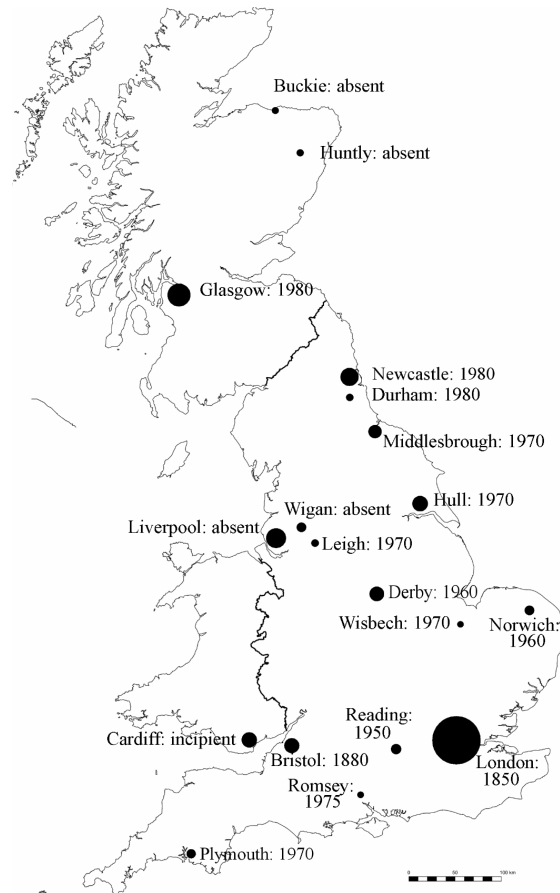
Figure 2.13: Spread of TH-fronting
Source: Kerswill (2003, p.236)

---

[18]The diffusion of TH-fronting is an interesting comparison, however, as a phonological innovation, it is not identical to PD-DROP.

Kerswill (2003) suggests that such a rapid spread may be ascribed to influence from "spoken media [...] which makes speakers more positively disposed towards the incoming forms they hear from the people they meet." (p.240).

The effect of media as a catalyst for the rapid uptake of a feature may be relevant to PD-DROP. The fact that PD-DROP may have already existed in the Midlands and North West does not preclude a subsequent, more recent proliferation of the structure, stimulated by a surge in usage from the South East in more recent history, and fuelled, in part, by use on social media. Such a scenario might explain the patterns found in Gopal et al. (2021). Indeed, the pre-existence of the feature in the path of such a diffusion event may have facilitated, rather than hindered, an increased uptake.[19]

What may have triggered this 'recent surge' in the use of PD-DROP? One possibility is the rise of Multicultural British English(es) over the past few decades.

2.5.2 Multicultural British English and PD-DROP

In Manchester and Birmingham, there exists the additional contemporary influence of what has been termed broadly as *Multicultural British English* (MBE) (Drummond, 2018; Fox et al., 2011). MBE mirrors *Multicultural London English* (MLE) (Cheshire et al., 2011) in that it shares a similar set of 'ethnic' input varieties. There is some debate in the field as to whether the development of MBE involved diffusion from MLE in the South East, or whether it emerged, to a greater or lesser degree, independently (Fox et al., 2011). However, as Fox et al. (2011) show, the diffusion of MBE features seems to be traceable to connected friendship networks between London and the Midlands.

In the context of PD-DROP, Bailey (2018b) suggests that there may be two forms: one which emerged out of MLE and an earlier form which has its roots in white working class East London and which since found its way into Kent. Given the discussion in the previous section, we can add to this that such an earlier form existed across multiple dialects of English throughout England. We know that input varieties to MBE, such as those from the Caribbean exhibit pervasive PD-DROP. Given that the extant 'Anglo' variety in London (Cockney) also had PD-DROP and was also an input to MLE, we have at least two

---

[19]However, in a general sense, the path of the diffusion of a feature may encounter resistance where it reaches locations which posses alternative variants that compete within the same envelope of variation. This is likely the case where PD-DROP finds competition from D-DROP in Yorkshire and P-DROP in the North West. As we will see, this appears to be the case, with the spread of PD-DROP is buffeted in these locations.

input varieties with the feature. Additionally, a similar set of inputs is likely in the Midlands, given both the pre-existence of 'native' PD-DROP and a similar contact scenario.

The idea that there may be an MBE variant of PD-DROP is interesting in terms of Gopal et al.'s (2021) diffusion model. MLE lexis, at least, is thought to have spread along similar route from London through the Midlands. Ilbury, Grieve, and Hall (2024), investigate the geographical diffusion of MLE lexis from London, based on a large Twitter corpus from 2013. They find strong evidence for diffusion of lexis to the Midlands, but less so for Manchester. Ilbury et al. (2024) put this discrepancy down to the earlier time period of their Twitter data, suggesting that MLE lexis had not yet spread to Manchester in 2013, but is in the process of doing so. Whilst the diffusion of lexical items is not necessarily equivalent to syntax, it is clear that some degree of diffusion is likely taking place, and it is important to consider this as part of the picture for PD-DROP.

Summary

There is frequently a tension in the study of *Language Variation and Change* between internally driven change and change which results through diffusion via contact and subsequent intergenerational transmission. Whilst it is often the case that analyses tend to become polarised in favour of a single explanation, the reality is likely less neat. Labov (2007), distinguishes two principal mechanisms through which language changes, namely whether a given feature is *transmitted* via an "unbroken sequence of native-language acquisition by children" (p.346), or, whether a feature *diffuses* between adult speakers in a population via various types of social contact. The difference in learning capacity between children and adults is central, as well as the level of exposure to a variety.

It seems likely that a process of diffusion has been taking place in recent history, however, it is also the case, as we have seen that across the Midlands and North West, that PD-DROP existed, and was intergenerationally transmitted, prior to the time period when rapid diffusion is tenable, and that, in addition, prepositions and determiners have been undergoing internal processes of reduction. I discuss this in the coming sections.

2.5.3 Folk-linguistic commentary

It is interesting that PD-DROP is often recognised by speakers as belonging to the local vernacular, which can be seen in folk-linguistic commentary. PD-DROP has garnered significant public attention in recent times. Commentary often reflects a sense of the longevity of PD-DROP in London, as reported in a comment on a blog post on the topic (Davis, 2014):

> Hi, just to say that adult members of my family (North London) were using this construction in speech at least 25 years ago. 'I'm going bingo', 'He's going football' etc were all pretty common. I don't know whether they actually used this in writing, too, but it was certainly characteristic of speech.

Figure 2.14: Blog comment on PD-DROP in North London
Source: `https://languagejazz.wordpress.com/2014/02/06/`
`im-going-shop-preposition-dropping-in-british-youth-dialects/`

Here, the association of PD-DROP with speech practice, more than with writing reflects what we have seen in the historical record and in the literature: PD-DROP is securely a feature of spoken English, not simply a social media artefact. The connection with London speech is not surprising, given what we already know about the present-day distribution of the structure. Meanwhile a sense of northern ownership of PD-DROP and a folk-linguistic sense of its connection to DAR is echoed in another response to the same blog post.

> Some of these constructions sounds more like a written form of Definite Article Reduction which is a common and old feature of loads of Northern English dialects. Especially that Leicester tweet, which has DAR in the regiolect. I would wonder if the DAR London tweets were not made by Northern transplants, or else were typing errors.

Figure 2.15: Blog comment on PD-DROP in the North of England
Source: `https://languagejazz.wordpress.com/2014/02/06/`
`im-going-shop-preposition-dropping-in-british-youth-dialects/`

The sense of Northern ownership expressed in this comment is telling, in the sense that PD-DROP at least does not carry an association with London speech, for some speakers. Likewise, the indication that PD-DROP may be perceptually similar to DAR is revealing.

Returning to the connection between PD-DROP and DAR, there may be some indication of a sociolinguistic motivation for a development from DAR →PD-DROP. DAR is frequently stereotyped as older, rural or regional speech practice

which is reflected in the representation of DAR in social media writing — *t'pub* — where its use is frequently parodied. There are numerous examples on Twitter, a large proportion of which are metacommentary, or imitating localised speech (see §4.6.2 in the Twitter results, page 230).

(35) a. 'tek yer coat off inside or you won't feel't benefit' and 'eee well al go t't foot of our stairs' aka my grandma's most used sayings

     b. true, I don't want to get airs and graces I might have to take t'coal out of t'bath!

Folk-linguistic accounts are useful in that they give a sense of first-hand speaker familiarity with the phenomena. The possible connection between DAR and [P/D-DROP] variants may, in fact, have considerable validity. I explore this in the next section.

2.5.4 [P/D-DROP] and definite article reduction

It is possible that variants of [P/D-DROP] emerged independently via processes of internally driven change. D-DROP, where the determiner is absent (also termed *definite article deletion*), and related *definite article reduction* (DAR) are well-known features of Yorkshire dialects (Rupp, 2007; Rupp & Tagliamonte, 2019). What I term D-DROP in this thesis—that is, tweets written without the determiner present—likely subsume cases where authors are producing what would be, in their speech, something closer to DAR.

(36) are you going to ~~the~~ cinema later?

There is no extant research connecting DAR with [P/D-DROP]. Perhaps, most obviously, a link between DAR and [P/D-DROP] has not been made because the focus has been on the dropping of the preposition, while DAR is, as the moniker suggests, on the reduction of the determiner. In fact, no studies on [P/D-DROP] have, to date, included the fourth logical variant; D-DROP in their investigation. Semantically, and functionally, however, D-DROP is identical to P-DROP, PD-DROP and NO-DROP and as such, all four variants constitute players in the same *envelope of variation*. Whilst it is clear that each variant has a distinct historical trajectory and each likely carries distinct social associations, they are nevertheless presented to the contemporary speaker as semantically equivalent.

The literature in the North West has focused on the syntactic properties of the P-DROP, in the city of Liverpool (Biggs, 2018) and in nearby Ormskirk (Myler, 2013). Linguistic investigation into [P/D-DROP] has focused on the structural distinction between P-DROP and PD-DROP and with the possible overlap with ditransitive structures where, for some speakers, the preposition may be absent, following Haddican's (2010) work in Manchester, as discussed.

DAR in the SED

There are a wide range of recorded forms of DAR in the SED (as recorded by Barry (1972)): /t/, /ð/, /θ/, /ðə/, /x/ (=/0/), shown in Figure 2.16.
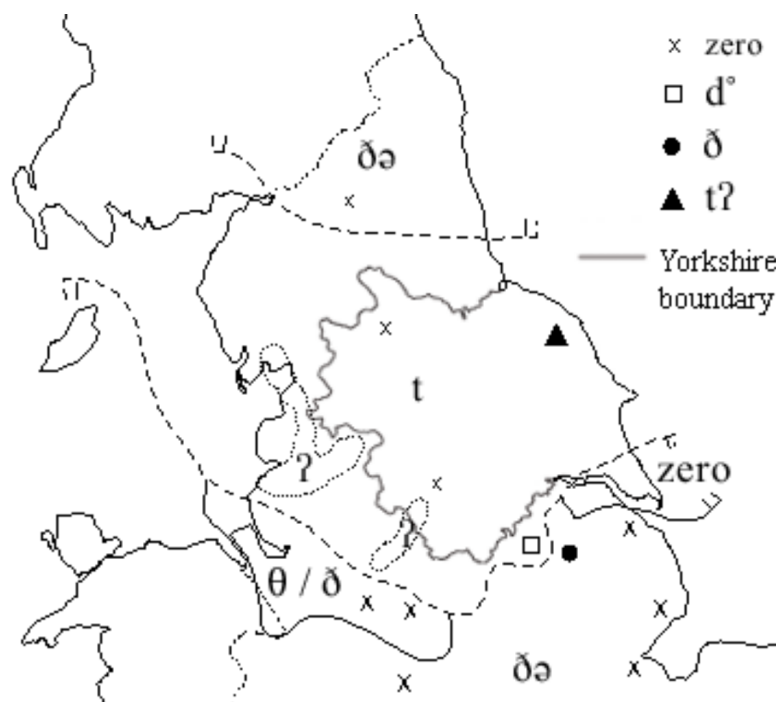
Figure 2.16: DAR before a consonant.
Source: Barry (1972, p.175)

It is relatively unproblematic to draw connection between DAR and D-DROP (*go to N*), where DAR is phonologically /Ø/ or where the determiner is deleted entirely as reported in North and East Yorkshire (Rupp & Tagliamonte, 2019). In fact, in many cases, D-DROP *is* straightforwardly definite article deletion.

(37)   DAR-Ø

    a.   going to/t' Ø shop

    b.   heading to/t' Ø pub

The exception to this is that D-DROP is also characteristic of *telegraphic writing*, which is motivated by communication economy, rather than reflecting local speech practice.

(38)   telegraphic writing

    a.   went to pub, then shop, now home
    b.   heading to beach, working on tan

In terms of variant frequency, however, we expect that *telegraphic writing* will not be localised to any particular region and will be relatively infrequent. We therefore would expect to find a low level of D-DROP across all regions with a tractable spike in use in those regions for which the definite article is reduced to Ø, or is NULL, in speech.

The SED reports DAR as fricated in the Cheshire and North Staffordshire region, which aligns with Ellis's (1889) description from 60 years prior:

> the normal form of the article is voiceless th, which produces an audible hiss without any admixture of voice, as th)maan , th)dug, th) a's, the man , the dog, the house (sw.La.) .

This would result in [ð/θ] areas as follows:

(39)   DAR-ð

    a.   going t' ð'pub
    b.   heading t' θ'shop

However, as the preposition is also frequently reduced or dropped (recall the examples in (22) from Ellis (1889)), we end up with either two phonologically weak cliticised elements, which are readily assimilated to the reduced determiner, or simply the reduced determiner, where the preposition is dropped. This results in:

(40)   DAR-ð with assimilated or dropped preposition

    a.   going ð'pub
    b.   heading ð'shop

In Lancashire (including present-day Manchester), the determiner is recorded as glottalised:

(41) DAR-ʔ

    a.   going t' ʔ'pub

    b.   heading t' ʔ'shop

With the preposition either dropped or assimilated, this renders as:

(42) DAR-t with assimilated preposition

    a.   going t'pub

    b.   heading t'shop

Whilst we find a zero article in much of Yorkshire.

Interestingly, Lodge (2010) reports DAR-θ as possibly 'extinct':

> "The [θ]-only dialects may be extinct by today, though they certainly existed between fifty and a hundred years ago." (Lodge, 2010)

We will see that DAR-θ is in fact not extinct, evidenced by its continued use in the Twitter data, and that it likely evolved into what is now identified as P-DROP.

Importantly, Biggs's (2018) prepositional element x only corresponds to overt *to* and *at*, and not to other prepositions such as *for* or *on*. This poses a problem for a DAR-derived approach, where we might expect to find P-DROP with a broader range of prepositions, given that DAR is frequent with a range of prepositions, as shown in Figure 2.17.

However, interpreting P-DROP as involving an element like Biggs's (2018) x, which is able to independently license the GOAL, rather than Myler's (2013) more restricted TO, which requires incorporation to the verb, is potentially more compatible with DAR. A question then is: do we find P-DROP and PD-DROP with other prepositions in the North West? As we will see, this is indeed the case.

Summary

D-DROP (*go to N*) likely has the most direct link to the /t/ variant of DAR, particularly when written in social media interaction. If this is the case, we would
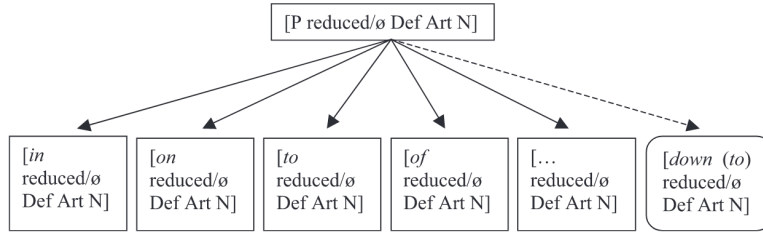
Figure 1. *Licensing of [*down (to)t'*/ø Def Art N] by the constructional schema [P reduced/ø Def Art N]*

Figure 2.17: Source: Hollmann and Siewierska (2011)

expect to see a correlation between D-DROP use on Twitter and traditional dialect maps showing the distribution of DAR, such as the Survey of English Dialects (figure 2.16). Likewise, if P-DROP and PD-DROP correspond to a re-analysis of DAR-ð and DAR-Ø, respectively, we would also expect to find some degree of geographical overlap between Twitter and the SED.

A distributional correlation between DAR and P-DROP may not entail that the two phenomena are structurally related. It may simply be that the existence of DAR results in ambiguities which promote the subsequent emergence of P-DROP. This said, it is worth considering if there may be structural consequences of such a link.

2.6   Traditional spoken dialect corpora

> "The linguistics of the twentieth century has been the linguistics of scarcity of evidence."
>
> Sinclair (1997, 27)

Corpus studies of syntax encounter the problem captured by the well-known maxim: 'absence of evidence is not evidence of absence'. However, formal approaches to syntax, which usually rely on acceptance judgements as evidence, run the risk of missing valuable data because they do not always know what is possible in a variety without corpus data. For these reasons, an approach which combines both corpora and acceptance judgements is preferable.[20]

---

[20]A note here that the transcriptions for BNC and FRED are partially open to transcriber interpretation and the possibility of error. Whilst unlikely to have a dramatic effect on the results presented in this thesis (both primary and secondary) the possibility data inaccuracy should be

2.6.1 British National Corpus (BNC)

The British National Corpus (BNC) is a 100 million word corpus of spoken and written British English. It is one of the most significant and widely used corpora for British English. The present thesis draws on data from the spoken component of the *BNC XML Edition* (an updated version of the original 1994 corpus), which constitutes approximately 15% of the corpus. The spoken texts include both formal and informal contexts, such as meetings, conversations, and interviews representing speakers from different regions, age groups, and social classes within the UK. Whilst geographical information is provided, it is not always present and fine-grained resolution is lacking, at least when looking for syntactic variables where the low frequency of results means that it is necessary to pool data from into wider areas, such as North, Midlands and South (Gerwin, 2014).

The distribution of texts is broadly reflective of the population distribution in the UK, with a higher proportion represented in the more populous South, than in the North and Midlands.

	texts	w-units	%	s-units	%
Unknown	35	448458	4.30	27496	2.64
South	311	4687877	45.03	457726	44.09
Midlands	213	2492236	23.94	240306	23.14
North	349	2781280	26.71	312552	30.10

Table 2.5: Distribution of textual material by region
Source: `http://www.natcorp.ox.ac.uk/docs/URG.xml?ID=BNCdes`

For present purposes, a quantitative comparative geographical analysis is not attempted. Rather, indicative examples of particular structures are considered in the context of the geographical distribution found in the Twitter data.

2.6.2 Freiburg English Dialect Corpus (FRED)

The Freiburg English Dialect Corpus (FRED) focuses specifically on morphosyntactic variation in British English. The project was produced by the research group 'English Dialect Syntax from a Typological Perspective' at Freiburg University, within the theoretical framework of functional typology. This has, at its root, the notion that language-internal variations can be integrated into global patterns of linguistic variation. The project was a major undertaking

---

borne in mind.

with approximately 2.5 million words were transcribed from oral history interviews. The result is the capturing of a wide range of dialectal variations from different regions.

As such, FRED is a valuable resource for the study of English dialects. Regions were broadly categorised as Southwest, Southeast, Midlands and North. These regions are shown in the map in Figure 2.18.

Figure 2.18: FRED regions

Within these broader categories, individual counties are represented. Table 2.6 shows the breakdown of the distribution of materials by county.

The distribution of texts is broad. It should be noted, however, that each county is not equally represented relative to population size. London, for example, represents 7.4% of the materials, while Kent represents 15.3%. These discrepancies are mitigated when comparing the relative rates of linguistic variants within a single variable, which inherently provide a normalising 'yardstick' for each location.

County	Dialect Area	Running Words	% of Material
Cornwall	SW	26,535	2.6%
Devon	SW	79,870	7.9%
Oxfordshire	SW	13,801	1.4%
Somerset	SW	69,321	6.9%
Wiltshire	SW	75,336	7.4%
Kent	SE	155,192	15.3%
London	SE	74,856	7.4%
Middlesex	SE	30,595	3.0%
Leicestershire	Mid	2,341	0.2%
Nottinghamshire	Mid	150,194	14.9%
Durham	N	26,507	2.6%
Lancashire	N	139,845	13.8%
Northumberland	N	27,777	2.7%
Westmorland	N	21,304	2.1%
Yorkshire	N	51,522	5.1%
East Lothian	SCL	28,985	2.9%
Midlothian	SCL	21,068	2.1%
West Lothian	SCL	16,347	1.6%

Table 2.6: Distribution of textual material by county in FRED
Source: `https://varieng.helsinki.fi/CoRD/corpora/FRED/basic.html`

The overall size (2.5 million words) is large, by traditional standards. However, due to the relative scarcity of specific syntactic phenomena in speech compared to phonological features, total counts for variants by location can be quite low. It is therefore often necessary to use broader regions when comparing relative rates.

The present work draws on data from both FRED and BNC, but takes a more qualitative approach, examining the occurrence of individual instants of different variants, rather than attempting a quantitative analysis.

2.7 Summary

1. Pronominal ditransitives are geographically variant, though the precise distribution is still not clear. For the present study, this need for more precision is particularly in the North West, where Stevenson (2016) showed that Liverpool is $_\text{P}$GTD rather than $_\text{P}$TGD dominant. This is confirmed in the data from the *Our Dialects project*, which show only 27% of Liverpool respondents report using $_\text{P}$TGD compared to 60% in Manchester. The question of the which constitutes the dominant variant in the North East, Scotland and Northern Ireland is also pertinent.

2. There is, as yet, no data on the relative geographical distribution of pronominal ditransitives by verb type, such as the *give*-type, *refuse*-type and Latinate verbs, as documented in Levin (1993) and used as diagnos-

tic criteria to investigate the underlying structure of TGDs in Manchester (Haddican, 2010) and Liverpool (Biggs, 2018).

3. Haddican (2010) finds that whilst the majority of speakers in Manchester do not accept full-DP ditransitives, a small subset does. Biggs (2018) finds that a majority of speakers in Liverpool accept them, including with verbs involving *transfer of possession* (*give, send*). It is not clear, however, the extent to which the grammar described by Biggs (2018) exists strictly in Liverpool itself, or whether it is not located in any particular place.

4. South East PD-DROP is characterised by Bailey (2018b) as follows: "[t]he determiner is obligatorily absent, the argument must be a directional GOAL, the verb must be semantically weak *come* or *go*, and the location must be familiar, anaphoric or a place name" (Bailey, 2018b, p.48). This is distinct from the analysis provided by Hall (2019) for London PD-DROP, which he describes as involving a *radical absence* of both preposition and determiner.

5. While Gopal et al. (2021) collapse P-DROP and PD-DROP into one set, [P/D-DROP] is treated in this thesis as a variable comprising P-DROP, PD-DROP, D-DROP and NO-DROP, each holding equivalent semantic interpretation. However, each '*variant*' is considered to be structurally distinct, likely with differing geohistorical trajectories. Additionally, *within* each variant, there is further variability. Bailey (2018b) reports that there are "likely two 'types' of preposition drop in London"; one emerging from MLE, and an older form "common to white working-class speakers". However, given that East End working-class varieties of London English (Cockney) are input varieties to MLE, it seems likely that the innovation occurred in this area, and subsequently spread to Essex, and Kent.[21]

6. Dating the innovation of PD-DROP is problematic, though Bailey (2018b) suggests that it is not older than a "few decades" as she finds that it is not accepted by older Kent speakers. However, we have seen substantial evidence from the historical record that PD-DROP is substantially older than this, dating to at least the 1730s in London. The variant's arrival in Kent, the area to which Bailey (2018b) refers, may well have been considerably more recent, perhaps during the period of counter-urbanisation

---

[21]This said, there is some overlap between P-DROP and PD-DROP, at the surface level, where the full form (NO-DROP) lacks a determiner, such as with proper nouns ("going Chessington tomorrow")

out of London in the second half of the 20th century. Alternatively, there may be multiple sites of innovation, if a tendency towards dropping of prepositions, in these contexts, is driven by language-internal pressure. The fact that P-DROP is found in Greek, Italian dialects, Kiezdeutsch, Creoles, (and AAL). indicates that this may be the case. Of course, standard English displays some P-DROP with specific lexical items like *home*.

7. Gopal et al.'s (2021) of PD-DROP as having diffused from London to the North West is complicated by the fact that there are historical attestations across the country at a time when rapid diffusion seems unlikely. Instead, there seems to be a viable mechanism by which PD-DROP, P-DROP and D-DROP may have developed locally out of various manifestations of DAR.

Mapping syntactic alternations using Twitter data

Rather than moving towards a single unified "netspeak" dialect, language evolution in computer-mediated communication reproduces existing fault lines in spoken American English.

(Eisenstein, 2013, p.1)

3.1 Introduction

The current chapter details the process of creating an atlas of syntactic variation in the UK using Twitter data, focusing on the set of structures introduced in §1.2, known to be variant in their syntactic realisation.[1][2][3] Three main Twitter corpora are employed, the first two are variable-specific ([DatAlt] and [P/D-Drop]) and use tweets based on GPS and other location data retrieved by Twitter (geo.place_id). The geo.place_id-based corpora cover the whole of

---

[1] See http://nwdialectatlas.uk for a working prototype of the atlas.

[2] A number of other structures and lexical items were mapped in the process of conducting this research which may be added to the atlas at a future point, but the focus for the current thesis is on pDit and P-Drop

[3] The first instantiation of the atlas uses Twitter data, but will include other data sources such as text and speech corpora and, as described in Chapter 5; survey data.

the UK and Ireland, with the aim of finding large-scale patterns of distribution. The third corpus, focused on the North-West, is based on user-reported location or place of origin (where they "grew up") (PO). Having both forms of location data offers a useful point of comparison, and feeds RQ(1).[4]

(1) How reliable are geotagged Twitter data? To what extent do they correlate with Twitter data located based on user-reported place of origin?

The work follows several recent lines of enquiry which focus on the mapping of syntactic variation, either by using pre-existing corpora of speech (FRED, BNC) (Gerwin, 2014; Siewierska & Hollmann, 2007), or through the collection of new task-specific data using a combination of corpus compilation and surveying such as was implemented in the *Syntactic Atlas of the Dutch Dialects* (SAND), the *Scots Syntax Atlas* (SCOSYA) and *Survey of Anglo-Welsh Dialects* (SAWD).[5]

The current project also follows a growing body of recent research that draws on Twitter data to map linguistic variation (Eisenstein, 2013; Grieve, 2016; Jones, 2015) and is contemporary with a more recent wave of studies that look specifically at syntactic variation using Twitter (Gopal et al., 2021; Stevenson, 2016; Strelluf, 2019; Willis, 2020).

The resulting Twitter data are then compared to data collected by studies that use synchronic corpora of spoken English (Gerwin, 2014; Siewierska & Hollmann, 2007), historical corpora (De Cuypere, De Coster, & Baten, 2014; Yáñez-Bouza & Denison, 2015) as well as acceptance surveys employed in studies that have the establishment of the formal properties underlying the locus of syntactic variation as their focus (Biggs, 2018; Haddican, 2010; Hall, 2019; Myler, 2013).

The study of syntactic alternation phenomena is an obvious use-case for Twitter data. First, the quantity of the data provides statistically robust results with fine geographical resolution. Second, as a written medium, word-order variation is more transparent than phonological variation. Additionally, syntactic alternations offer a closer approximation to the sociolinguistic variable than lexical variation, where it is often not obvious that different lexical items are semantically equivalent.

Whilst the main data source for the atlas is Twitter, a key motivation for the atlas is that it should offer a *starting point* for further study using a range of

---

[4]Breakdown of the corpora is detailed in section §3.4.
[5]These dialect atlases are discussed further in chapter 5

methods and data sources. Ultimately, it is envisaged that a full syntactic (and potentially phonological/lexical) atlas of British English may be established over time, building on this initial foundation — and perhaps combining with other projects. A principal goal, then, is to provide a framework onto which more data may be added. The survey data presented in Chapter 5 demonstrates a case study of how such an expansion may proceed in the North West.

3.2   On the nature of Twitter data

Social media, such as those found on Twitter, occupy a distinct position in terms of linguistic data, lying somewhere between spoken and written language. The kind of interaction exhibited on social media follows a lineage which can be traced back to the early days of computer-mediated communication in the late 1980s, with services such as Internet Relay Chat. Ferrara et al. (1991, p.9) dubbed this kind of interaction "interactive written discourse (IWD), the written language occurring in simultaneous terminal-to-terminal typed dialogues". Since the introduction of smartphones, users have developed close relationships with their devices, and this kind of written interaction is conducted routinely across multiple applications. As Deumert and Lexander (2013, p.536) put it, "texting is a person-to-person technology and the phone itself is perceived as a deeply personal device, carried on the body throughout the day, and close to one's bed at night". Shortis (2015) describes the loosening of constraints, freed from normative regulation, "de-regulation", followed by a "viral re-regulation" of communicative practice in SMS text messaging (Shortis, 2015, p.239). In place of the top-down imposition of rules, linguistic practice is negotiated between users.

> "Novel spelling and script use may be modelled and virally in the course of localised interaction, rather than by processes authoritatively legitimated and relayed from above by codification and programmatic instruction." (Shortis, 2016, p.489)

It is in this context that we should view the proliferation of vernacular expression on social media. The use of local dialect features is conditioned by novel and localised norms of written social communication, which inherently draws from local speech practice. Speaker identity, insofar as it finds expression through local vernacular, is fundamental here in the negotiation of new

norms of written communication.

There is a sense in which users on social media are self-transcribing how they would speak in face-to-face interaction. This sense is evidenced in Eisenstein (2013) who shows that "social media displays influence from structural properties of the phonological system" (p.1). In an analysis of tweets where users simulate consonant cluster reduction (*just → jus*) and th-stopping (*with → wit*), he finds that both are "significantly influenced by the phonological context".

Further, these variations are shown to map to known regional patterns in US English (Eisenstein, 2018). Jones (2015) has shown that orthographic representations of phonological features, such as phonetic respelling on social media, can serve as a proxy for phonological variation, aligning with expected geographical patterns in African American English.

Compared to more traditional sources such as transcribed spoken corpora or informal personal letters, Twitter data offer both scale and immediacy. Transcriptions of spoken language—while more directly representative of speech—are resource-intensive to collect. Informal letters, though revealing of colloquial written norms, and a crucial component of historical linguistic research, are increasingly rare and constrained by genre conventions. In contrast, Twitter provides access to a very large and diverse set of real-time language data, encompassing a wide range of dialects, registers, and social contexts. Crucially, metadata—such as geolocation, timestamps, and interaction patterns—also make it possible to situate language use socially and geographically in ways that are difficult with traditional corpora.

3.2.1 Public facing versus conversational tweets

Twitter messages are not all of the same type. They can be broadly divided into two categories: *One→many* or *one→few* (Page, 2012). The distinction here is crucial. In the first case; *one→many*, tweets serve the function of communicating to a general audience. In this sense they are more akin to *broadcast* media. In the second case; *one→few*, they serve the function of interpersonal communication. Within this, Pavalanathan and Eisenstein (2015) show that in conversational dialogues, the frequency with which local variables are used varies depending on the size and identity of their intended audience.

Nevertheless, it is where the intended audience is small, often to just one

other user, that we find the kind of interaction that can be characterised as conversational, and where we encounter the negotiated re-regulation of written norms conducive to the expression of local linguistic features. Following an initial tweet, which serves as a topic opener, a series of interactions may ensue, often in quick succession. As we will see (§4.3), it is in these kinds of messages that we tend to find the majority of non-standard language use.

This can be seen in the example shown in Figure 3.1 of a conversational exchange between two Twitter users.
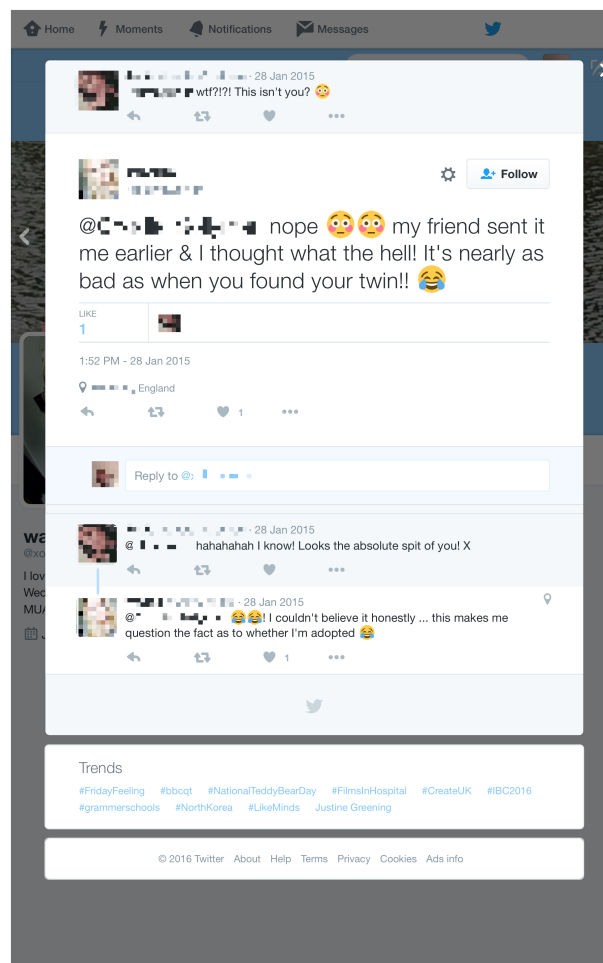
Figure 3.1: Example of a conversational exchange on Twitter.

In this example, the opening tweet refers to an object (in this case a photograph), asking a direct question to another user. The other user responds,

104

referring to the object with the pronoun *it* ("my friend sent it me").

An overall measure of the extent to which tweets in a given dataset are of one or the other category is the relative rate of messages that are *in reply* to another user. Eisenstein (2018, p.370) reports that his dataset of 114 million geotagged messages involved "more than 40% of messages... addressed to another user". The relative rate of *in reply* versus *not in reply* depends to some extent on the nature of the corpus. For instance, Stevenson's (2016) corpus of pronominal ditransitives with the verbs *send* and *give* ("sent it me", "sent me it" etc.) shows a ≈ 70%-30% split between *in reply* and *not in reply*. This skew reflects the pragmatic status of pronominal ditransitives (and the use of pronouns more generally), in anaphoric reference to a previously established referent as being more likely to occur in conversation than broadcast texts.

The use of Twitter for the purpose of conversational exchange, as we will see, has reduced over the lifetime of the platform. This has, in turn, resulted in a reduction in the occurrence of the kind of non-standard structures that are the focus of the current project. The net effect, however, when the ten-year time period is taken as a whole, is a more balanced corpus, which arguably better reflects the range of communication styles that make up modern English.

Tweets are still written language

Whilst conversational language use on Twitter frequently approximates features of spoken language, it is important to acknowledge that tweets are, of course, not actual speech, and the medium has inherent limitations for certain types of linguistic analysis.

Despite the spontaneity associated with tweets, they differ from speech in that users generally have more time to compose, edit, and curate their linguistic output. This introduces a degree of planning and self-monitoring not typically present in real-time conversation. Whilst we have seen that phonological information may be gleaned from respellings, it is frequently not so transparent. This is the case, for example, with DAR. Whilst, as we will see, users may explicitly represent DAR orthographically, such as in 'going t pub', it is not clear whether a form of DAR is present, but simply not spelled out, in 'going pub'.

Twitter and other social media data, then, should be viewed as complementary to traditional spoken corpora. The position taken in this thesis is that these data, and the detailed maps that they enable, provide researchers with a novel

way to find particular locations of interest, where there are clear distributional boundaries, which can be compared to traditional data. These areas of interest can then be more deeply inspected using additional methodologies, such as the survey presented in chapter 5. In this way, social media data offer a powerful new tool for researchers alongside traditional methods.

3.2.2 Twitter agreement with traditional methods

In addition to the studies mentioned above, there have now been a number of research projects which show a close alignment in the geographical distribution of Twitter data to that found using traditional methods.

For example, Willis (2020) shows that a corpus of Welsh Twitter data aligns quite well (though not perfectly) to the Siarad Corpus (a corpus of Welsh-English bilingual speech), in an analysis of the distribution of the innovated second person pronoun *chdi*. Meanwhile Strelluf (2020) shows that NEED + *past* (*the car needs washed*) patterns with pre-established dialect boundaries in Britain and the US.

Grieve, Montgomery, Nini, Murakami, and Guo (2019), comparing regional lexical variation on Twitter to data from the BBC voices project, showed a broad agreement between the two sources. They suggest that this offers additional evidence that Twitter data align with transcribed audio data. This, they argue, "licenses the use of Twitter corpora for general inquiries into regional lexical variation and change" (p.1).

3.2.3 Sample population

Research into the Twitter user-base has reported that it is biased towards a younger, more middle-class subset, than the general population (Morgan-Lopez, Kim, Chew, & Ruddle, 2017). However, whilst these are important considerations when looking at sociolinguistic variation and should be taken into account in any analysis, it is not *certain* the extent to which social class is, or is not, evenly represented. Additionally, when focusing predominantly on geographic variation, rather than sociolinguistic variation, the question of social class is not as critical. In terms of gender, reports generally agree that there is no significant difference; Eisenstein (2018) reports a more or less 50/50 split between male/female users.

The most pressing issue is that of user age. The age profile of Twitter users

has changed significantly over the past ten years: this can be seen clearly in the difference between Figure 3.2 and Figure 3.3. Unfortunately, the age ranges for each group in the presented data are different between the two charts, but the change is nevertheless clear. In 2013, 43% of users were in the 10-19 bracket compared to 2.4% in the similar 13-17 bracket in 2023. Meanwhile, older groups are clearly more represented in 2024. This is particularly problematic for Twitter studies which take a slice of data from one point along this time period.
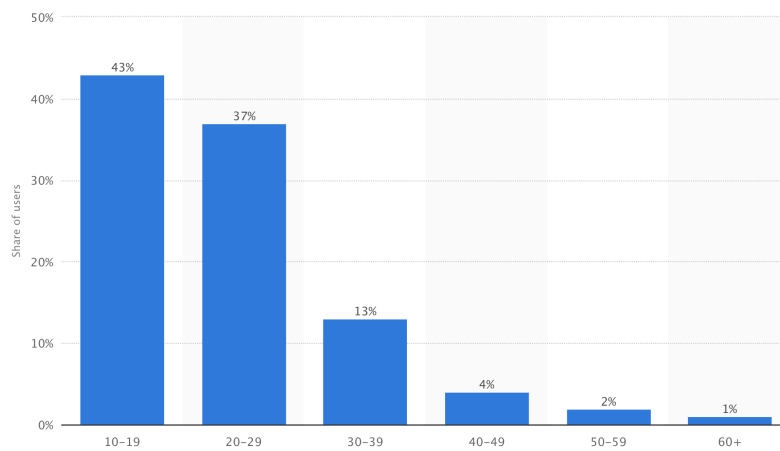
Figure 3.2: Age distribution of Twitter users worldwide, 2013
Source: Statista (2024)
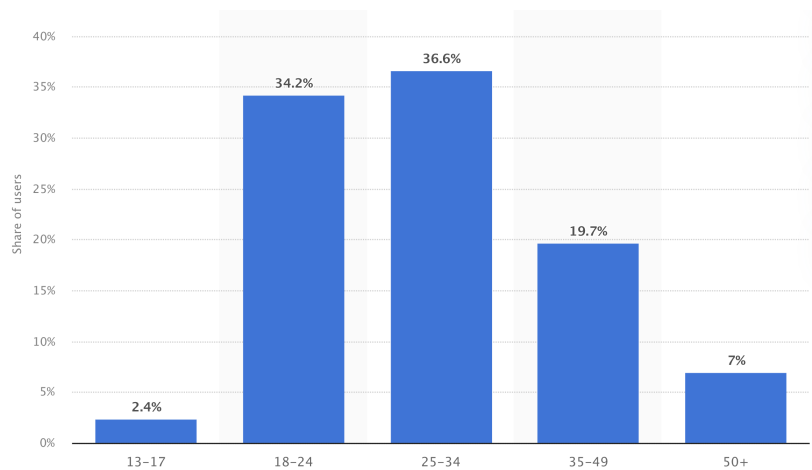
Figure 3.3: Age distribution of Twitter users worldwide, 2024
Source: Statista (2024)

These age differences are, in part, likely due to Twitter users themselves having aged, with relatively little uptake in new, younger users, possibly as a result of competition from rival services such as Tiktok (Grondelaers et al., 2023). The 10-29 groups in 2013 are in the 18-40+ groups in 2024.

For the UK specifically, data from 2023 show a more or less even split between *Gen Z* (born 1995-2012) with 23% and *Gen X* (born 1965-1979) with 29%. *Millennials* (born 1980-1994) have the largest share at 40% and *Baby Boomers* (1946-1964) the smallest at 8%, as shown in Figure 3.4.
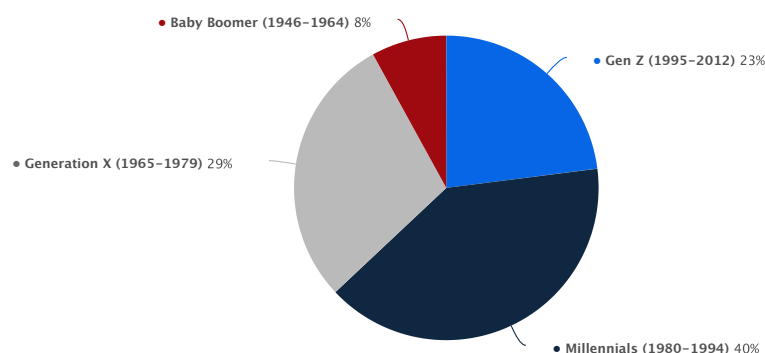
Figure 3.4: Statista survey. Response to the question: *What social media platforms do you use regularly?* 24,176 responses.
Source: Statista (2024)

This puts UK users as a little older compared to the worldwide age distribution in 2024 presented in Figure 3.3.

A changing age demographic may initially raise concerns, however, it may actually be useful for the present study. A ten-year corpus arguably offers a more balanced sample overall, to one that focuses on a period which may be skewed towards younger users. The effect of the changing age profile is documented in Grondelaers et al. (2023) who show that the way Twitter is used, as a platform, has itself changed since its inception, with an increase in registers associated with the standard language, and a corresponding reduction in informal styles. They attribute this change primarily to the change in the age demographic of Twitter users.

This change in how Twitter is used is confirmed in the data used in the current project. These results, and their implications are discussed in §4.3. As well as a changing age demographic, I show that the switch to more standard registers appears to be linked to a trend away from the use of Twitter for

conversational exchange, to more public-facing messages.

A note on the longevity of Twitter data and alternative sources

Since the acquisition of Twitter by Elon Musk in 2022, there have been some concerns over the length of time for which Twitter data will be accessible. In early 2023, Twitter, rebranded 'X', announced a new marketing model for access to data that would dramatically increase the cost of accessing data, putting it largely out of the reach of most academic research projects.

This does not directly affect the current project, though it may impact follow-up research. However, there are several reasons why these concerns are not as serious as they may first seem. First, the data already gathered provide a snapshot of the present time (mostly the decade 2010-2020) that will continue to be useful indefinitely.

Second, the principle that social media data, in general, are a useful source of dialect data is clear - it does not need to be Twitter data, as such, it just happens that Twitter had been the most open with its data. There are other possible data sources: TikTok and Reddit are immediate possibilities, as well as YouTube, which offers audio recordings as well as text comments.

3.2.4 A note on the ethics of using Twitter data

There has been some ethical concern raised on the use of Twitter data for linguistic research. Whilst Twitter data are publicly available, meaning that they are available 'on the open web', without the need for a login to access the data, and Twitter users engaged on the platform under Twitter's agreement which makes clear that interactions are not private, this does not necessarily entail that users anticipated, or are content with, their outputs being used for academic research. There is a clear distinction to be made here in how such data are used. Where the majority of data are used in aggregate, to count the relative rates of syntactic structure, an individual user, and their personal data is entirely hidden in the statistics. This use of Twitter data is uncontroversial.

Potentially more problematic, is where individual tweets are reproduced as examples, as they are in this thesis. Whilst specific user names and handles can be removed, the body of the text can be searched using Google, often revealing the original tweet, and thus the user who produced it. However, the fact that a tweet is still available, and searchable in this way, arguably is the responsibility

of either Twitter, or the user in question. There are cases where a user may have removed a given tweet, or their entire account, since it was published in a given academic publication. In these cases, the text of the tweet will no longer be traceable back to the original user using a Google search.

A final point, as will be described in the following section, is that all of the Twitter data used in the current work were retrieved using Twitter's Academic Research API, which was specifically made available for the purpose of academic research of this nature and, as such, is sanctioned by Twitter for this use case.[6]

3.3 Methodology

3.3.1 Accessing Twitter data

The basic retrieval and processing of Twitter data can be a relatively straightforward process. It is possible without any 'coding' using one of several applications that offer Twitter data gathering through a point-and-click interface, such as FireAnt (Anthony & Hardaker, 2022).

However, the methods used in the current project did require a degree of scripting and text processing. There are advantages to this, such as the ability to iterate over many searches, incorporating an exhaustive list of possible word order combinations, something not possible using an application like FireAnt.

Access to Twitter data has changed significantly in the years since it was first made publicly available. For most of this time, approximately 1% of the total volume of Twitter traffic was accessible via the API at no cost, with paid tiers allowing access to progressively greater quantities, up to the full historical archive. Full access was prohibitively expensive for individual academic projects, though some institutions have set up ongoing accounts with Twitter with full access to cover research programs (such as Media lab at Sheffield University).

This changed at the start of 2021 with the introduction of the Academic Twitter API (AcTw-API) which gave full access to academics for non-commercial use free of charge, subject to per-project approval by Twitter. As a result, the method described in detail here is using AcTw, the one ultimately used for this project, with a brief mention of previous and alternative methods. Alter-

---

[6]Approval for the use of Twitter data was also provided by the ethics board at the University of York, which has its own guidelines for the use of social media data.

native methods, upon which a large part of this project was conceived, differ only marginally, but differences will be noted where relevant. By far the most consequential change introduced, of course, is the increase in quantity of data permitted by AcTw-API.

Academic Twitter R package

The Academic Twitter R package (AcTw-R) (Barrie & Ho, 2021) was adapted from a previous R package (TwitteR) to access the AcTw-API. All of the Twitter data presented in this thesis was retrieved using a modified version of AcTw-R to access AcTw-API.
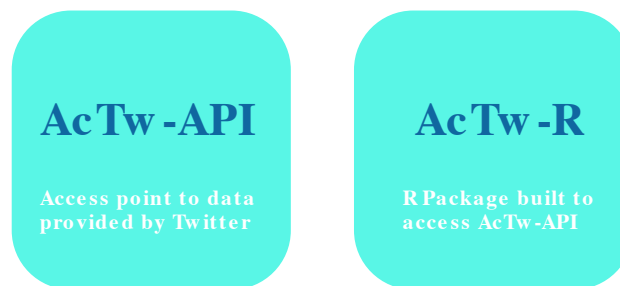
Figure 3.5: Academic Twitter API and R package

The primary modifications made to AcTw-R were:

1. Allowing for the retrieval of very large amounts of data using multiple complex queries, speeding up the process of extraction and resuming if there was an error.

2. Saving an index of place names and associated geo.place_ids.[7]

Within the AcTw API, there are number of ways to go about getting data. These are outlined in this section. The main difference between the methods discussed here hinge on the extent to which data are filtered on the Twitter server prior to download.

---

[7]Previously for this project, data were gathered via other means, such as the web scraping tool in Python (TWINT) (Pielco & Haccer, 2017), and a tool that ran via a Google Spreadsheet (TAGS) (Hawksey, 2018). I will not discuss these tools here as the current method is the one officially sanctioned by Twitter, and provides the most complete datasets.

Download then search

In terms of downloading data, the most expensive method, that I am calling *download then search* (DtS), is to start by downloading all tweets in a given time period, producing a substantial amount of data that can then be processed and searched locally. The obvious advantage to this method is that the data are then secured and can be coded, tagged and queried at will. The problem is that the sheer quantity of data provided rapidly becomes prohibitively expensive both in terms of actual cost/time for retrieval, and in terms of the time and computing power required for processing and coding. For this reason, most Twitter dialect projects that have used this method cover a maximum of one year. Even the new Academic Twitter API, which permits free access to the historical archive, has a limit of up to 10 million tweets per month. Given that one year of data will run into hundreds of millions of tweets, it is not feasible to retrieve more than a few months in this way.

Search then download

An alternative to the DtS method is to try to achieve as much of the filtering of data on the Twitter server prior to download, such that only relevant strings are retrieved (Search then Download - StD). This means that a much longer timespan can be covered within the search quota. To achieve this, a long list of possible strings for a given structure was generated and the API was queried sequentially only for those strings. The big advantage to using this method is that it made possible the retrieval of a full *ten years'* worth of data (2011-2021), whilst only using a fraction of the 10 million tweet quota in a given month, resulting in not only more results, but a usable timespan for looking at certain changes over the period (see §4.3). Whereas the DtS method may result in very large corpora, the total corpus for the StD approach is essentially all geocoded tweets sent in the past ten years for the selected locations in the UK and Ireland.

The obvious issue with using StD is the lack of flexibility once the data are retrieved. The use of an automatically generated list of thousands of possible combinations of strings to supply the search API mitigates this to some degree. Further, in the North West, an unrestricted corpus was downloaded, allowing this flexibility in the key area of interest for the current study.

3.3.2 Creating variable-specific corpora for [DATALT] and [P/D-DROP]

Following the *search-then-download* method, two corpora were created by querying the Academic Research API for a long list of possible strings:

1. **pDit-corp**: built from strings for pronominal ditransitives

2. **pDrop-corp**: built from strings for preposition-determiner dropping

I detail this method in the coming sections.

The GREP file

A file (GREP file) was generated with all possible iterations of each structure and combinations of relevant parts of speech, such that each pattern occupied a single row, and each column was used for adding coding information such as *variant, noun, pronoun, verb,* etc. The GREP file served three main purposes:

1. To supply the search API with relevant search patterns

2. To provide patterns for the KWIC function, for concordancing of raw data

3. To apply coding information for each match

Pronominal ditransitives

Pronominal ditransitives where both THEME and GOAL are pronouns are generated by combining a verb (table 3.2) with 'it' and recipient first-person pronoun (*me, you, him, her, them, us*) as follows:[89]

Type	Structure
TGD	VERB + it + PRO
GTD	VERB + PRO + it
PDAT	VERB + it + to + PRO

Table 3.1: Pronominal [DATALT] generation

The R code generates a data-frame containing all possible word combinations. The resulting strings are then used to query the Twitter API sequentially. In total 3036 alternating strings were queried; a subsection is shown in Figure 3.6.

---

[8]In the GREP file, the theme is always *it* with pDit. This means it will not capture marginal cases involving THEME person pronouns ("gave him her"), in the context, for example, of giving a male dog to a female recipient, though the frequency of such structures are low.

[9]Full R code in appendix.

Type	Verb list
Manner of communication	shout, yell, whisper, mention
GTD pref.	ask, deny, envy, refuse
PDAT pref.	address, convey, contribute, credit, delegate, deliver, demonstrate, describe, dictate, dispatch, display, distribute, donate, elucidate, exhibit, express, explain, forfeit, illustrate, introduce, narrate, portray, recite, recommend, refer, reimburse, restore, return, sacrifice, submit, surrender, transfer, transport
Verbs of communication	email, fax, text, txt, tweet, message, msg, pm
Messaging	quote, read, show, teach, tell, write
Throwing	pass, throw, flick, fling, chuck, bung
Carrying	carry, kick, pull
Sending	forward, fwd, hand, mail, post, send, ship
Bringing	bring, take
Future	assign, award, leave, offer, owe, promise
Giving	feed, give, lease, lend, loan, pay, refund, rent, repay, sell, serve, charge, wire

Table 3.2: Verb types included in search, based on (Levin, 1993)

give	me	present	dit	give	"give it me"	"give me it"	"give it to me"
give	you	present	dit	give	"give it you"	"give you it"	"give it to you"
give	him	present	dit	give	"give it him"	"give him it"	"give it to him"
give	her	present	dit	give	"give it her"	"give her it"	"give it to her"
give	us	present	dit	give	"give it us"	"give us it"	"give it to us"
give	them	present	dit	give	"give it them"	"give them it"	"give it to them"
lease	me	present	dit	lease	"lease it me"	"lease me it"	"lease it to me"
lease	you	present	dit	lease	"lease it you"	"lease you it"	"lease it to you"
lease	him	present	dit	lease	"lease it him"	"lease him it"	"lease it to him"
lease	her	present	dit	lease	"lease it her"	"lease her it"	"lease it to her"
lease	us	present	dit	lease	"lease it us"	"lease us it"	"lease it to us"

Figure 3.6: Section of pdit GREP file

Preposition-Determiner dropping

Preposition/determiner dropping [P/D-DROP] has four logical variants represented in Table 3.3:

Variant	Structure
PD-DROP	VERB + N
P-DROP	VERB + the + N
D-DROP	VERB + to + N
NO-DROP	VERB + to + the + N

Table 3.3: [P/D-DROP] generation

114

Hall (2019) reports three main verbs used in PD-DROP: *going, coming, heading*, while (Myler, 2013) describes a slightly wider set of verbs for P-DROP, including *pop, nip, jog.* The set of viable nouns are similarly restrictive, limited to familiar or institutional places (e.g. *pub, shop, gym, centre*).[10]

However, it seems that there is quite a lot of potential here for variability, in terms of the range of verbs and destination nouns that are permitted, particularly for P-DROP. For this reason, a longer list of both verbs and nouns were included in the GREP file. These are shown in Table 3.4.[11]

Verbs	Nouns
came, come, coming, go, going, gone, headed, heading, went	airport, bank, barbers, beach, bus stop, cafe, cinema, doctor, doctors, football, garage, gym, hairdressers, kitchen, library, movies, office, post office, pub, seaside, shops, station, supermarket, swimming pool, theatre, zoo

Table 3.4: Verbs of motion and destination nouns included in [P/D-DROP] search

The verb *head* was originally included in its root form, but later excluded as it produced too many false positives with PD-DROP, where *head* can act as an adjective modifying the noun such as *head office*.

(1) a. I am going to head office

It is often not possible to disambiguate between the two.[12]

The search string

Below is a code snippet, simplified for the purpose of exposition.[13]

```
for (s in 1:nrow(grep)) {
  q<-paste0(tw_search,collapse = ' OR ')
  query <- paste0('(',q,')', ' has:geo -is:retweet
  (place_country:GB OR place_country:IE)')
```

---

[10]*Nip* and *pop* were not included in the UK-wide corpus (pDrop-corp). This is unfortunate, as it would be useful to be able to quantify the rate of other verbs with PD-DROP in London against in the Midlands and North West, in the light of the question of diffusion vs multiple innovation. They are, however, available to search within NWatlas-corp, along with other verbs.

[11]The original search actually included a greater range of verbs such as *walk/drive.* Some of these results are used in the analysis, however, there were too many false positives for effective quantificational analysis.

[12]In hindsight, it may have made more sense to exclude only instances of *head (to) (the) office.* I chalk this to future work.

[13]Full code is in the Appendix.

```
}
```

Essentially, the code goes through each row of the GREP dataframe (see above) which contains all possible combinations of given parts of speech. The code extracts the string from each row and, with it, builds an individual query to send to the Twitter API.

This query says "find tweets with the given string (q), that have geographic metadata (has:geo), is not a retweet (-is:retweet) and is within Britain or Ireland".

Retweets are excluded for the obvious reason that they repeat a previous tweet, and their inclusion would result in the arbitrary repetition of strings and the over-representation of the structure they contain.

Geographic data

Tweets retrieved with the argument 'has:geo' include not only GPS encoded tweets, but also tweets that are enriched with geographic data that were identified by Twitter's algorithm with given locations. Each location is assigned a *geo.place_id*; an alphanumeric string which corresponds to a *bounding box*, defined by four geographical points. The size of each bounding box varies, but usually corresponds to a settlement: a town, city or conurbation. Twitter's algorithm assigns tweets to a geo.place_id based on the GPS data provided by the user's mobile device at the time it was sent, or other data provided by the user.

(2) Description of Profile Geo enrichment from `https://docs.x.com/x-api/enterprise-gnip-2.0/fundamentals/data-enrichments`

  a. The Profile Geo enrichment attempts to determine the best choice for the geographic place described in the profile location string. The result may not be accurate in all cases due to factors such as multiple places with similar names or ambiguous names.

  b. If a value is not provided in a user's profile location field (actor.location), we will not attempt to make a classification.

This means that tweets with geo.place_id metadata are more abundant than those that are only GPS coded: all tweets with GPS location also have geo.place_id metadata, but not all tweets assigned a geo.place_id have GPS

data. Previous Twitter studies have tended to focus on GPS data alone. This can be useful, for example, when the goal is to find GPS points which fall within predetermined geographic regions, known as the *point in polygon* (PIP) technique. This was the case in Baxter and Stevenson (2024), which needed to fit Twitter data to US census areas to compare the distribution of a syntactic feature of African American English with the population distribution of African Americans.

The present work instead leverages the more abundant data provided by geo.place_id and focuses on the relative distribution found in the settlements they identify.

Building an index of geo.place_ids

It is possible to search Twitter for tweets from a given list of geo.place_ids. However, Twitter does not provide an index of geo.place_ids with corresponding place names and GPS locations. Instead, an API call is available to request up to 20 place_IDs for a given LAT/LON every 15 mins, so it would theoretically be possible to build an index for a country the size of the UK over a period of a few months this way.

However, it is possible to build such an index more quickly: each time a tweet is retrieved, the location metadata are all present. To do so, lines were added to the script that saved Place ID, GPS point(s) and place names to a separate file, for every tweet. Subsequently, the script would first check if the place ID had already been recorded to the index. If it had, the geographical information would be taken from file, if not, the new place would be added to the file. As a result of this process, the Place IDs of 8740 places were recorded in the UK and the Republic of Ireland.

Having this Place ID index is useful as it allows for the precise retrieval of tweets from (a list of) individual places. So, for instance, it would be possible to define a geographical area and search only Place IDs that fall into that area. Whilst the API does allow for setting a LAT-LON point and radius of up to 15 miles that would fetch all places falling within that circle, once the Place IDs are known, there is no limit on the number of geo.place_id.s that can be included in a given series of searches, and therefore the size of the area that can be defined in this way.

Retrieving overall counts

The Academic Research API is able to provide the total number of tweets that would be returned for a given search query without the need to gather the full tweets and metadata. Using the Place ID index just described, it is possible to build a script that finds the total number of GPS-encoded tweets available for each Place ID, and add this to the index. The process takes a relatively long time as each request to the API returns counts for 30 days, and there is a limit of 300 requests per 15-minute interval. This means for the 10 years between 2011 and 2021, counts for roughly 2.5 places can be retrieved every 15 minutes. At this rate, when left running continuously, a script would take about 40 days to get count data for all 8740 places in the Place ID index. The script can be divided between multiple access tokens, however. With two access tokens the time is halved (so, 40 days becomes 20 days).[14] Requests for count data do not count against other API limits, so it is still possible to use the same access tokens to run conventional, full-data queries.

Whilst time-consuming, getting count data in this way is worthwhile as it allows for the tracking of the relative frequency of a given query against the total for a given place. This is particularly useful when looking at linguistic phenomena that do not neatly fit into an *envelope of variation* such as lexical innovations, and some syntactic phenomena. In this way, it is possible to track the diffusion of an innovation between places over the ten-year period in question. In the current project, this is shown with the tracking of subtypes of preposition-dropping, but can be easily adapted to (and is arguably more suited to) searching the chronological development and geographical diffusion of lexical innovations.

Additionally, retrieving overall counts it is possible to estimate the overall corpus size from which data are gathered when using the StD approach. This allows for a comparison, in terms of scale, to other corpora. For pDit-corp and [P/D-DROP], the overall size is calculated at approximately 550 million tweets, which, based on mean tweet length of 19.5 words, equates to a corpus size of 9.8 billion words.

---

[14]Scripts are available to view in the appendix.

Merging locations

As mentioned, geo.place_id identifiers correspond to a *place name* and a *bounding box* defined by four latitude/longitude points. For instance, *Manchester* is assigned the *geo.place_id*: $6e2453d6a2f968cb$, corresponding to the points [-2.3199, 53.3436, -2.1470, 53.5703]. For the purpose of mapping locations, the centre point for each bounding box was taken. In the Manchester example, this is [LON: -2.23345, LAT: 53.45695].

However, over the course of the ten-year period from which Twitter data were gathered, Twitter assigned new *geo.place_ids*, sometimes corresponding to slightly different bounding-box coordinates, resulting in correspondingly different centre points.

geo.place_id	location	lon	lat
6e2453d6a2f968cb	Manchester, England	-2.2334500	53.44235
315b740b10848f6	Manchester, England	-2.2334500	53.45695
90045a310627f75c	Manchester, England	-2.2334500	53.44235

Table 3.5: Sample data for Manchester, England

When counting the rates of variants per location, each distinct geo.place_id is counted, which is not what we want. It was therefore necessary to merge locations which share the same name. However, sometimes we get cases where two distinct places share the same name.

geo.place_id	location	lon	lat
6d74aec90800ab2e	Melton, England	-0.8553500	52.81175
02a838a6d3adf4b6	Melton, England	1.3358000	52.10860
49b9d8d8b4ff4a5e	Melton, England	-0.5276000	53.72425

Table 3.6: Sample data for Melton, England

In other cases, Twitter assigns a different name to the same location:

geo.place_id	location	lon	lat
71c8eb57c400c9b6	Saint Helens, England	-2.72985	53.44720
7aac76d1d0fc2dc0	Saint Helens, England	-2.70085	53.45840
402b6f5dbfc8d78c	Saint Helens, United Kingdom	-2.70085	53.45840
3b5171397189f0a5	St. Helens, England	-2.70085	53.45840

Table 3.7: Sample data for Saint Helens, England

Two functions were written in R to first standardise the names of places and then merge locations. If a place has the same name, the merge function checks to see if the latitude and longitude coordinates are within a given range

of each other, if they are similar, the places are merged, if they are distinct enough that they refer to different places, they are kept separate.

A *precision level* was set to determine what constituted places having similar-enough coordinates to constitute being identified as the same place, essentially the number of decimal places to take into consideration. Setting the precision level to 2 decimal places was enough to successfully identify and merge the appropriate places and variable counts.

3.3.3 R-functions for concordancing and POS-tagging Twitter and other corpora

A series of *functions* were developed in R that could be applied to both Twitter data and traditional corpora. A function is essentially a series of automated steps that can be applied to a given input, with a resulting output, like a recipe.

A data frame or a folder of text files can be given as an input, with the output as a tagged and concordanced data frame. The steps applied in the function are as follows:

1. Apply the KWIC function, from the *quanteda* package (Benoit et al., 2018), using a defined 'pattern' to set as the *keyword* to concordance.

2. Apply the POS-tagger.

3. Bind the data together with the coded GREP file (this step was made optional).

Building a simple function such as this was useful as it allowed for consistently formatted outputs from a variety of input sources, including both Twitter data, and data from FRED. Data from FRED were provided as unprocessed text files by Benedikt Szmrecsanyi. Data from BNC are already tagged, and were searched using Lancaster's BNC web interface *QCP-web*.

3.3.4 Removing false positives

Ditransitives: pDit-corp

A number of issues arise with false positives due to the fact that the Twitter API does not recognise intervening punctuation. This means that, for example, the search string in (3a) will return results such as in (3b).

(3)　　a.　　shown it you

b. @baldywully They definitely haven't <u>shown it. You</u>'d expect [...]

This is quickly resolved by filtering out results using search that is sensitive to punctuation. Here, a simple GREP search in R is sufficient.

Person pronouns as subject of a new clause

More problematic cases arise where there are clause boundaries that are unmarked by punctuation and where grammatical case is not explicitly marked as with *you, it* and *her*. These situations are discussed in the following sections.

For the TGD, a problem arises with the second person pronoun *you* which is homonymous in both dative and nominative case, so it may act as the subject of a new clause.[15].

(4) a. If you don't <u>give it you are</u> 100% not gonna get it sorry
    b. which ever way you address it you are a terrorist

You **as part of a vocative expression**

Second person pronouns may occur as part of a nominal vocative expression where the recipient is not present (*give it, send it*). In these cases, it is possible to have *you* immediately following, with a descriptive element targeted at the speaker's interlocutor, functioning as a vocative term of address, often an expletive.

(5) a. Weren't a pen but just <u>give it you cuntt</u>
    b. put your body away <u>I envy it you size 4 slut</u>
    c. Just ask it you fucking dickhead!
    d. please could you bring it you star
    e. @DownloadFest in a few weeks bring it you legends

These kinds of discourse markers would standardly be separated by a comma in writing, as in:

(6) bring it, you star

To extract these examples from the corpus, an exclusion list was created. Tweets coded as TGDs with the GOAL pronoun *you* and followed by a POS tagged as a

---

[15]Whilst this kind of ambiguity is an issue here, it can offer potential scope for psycholinguistic study using self-paced reading tasks. This is discussed in §6.4.

NOUN or ADJ + NOUN were filtered and manually inspected.

Out of the 657 results that matched this query, many were cases where 'tomorrow' or 'today' were incorrectly identified as NOUNS instead of ADV. The query was re-run excluding a list of these 'time adverbials' (*today*, *tomorrow*, *yesterday*, *tonight*, etc.) following the key phrase. This reduced the exclusion list by 239 tweets.

Additionally, discourse elements such as 'babe, babes, bro, man', which followed the key phrase (e.g. 'give it you man') were permitted. Finally, nominals which are selected by the pronoun to form an NP such as 'guys, lot' (e.g. 'you guys', 'you lot') were removed from the exclusion list.

This left 373 tweets to be coded manually.

Pronouns as anaphoric appositive

In colloquial/informal speech, appositive anaphoric pronouns serve to emphasise the subject. This is reflected in some tweets, such as:

(7)   a.   I've never denied it me I'm too honest
         b.   ruthless that goat ill kick it me haha

These were manually removed.

Pronoun *it* as subject of a new clause

For the GTD, where we are looking for strings that end in accusative *it* (where *it* denotes the THEME), we run into the problem where, in the absence of punctuation, *it* may be the nominative subject of a new clause.

Where *it* is nominative, it will always be followed by a verbal form (standardly conjugated to agree with the subject with -s), or with copula be (*is*), or in cases such as (8), where the pronoun *it* is followed by a copula (is) or a modal, as in (8a)

(8)   a.   Yes when <u>they ask you it is the rules for a interview</u>
         b.   then <u>asked me to show him it worked</u>

Where *it* is a subject, it is usually followed by auxiliary modal or copula verbs, but may also be followed by a dependent verb in a subordinate clause, as in (8b).

Finding these false positives is therefore relatively straightforward. GTD strings that are followed by a part of speech that is tagged as AUX, were filtered and inspected, see Figure 3.7.

```
exlp<-'^(tomorrow|tomorow|tomoz|tomorra|yesterday|tonight|
    today|
monday|tuesday|wednesday|thursday|friday|saturday|sunday|
    first|
last|soon|sooon|now|later|next|shortly|lunch|ages|months|
    often|
everyday|every|quick|quickly|direct|directly|because|cos|
    coz|
drunk|with|wiv|back|all|over|babe|babes|babs|bud|darling|
    lot|
guys|bro|man|pal|mate|son|fam|blud|blood|lad|lad|love|dude|
    baby|
hon|bruvs|bruv|cos|x|xx|xxx|xxxx|xxxxx|did you|didn\`t you|
    did ya|
didn\`t ya|didnt you|didnt ya|free|safe|
please|plz|pls|thanks|thx)'


gtdFP<-d[d$variable=='gtd' & (grepl('^(VBZ|VBD|VBP)', d$
    xpos) |
grepl('^AUX', d$upos)) & (!grepl('(who|whoever|who ever|
    that|which)
$',d$pre) & !grepl(exlp,d$post,ignore.case = T)),]
```

This code extracts tweets containing the pGTD variant where it is immediately followed by a lexical or modal verb or preceded by a relative pronoun. On its own this would be overly aggressive: it captures the majority of false positives, but also captures instances which are not. The *exlp* is a white list of permitted tokens following the keyword phrase.

Out of 29351 tweets initially coded as GTD, 4831 tweets were identified as false positives in this way. The remaining 24970 tweets were spot-checked for further false positives not captured. Out of 200 tweets, there were no examples of further false positives.[16]

---

[16]Of course, there are still likely some false positives in the data, though based on this spot

keyword	post	pattern	upos
passed him it	was because he stopped for	passed him it	VERB, SCONJ, PRON, VERB, ADP
passed him it	was like as if he	passed him it	VERB, SCONJ, SCONJ, SCONJ, PRON
passed me it	spilled some of my coffee	passed me it	VERB, DET, ADP, PRON, NOUN
passed me it	was 6.30 am mind	passed me it	AUX, NUM, NOUN, NOUN
passed me it	would be a big thumbs	passed me it	AUX, AUX, DET, ADJ, NOUN
passed me it	was the grim reaper driving	passed me it	AUX, DET, ADJ, NOUN, NOUN
passed me it	would have been nice if	passed me it	AUX, AUX, AUX, ADJ, SCONJ

Figure 3.7: Filtering false positives: GTD-it followed by AUX|VERB

Typos

Finding typos is particularly difficult. For the most part, the frequency of typos is small enough that it does not have a bearing on the analysis, especially for verbs where overall counts are high. An exception is with the verb *ask* with the $_p$GTD. There are a number of instances where *if* is mistyped as *it*, as in:

(9)   shall I actually <u>ask him it he</u> wants to come??

As overall counts for *ask* are relatively low, such instances are worth removing. All instances of $_p$GTD-*ask* were inspected, and in all (but one) cases they were immediately followed by a pronoun. Such cases were removed, and the results checked to verify that all removed tweets constituted offending cases.

Possessive pronouns: *her* and *me*

Third person pronoun *her* is homophonous between possessive and dative forms. This means that the search string "give it her", intended to return pronominal TGD, also returns examples where *her* is possessive.

Additionally, the pronoun *me* can denote an alternative dialectal spelling of *my*. This puts *me* in the same paradigm as possessive *her*.

Idiomatic phrases with *give*

There are also cases involving idiomatic phrases where the possessum is a superlative, as in (10), where the superlative describes the manner in which the subject conducts themself.

(10)   a.   give it her all

---

check, the total should amount to less than 1% of the data.

    b.   Ah sure gave it me best!

    c.   have to just try and give it me best shot

    d.   Gonna give it me all next weekend

    e.   send it her way

Additionally, superlatives may be modified by an adjective, as in:

(11)   a.   That was flawless. @luciejones1 <u>gave it her absolute all</u>

         b.   No matter what she <u>gives it her very best shot</u>

Akin to the superlative examples, it is also possible to have cases where possessive *her* selects for a nominal that describes the subject's manner of conduct.

(12)   a.   it sounds like she gave it her usual competent ride

         b.   She gave it her heart and soul and then

         c.   for the Birthday girl to give it her final finishing touch

There are also cases where possessive *her* selects for a nominal which denotes the subject's endorsement of the object denoted by the pronoun *it* such as *endorsement, blessing, approval* or *support.*

(13)   a.   phab ! ! @MylahMorales gave it her seal of approval

         b.   ..when this happened and gave it her support

         c.   Oprah gave it her stamp of approval

         d.   I'm sure the creator would give it her endorsement too

Removing possessive pronoun *her*

Results were filtered to exclude text following the key phrase, where the pronoun was *her* or *me* and the following text started with: 'all, way, best, everything, seal, stamp, endorsement, support, usual, approval, heart, absolute, blessing, nod'.

    Results were also manually inspected. A total of 345 false positives were identified for *her* and removed.

Removing possessive pronoun *me*

For pronoun *me*, the string 'g[a/i]ve it me all' was, in most cases mis-identified as a false positive, where it means 'gave it all to me'. The handful of cases

where it meant 'gave it my all', were identified manually, leaving a total of 55 false positives to be removed from the corpus.

GOAL noun phrases

There are cases where possessives *her* or *me* select for a nominal, where it forms a GOAL noun phrase, such as in (14). These can occur with any ditransitive verb. They should also not strictly be part of the pDit corpus. However, as they are examples of the TGD paradigm, it is desirable to code them as pDP for observation and treated separately.

(14) a.   give it her dad/dog/cat/friend
      b.   if I'm not there just <u>give it me mum or dad</u>

This may lead to ambiguity between possessive and GOAL pronouns especially in the imperative, such as in:

(15) a.   Think u shud give it me mate
      b.   well pick it up and give it me mate ;)
      c.   if you get one mail it me mate
      d.   I'm very interested you've sold it me pal saying he can get
      e.   would give it me best mate.

In these examples, the GOAL noun *mate* could either be selected by possessive *me*, denoting the *mate* of the speaker, or, as in the examples in (5), function as a discourse marker referring to the speaker's interlocutor, standardly separated by a comma in writing, as in:

(16) give it me, mate

In (15e) *me,* could form a possessive noun phrase with *best mate,* where *best mate* is the recipient of the object *it.* Alternatively, it could be an example, as in (17), where *best* describes the manner of conduct, and *mate* is again a discourse marker:

(17) give it me best, mate

Modified GOAL nouns

Of course, GOAL nouns may be modified by an adjective, such that the GOAL forms an adjectival phrase, as in (18).

(18)   <u>fed it her little dog</u> !

"Give it her" occurs 1087 times in the corpus, the results were manually coded to exclude these false positives.

Preposition/determiner dropping: pDrop-corp

For [P/D-DROP], there were also a number of false positives to be removed.

how come

When *how* precedes the verb *come,* it forms a question meaning *how is it that...* as in:

(19)   how come the pub was shut with the lights on?

There were 42 examples that followed this pattern, all were removed.

going NOUN mad

There are a small number of false positives with PD-DROP, the verb *go*, followed by an adjective which modifies the noun, such as in:

(20)   London's gone football crazy

(21)   Today I've just done boot camp and at lunchtime I'm doing another class. Think I've gone gym crazy

(22)   need to do it quick, the fixture list for new football season comes out on Wednesday. Twitter will go football mad for the day

(23)   Our @Swim_England guidance on going beach ready to the pool means popping your swimming costume on under your clothes after showering at home. It didn't mean flocking to a crowded beach.

(24)   I've gone cafe crazy today

(25)   is it really nice weather everywhere but here because everyone seems to be going beach mad???

The corpus was searched for PD-DROP strings with variants of *go*, followed by *mad, crazy, bonkers, wild, nuts*. 94 were identified and removed. It is possible, of course that there are some examples with other adjectives, but on manual inspection, none were found.

going beach shopping

There were a number of cases with PD-DROP where the goal noun is actually an adjectival complement of a following gerund such as *shopping*, a structure in which the preposition is obligatorily absent in standard English.

(26)   I'm going holiday shopping

(27)   we went pub crawling

This is distinct from structures where a gerund follows the noun, but acts as an adjectival complement of the noun to form a noun phrase, such as in:

(28)   I'm going football training

Problem verbs: *deny*

Deny can have a different semantics: *prevention of possession* versus *disclaiming an action*.

(29)   I'm at work and they're denying me it until I go to chester/the wirral (Prevention of possession)

(30)   You cheated why do that and then try to deny it to me I was going to find out in the end (Disclaiming an action)

going to football

The string *going to football* is problematic in terms of fitting into a consistent envelope of variation. It is ambiguous between:

(31)   a.   going to [football practice|training] (absent D in standard English)
        b.   going to [the] football (to watch the game)

The result is that D-DROP may be over-represented in some areas where the meaning is (31a), where *football* is a shortening of *football practice*, rather than

(31b), which would correspond to D-Drop.

Analysis of the results showed that the majority of cases of apparent D-Drop with *football* as a GOAL noun, were not comparable with the other variants. It was decided to remove all occurrences of *football* across all variants, which amounted to just under 10,000 tweets.

D-Drop and telegraphic writing

It is common to drop the determiner in *telegraphic* writing, a standard form of shortened writing which tends to drop semantically weak elements in favour of brevity, rather than reflecting traditional (Yorkshire) dialect speech. Some instances of D-Drop in the dataset may be examples of this, particularly in the south.

Again, it is hard to remove such examples systematically, and they are generally at a low level, and likely have a comparable probability of occurring regardless of location. With D-Drop we are interested in where there is a substantially increased proportion of the variant, over and above any background noise. As we will see, the results show clearly the areas where D-Drop is represented.

3.3.5 Creating an unrestricted corpus for the North West: NWatlas-corp

For the North West region, an unrestricted corpus was created, based on where a user once tweeted that they had "grown up in" or "being from" a given place. There are two main advantages to the creation of such a corpus.

1. As it is unrestricted, meaning tweets are downloaded regardless of their content, it may subsequently be searched for patterns not included in an initial search. Using POS-tagging, the door is open to finding more complex structures that may otherwise have been missed.

2. Because it is based on where users reported having grown up, rather than on where they happened to be when they sent a tweet, linguistic output is potentially a better approximation of associated local practice.

The creation of NWatlas-corp used the following protocol:

This protocol resulted in a corpus of approximately 7.8 million tweets for the North West, labeled by the location that users once tweeted that they were originally from.

129

Tweets were retrieved from the API that mentioned having *grown up in* [PLACE], being *born and bred in* [PLACE] or *from* [PLACE], where [PLACE] was a list of locations in the North West.

The search query **did not** request that tweets contained Twitter's geo.place_id information. This greatly increases the number of tweets downloaded, as only a fraction of tweets have this metadata.

Place mentions were disambiguated to ensure that they referenced locations that were truly in the North West of England (e.g. Manchester, UK - not Manchester, NH)

A list of user IDs was compiled for each location, with a limit of 2000 users per location.

For each user ID, up to 5000 tweets were downloaded and labeled with the corresponding place name.

Figure 3.8: Protocol showing steps taken to create NWatlas-corp

3.4 Overview of Twitter corpora

For reference, this section outlines the main corpora. There are three main corpora. The first two (pDit-corp and pDrop-corp) use geocoded data and are limited to a wide range of permutations of ditransitives and preposition dropping. Being limited to specified strings carries the advantage that it is able to cover the whole of the UK and Ireland for the entire ten-year period (2011-2021).

The third corpus (NWatlas-corp) is based on where users reported having 'grown up' at some point on Twitter. NWatlas-corp is not restricted to any word or string, gathering all tweets associated with a given place, but is limited to a range of places in the North West (Birkenhead, Cheadle, Chorley, Eccleston, Formby, Garston, Halton, Haydock, Knowsley, Leigh, Liverpool, Manchester, Ormskirk, Runcorn, St. Helens, Warrington, Widnes, Wigan). NWatlas-corp covers the same time period as pDit-corp and pDrop-corp (2011-2021).

For pDit-corp and pDrop-corp, the number of *hits* is the total number of matching strings, with false positives removed. The corpus size is calculated based on the total number of tweets sent for the locations included during

Corpus	Description	Location	Hits	Corpus size (tweets)
pDit-corp	Strings for $_p$GTD/ $_p$TGD/ $_p$PDAT	UK/IRE	133,160	≈550m
pDrop-corp	Strings for P-DROP/ PD-DROP/ D-DROP/ NO-DROP	UK/IRE	289,107	≈550m
NWatlas-Corp	Up to 5000 unrestricted tweets from users who mentioned being 'from' or 'grew up in' a given place (up to 2000 per place)	North West region	unrestricted	≈7.8m

Table 3.8: Breakdown of Twitter corpora

the 2011-2021 period, using Twitter's Count API. As described, the Counts API allows for the retrieval of the total number of Tweets that are available on Twitter's servers for a given search query. Using this, the total number of Tweets available for a given geo.place_id could be retrieved. However, following the *Search then Download* method, only the matching strings from the GREP file were downloaded from Twitter.

For NWatlas-corp, the entire corpus was downloaded comprising up to 5000 tweets per user and up to 2000 users per location. Because NWatlas-corp is unrestricted, it allows for a more detailed analysis of variability beyond the set strings used in pDit-corp and pDrop-corp. Using the two corpus types allows us to get both a macro view of variation at the level of the UK and Ireland, and a micro view of region-specific variation in the North West. Additionally, the two methods for getting using location information can be compared, which is a useful case-study for the veracity of geo.place_id data as a proxy for user place of origin. I present these findings in the next chapter.

Results and analysis

> "[A]pparently even saying stuff like "give it me" "pass it me" is manc 🖐 🖐 🖐 were supposed to say "give it TO me" kmt that's too long man 😳"

<div align="right">Manchester tweet</div>

4.1 Introduction

This chapter presents data from the three Twitter corpora and sets these alongside data from FRED and BNC. The chapter is structured as follows:

1. I start with a comparison of the Twitter corpora, addressing the question of the reliability of using *geo.place_id* versus *place of origin* to locate users.

2. I then look at the change in the nature of Twitter data over the ten-year period from which they were gathered and the consequences for the analysis.

3. I look at the distribution of [DATALT] and [P/D-DROP] in turn, starting with:

   (a) A broad geographical overview, based on the set strings in the geo.place_id corpora; pDit-corp and pDrop-corp. Here I use cluster

analyses to identify the spatial distribution of variants, and present possible socio-historical interpretations.

(b) Then I take a closer look at each variable with a wider range of component parts of speech in NWatlas-corp. Here, I leverage a combination of part-of-speech tagging and concordances.

4.1.1 Research questions addressed in this chapter

Data reliability

I begin the chapter by addressing the question of data reliability:

RQ (1) How reliable are geocoded Twitter data? To what extent do they correlate with Twitter data located based on user-reported place of origin? How have Twitter data changed over time?

Here I compare the distribution of both variables ([DATALT] and [P/D-DROP]) between the geocoded corpora (pDit-corp/pDrop-corp) and the *place of origin* corpus (NWatlas-corp). For the second part of the question, I then look at the rate of change in each variant over the ten-year period.

Geographical distribution

Of the research questions posed in §1.8, the atlas most obviously addresses:

RQ (2) What is the (comparative) geographical spread of variants of (a) the pronominal ditransitive and (b) preposition/determiner dropping?

Here, I approach each variable in turn, and subsequently look at the comparative spread in §4.8. The maps function initially as exploratory devices, where it is possible to look for patterns, some of which may be unexpected. Once we have established the geographic variation that exists, we are confronted with the question of why we find the distributional patterns that we do:

RQ (3) What historical factors explain the geographical distribution?

Here, I address each variable in turn, starting with [DATALT].

[DatAlt]

The distribution of the pronominal ditransitive (_it me/_me it/_it to me) is robust, and likely has considerable time-depth. Regarding the overall distribution of the ditransitive, I visit the topic broadly, pointing to areas that are fruitful for further study. This is due to the fact that there are simply too much data to offer a comprehensive account of all phenomena. I briefly touch on the likelihood that such distributions may be explained by historical migrations and language contact as far back as the Danelaw in the East and North East, an idea first suggested by Gast (2007), or Norman influence in the South (Tagliamonte, Durham, & Smith, 2014).

The focus, for the discussion of ditransitives is the question of the extent to which the TGD is underlyingly a GTD, or a PDAT with a dropped preposition:

RQ (4) What quantificational evidence can be brought to bear on the TGD-as-GTD question?

Here, I look at the geographic and structural distribution patterns across a range of verb types. Following this, I look in closer detail at the occurrence of ditransitives with full-DP objects in NWatlas-corp, again comparing different verb types. Here, the comparison is between *motion* ditransitives (*bring/take*) and *transfer* verbs (*give/send*), with each type showing distinct distributions. The analysis of ditransitives in NWatlas-corp naturally leads into the question of the geographical distribution of [P/D-Drop] phenomena, which is the topic of the following section.

[P/D-Drop]

Regarding the geographical distribution of the [P/D-Drop] variable, I go into some historical depth, addressing the question of the origins of P-Drop (*go the N*), D-Drop (*go to N*) and PD-Drop (*go N*). The distribution of [P/D-Drop] is more complex than [DatAlt]— some variants are more established, historically, than others. The current understanding is that PD-Drop (*go N*) is a relatively recent development, thought to have emerged in London (Gopal et al., 2021; Hall, 2019), with a possible concurrent development in Kent (Bailey, 2018b). The origins of P-Drop (*go the N*), meanwhile, are not clear, but are likely to have been independently innovated in Liverpool and its use is attested use by older speakers in the OLIVE corpus of Liverpool English (Watson, p.c.).

134

D-DROP (*go to N*), on the other hand, has a much longer history that is likely related to definite article reduction (DAR) (Rupp & Tagliamonte, 2019). As we will see, the extent to which both D-DROP and P-DROP align with the distribution of DAR, suggests a close connection, the exploration of which forms a sizeable part of the current chapter.

RQ (5) To what extent do [P/D-DROP] variants align with DAR distribution?

In relation to a possible connection with DAR, I consider the question of the extent to which PD-DROP originated in London and subsequently diffused across the country, or whether it may have been independently innovated in multiple locations:

RQ (6) Was [P/D-DROP] independently innovated in multiple locations, or did it spread from one point of innovation in the South East of England?

For PD-DROP, the geographical distribution does appear, as Gopal et al. (2021) demonstrate, to follow what we would expect to see if it had recently diffused north from the South East via a gravity-based model. However, Gopal et al. (2021) collapse PD-DROP and P-DROP into one dataset which masks the distinct distribution of P-DROP in the North West. It is clear, however, from recent literature that each variant is structurally distinct, and likely developed independently in the South East and North West, respectively.

Whilst I show that PD-DROP is well attested as far north as Manchester, it is also attested across Ireland and the South West of England. In the Midlands and North of England, there is evidence that it may also have a connection with DAR. Meanwhile, P-DROP seems to have originated — probably earlier — in the Liverpool area, exhibiting a particular localised spread, which I suggest may have been set in place during the time that Liverpool was emerging as an industrial port.

The current distribution of variants in the North West creates an interesting dynamic, where the towns in Manchester's sphere of influence (SOI)[1] such as Bolton, Bury and Wigan seem to have followed suit in preference for PD-DROP over P-DROP. Meanwhile, Liverpool appears to have its own SOI that extends to the border of Manchester's, including Warrington and St.Helens, but also as

---

[1]The term 'sphere of influence' is usually used with reference to the political and economic influence of a state on surrounding smaller states, but is adopted here to describe the sociolinguistic influence of a city on surrounding towns. It is equivalent to the term *functional zone*.

far as Stoke-on-Trent and even parts of the Black Country and Dublin.

Finally, in the North West, as we will see, [DATALT] and [P/D-DROP] exhibit distinct spatial distributions. Here, I compare the relative geographical distributions of both [DATALT] and [P/D-DROP] and, following Biggs (2018), look at where the two phenomena intersect:

RQ (7) Is there a likely location for the grammar described by Biggs (2018), given the distribution of the Twitter data?

The intersection of TGD ditransitives and P-DROP phenomena are of special interest for the survey (presented in the following chapter), as it is in these locations that we are most likely to find evidence for TGD-as-P-DROP.

4.2 Comparison of corpora

4.2.1 Comparing to previous corpora

I start with a brief comparison of pDit-corp to the results extracted from other corpora, by previous studies, which targeted the dative alternation. This is useful for getting a sense of the scale of the data available in the present corpus, and for cross-checking the overall relative rates of variants.

The overall UK/IRE counts for pDit-corp are presented in Table 4.1, compared to the results from a previous Twitter corpus (TAGS-corp (Stevenson, 2016)), FRED and BNC as reported by Gerwin (2014). TAGS-corp used data exclusively from GPS-encoded tweets from November 2014 to March 2016, which provided a useful increase in geographical detail over traditional corpora.[2] The present pDit-corp, which uses geo.place_id and drew data for a wide range of verbs from 2011-2021, greatly increases the amount of data available, allowing for subsetting at the level of small towns, and the revealing of region-internal linguistic boundaries.

Aside from the obvious increase in the scale of the data, overall relative percentages are comparable between the corpora.[3]

The fact that the relative proportion of variants is equivalent between Twitter data and traditional spoken corpora supports the frequently made observa-

---

[2]Additionally, TAGS-corp only used the past tense forms of *give* and *send*.

[3]I present here a comparison only between pDit-corp and other corpora as there are no equivalent corpora for [P/D-DROP] with which to compare pDrop-corp. Whilst I do run a number comparisons between [P/D-DROP] variants in pDrop-corp to examples in FRED and BNC, a systematic quantitative treatment of traditional corpora was not undertaken. I leave this to future work.

Corpus	%TGD	%GTD	%PDAT	Hits	Corpus size (words)
pDit-corp	13	23	64	133,160	≈9.8 billion
TAGS-corp	19	26	55	1,407	≈107 million
FRED (spoken)	10	15	75	140	2.5 million
BNC (spoken)	14	30	56	410	3.3 million

Table 4.1: Percentages and totals for pDit-corp, compared to TAGS-corp (Stevenson, 2016), FRED, and BNC. See appendix for a full Table of counts by place (pDit-corp filtered to a minimum of 15 per location and maximum of 5 tweets per user).

tion that linguistic production on social media, while written, actually mirror speech practice in a way that makes it particularly useful for dialectological analysis. Recall Biber's (1991) finding that pronominal ditransitives occur rarely in standard writing (Figure 2.3). Overall parity with traditional corpora is also reassuring in the context of lingering doubts over the validity of Twitter as a valid resource for the study of dialect variation.

4.2.2 Comparing pDit-corp/pDrop-corp with NWatlas-corp

Here, the two methods for gathering location data (geo.place_id and place of origin) are compared. This sheds light on the question:

(1)   How reliable are geotagged Twitter data? To what extent do they correlate with Twitter data located based on user-reported place of origin?

Answering this question is important, given that the majority of Twitter-based research uses exclusively geo-tagged data. The exception to this, and the method of using place of origin on which NWatlas-corp is inspired, is work by Willis (2020) and Gopal et al. (2021). However, to-date, no study has compared the two methods. If geo-tagged data are found to correlate reliably with *place of origin* data, this would offer some reassurance to the validity of the former method.

pDit-corp

Table 4.2 compares the percentages for each variant between NWatlas-corp and pDit-corp. For consistency in this comparison, NWatlas-corp was searched for the same set of strings that were used to create pDit-corp.

Plotting pDit-corp data against NWatlas-corp data (Figure 4.1), we can see that, overall, there is a strong correlation between the two types of location data. It is apparent that there is a slight preference towards the standard PDAT

| | **NWatlas-corp** | | | | **pDit-corp** | | | |
Location	$_p$TGD	$_p$GTD	$_p$PDAT	Total	$_p$TGD	$_p$GTD	$_p$PDAT	Total
Birkenhead	8.2	29.7	62.1	380	16.4	28.5	55.1	162
Cheadle	35.4	6.2	58.4	161	37.6	9.4	51.8	84
Chorley	35.5	12.4	52.1	307	44.4	6.3	46.3	122
Garston	1.6	21.3	77.1	61	3.1	34.4	54.7	59
Halton	7.8	23.1	69.2	39	41.5	22.3	36.2	98
Knowsley	9.3	38.7	52.0	75	5.8	45.6	42.7	97
Leigh	17.9	14.3	67.8	321	18.5	18.5	62.9	27
Liverpool	7.2	29.3	63.5	2128	10.3	35.8	46.9	1304
Manchester	26.1	10.7	63.2	1948	32.2	11.7	52.1	1840
Ormskirk	23.1	32.6	44.3	117	26.3	36.8	57.9	19
Runcorn	21.7	27.2	51.1	387	22.5	20.0	57.5	39
St. Helens	24.4	12.8	59.7	651	36.4	22.2	41.4	100
Warrington	26.8	12.3	60.9	928	43.4	13.4	43.1	182
Widnes	20.2	16.5	63.3	315	56.7	20.0	23.3	46
Wigan	25.4	11.9	62.7	1526	44.6	10.9	44.6	292

Table 4.2: Percentages of variants for each location in NWatlas-corp and pDit-corp, with totals. The same search was used for each corpus.

in the *place of origin* data, with all but one point falling below the reference line. Conversely, there appears to be a slightly higher incidence of TGD for places in the *geo.place_id* dataset. The differences are interesting, though relatively marginal. In general, however, these results show that both measures of user location are comparable.
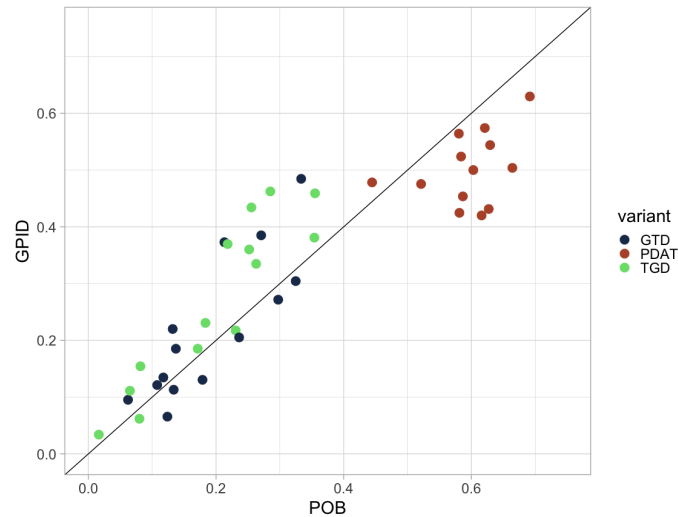
Figure 4.1: [DATALT], geo.place_id plotted against place of origin/P.O.B. Correlation: 0.8906833

pDrop-corp

Results for [P/D-DROP] are similarly comparable between NWatlas-corp and pDrop-corp.

Location	NWatlas-corp					pDrop-corp				
	pdD	pD	dD	noD	Total	pdD	pD	dD	noD	Total
Birkenhead	6.4	48.3	4.0	41.3	698	4.5	56.9	6.4	32.3	533
Cheadle	17.1	9.1	4.4	69.5	298	22.4	3.5	4.8	69.3	313
Chorley	14.8	6.2	8.3	70.7	661	24.9	10.3	6.5	58.3	398
Formby	2.4	33.2	4.7	59.7	211	1.7	33.3	8.3	56.7	60
Garston	3.8	29.1	6.3	60.8	79	5.9	45.2	10.1	38.8	188
Halton	3.3	15.0	18.3	63.3	60	9.1	68.5	2.2	20.3	276
Haydock	6.0	44.0	8.0	42.0	50	10.3	43.6	2.6	43.6	39
Knowsley	2.7	62.0	5.3	30.0	150	8.2	69.1	4.7	18.1	343
Leigh	22.8	7.6	6.1	63.5	855	13.9	11.6	5.8	68.6	86
Liverpool	6.5	39.4	5.2	48.9	2925	9.8	48.7	5.2	36.4	3512
Manchester	15.2	4.1	7.0	73.7	3079	25.5	4.8	6.1	63.6	6226
Ormskirk	2.7	45.0	4.7	47.7	149	6.9	34.3	2.2	56.7	111
Runcorn	4.9	55.1	3.3	36.8	799	10.6	50.0	3.6	35.7	154
Warrington	8.1	35.0	3.4	53.5	1963	8.9	31.5	5.9	53.7	1263
Widnes	5.9	58.1	3.2	32.8	808	11.4	48.4	2.2	38.1	183
Wigan	16.4	21.1	5.5	57.0	3262	11.0	26.8	5.7	56.4	958

Table 4.3: Percentages of drop categories for each location in NWatlas-corp and pDrop-corp, with totals
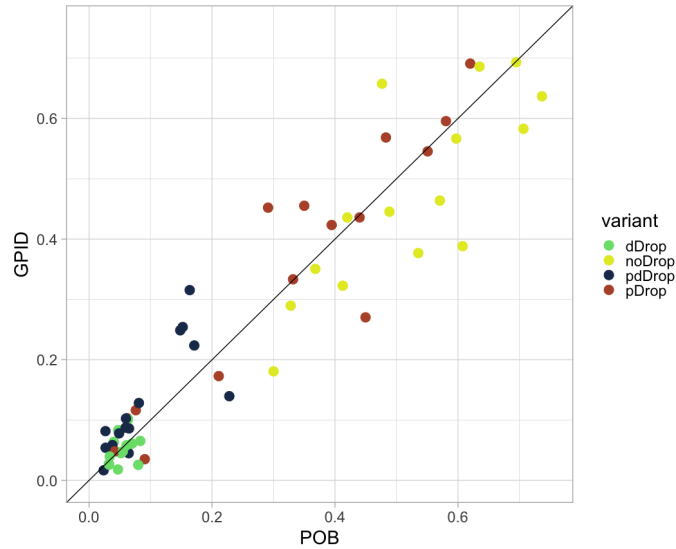
Figure 4.2: [P/D-DROP], geo.place_id plotted against place of origin. Correlation: 0.9440272

There is again a strong correlation between the results for the *place of origin* data and the *geo.place_id* data with [P/D-DROP]. In terms of raw counts, the method used to gather data for NWatlas-corp— up to 2000 users per place and up to 5000 tweets per user — retrieved a comparable number of *hits* per

location, when the same search criteria were used.[4]

Summary

A frequent challenge against the use of Twitter data for dialectal research is with its use of tweet-geolocation as a proxy for a speaker's (socio)linguistic roots. Dialect study is naturally interested in where a speaker spent the majority of their upbringing, rather than where they happen to be located at a given time. Nevertheless, the data presented here show that tweet-geolocation does appear to function adequately as such a proxy. This is supported by the fact that numerous investigations which use this method have shown that variation appears to correlate with that gathered using traditional methods.

Further, as we will see, pDrop-corp and pDit-corp, which use Twitter's *geo.place_id* metadata, display robust and consistent geographical variation which is hard to explain if tweets sent from a particular location are not broadly reflective of the language use in that location.

4.3 Twitter data: change over time

Twitter studies have typically drawn on data produced between 2013 and 2016, when GPS metadata were most plentiful. Gopal et al. (2021) is one exception to this trend, drawing data from 2017-2019. Another exception is Grondelaers et al. (2023) who examine Twitter data between 2011-2019. The present study extends this window from 2011 to 2021, using both geo.place_id data (for pDit-corp and pDrop-corp) and place of origin data (for NWatlas-corp). Drawing on data from a longer time span is advantageous as it both increases the amount of data available, and offers the possibility of investigating change in the output of Twitter users over time.

If the nature of Twitter data were consistent over time in terms of register and population sample, it may be possible to track the evolution of variants within and across locations. It might be expected, for instance, that the rate of PD-DROP would increase over time, particularly in northern areas (or at least not decrease), if it were in the process of diffusing from London, following Gopal et al. (2021).

---

[4]For the purpose of consistency in these comparisons, the same GREP file was applied to NWatlas-corp as was used for pDit-corp and pDrop-corp, using a long list of predefined strings (see §3.3.2). For later analyses of NWatlas-corp, presented in the coming sections, a more nuanced filtering of the data was employed.

This is, however, not what we find. As Figure 4.3 shows, the relative frequency of PD-DROP peaks in 2013, then falls markedly over the rest of the ten-year period. Standard NO-DROP is the inverse of PD-DROP, increasing during the same period. D-DROP actually increases overall, while P-DROP stays at a relatively constant $\approx 5\%$.[56]
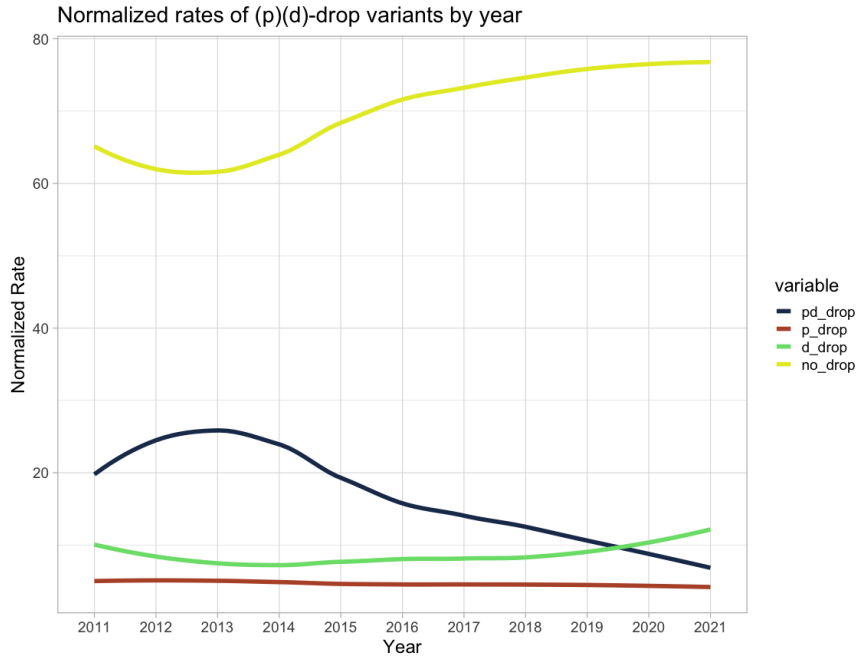
Figure 4.3: Change in relative frequency of [P/D-DROP] variants by year.

These results closely mirror those found by Grondelaers et al. (2023) in their study of the non-standard use of object pronoun *hun* in subject position in Dutch-language tweets, reproduced in Figure 4.4.

Relative rates of the non-standard drops dramatically after 2013, going from a majority variant in 2011-2013, to a minority variant by 2016/2017. The authors attribute this decline to an 'exodus' of young Twitter users and a "changing of the guard" towards an older age demographic which engenders the use of more standard registers. Interestingly, they ultimately link the use of the non-standard form not exclusively to younger age groups, but to "youngish informal self-stylization by tweeters of any age" (Grondelaers et al., 2023, p.242), and that it is this kind of stylisation that has declined on Twitter.

---

[5]The extremes of the shift in the [P/D-DROP] data are shown in the difference in the results plotted geographically between 2013 and 2021, in Figure B.1, in the Appendix.

[6]See appendix for a full breakdown of counts over the ten-year timespan.
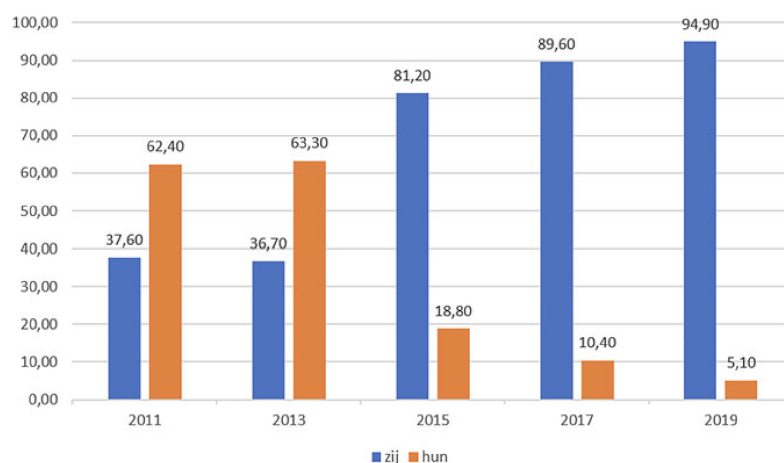
Figure 4.4: Relative frequency of standard *zij* and nonstandard *hun* as a function of year.
Source: Grondelaers et al. (2023, p.237)

Assuming that the rate of PD-DROP in actual speech has not decreased, I make some suggestions as to why we might see this apparently global decline in non-standard use on Twitter.

1. Following Grondelaers et al. (2023), the changing age demographic is likely to be playing a substantial role: as shown in Section 3.2, the age profile of Twitter users shifted markedly between 2013 and 2024 towards older speakers. This may result in *age grading,* the well-documented effect of a shift in the rate of standard/non-standard forms as speakers move through different life stages.

2. Increasing sophistication of grammatically aware auto-correct.

3. A cultural shift in the way that Twitter is used as a social media platform. It may be that, in its earlier years, Twitter functioned as a more interpersonal social network, with users interacting as if they were 'chatting' in the way that they might if they were using direct messages over WhatsApp or Apple Messages. Latterly, Twitter may have found a different role in the social media landscape, associated more with public or professional engagement, which engender use of more standard registers.

4. Competing platforms, such as TikTok, likely played a significant role in capturing younger audiences, and offered a space for less formal exchange (a point also made in Grondelaers et al. (2023)).

5. A change in the way that Twitter assigned location data. After 2015,

142

the amount of data tagged with precise GPS location dropped to near zero, being replaced with location data derived from user-entered data by Twitter's *Profile enrichment* algorithm. It is unclear why this might have the effect of increasing the relative rate of standard variants, however. Critically, we see a similar shift towards standard forms in the NWatlascorp which uses a consistent measure of user location (based on where a user reported being from).

The shift from non-standard to standard language use poses a problem for our analysis in terms of representativity. Where do we draw the line, chronologically? Which time period is more representative of actual speech practice? This is, of course, not just a problem with Twitter, but any corpus study, which ideally should seek a balance of text types.

However, it is clear from the data that, for a given time-period, the results are consistent over geographical space: between neighbouring locations, the relative rates of variants change little across broad areas (and when they do, they do so unilaterally, indicating clear isoglosses). There is also consistency over several years - from 2012 to 2015, the distributional patterns change little.

If my suspicion is true that part of the explanation for a shift towards more normative language use is a move away from the use of Twitter for conversational messages, then we might expect to see a relative decrease in the rate of tweets that are in response to another tweet for the variants in question. And this is indeed what we find: Figure 4.5 shows that where tweets are *not* in reply to another tweet, the rate of change is more modest. Tweets that *are* in reply appear to be driving the increase in PD-DROP (blue) and corresponding decrease in NO-DROP (yellow) in 2013.

The chart shows clearly that the *is reply* status does not have a significant bearing on the rate of P-DROP (*go the N*: red) at any time period, perhaps a first indication of its distinct sociolinguistic and pragmatic status. The fact that P-DROP is locally maintained in spite of the global shift towards NO-DROP, indicates that it—in part—belongs to an established, localised standard register, defined by regional norms. As we will see, this maps to what we find on closer inspection of the use of variants: P-DROP is found across registers, including informal styles, but also impressionistically higher-register linguistic outputs, involving abstract nouns, which are unlikely to be permissible with PD-DROP, following Hall (2019):
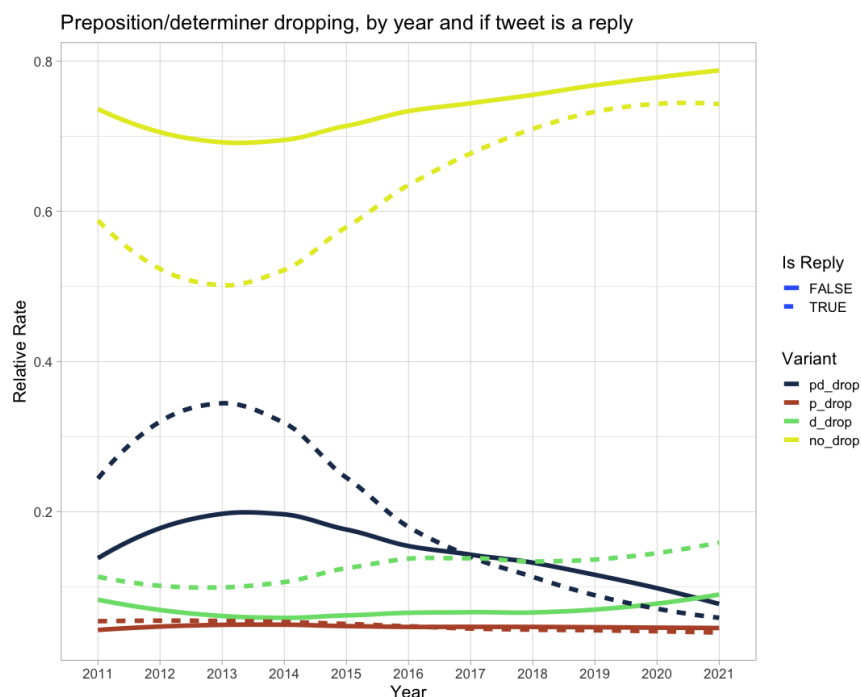
Figure 4.5: Change in relative frequency of [P/D-DROP], including whether a tweet was a reply to another tweet. The dotted lines represent the proportion of tweets that are in reply to another tweet, for a given variant.

(1) a. I have come the conclusion that..

b. *I have come conclusion that..

The maintenance of P-DROP in the face of the reduction of PD-DROP is further indication of its structural distinction. Interestingly, the rate of D-DROP is consistently higher when *in-reply* than when *not-in-reply*, and shows an overall increase in use. It is not clear, at this point, what is underlying this increase, but it may be due to the fact that D-DROP is characteristic of telegraphic speech (which is a feature of standard varieties), as well as reflective of regional speech practice.

Pronominal ditransitives

Regarding pronominal ditransitives, Figure 4.6 shows that both TGD and GTD undergo a similar reduction in frequency relative to PDAT, though the difference is not as pronounced as with [P/D-DROP]. ₚTGD (*it me*) peaks at around 15% in 2013 and drops to around 10% by 2017, where it appears to stabilise.
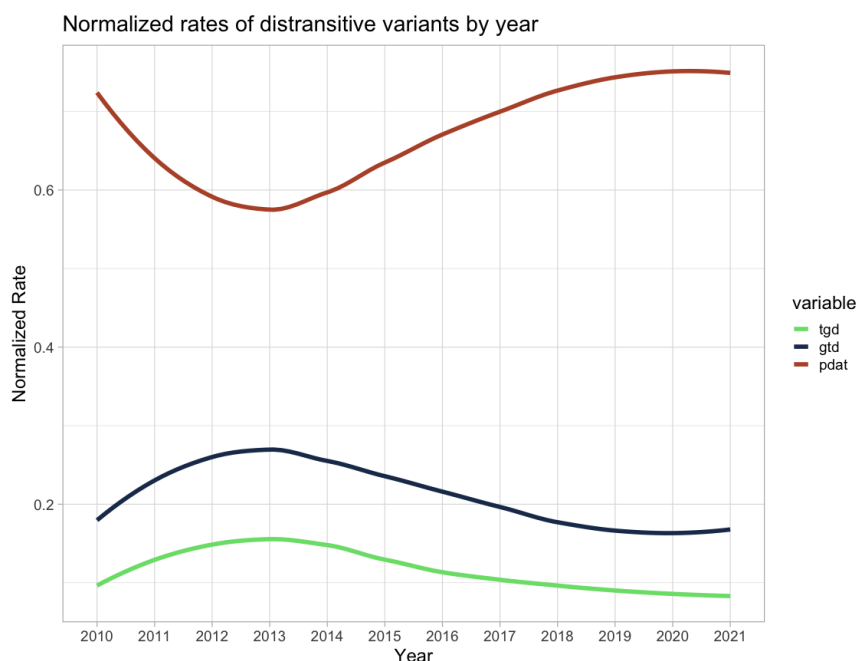
Figure 4.6: Change in relative frequency of pronominal ditransitive variants by year.

The relative stability of ₚTGD (_it me) indicates that it is, like P-DROP (*go the N*), locally not associated with informal or non-standard registers. The reduction in ₚGTD (_me it) is perhaps surprising, given that it is not typically considered non-standard, though this may be explained by register variation; ₚGTD while not non-standard, is arguably of a more informal register.

When we split the data by whether a tweet was *in reply* to another tweet (Figure 4.7), we again see that, for *not-in-reply* tweets, the relative rates for ditransitive variants remains markedly more stable, showing a slight decline in ₚTGD (_it me) and ₚGTD (_me it), and corresponding increase in ₚPDAT (_it to me), over the ten-year period.

Meanwhile, the rate of ₚTGD and ₚGTD is inflated through 2012-2014 for *in-reply* tweets, but drops to near parity with *not-in-reply* tweets by around 2017. This indicates that the increase in ₚTGD and ₚGTD shown in Figure 4.6 is largely driven by *in-reply* tweets. This fits the picture for pronominal ditransitives, that pronouns, as anaphora, necessitate an object to which they refer and are therefore more likely to occur in speech. It also demonstrates a pragmatic alignment between ₚTGD and ₚGTD, fitting the assertion in the literature (Haddican, 2010) that they are underlyingly akin to each other. Furthermore,
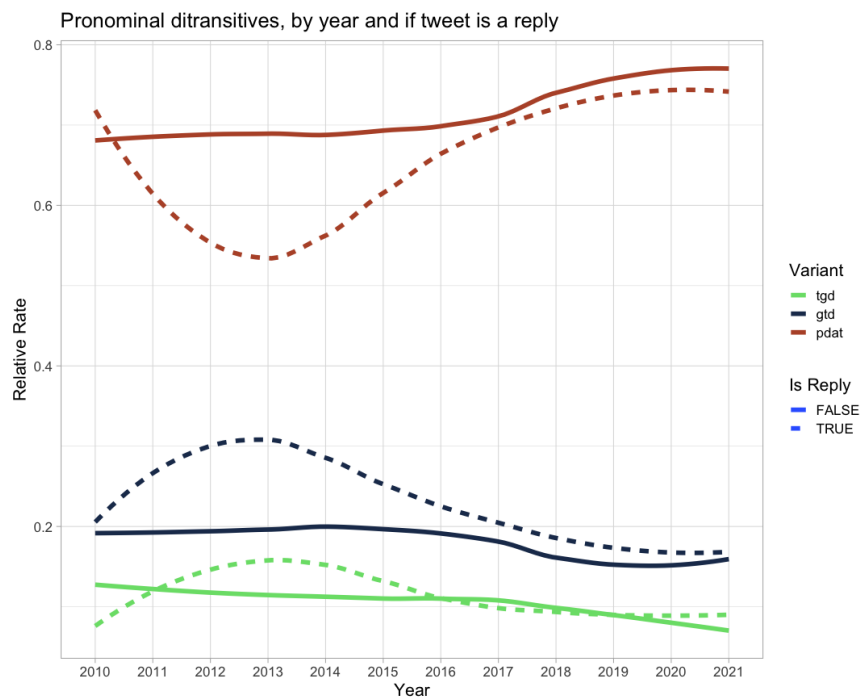
Figure 4.7: Change in relative frequency of pronominal ditransitives, including whether a tweet was a reply to another tweet.

it seems likely that specific pronouns will also have a bearing on relative rates of pronominal ditransitive variants: first and second person pronouns are more likely to be used with reference to interactional participants than third person pronouns, more likely to be *in-reply*, and therefore more likely to account for a greater proportion of the increase in DOC variants.

Summary

The changing nature of Twitter data over time is a reminder that we are dealing with dynamic and shifting data. Any linguistic study necessarily takes a subset of the linguistic output for any given time and place and it can not be said, for certain, the extent to which a given subset is truly representative of the population as a whole. We are left with a difficult methodological decision: should we focus on data drawn from Twitter during its more non-standard-leaning incarnation, or include the whole range? The decision taken here is the latter.

The data presented in the coming sections are for the entire ten-year period. Taken as a ten-year window, we are arguably seeing a more balanced snapshot

146

of digital production by a larger subset of the population with a greater age-range, and including an increased range of registers.

Crucially, the view taken here is not that PD-DROP or $_\text{P}$TGD are in decline, rather they have become less represented on Twitter. It is likely that if data from other social media were collected, such as private Whatsapp and Apple Messages, we would see a continuation of its spread. Clearly, this is hard to test empirically, though such methodologies have been undertaken (Dorantes, Sierra, Donohue Pérez, Bel-Enguix, & Jasso Rosales, 2018) (also see Shortis (2015); Tagg (2009) with SMS text messaging).

4.4 Mapping the dative alternation

We start with the mapping of single variants. These initial maps provide a first impression of the distribution of each variant.

Pronominal THEME-GOAL ditransitives

The spatial distribution of pronominal THEME-GOAL ditransitives (_it me) is clearly dominant in the area expected, given the distribution reported in the SED and the Dialect App (Britain et al., 2018) (recall Figures **??** and 2.5), as well as in the Our Dialects project (MacKenzie et al., 2022). The distribution also, unsurprisingly, mirrors that reported in previous Twitter corpora (Gopal et al., 2021; Stevenson, 2016), though supplies additional detail.
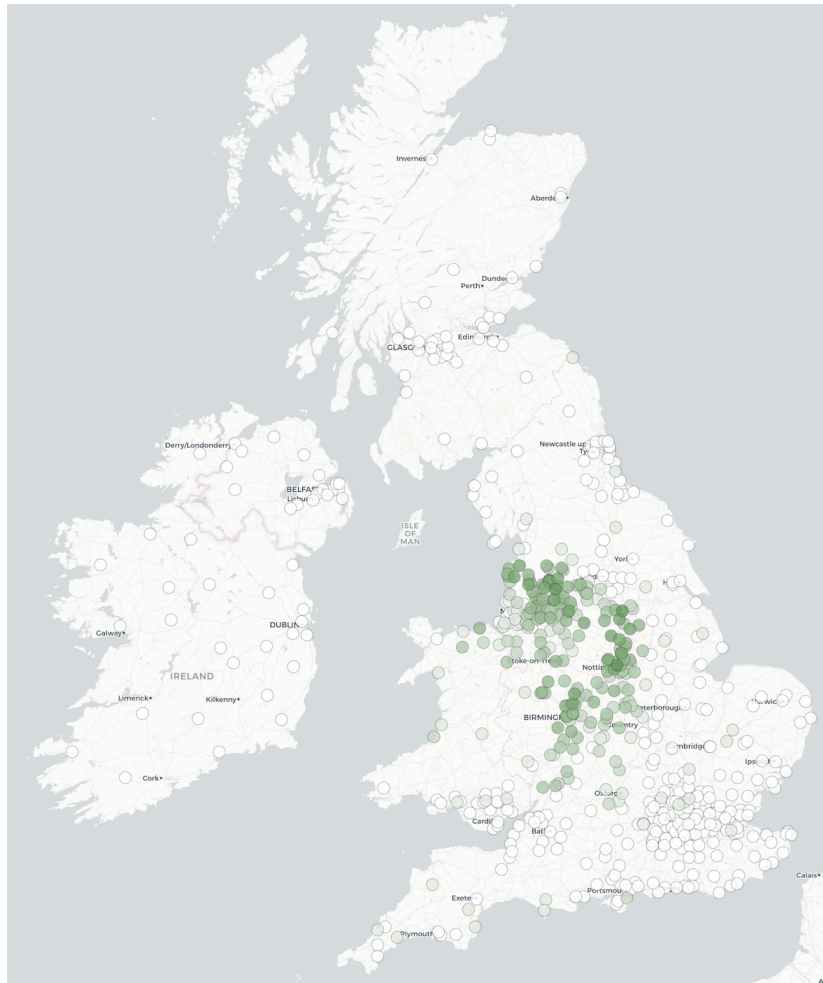
Figure 4.8: Pronominal THEME-GOAL ditransitives on Twitter 2011-2021
Darkest green ≈ 60%, Mid green ≈ 30%, Lightest green ≈ 10%, White ≈ 0%

The new data presented here shows a fine-grained gradation at the fringes of the most dominant places. This gradation, in places south of Manchester and in the area between the East and West Midlands, is starkly contrasted by an apparent lack of gradation on the Eastern and Southern borders of the TGD area. Here, instead, we see a sharp border, where usage rates fall off dramatically, going from dark green (50%) to white (close to 0%), over the space of only a few miles.

Pronominal GOAL-THEME ditransitives

The distribution of pronominal GOAL-THEME ditransitives (_me it_), is, in the North of England, almost a relief of the imprint of the distribution of ₚTGD.

148
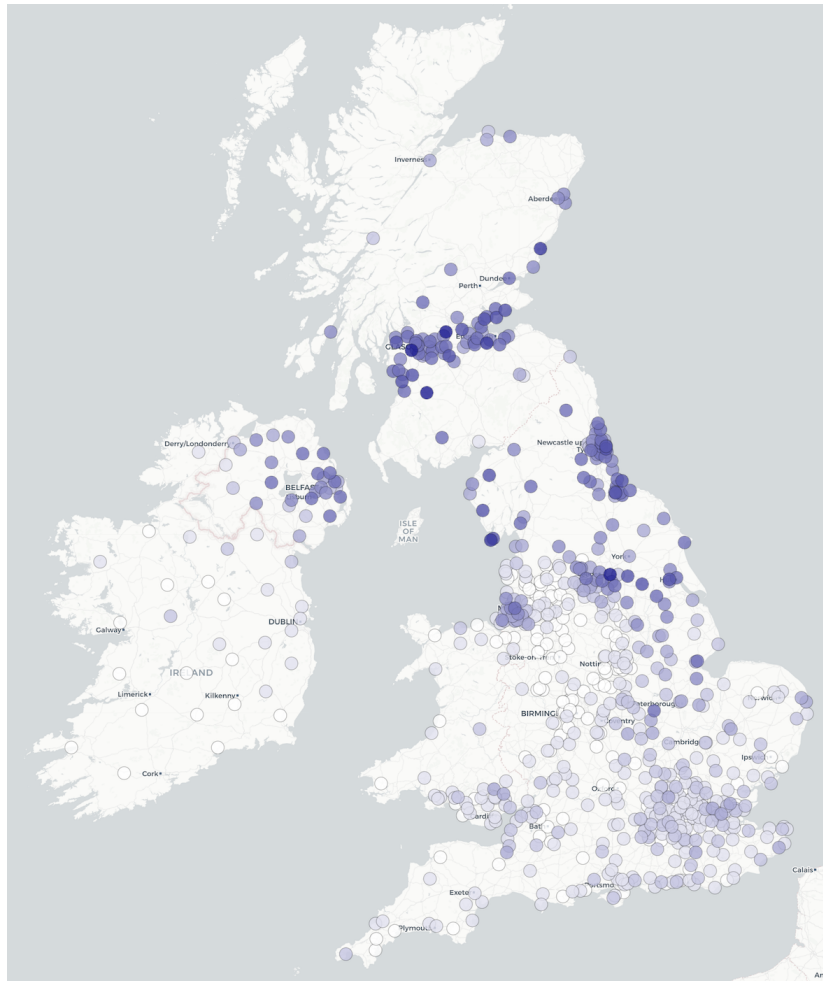
Figure 4.9: Pronominal GOAL-THEME ditransitives on Twitter 2011-2021
Darkest blue ≈ 60%, Mid blue ≈ 30%, Lightest blue ≈ 10%, White ≈ 0%

We can clearly see the same sharp border in the East, where places such as Bradford, Leeds, and Doncaster strongly favour ₚGTD (_me it_). Attention is also now drawn to the North West; Liverpool shows up as a small island of ₚGTD dominance. The rest of the ₚGTD-dominant distribution stretches across the East and North East of England, through Northumberland, with some of the heaviest use in Scotland and Northern Ireland.

In Ireland, we see a dramatic distinction between Northern Ireland and the Republic of Ireland in the geographical south. A continuity between Scotland, the North East and Northern Ireland is particularly interesting, and reflects historical migration patterns, a point I return to in the discussion of [P/D-DROP] distribution. We also see that ₚGTD is used quite widely in the South of England

149

and Wales, though here it represents a minority variant.

The identification of $_p$GTD (_*me it*) as dominant in Scottish English is reported as far back as 1800, in — where its use is discouraged as a 'vulgar Scotticism'. The fact that it was identified by grammarians is an indication that it was likely frequent in England too, at that time.

> Give me it, show me it; Sc.—Give it me, show it me. The former
> is Scotch, the latter English. (Mitchell, 1799, p.54), cited by Yáñez-
> Bouza (2016b, p.156).

As I will discuss, the presence of $_p$GTD in North East England, Scotland, and subsequently Northern Ireland, likely has deep-historical roots, which is echoed in the distribution of D-DROP, and in Strelluf's (2020) maps of NEED + *past* (as in *the car needs washed*).

Pronominal prepositional datives

The pronominal prepositional dative (_*it to me*, Figure 4.10), is the most widespread of the three variants, with no locations showing an absence of the structure.

This is no surprise, given its status as the present-day standard form. We can see that it is most dominant in the South of England, and in the Republic of Ireland. There are also notable pockets of dominant $_p$PDAT use scattered throughout England, including within broader regions that are broadly $_p$TGD or $_p$GTD, such as the Midlands and Scotland, respectively. There additionally appears to be an urban factor driving an increased use of $_p$PDAT, with clusters of darker red points ($_p$PDAT →80%) around urban centres such as Manchester, Leeds, Nottingham and Birmingham.

The dominance of $_p$PDAT in the South of England, and the Republic of Ireland does fit geographically with some suggestions that have been made regarding a Norman origin of the preposition variant (Tagliamonte, 2014, p.299), a point to which I will return.

4.4.1 Cluster analysis

Whilst presenting geographic distribution by individual variant is helpful to simplify the visualisation and see the gradation in use between locations, *K-means cluster analysis* (part of the *Stats* package, (R Core Team, 2013)) offers
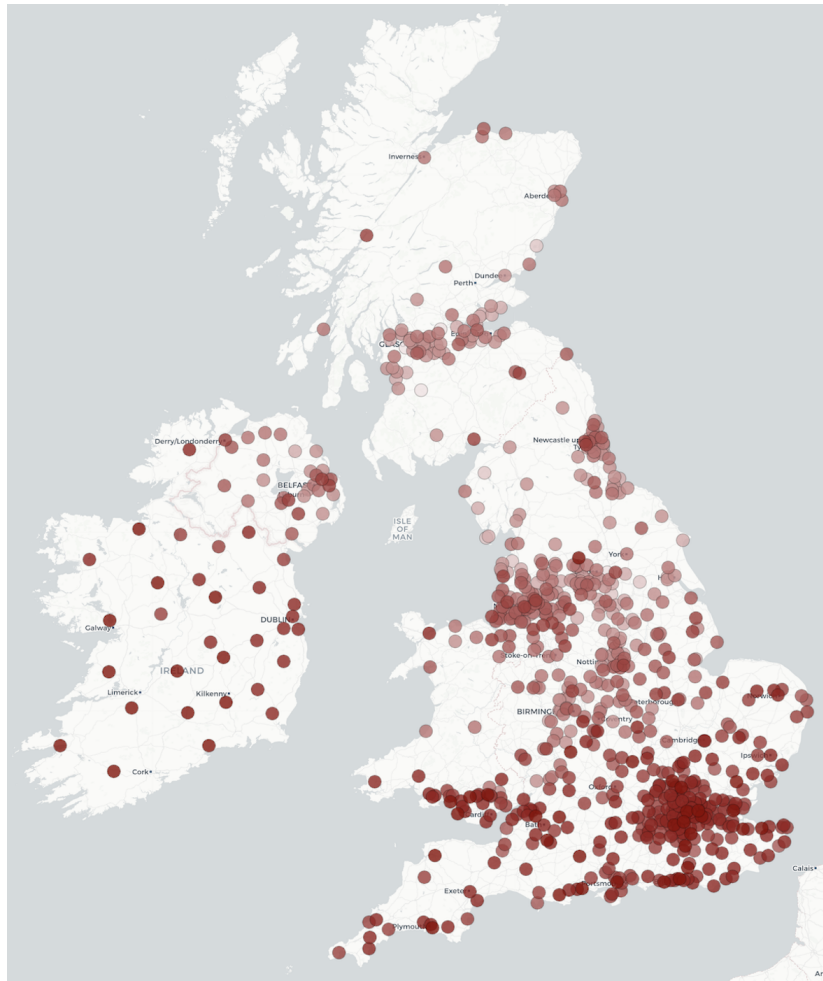
Figure 4.10: Pronominal prepositional ditransitives on Twitter 2011-2021
Darkest red ≈ 90%, Mid red ≈ 60%, Lightest red ≈ 30%

a way to group locations which pattern together by the relative frequency of all variants. Crucial to K-means analysis is determining the optimal number of clusters into which data should be grouped. Ideally, the number of clusters should be as small as possible while maximising uniformity within each cluster. Often, the optimal number of clusters is not clear and there is some leeway for discretion. A smaller number of clusters may be helpful for visualising broader patterns and eliminating noise in the data, whilst adding more clusters may offer greater nuance. There is however a clear point of diminishing returns. With too many clusters, the *centre* of each cluster may be too similar to neighbouring clusters.

A starting point for determining the optimal number of clusters is to use an

elbow plot. An elbow plot essentially involves running the clustering algorithm multiple times, for a given range of test cluster values, and then plotting each against the total *within-cluster* sum of squares. With a single cluster, the *sum of squares* value is equal to the dataset as a whole, but decreases sharply as clusters are added. The point at which this rate of decrease levels off, gives an indication of the optimal number of clusters required to effectively group the data.

This can be seen in Figure 4.11: after three clusters, there is a relatively modest decrease in the *sum of squares* for each cluster. This indicates that three clusters are likely optimal for this data set, which fits with the impressionistic interpretation of the single-variant maps in the previous section.
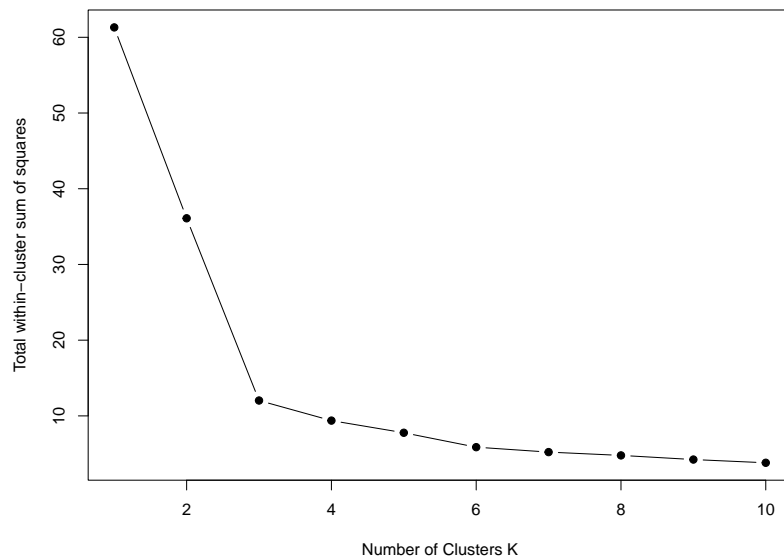
Figure 4.11: Pronominal ditransitives, elbow plot.

Taking three clusters as our starting point, we can then examine the *Centres* for each cluster. *Centres* represent the mean value for each variant against which data points are measured. centres for the pronominal ditransitive data are presented in Table 4.4.

Clusters 1 and 3 can be characterised as high-$_p$GTD and high-$_p$TGD, respectively, each with around 45-50% usage frequency. Both clusters have equivalent rates of $_p$PDAT at around 45%. Notably, cluster 1—the high $_p$TGD cluster—also shows around 10% $_p$GTD, reflecting what has been suggested in previous

Cluster	pTGD %	pGTD %	pPDAT %
1	4.5	50.5	45.0
2	6.1	16.0	77.9
3	44.5	11.4	44.1

Table 4.4: Breakdown of centres for each cluster:
pronominal ditransitive, three clusters

work; many speakers who use $_p$TGD may also use $_p$GTD. Speakers who use $_p$TGD and $_p$GTD interchangeably are covered in §A.

Cluster 2—high $_p$PDAT—increases, by an additional 32%, the *baseline* 45% found in the other clusters, with very low levels of $_p$TGD and moderate $_p$GTD use. Cluster 2 mirrors the end state found in Yáñez-Bouza and Denison's (2015) historical investigation into the relative frequency of pronominal ditransitives (see Figure 2.4 on page 60).

We can further examine the appropriateness of using a given number of clusters (in this case three), by plotting the clustered data as a *silhouette* plot (produced using the *silhouette* in *cluster* package (Maechler, Rousseeuw, Peter, Struyf, Anja, Hubert, Mia, & Hornik, Kurt, 2025)). In the silhouette plot, each data point (location) is drawn as a line on a range of *Silhouette width* (s_i), from -1 to +1. An s_i of 1 corresponds to an exact match between a data point and the centre values for that cluster. The closer a data point is to 1, therefore, the better its *fit* is to its cluster. A higher number of data points with s_i values approaching 1 is therefore desirable, with as few as possible near zero.

Figure 4.12 shows the silhouette plot for the pronominal ditransitive data set. The mean s_i value for each cluster is presented on the right of the plot, following the number of data points in that cluster. We can see that the mean s_i is about 0.6 for each cluster (0.63 combined), which denotes a good fit.

Finally, it is helpful to plot clusters on a scatterplot. Doing so allows us to visualise the degree of variance for each cluster. In 4.13, we can see that there is a good degree of separation between clusters. Within each cluster, there is quite a range: $_p$TGD ranges from around 20% at the low end, and 60% at the high end. There is less variability with the $_p$GTD cluster which ranges from 30% to 60%. Note too, that there is some degree of overlap at the fringes (indicated by the dotted oval). For instance, some TGD clusters have an almost equal ratio of $_p$GTD to $_p$TGD at around the 20% level. Importantly they are nevertheless still characterised as TGD due to the fact that 20% for $_p$GTD is still relatively low compared to other locations that are diagnosed as $_p$GTD-dominant.

3 clusters C_j
$j : n_j | ave_{i \in C_j} s_i$

1 : 535 | 0.64

2 : 284 | 0.55

3 : 221 | 0.53

0.0   0.2   0.4   0.6   0.8   1.0

Silhouette width s_i

Average silhouette width : 0.59
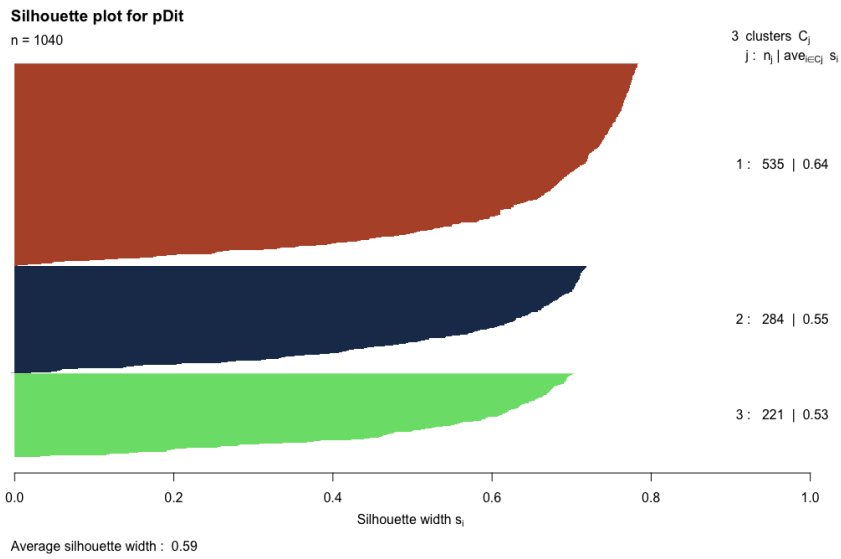
Figure 4.12: Pronominal ditransitives silhouette plot: three clusters.

Relative rates of TGD/GTD by location, coloured by cluster
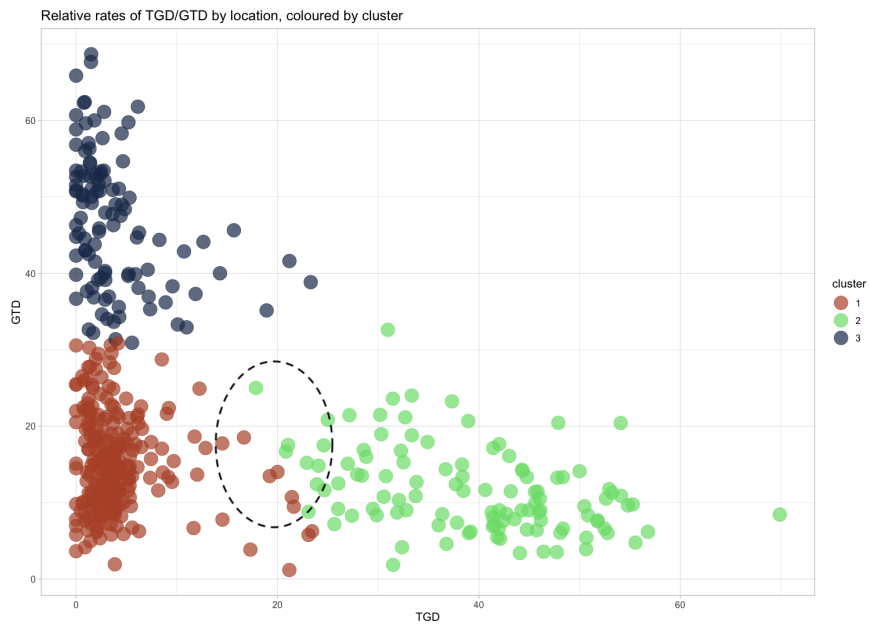
Figure 4.13: Scatterplot showing the distribution of clusters for pronominal ditransitives

Having established that three clusters offer a good fit for the data set, we can now map each location, colour-coded by cluster. The resulting map is shown in Figure 4.14.

We can now immediately see the clear geographical separation between
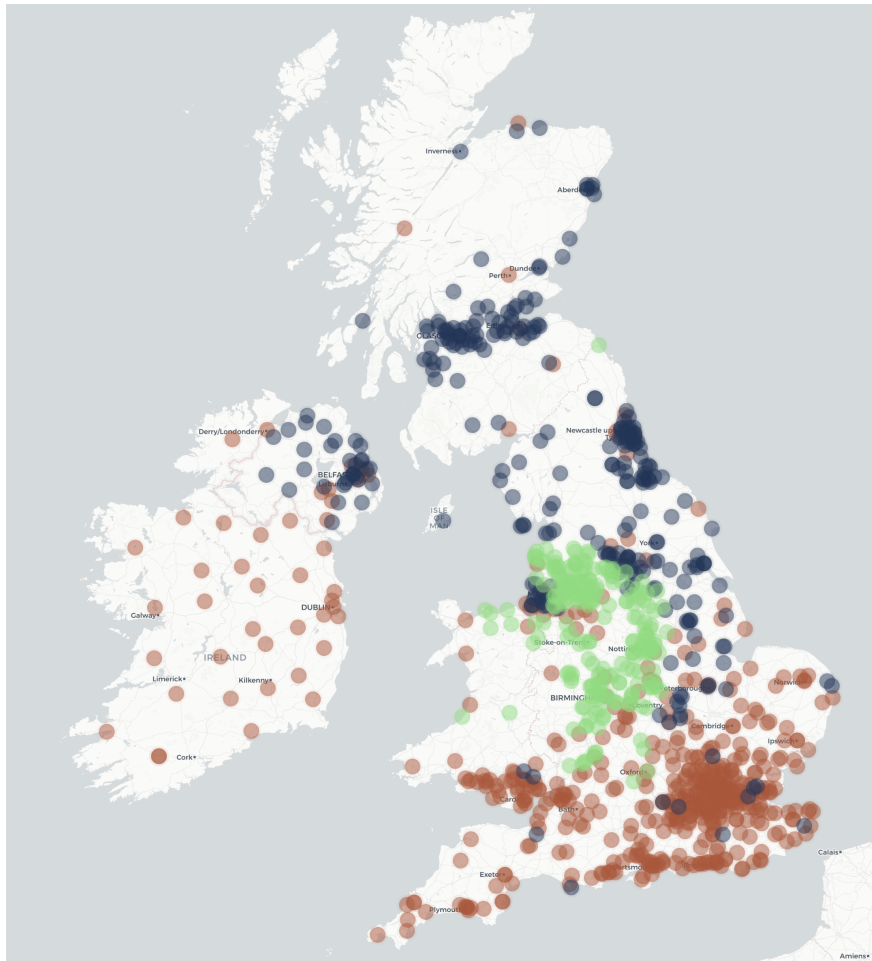
Figure 4.14: Pronominal ditransitives cluster map (K-means). Three clusters.
Interactive map available at: `bit.ly/3wYMWog`

the three variants. ₚGTD (_me it) is *dominant* in the East, Cumbria, North East, Scotland and Northern Ireland, with a pocket in the Liverpool area. ₚTGD (_it me) is *dominant* in the Midlands, South Yorkshire and North West. ₚPDAT (_it to me) is dominant across the south of England and the Republic of Ireland.

If we overlay the SED map (presented earlier in Figure 2.5 on page 64), we can see a close correspondence between the ₚTGD and ₚGTD areas on the two maps, with some exceptions. There is one notable difference in the distribution of ₚTGD: the SED shows a finger extending across the north of London, which is not present in the Twitter map. Additionally, the ₚGTD area extends further south in the SED, covering the areas around London.

The Twitter map shows the whole of this area as ₚPDAT-dominant. It is
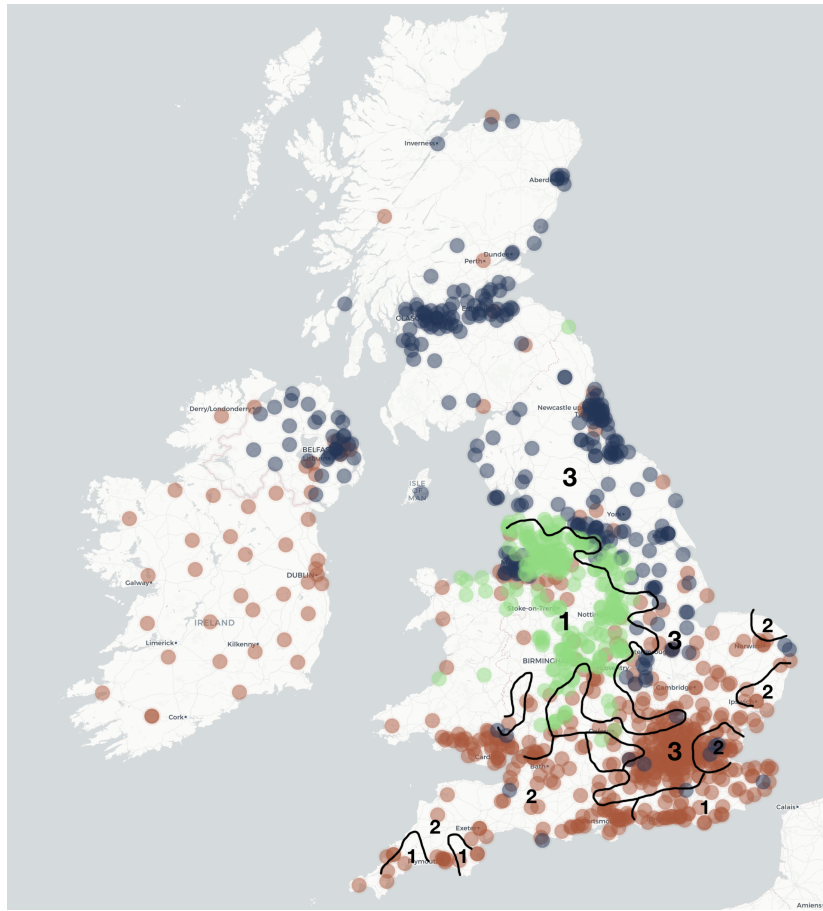
Figure 4.15: Pronominal ditransitives cluster map, overlaid onto SED map.
Reproduced from Kirk et al. (2014).
Confusingly, the numbers here **do not** correspond to the cluster numbers. 1 = TGD, 2 = PDAT, 3 = GTD

tempting to say that this reflects the increasing spread of ₚPDAT in the South. This may account for some of what we see, however, we should also bear in mind that the SED data were drawn exclusively from rural locations, while the Twitter cluster data cover both urban and rural locations and show places that are dominant in a given variant. We saw in the single variant map for ₚGTD (Figure 4.9), that ₚGTD is still present in these areas, but not dominant. The differences between the two maps should therefore be viewed with these distinctions in mind.

There is a significant retreat in the TGD area compared to Kirk et al.'s (2014) SED map. This is particularly apparent in the updated map provided by Britain et al. (2018), shown again in Figure 4.16.

156
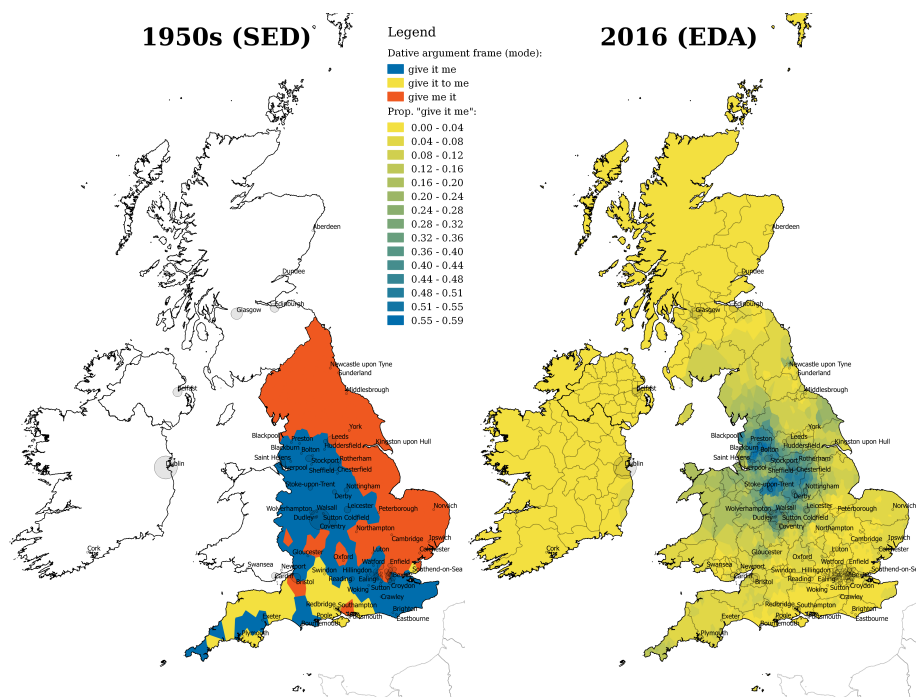
Figure 4.16: Pronominal ditransitives in the Survey of English Dialects
From Britain et al. (2018, p.88)

Here we see that the Twitter distribution closely matches that found in the English Dialects App (EDA). The SED maps show $_p$TGD spreading through to the South East, including Kent while both Twitter and EDA maps show a more restricted area. This fits with Yáñez-Bouza and Denison's (2015) finding that $_p$TGD use has declined relative to $_p$PDAT, at least in the South East.

157

4.4.2 The broader distribution of ditransitives

In this section, I provide some brief discussion regarding the overall distribution patterns for the UK and Ireland. Given the amount of data available, there is substantial room for further investigation that is beyond the scope of the current project.

The geohistorical distribution of GTD

The map shown in Figure 4.17 was generated in R by mapping UK place names that end in suffixes known to be of Scandinavian origin. Doing this provides an indication of the extent of Norse settlement to which we can compare to distribution of pGTD.
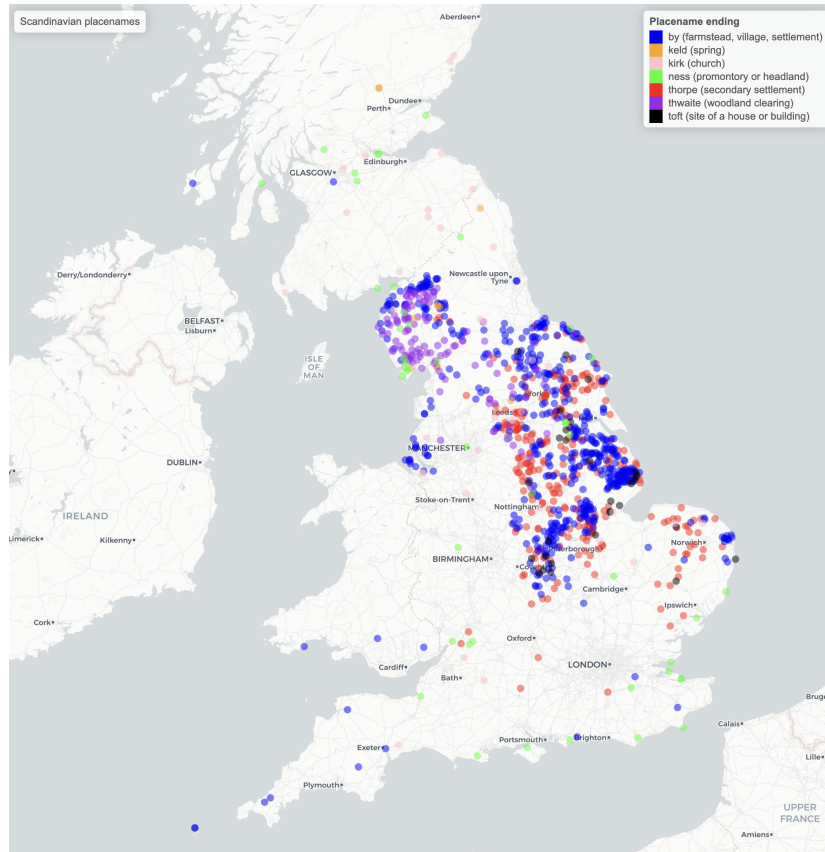
Figure 4.17: Map of place names with Scandinavian suffixes, generated using R script.

If we compare the pGTD-dominant cluster (cluster 3: _me it_) to the map of place names with Scandinavian suffixes (PNS), presented in Figure 4.17, we can see a striking correspondence in England. This follows Gast's (2007) proposal

that the pronominal form of the GTD may have emerged as a result of contact with Old Norse. Gast (2007) based his inference on the distribution of $_p$GTD as found in the SED, which did show an apparent overlap with the known extent of the Danelaw. Gast (2007) additionally reviews the fact that $_p$GTD was the word order in Old Norse, as well as in present-day Norwegian and Icelandic. Whilst the claim is speculative (as Gast acknowledges), it is nevertheless worth considering the idea in the light of the new data provided by pDit-corp.

This comparison is presented in the map shown in Figure 4.18, which shows the places tagged as belonging to the GTD dominant cluster in blue, overlaid onto places which have Scandinavian suffixes, in light grey.
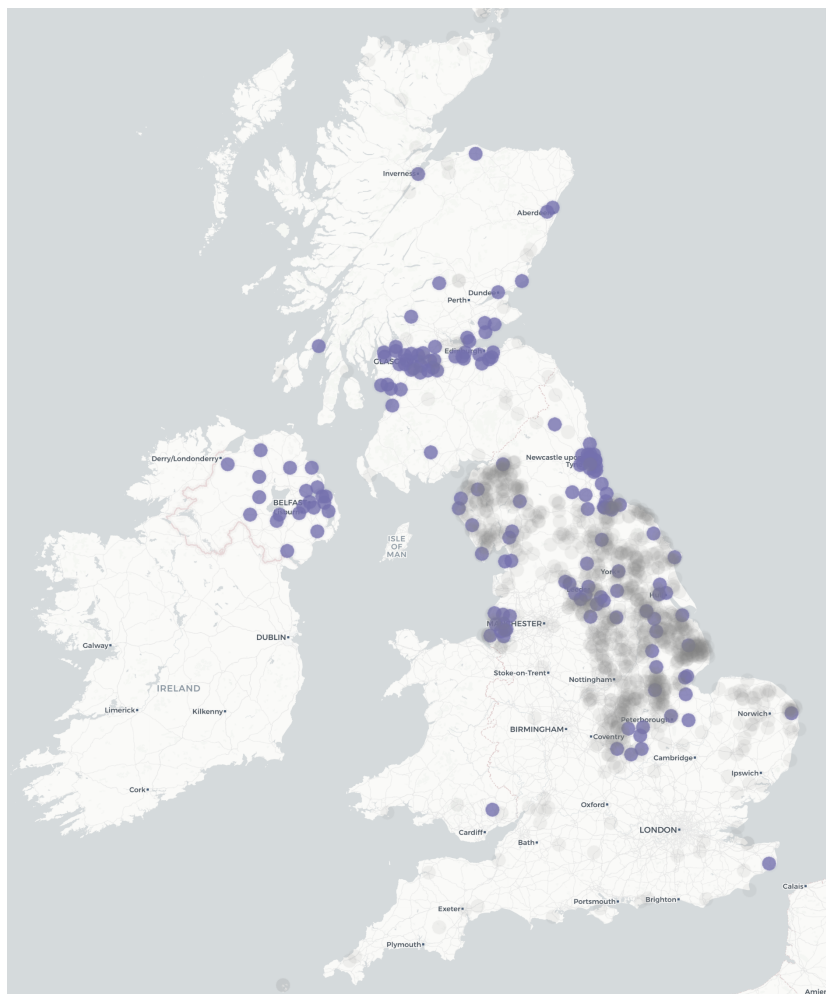
Figure 4.18: GTD dominant cluster (blue), overlaid onto the map of place names with Scandinavian suffixes, presented in Figure 4.17, here shown as light grey.

The degree to which $_p$GTD dominant clusters align the PNS area in England is quite striking. There is, in fact, a closer correspondence with the Twitter data than with the SED data. In particular, the southern edge of the $_p$GTD-dominant area aligns more closely with the PNS area than the $_p$GTD are on the SED. Focusing on $_p$GTD dominance, relative to the other two variants, allows us to see where it is most deep-rooted.

The areas where there is low correspondence between $_p$GTD-dominant locations and the PNS area; the North East, Scotland and Northern Ireland, may be explained by the fact that English did not make its way to these areas until several centuries later. Old Northumbrian English, which is thought to have led to the development of Scots, would have seen substantial Norse contact prior to migrating north to Scotland. As Johnston (1997, p.52) remarks: "it is likely that the modern varieties [of Scots] descend from a radically restructured, Norse-influenced Northumbrian going back to the Danelaw proper".

The *Dictionary of the Scots Language* (DSL) describes a similar narrative:

> the native tongue of [...] immigrants [to south-eastern Scotland] of lesser rank was a variety of Northern English heavily influenced in pronunciation, vocabulary and grammar by the Old Norse language brought to northern and midland England by Viking-era invaders and settlers. (Source: `https://dsl.ac.uk/about-scots/an-outline-history-of-scots/`)

The data and narrative presented here lend support to Gast's (2007) proposal, however, further research is needed to verify the link.

The distribution of $_p$PDAT

De Cuypere (2014) reports that the to-dative was already in common use in Old English with verbs involving *communication* (*speak, say, utter*) and verbs involving *motion* towards a directional GOAL (*send, take, bring*). In OE, it was specifically with verbs involving *transfer of possession* to a human recipient, where the preposition was not used.

Interestingly, what I have been referring to as $_{DPDP}$TGD, that is the preposition-less THEME-GOAL order with verbs involving *transfer of possession,* as in (2), was perfectly possible in Old English.

(2)   he gave the letter the bank (OK in OE, *in ModE)

De Cuypere (2014) is in agreement here with Allen (2006, p.214), who makes the claim that it was the rise in the use of the preposition with the THEME-GOAL order that reduced the requirement for the preposition-less variant.

It may be that the expansion of the *to-dative* (PDAT) to verbs involving *transfer of possession* did occur, at least in part due to accommodation with the Norman French pattern (Tagliamonte, 2014), given the predominance of the variant in the south. However, there is an additional, internally driven motivation to adopt the preposition, following the loss of case, and a resultant need to disambiguate the intended THEME or GOAL. A proper analysis of this line of inquiry would involve a full treatment of historical texts, which is unfortunately beyond the scope of the current project. For now, I again leave this potentially fruitful line of inquiry to future work.

Yorkshire and the East Midlands

In the single-variant maps, we saw a clear boundary running North to South, between Doncaster←→Sheffield and Leeds←→Halifax/Huddersfield. Here, in the cluster map, the boundary is made more explicit.
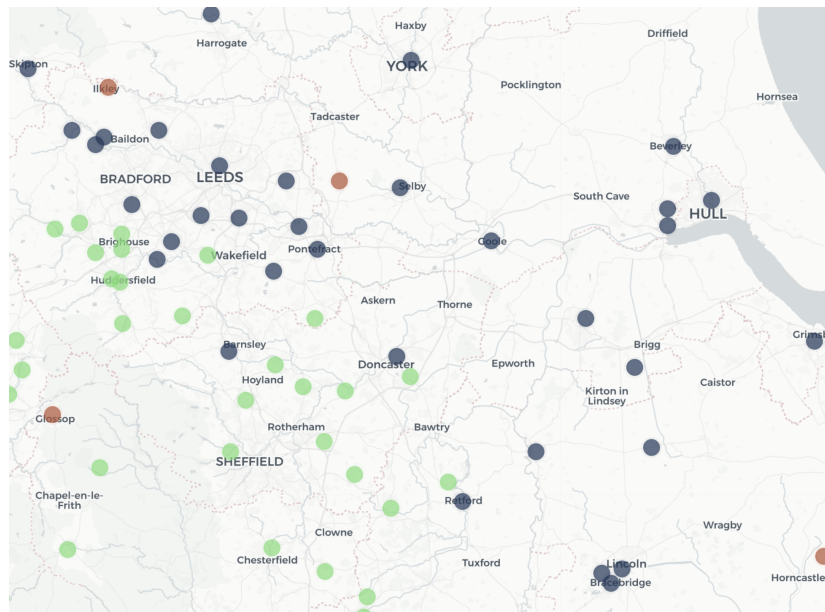
Figure 4.19: Detail of Yorkshire and the East Midlands.

These are locations that are no more than 3 or 4 miles apart. What is underlying this boundary is unclear, though from the discussion on the possible

connection between ₚGTD and Old Norse, it could be that it is longstanding. Of course, there are myriad sociolinguistic factors to consider here too. I present this boundary as a fruitful site for future research, and more broadly to demonstrate the potential use of Twitter dialect maps to identify such places.

4.4.3 Scotland, the North East and Ireland

As previously noted, there appears to be a continuity in the distribution of ₚGTD across the North East, Scotland and Northern Ireland. I discussed the possible deep-historical connections here, in relation to Gast (2007), and the possible source of ₚGTD. Here I look more closely at the connection.
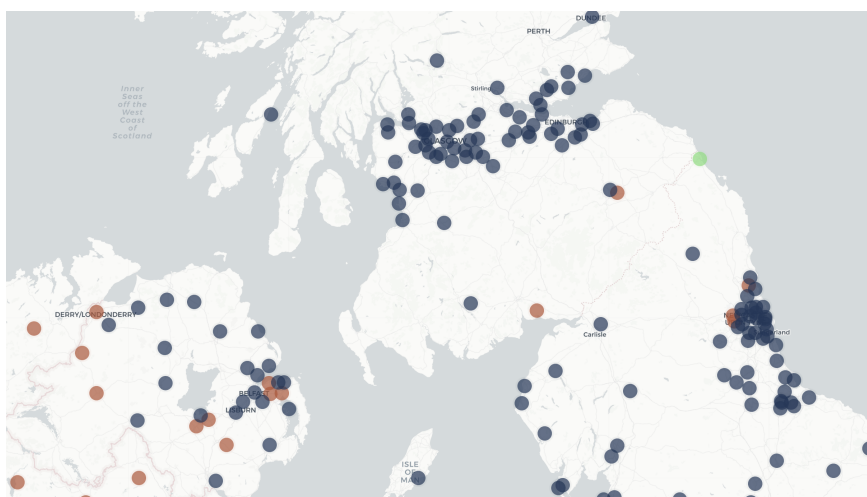
Figure 4.20: Detail of North East England, the Central Belt of Scotland, and Northern Ireland.

Figure 4.20 provides detail from the cluster map for [DATALT]. The presence of ₚGTD in Northern Ireland is no accident, and can be attributed to the settlement of Scots in the 17th century. This settlement pattern is shown in Figure 4.21, from Robinson (1997), which shows the relative distributions of English and Scottish settlement in Ulster c.1630.

As mentioned, the continuity in the distribution of ₚGTD between Scotland and Northern Ireland sets the date for high ₚGTD use in Scotland back to at least the 17th century. As we saw, such high rates of ₚGTD use in this period do register in Yáñez-Bouza and Denison's (2015) plot of the relative rates of pronominal ditransitive variants over the past several centuries (see Figure 2.4).

Looking more closely at Ireland, we see a dramatic distinction between
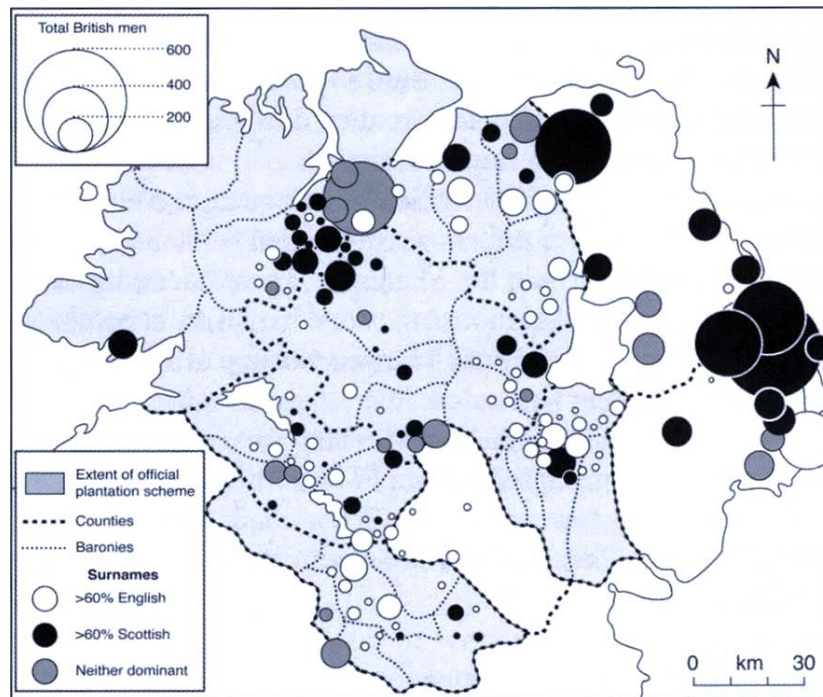
Figure 4.21: English and Scottish Settlement in Ulster (based on numbers and surnames recorded in the muster rolls), c.1630
Source: `https://www.ulsterscotsacademy.com/research/gregg/mapping-ulster-scots.php`
(Robinson, 1997)

Northern Ireland and the Republic of Ireland. The patterns here mirror Strelluf's (2020) findings for NEED + *past*, which showed a similarly stark distinction between the two places.

Whilst Northern Ireland patterns with Scotland, with $_\text{p}$GTD as the dominant variant, the Republic is $_\text{p}$PDAT dominant, in common with the south of England. Further, referring again to Robinson's map in Figure 4.21, we see that the places with high $_\text{p}$GTD rates within Northern Ireland, correspond quite neatly to those marked as predominantly Scottish, rather than English settled areas.
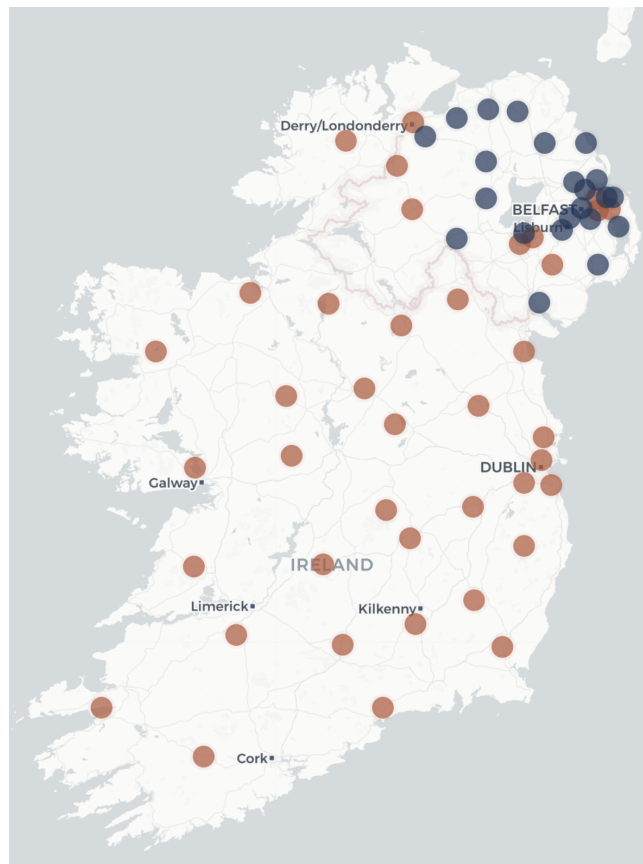
Figure 4.22: Detail of the pronominal ditransitive cluster map, showing the island of Ireland.

The North West

Focusing on the North West, we can identify a clear boundary between Liverpool and Manchester. The substantial variation evident here is one of the reasons that it is chosen for further investigation in the current study, discussed in the coming pages and the following chapter.

Liverpool is predominantly $_\text{p}$GTD (blue), and Manchester predominantly $_\text{p}$TGD (green).[7] The fact that $_\text{p}$TGD is not dominant in Liverpool is interesting given that it is found to be well accepted there in Biggs (2018).

It is notable that there are a number of red points in and around Liverpool, and south of Manchester. These points correspond to $_\text{p}$PDAT-dominance, which puts them in the same cluster as the South. However, the red points in the North West are not identical, in terms of the relative rates of variants, to those in the South: North Western points (apart from around Liverpool) contain higher rates of $_\text{p}$TGD, as might be expected, given the predominance of $_\text{p}$TGD in the region.
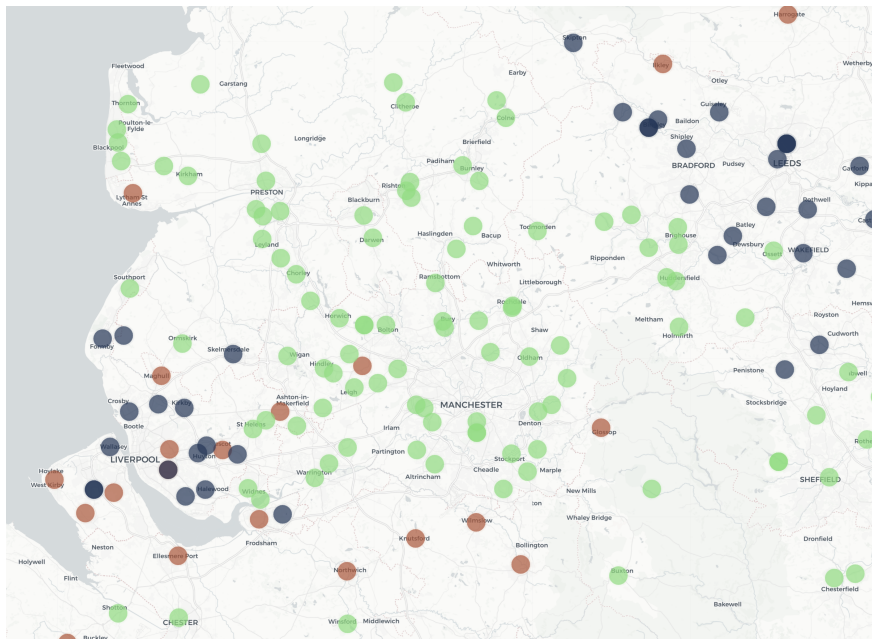
Figure 4.23: Pronominal ditransitives cluster map. Three clusters, detail North West.

Red points around Liverpool, however, show higher levels of $_\text{p}$GTD. These

---

[7]Here I am showing only the maps based on geo.place_id data. The place of origin data are used for a closer analysis of structures that were not picked up in the search strings for the geo.place_id data.

effects are the result of the clustering algorithm classifying based on similarly-high rates of ₚPDAT. Increasing the number of clusters, or employing a different clustering technique, such as the Ward Method, may tease out these differences.[8]

Interestingly, a preference for ₚGTD and ₚPDAT mirrors that found in Ireland, as we will see, which may follow the migration connections between Ireland and the North West in the 19th century, as discussed in §1.3.1.

The fact that ₚGTD is so localised to Liverpool and its immediate environs suggests that it did not survive transplantation to the wider region during the period of de-urbanisation in the second half of the 20th century. Instead, it remains peculiar to Liverpool in the North West.

A closer examination of the distribution of [DATALT] in the North West, and how it compares to [P/D-DROP] is presented in the coming pages.

4.4.4 The status of individual verbs

This section addresses:

RQ (4) What quantificational evidence can be brought to bear on the TGD-as-GTD question?

Theoretical syntax often discounts the relevance of structure frequency for good reason. Just because a structure is not frequent in a given corpus does not entail that its acceptability is necessarily degraded. This point is well-illustrated by Adger (2019, p.1) who makes the point that for any sentence beyond a few words, its likelihood of appearing in any corpus diminishes to near-zero.

However, where we are looking at very short phrases (e.g. *explain it them*), and the data are particularly substantial, it is possible to make some inferences in situations where we would expect the availability of a given string to be frequent, but find that we do not. I present such data in the following sections.

Verbs which favour PDAT

This is particularly apparent with many verbs of Latin origin. Out of 3059 examples of pronominal ditransitives with the verb *explain*, there are 22 ₚTGDs

---

[8]I experimented with different cluster levels and the Ward Method. Indeed, doing so did result in more subtle distinctions. However, with more clusters, it became increasingly difficult to visualise the data. Due to space limitations, and in the interest of maintaining relative simplicity, I decided to stick with K-means, and three clusters.

and just 1 $_p$GTD; the rest are $_p$PDAT.

(3) $_p$TGD-*explain* (n = 22)

    a.   although <u>my mother did explain it me</u> when I asked her

    b.   ppl cannot rest until <u>somebody explains it them</u>, no matter how unreasonable

    c.   Why is it called a fry up? Why not be called fry down? <u>Explain it me</u> thanks

    d.   What are you talking about, <u>I have explained it you</u>, don't slander.

(4) $_p$GTD-*explain* (n = 1)

    a.   well ill get maggie to <u>explain me it</u> tomorrow

This is expected, following Levin's (1993) categorisation of ditransitive types (latinate verbs are considered only viable with PDAT) and supports the diagnostic used by Haddican (2010) to assess whether $_p$TGD is underlyingly more akin to $_p$GTD or to $_p$PDAT, with an elided preposition: if $_p$TGD is $_p$PDAT with an elided preposition, we expect $_p$TGD to be permissible for verbs which disfavour $_p$GTD.

The dominance of PDAT with *explain* can be seen in Figure 4.24.

Whilst a lack of attestation in a corpus does not necessarily entail that a structure is unacceptable to speakers, if $_p$TGD were $_p$PDAT with an elided preposition, we might expect to find a greater number of examples for a verb like *explain* which strongly favours $_p$PDAT. It is nevertheless informative that there are *some* instances of $_p$TGD and just one instance of $_p$GTD (even in $_p$GTD dominant regions), suggesting that:

1. $_p$GTD-*explain* is degraded for most speakers (or ungrammatical)

2. $_p$TGD-*explain*, where it does occur, is (produced by speakers for whom it is) underlyingly $_p$PDAT, and they are dropping the preposition

3. Or, for speakers who use $_p$TGD with other verbs, they apply the same rule to latinate verbs by analogy

This is also consistent with Haddican's (2010) results in Manchester, which showed a significant degree of inter-speaker variation: for some survey respondents, $_p$TGD did appear to involve a dropped preposition.

We additionally find the same pattern with the verb *mention*.
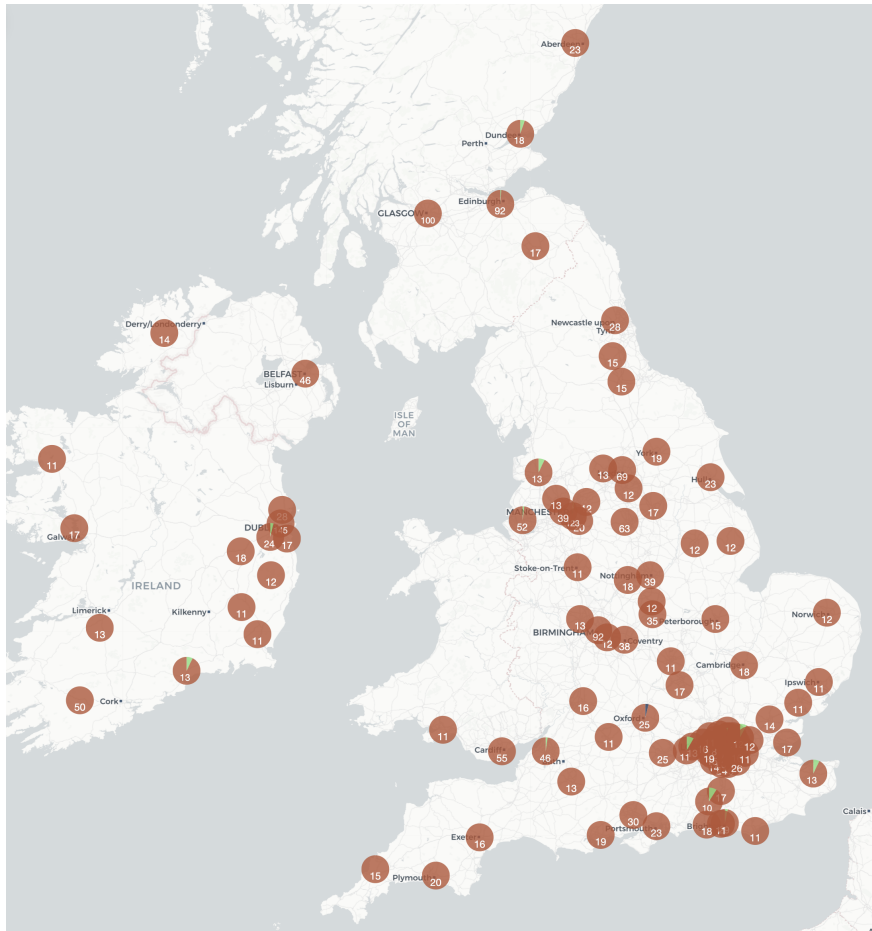
Figure 4.24: Map showing near categorical dominance of PDAT with *explain*

ₚPDAT-preference is not exclusive to latinate verbs: ditransitive *post* follows the same trend as *mention* and *explain*, though ₚGTD is also found. We see categorical dominance with ₚPDAT in most places, though we see a few cases of ₚTGD in Sheffield and Manchester.

donate and *deliver* also show categorical preference for ₚPDAT, though there are too few results overall to map geographically.

As Table 4.5 shows, of the verbs which are most frequent with ₚPDAT, the top seven are latinate:[9]

The fact that latinate verbs on Twitter follow the pattern predicted by Levin (1993), favouring ₚPDAT, is reassuring. If Twitter data were biased towards shortening, the use of *netspeak*, then we might expect this to apply across the

---

[9]See appendix for full list.

168
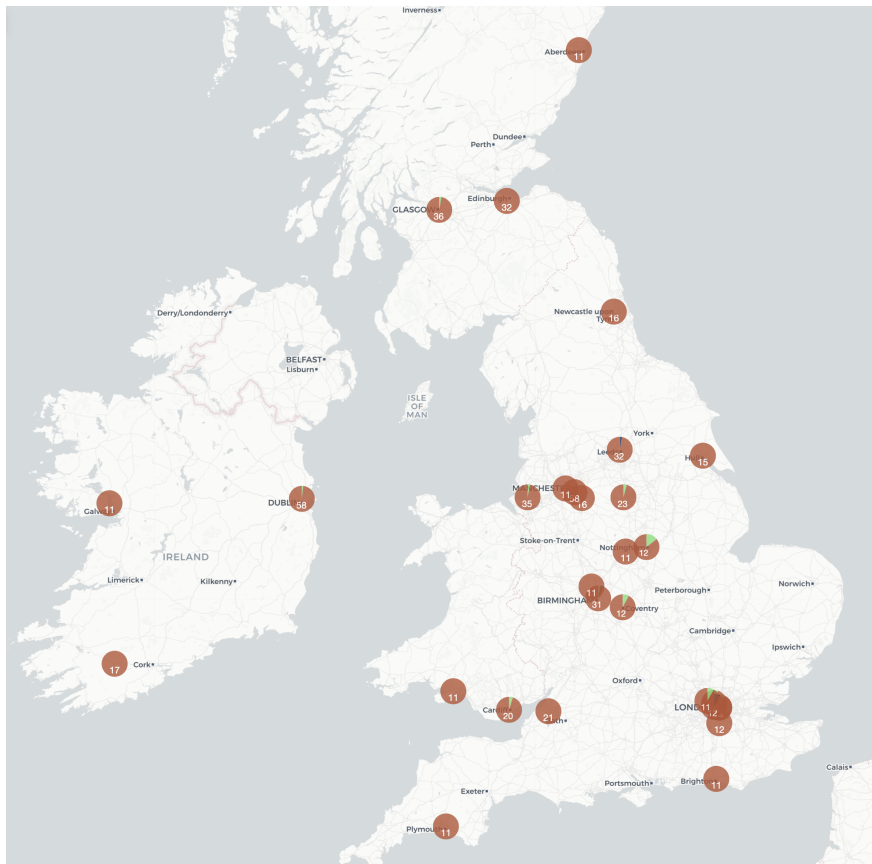
Figure 4.25: Map showing near categorical dominance of PDAT with *mention*

Verb	ₚTGD	ₚGTD	ₚPDAT	**Total**
explain	0.6	0.1	99.3	6102
introduce	0.0	0.8	99.2	491
deliver	0.8	0.6	98.6	1154
address	1.5	0.0	98.5	131
describe	1.1	0.5	98.4	440
donate	1.0	0.7	98.3	299
mention	1.4	0.3	98.3	2512

Table 4.5: Relative frequency of [DATALT] variants, by verb, ordered by ₚPDAT
Top seven are all latinate verbs.

board.

Verbs which favour TGD/GTD

Traditionally, verbs such as *refuse* or *deny* are found to favour ₚGTD/ₚTGD over ₚPDAT, and as such are useful as a diagnostic in determining whether TGDs are underlyingly GTDs (Haddican, 2010; Haddican & Holmberg, 2012). However,
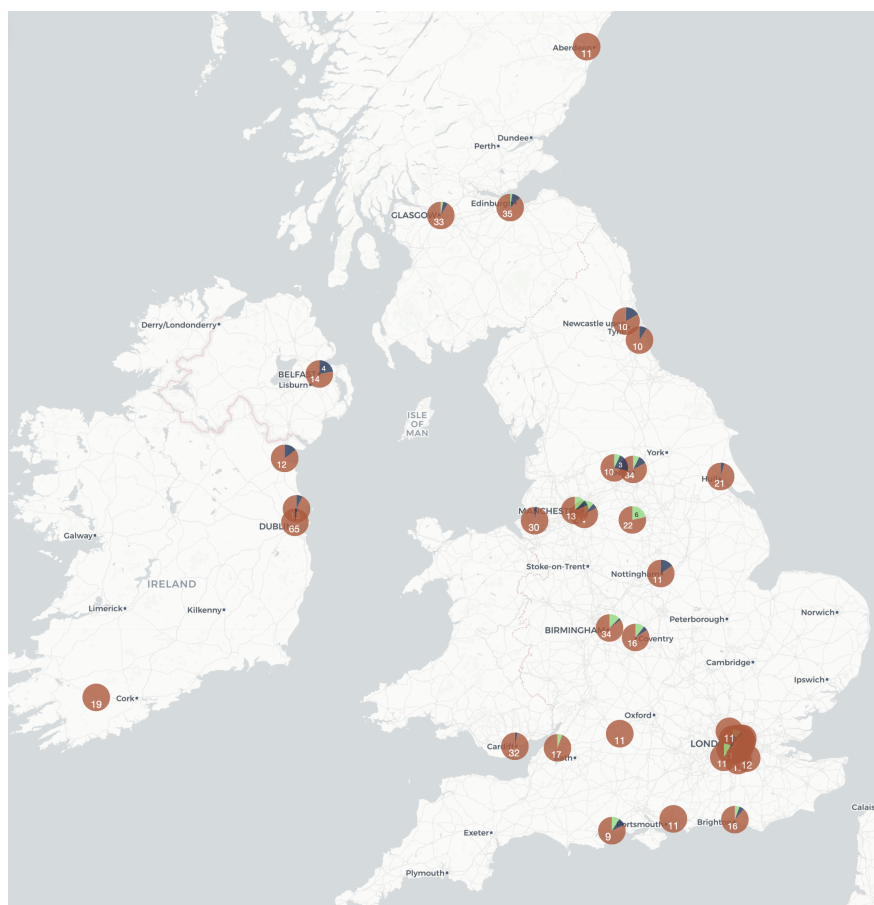
Figure 4.26: Map showing near categorical dominance of ₚPDAT with *post*

whilst *refuse* type verbs are less degraded with DOCs than PDAT, they are still found to be only marginally accepted. In pDit-corp, there were found to be infrequent.

The verb *text*,[10] with the ditransitive meaning to *text something to someone* shows competition between ₚGTD and ₚPDAT in the South and near categorical dominance of ₚGTD in the North East and Scotland. In the Midlands and North, the picture is mixed: a substantial number of locations appear to show a two-way competition between ₚTGD and ₚGTD, while some, particularly in the Midlands, show competition between ₚPDAT and ₚTGD. It is difficult draw any firm conclusions here, particularly as overall counts are low in some places. However, it would appear that there is some degree of distributional alignment between ₚTGD and ₚGTD.

---

[10]In the corpus, *text* includes the respelling *txt*.

170

It is noteworthy here that Liverpool, as expected, is $_p$GTD-dominant.
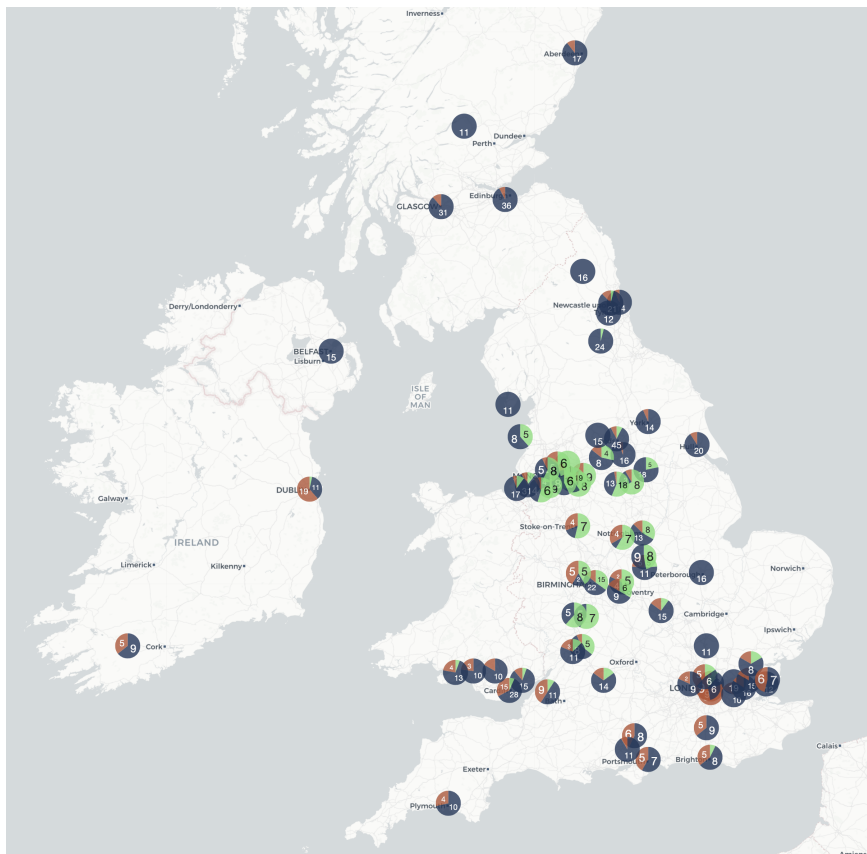
Figure 4.27: Map showing near bivariate competition between $_p$GTD and $_p$TGD in parts of the Midlands and North West with *text*

In Table 4.6, verbs are again ordered by the relative rate of $_p$PDAT, from highest to lowest, for verbs where $_p$PDAT constitutes less than 70% of the variation. This allows us to view the nature of the competition between $_p$GTD and $_p$TGD. As $_p$PDAT frequency declines, we do not generally see a corresponding reduction in $_p$TGD, as might be expected if it were underlyingly $_p$PDAT. There are, however, exceptions to this rule: *ask* is a case in point, uncommon with $_p$PDAT, save for particular contexts (5), and $_p$TGD is also substantially degraded.

(5)   a.   A <u>question</u> I will never answer because, firstly, I didn't do it, and secondly I was <u>asking it to you</u>.

        b.   So I don't know how to answer that <u>question</u>...or why you're <u>asking it to me</u>.

        c.   Hi, we want Jaden's fans to tweet <u>questions</u> for him - you could get

the chance <u>to ask it to him</u> on Skype.

It is surprising that there are any instances of pPDAT-*ask*; it seems quite ungrammatical, but there are apparent uses, usually where the pronoun *it* is directly anaphoric to *question*.

Verb	pTGD	pGTD	pPDAT	**Total**
loan	6.1	24.9	69.1	181
give	13.8	20.4	65.8	59867
feed	9.4	27.4	63.3	1037
send	13.6	23.5	62.9	37764
mail	8.1	30.6	61.4	991
offer	15.4	25.2	59.4	1333
lend	20.1	32.7	47.1	2420
deny	14.0	43.4	42.7	143
pm	8.5	55.5	36.0	317
show	10.2	56.4	33.4	6905
promise	17.8	51.7	30.5	118
message	12.9	59.3	27.8	1669
pay	25.4	49.8	24.9	209
text	15.5	67.1	17.4	3661
ask	6.4	84.0	9.6	156

Table 4.6: Relative Frequencies of Syntactic Variants for Verbs, ordered by pPDAT, where pPDAT < 70%

If pTGD were straightforwardly pPDAT with a dropped preposition, we might expect to find cases with low pGTD, but higher pTGD, but this is not generally what we find.

Classifying verbs

Favouring pPDAT or pTGD: *credit* and *leave*

Bring behaves a little differently. In most places pPDAT is preferred, yet it is quite common with pTGD.

Credit, as a ditransitive, has two senses: *appraisal* (giving someone credit of ownership/action) and *transfer of possession* (sending money, for example). *Credit* bucks the trend that we have seen above, favouring both pPDAT, and pTGD, with pGTD is *almost* unattested with the sense of *appraisal*).

The vast majority of examples in pDit-corp were of the *appraisal* type.

(6)   *appraisal*-PDAT

  a. he just read my tweet out and <u>credited it to you</u>!

172
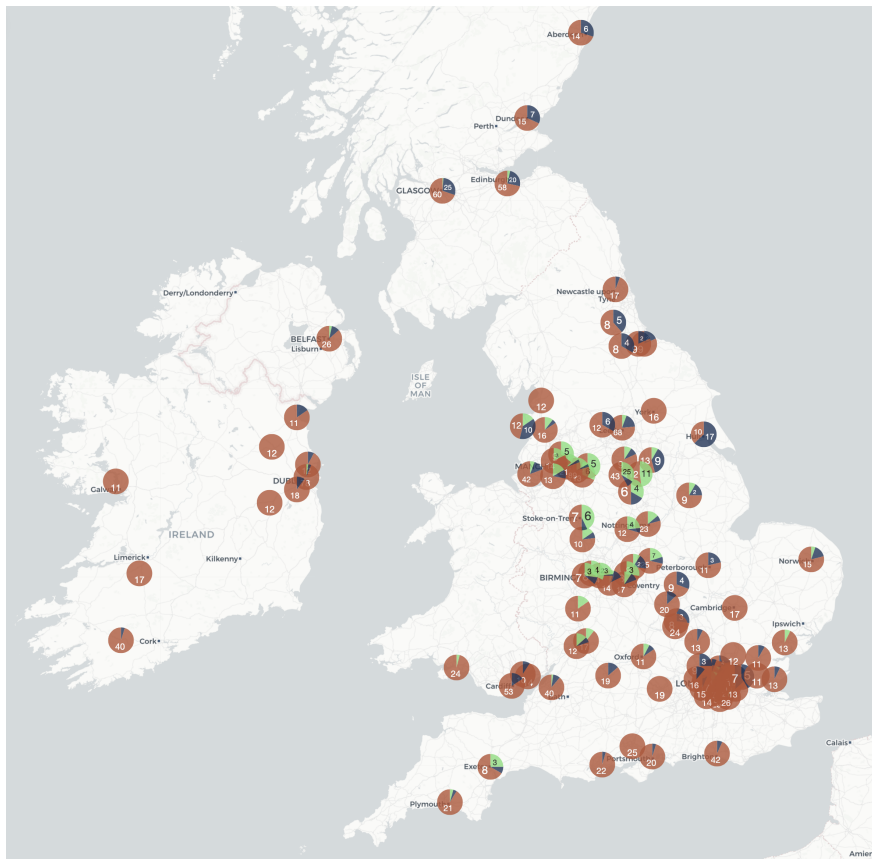
Figure 4.28: *bring* is categorically ₚPDAT in much of the south. In the Midlands and North West, 25% ₚTGD, in the North East and Scotland, 25% ₚGTD.

    b.   I wish I could remember which comedian made that joke so <u>I could credit it to them</u>

(7)   *appraisal*-TGD

    a.   of course you can use it, please <u>credit it me</u> though

    b.   We apologise for not crediting your video if you have copyright. We can take it down or <u>credit it you</u>?

There is one example of ₚGTD-*appraisal*, though, it seems that for most speakers, use in this sense would be ungrammatical.

(8)   *appraisal*-GTD

    a.   Cheryl stole that girls photo but <u>credited her it</u>

(9)   *transfer*-GTD

a. Rang them this morning, they <u>credited me it back</u>. Reinvoiced me £47 +

This, of course would make a good diagnostic on the status of _pTGD. It seems that, for ditransitives which are almost exclusively _pPDAT, there is some room for speakers to drop the preposition and maintain grammaticality, in areas or speech communities in which TGD is otherwise licit. Notably, the availability of P-DROP/PD-DROP with verbs of motion does not seem to have a bearing on the availability of _pTGD with verbs that favour PDAT.

Leave

Ditransitive *leave* also has several readings. It may indicate a *transfer of possession* (as in leaving a will to/for someone), where it has a benefactive reading. Here, *leave*, may occur as TGD or GTD and intersects with _pPDAT, where the preposition is *for*, as in *leave it for me = leave me it*.

(10)   *Transfer of possession*

    a.   Surely you must <u>leave it them</u> and just put the 🎄 wrapping paper around it all?! 😆

    b.   I had a Vespa, Grandad <u>left it me when he died</u>..

    c.   <u>left it me</u> when he passed away last December

    d.   it was lovely of previous owners to <u>leave us it</u>

    e.   Dad made me a coffee and <u>left me it</u> in a flask

However, *leave* may also have an idiomatic reading as *deferral of responsibility*. In this case, it is only standardly available with the preposition *to*. It is ungrammatical as _pGTD:

(11) **I will leave them it as professionals

However, it is attested as _pTGD, suggesting a P-DROP analysis in these cases.

(12)   *Deferral of responsibility* Predominantly with *to*, likely TGD-P-DROP

    a.   I said earlier, I hope universal run it and <u>leave it them as professionals,</u> and the club have nowt to do with it

    b.   You old men need to accept ur age and your youth days are gone. <u>Leave it us young generation</u>

174

    c.   Can I <u>leave it you & Ian Hislop to decide</u> who is the really big banana around here?

    d.   I'll leave it you to tell Miss P ;)

Additionally, *leave* has a stative reading, where it is only standardly possible with prepositions *at* or *in*.

(13)   *I left her room it

But again, it is possible as $_p$TGD:

(14)   *Stative*

    a.   ye if it can't go in is car,will <u>leave it her mams</u> on the green (AT)

    b.   rememering when i made jelly with georgia and we <u>left it her room</u> overnight (IN)

These examples are similar to those with the verb *stay*, reported by Biggs (2018) as licit in Liverpool:

(15)   I stayed John's house last night.

As we will see, examples with *stay* are accordingly attested in NWatlas-corp (example (85), p.243), and accepted in the survey (reported in the next chapter).

Summary

We have seen here that, broadly, frequency distributions by verb support the notion that $_p$TGDs are underlyingly akin to $_p$GTDs. However, there are some data which suggest that the opposite is true for some speakers. Some readings of *leave*, as we have seen, contradict the trend. *Explain* is absent with $_p$GTD and almost categorically $_p$PDAT, but is attested, albeit infrequently with $_p$TGD. We find the inverse pattern with *ask*, which is disallowed with $_p$PDAT, save for particular discourse contexts. Here, $_p$GTD is overwhelmingly the dominant choice, and we do not see an equivalent increase in $_p$TGD.

Overall, this is in line with Haddican's (2010) findings that for most speakers (amongst his Manchester survey respondents), $_p$TGD behaved 'like a true double object construction', but there were some speakers for whom the opposite pattern was observed.

Comparing the relative frequency of [DATALT] verbs by variant

As discussed, if $_p$TGD and $_p$GTD are underlyingly DOC, then we would expect to find that, as the rate of $_p$PDAT decreases, by verb, the rate of both $_p$TGD and $_p$GTD would increase at the same rate. If $_p$TGD were underlyingly a $_p$PDAT with a dropped preposition, we would expect to see higher rates of $_p$TGD rates when $_p$PDAT is high.

Figure 4.32 plots the relative rates of [DATALT] variants by verb, ordered by $_p$PDAT.
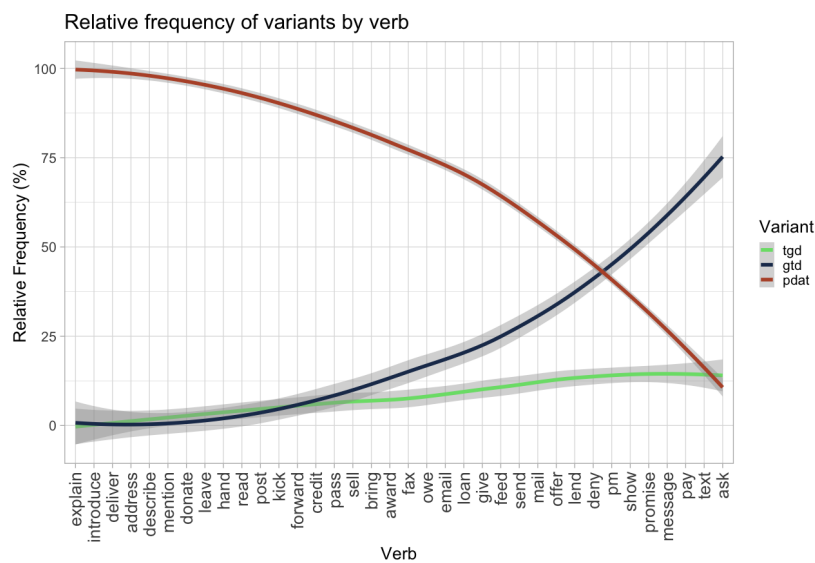
Figure 4.29: Relative rates of [DATALT] variants by verb, ordered by $_p$PDAT. Verbs limited to minimum 50 total hits. See Appendix for full list of verbs.

On the left of the plot, we find verbs which are categorically $_p$PDAT (such as *explain*), on the right, we find verbs for which $_p$PDAT is lower than 10% (such as *ask*).

We can see that the rate of $_p$GTD and $_p$TGD do, indeed, increase together as $_p$PDAT decreases. $_p$GTD rates are higher overall, which is expected, for the UK and Ireland as a whole, as $_p$GTD users constitute a higher population. We also do not see verbs where there is a high rate of $_p$TGD, but low $_p$GTD, which we might expect, if $_p$TGD were underlyingly $_p$PDAT. This shows the general trend, however, and does not show individual variability: it is likely still the case that, for a minority of speakers, $_p$TGD *does* involve P-DROP, particularly with certain verbs, discussed above.

176

Relative frequency by cluster

As we have seen, there is considerable regional variation in the relative rates of [DATALT] variants. The overall rates for verbs, presented in Figure 4.32, show higher rates of $_p$GTD. If we focus on the relative rates by regional cluster, we can better see the competition between variants.
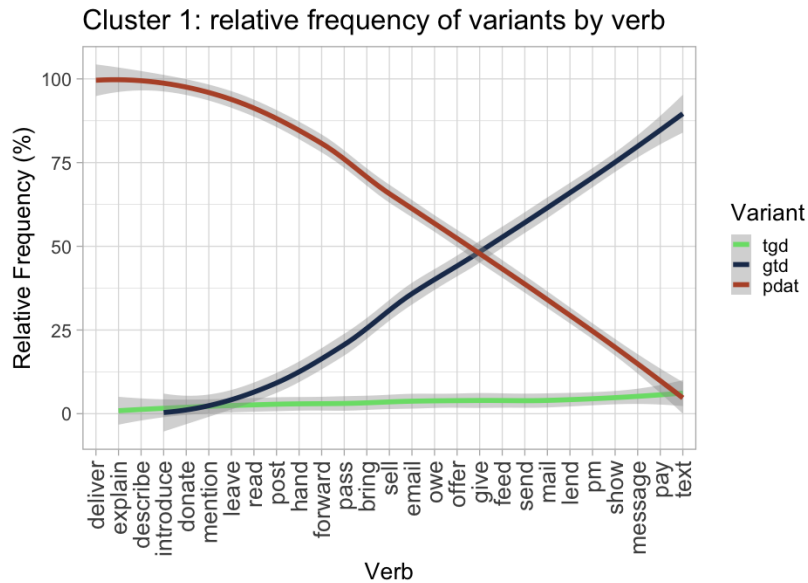
Figure 4.30: Cluster 1 ($_p$GTD-dominant): Relative rates of [DATALT] variants by verb, ordered by $_p$PDAT. Verbs limited to minimum 50 total hits.

Cluster 1, which represents locations that are $_p$GTD-dominant, shows a clean exchange between $_p$PDAT and $_p$GTD. This offers a useful benchmark for scaling which ditransitive verbs most favour the double object construction.

Cluster 2 shows $_p$TGD (_it me_) as consistently higher than $_p$GTD (_me it_), for most ditransitives, which is expected, given that cluster 2 represents places that favour $_p$TGD overall. It is again at the extreme ends of $_p$PDAT frequency that we see a closing of the gap between $_p$TGD and $_p$GTD. Where $_p$PDAT is categorical, $_p$TGD and $_p$GTD are correspondingly at zero, with both rising gradually as $_p$PDAT gives way. $_p$TGD is most prevalent for traditional alternating ditransitives such as _give_, _send_ and _offer_.

Interestingly, for novel ditransitives _message_ and _text,_ the gap is closed between $_p$TGD and $_p$GTD. This could be an indication of a waning productivity of the TGD.
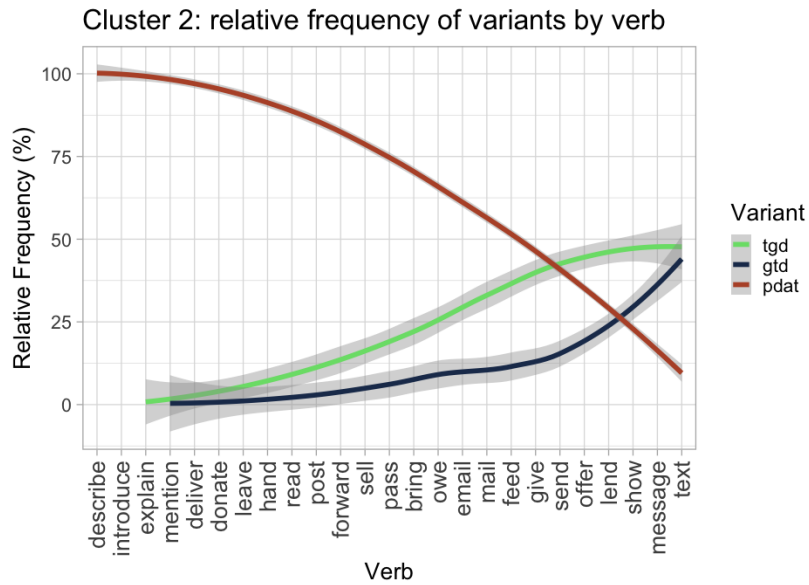
Figure 4.31: Cluster 2 ($_p$TGD-dominant): Relative rates of [DATALT] variants by verb, ordered by $_p$PDAT. Verbs limited to minimum 50 total hits.
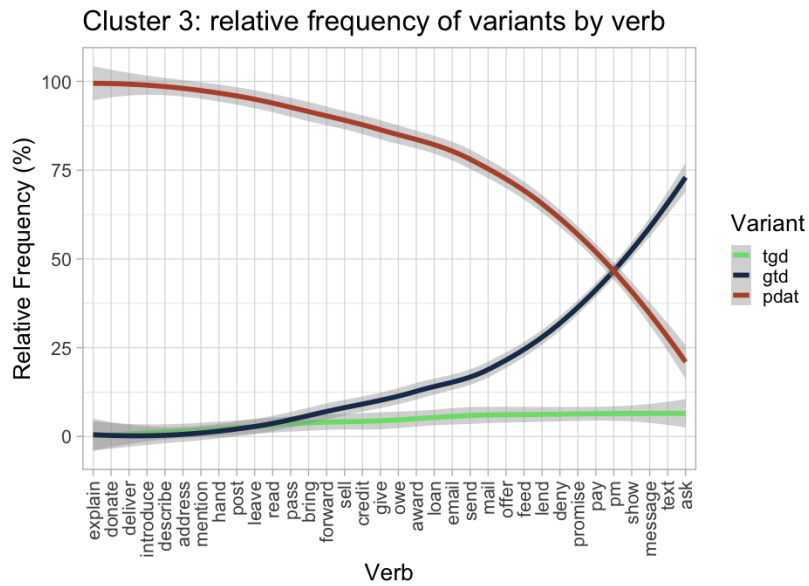
Figure 4.32: Cluster 3: ($_p$PDAT-dominant): Relative rates of [DATALT] variants by verb, ordered by $_p$PDAT. Verbs limited to minimum 50 total hits.

4.4.5   TGD with quantifier *all*

$_p$TGD (*it me*) exhibits an interesting ambiguity when followed by the quantifier *all*. In some cases, *all* scopes over the THEME, with the meaning "*give all of*

178

something to someone". Other cases, *all* scopes over the GOAL, meaning "*give something to everyone*".

Compare:

(16)   a.   Saw this & had to <u>send it you all</u> @adrianslowey @pickupmotor-spor @PorscheSauce

       b.   she <u>sent it us all</u>, not you individually.

       c.   Just get kids nowadays who think they're invincible and think <u>the world owes it them all</u>

(17)   a.   it may take 2-3 months before they <u>pay it you all back</u>

       b.   you can't expect god to just work miracles and <u>give it you all</u> on a golden plate

       c.   hope your okay honey! <u>Text it me all</u> when ur up tomorrow :) xxxxx

In the first cases, it seems that TGD is behaving as though it were a PDAT:

(18)   a.   she sent it to us all, not you individually

       b. ?*she sent us it all, not you individually

In the second cases, it seems that the TGD is behaving like a GTD, equivalent to:

(19)   I gave you it all

4.4.6   Insights from NWatlas-corp

Where pDit-corp and pDrop-corp were created by querying the Twitter API for set search strings, NWatlas-corp is unrestricted. This has the advantage that it is possible to more deeply investigate the status of structures with a greater range of parts of speech. For instance, the role of full-DP objects, the animacy of THEMEs and GOALs, alternative spellings, and with a greater range of verbs, may be investigated. This allows for a deeper understanding of underlying structure, the indication of the connectedness between P-DROP and [DATALT] for some speakers, and the most likely geographical location for particular grammars. In the following sections, I elaborate on these findings.

TGD with pronominal THEME and full DP GOAL

As we have seen the TGD is most frequent where both THEME and GOAL are pronominal. This is one reason why the set strings used to create pDit-corp focused exclusively on this structure, along with the fact that trying to predict all possible GOAL DPs was impractical.

Explain

We have seen that *explain* fits the expected pattern for latinate verbs in overwhelmingly occurring as a prepositional dative (PDAT). There is only one example with *explain* without the preposition present, it is nevertheless informative that it is possible.

(20) The commentators said it was that system so I had <u>to explain it the missus</u> then they didn't bother with it! (Liverpool)

As we have seen, however, there are a handful of cases with TGD-*explain* in locations without P-DROP, such as Sheffield, so it is not clear that P-DROP is a requirement, it may also be possible by analogy where other TGD verbs are frequent, as is likely the case in Sheffield.

Take/Bring

Interrogating NWatlas-corp, we find that cases where the THEME argument is pronominal (it) and the GOAL is a full DP (pDP) are also quite frequent.

Importantly, they are most frequent with ditransitive *take* and *bring,* where:

1. The GOAL is a location

2. The AGENT accompanies the THEME

3. The directional motion of the AGENT and THEME to the GOAL are standardly encoded by the preposition *to*

According to Levin (1993, p.135), *take* and *bring* are "considered the 'causative' counterparts of come and go". This semantic connection to verbs of motion; *come* and *go,* is important. It is here where we find the meaning expressed standardly by the preposition *to.* As with *go* and *come,* where the preposition is absent, there is a requirement for a mechanism to maintain its functional properties. For Myler (2013) this is achieved via the incorporation

of NULL TO to the verb, for Biggs (2018), it is via the availability of an innovated NULL functional element, ϰ.

The double object GOAL-THEME order in this case is not licit:

(21) a. *I took the game it
      b. *I took the game my bag
      c.   I took it (to) the game
      d.   I took my bag (to) the game

The majority of examples are in Liverpool and environs, which, as we will see, follows the expected geographical pattern for P-DROP with *come* and *go*.

(22) P-DROP with *take (n=35)/bring(n=5)*

      a.   i <u>took it the bar</u> and got a free vodka (Birkenhead)
      b.   Can't even <u>take it the game</u> (Birkenhead)
      c.   If they go to the bother of dumping on a road, why not just <u>take it the tip</u>? (Birkenhead)
      d.   So my mum bought my Uncle a birthday cake and <u>brought it the pub.</u>
      e.   yeah obvs I'm in so ill come for the iPad keyboard tomorrow and <u>bring it the pub</u> Tuesday night

P-DROP is also possible in the abstract sense conveyed by the phrasal verb *to put something to someone*, meaning to present an idea:

(23) That's where he'll be <u>putting it the big boss</u>

P-DROP is additionally not limited only the preposition *to*. There is one example for ditransitive *bring* where preposition *in* is dropped:

(24) You won the blue fabric wreath!! :-) will <u>bring it the morning!</u>

This is interesting as Biggs (2018) finds P-DROP to be available exclusively with *to* and *at.* As we will see in the main discussion on P-DROP, data show that the range is apparently greater than this.

Send

Send is often considered to involve directional motion. Where the GOAL is inanimate, *to* is standardly a requirement. In the following examples, a GOAL-THEME double object would be degraded for most speakers (**sent the papers it*) (Biggs, 2018; Haddican, 2010), likely due to the *principle of end weight,* which disfavours ending on a phonologically light element (Gast, 2007).

(25) with *send* (n = 3)

    a. Too late <u>sent it Royal Mail</u> (could be *to* or *by*)

    b. <u>send it the papers</u>. That's as good as finding Jesus in your toast. (Liverpool)

    c. imagine if they did <u>send it the customs</u> (Liverpool)

These examples suggest that P-DROP is additionally available with *send.* It is worth noting, however, as we saw, following Haddican (2010), that THEME-GOAL ditransitives of the *give/send* type, are acceptable in Manchester for a substantial proportion of speakers.

***take, send* and *bring* with animate THEME**

Where the THEME is animate, P-DROP is also common with *take, send* and *bring.* Such structures are reported in Biggs (2018); Myler (2013). These are accordingly attested in NWatlas-corp:

(26) Take/send

    a. Couldn't you <u>take them the fair or the zoo</u>? (Knowsley)

    b. I thought it was funny when they came here, <u>took them the disco</u>

    c. <u>Sent him the shop</u> to buy me an "orangu" yesterday.

    d. I Bet Gerrard used to <u>send you the shop</u> for a 10p mix

Additionally, P-DROP is possible with ditransitive *bring*:[11]

(27) Bring

    a. Thanks mate, glad we could <u>bring you the game</u>

    b. Can you <u>bring him the match</u> josh?

---

[11]Decontextualised, some of these examples may be ambiguous between TGD/GTD, and could offer useful structures to test speaker intuitions.

    c.   <u>Bring me the karaoke</u>!

    d.   ill <u>bring them the game</u> tonight lark! You in the squad?

This point of contact between P-DROP and TGDs is part of what feeds the extant analysis proposed for TGDs as underlyingly P-DROP, at least for those speakers who have both P-DROP and TGD in their grammar.

Full-DP THEME and GOAL

Animacy constraints

P-DROP is frequent where both THEME and GOAL are full DPs. Crucially, almost all examples are with *take*, with animate THEME and directional GOAL, and are found within the Liverpool sphere of influence. There were 18 examples in NWatlas-corp with animate THEME, and one example where the THEME was inanimate.

(28)   Animate THEME (n = 18)

    a.   Ok so I'm defo <u>taking the boys the park</u> tomorrow early

    b.   Enjoyed the choirs and singing, but after that <u>we took the kids the community centre</u> then headed home. (Birkenhead)

    c.   He's more Home Brew Hobby, had his kitchen redone divorced. Fox is more <u>taking the kids the zoo</u> on his access day divorced. (Birkenhead)

    d.   When <u>taking the dogs the park</u> seemed like such a good idea (Liverpool)

(29)   Inanimate THEME

    a.   Do you think I'd <u>take a rucksack the match</u> yer div (St. Helens)

In all but one of the above cases, the THEME is animate: a person/animal is being taken or sent to a place by an AGENT. This is semantically distinct from the TGD structures that we have seen where the THEME is inanimate. Where the THEME is animate, there is implied agentive motion. *Send*, *bring* and *take* are co-agentive in the sense that there is a subject performing an action, and a secondary subject who is taking part in the action.

    Whilst Myler (2013) reports that inanimate subjects are possible with intransitive P-DROP, it is worth noting that there is a strong preference for an-

imate subjects in terms of frequency, which carries to ditransitive P-DROP, where THEMES are much more frequently animate than inanimate. While the animacy of THEME is important, the animacy of the GOAL is critical.

Consider the distinction between:

(30) a. he brought the keys to his boss
      b. he brought the keys his boss
      c. he brought his boss the keys
      d. he brought the keys the park
      e. *he brought the park his keys

Where the GOAL is inanimate, such as a destination (the park), it is illicit in the standard double object construction (GTD) with *bring/take*. Where the GOAL is animate (his boss), it is fine with the GTD (30c), and likely wins out over the ditransitive P-DROP variant (30b) in production.

Of course, these animacy constraints are well known with ditransitives:

(31) a. *he sent France the letter
      b. he sent the letter to France

With send, there are two examples; one active and one passive:

(32) with *send* (n = 2)

      a. he'd lost his marbles and was trying <u>to send the dogs the shops</u> with his shopping lists!! (Liverpool)
      b. When really lickle man gets slaps on the head and <u>sent the shop by his boys</u> loooool

The examples with *send* are interesting. As mentioned, *send* can carry a *motion* meaning. This aligns it with the paradigm where P-DROP is available with verbs of motion. As we have seen, it is also attested where THEME is pronominal.

Send also carries a *transfer of possession* meaning, which it has in common with *give*. This has meant that it often gets put together with *give* in analysis. I turn now to a discussion of *give*.

184

Give

There are fewer examples with *give*. *Give* involves *transfer of possession*. In syntactic theory, it is often argued to have two components which are able to license separate arguments CAUSE-HAVE (e.g. Harley & Jung, 2015). *Give* is able to alternate between PDAT and GTD. However, where the THEME is pronominal, it is usually considered to be poorly accepted, unless it carries particular emphasis.

(33) a. they gave the attackers the weapon
      b. ?they gave the attackers it
      c. they gave it (to) the attackers

The question of whether P-DROP extends to *give* is important. Biggs (2018) claims that it is the availability of the functional element $\varkappa$, which performs the underlying function of overt *to* in verbs of motion, which additionally allows for *give* to occur with full DP complements in both active and passive structures.

Interestingly, the majority of cases with *give* are in the area identified as $_p$TGD-dominant in the cluster analysis: Manchester, Wigan, St. Helens, Leigh and Warrington, and less so in Liverpool.

(34) with *give* (n = 13)

      a. Could have <u>give it the rest of us</u>
      b. If you can't tell, <u>give it the attackers</u> as an advantage
      c. Don't <u>give it the press</u> lol
      d. have saved there money and <u>give it the food banks</u> imo
      e. That's what I would do for Great British Bake Off, Stick a pie in a barm and <u>give it the judge's</u>.

Haddican (2010) finds such structures to be acceptable by many of his Manchester respondents. Recall the plot in Figure 2.1: while TGDs with two pronominal objects are most accepted, TGDs with a pronominal THEME and full DP GOAL (pDP) are the second most accepted.

The availability of pDP in the North West is unsurprising. It is reported quite widely in the literature as being accepted both in Manchester and Liverpool (Biggs, 2018; Haddican, 2010). For Haddican's Manchester sample, pDP had

a mean acceptance rating of around 40%.[12] Structures where the GOAL is a name are also frequent:

(35)   a.   <u>Give it Joe Hart</u> at least then he can lead from the back, and let the others play without the shackles

     b.   don't <u>give it Olivia</u>

     c.   yeah will do, il <u>give it Sam</u> or someone this morning

Whilst the data for pDP structures are frequent, they are not frequent enough to provide a comparative geographical distribution in the way that it is for ditransitives with two pronominal arguments. However, it is clear that pDPs are widely available.[13]

TGD-*give* with two full-DP objects

TGD-give is not attested with two full-DP objects in NWatlas-corp. However, it is reported as accepted by some speakers in Haddican's (2010) Manchester sample. It is also reported as accepted by Biggs's (2018) sample in Liverpool.

(36)   From Biggs (2018, p.6):[14]

     a.   Mary gave the book the teacher. (OKLiverpool, *Manchester)

     b.   Mary sent the package her nan's. (OKLiverpool, *Manchester)

The fact that it is unattested may be explained by the fact that, with two full-DP objects, it is competing with the standard double object construction (GTD), which is essentially categorical in the corpus.

(37)   Standard double object construction (GTD)

     a.   Mary gave the teacher the book

     b.   The teacher was given the book

---

[12]Participants were asked to shade a bar representing between 0% and 100% acceptance rate, and 40% was the mean response to pDP structures.

[13]I did gather a separate dataset based on strings containing *give* and *sent*, with *it* as the THEME and a list of common GOALs (popular first names, *the post office*, *the bank* etc.). The results confirm that the structure is indeed widespread.

[14]It is notable that the majority of the sentences reported by Biggs (2018) as THEME-GOAL ditransitives with two full DP objects are with the verb *send*. There is only one example with the verb *give*. We have seen that *send* can play both *transfer of possession* and *motion* roles. It is found in Liverpool in the above examples. However, with *give*, the emphasis on a *transfer* reading may put it out of the bounds of the kind of preposition-drop that we find in the North West. I return to this in the next chapter.

However, as we will see in the survey data in the next chapter, *give* is *accepted* with full DP objects by many speakers in the area *between* Manchester and Liverpool. It may be that some speakers who accept DPDP structures with *send*, due to its property of conveying motion, carry over that acceptability to *give* by analogy.

THEME-passives

Biggs (2018) additionally cites examples of TPASS where the derived subject position is a full noun phrase as distinct to the grammar available in the North West, which she links to the wider availability of P-DROP.

(38)   From Biggs (2018, p.6):

   a.   The book was given the teacher. ([OK]Liverpool, *Manchester)
   b.   The package was sent her nan's. ([OK]Liverpool, *Manchester)

Such examples, Biggs argues, support Liverpool ditransitives as involving P-DROP, and as being linked with P-DROP found with motion verbs, at least for those speakers who possess this grammar. There were again no available examples in NWatlas-corp of TPASS-give with a THEME in a derived subject position, only one example repeated below for *send*:

(39)   When really lickle man gets slaps on the head and <u>sent the shop by his boys</u> loooool

There are two other attestations in NWatlas-corp, but both involve a CP head, and neither of them are from Liverpool speakers. This said, if it were the case that TPASS was derived via P-DROP, we might expect it to be more frequent in areas such as Liverpool, if P-DROP is productive there.

(40)   a.   They reported <u>the data that was given them</u>. Our government took no actions until March, even with the new data that was given to them. (Warrington)
   b.   Dunno mate, just <u>a photo that was sent me</u> (Manchester)

Again, the rarity of such structures in production does not necessarily mean that they will be degraded for speakers. In fact, TPASS is rare even with two pronoun objects, though still widely reported as acceptable (Haddican & Holm-

187

berg, 2012).

It is hard to verify the claim that P-DROP in verbs of motion (*go*, *come*, *bring* etc.) carries to verbs involving transfer of possession (*give*, *send* etc.) from the corpus data, given its rarity in production. We can, however, test its availability through the judgement survey. I return to this in the following chapter.

For now, we are left with the following questions:

1. What is the relative geographical distribution of [P/D-DROP]?

2. To what extent do variants of [P/D-DROP] align geographically with [DATALT]?

Identifying locations where there is both a high rate of ₚTGD and a high rate of P-DROP is informative, as we might expect these locations to be the most likely hosts for a grammar which derives TGDs via P-DROP. As we have seen, Liverpool itself is not, in fact, a strong candidate, as it strongly favours ₚGTD over ₚTGD. However, the ₚGTD area is tightly limited to Liverpool and its immediate environs; we do not need to go very far outside of Liverpool to find areas that are both ₚTGD dominant (shown in Figure 4.23, repeated below in 4.33), and as we will see, P-DROP-dominant. I return to this area of intersection between ₚTGD and P-DROP, and its importance, in §4.8.
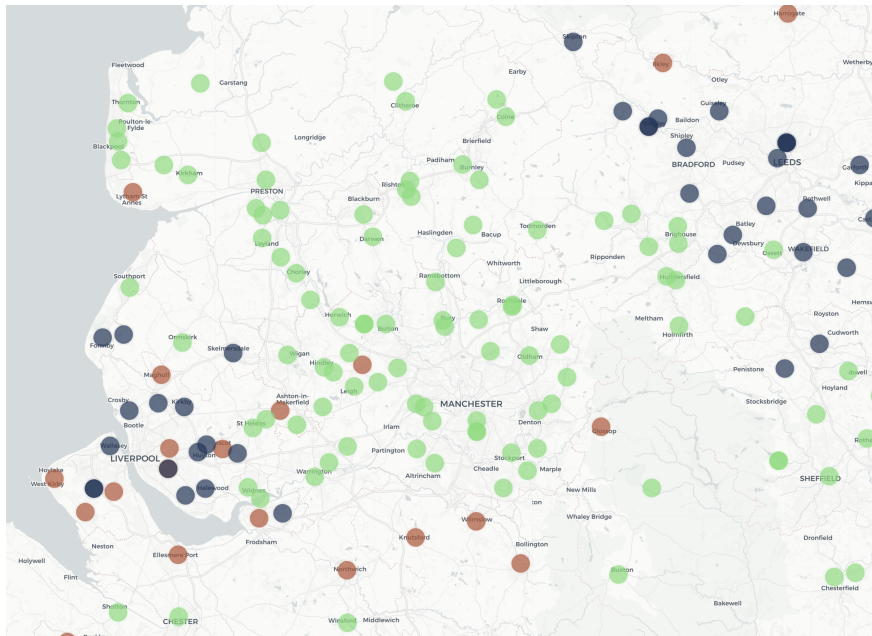
Figure 4.33: Pronominal ditransitives cluster map. Three clusters, detail North West.

188

Additionally, given the discussion on the possible origins of P-DROP, presented in the background as being in some way related to definite article (and preposition) reduction, the following questions are pertinent:

1. To what extent do variants of [P/D-DROP] align with the historical distribution of DAR?

2. What evidence is there that (some) variants of [P/D-DROP] are descended from DAR?

I present these data in the following section.

4.5 Preposition/determiner dropping

Section overview

I begin the discussion of [P/D-DROP], as with [DATALT], with the overall distributional patterns for the UK and Ireland, first as the relative rate of each variant, and second via a cluster analysis of all four variants, by location. Through an observational analysis of the distribution of clusters, I complicate the question of whether PD-DROP is originally a feature of London/South-East English which has diffused North West, as reported in Gopal et al. (2021), or if it may have been separately innovated in multiple locations.

Here, I assess the geographical alignment of [P/D-DROP] clusters with the distribution of DAR reported in the Survey of English Dialects. I additionally draw on data from traditional corpora (FRED and BNC), which show variants of [P/D-DROP], including PD-DROP to be geographically widespread and attested in older speakers. I then look at the structural properties of [P/D-DROP] variants in the North-West, and assess the extent to which they align with the characteristics reported for London and South East PD-DROP.

Returning to the question of TGD-P-DROP, I look for locations for which there is a geographic overlap between the two variants. This intersection will prove pivotal to the later analysis in the survey.

Mapping individual variants

I start with an overview of the variation by mapping the relative rates of individual non-standard variants of [P/D-DROP]. Here I give a brief description of the apparent variation, but I leave a more in-depth discussion to the subsequent cluster analysis.

Preposition dropping

Figure 4.34 shows the rate of P-DROP compared to the other possible variants, in geo.place_id locations across the UK and Ireland. The darkest shades correspond to $\approx 50\%$.
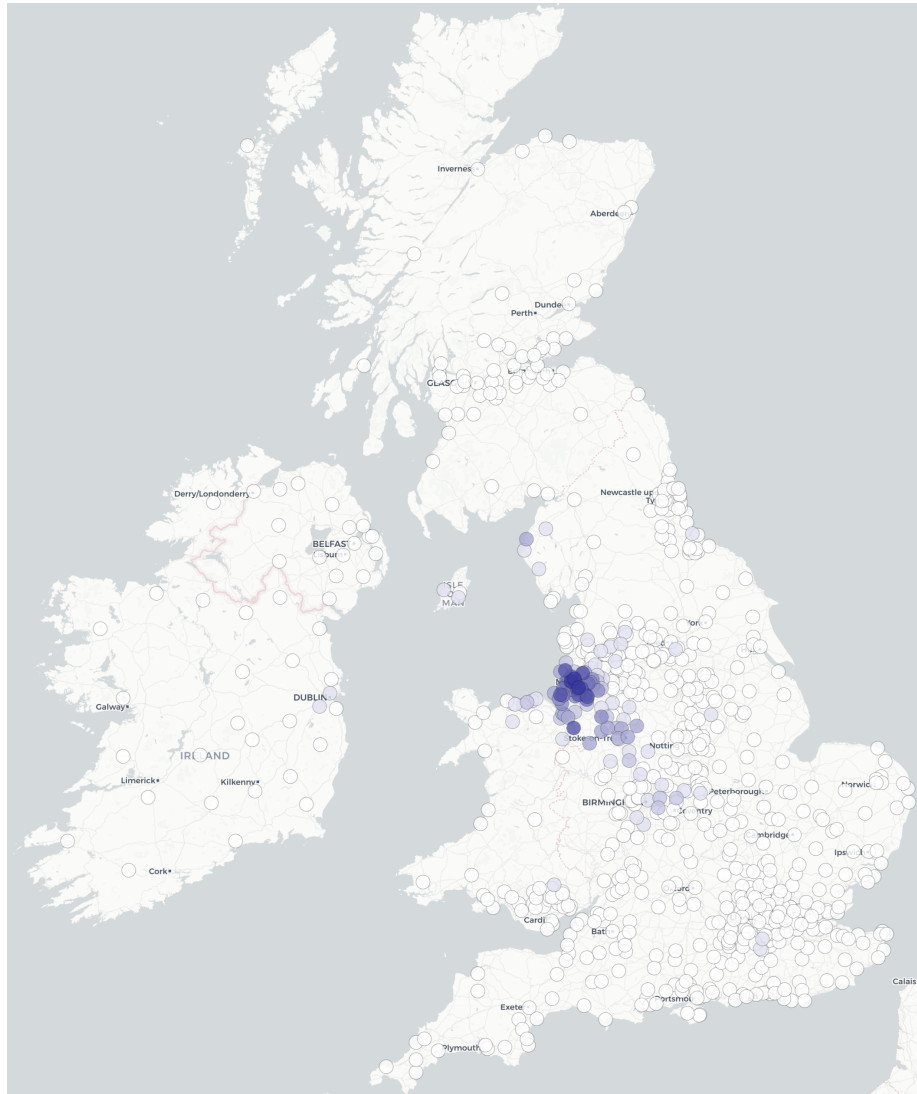
Figure 4.34: The distribution of preposition dropping on Twitter, 2011-2021

The distribution is localised to the North West, with substantial rates of use extending south and south east from Liverpool, towards Birmingham. There are additionally some notable points of high use across the Irish Sea in Dublin, and along the North Welsh and Lancashire coasts. I return to the local distri-

bution, and possible explanations in §4.5.2.

Preposition-Determiner dropping

It is immediately apparent from the map presented in Figure 4.35 that PD-DROP use is widespread across England with hotspots—representing $\approx 50-60\%$ use—in East London, Essex, and the East Midlands and high rates across the West Midlands, South Yorkshire, and the Manchester city region.
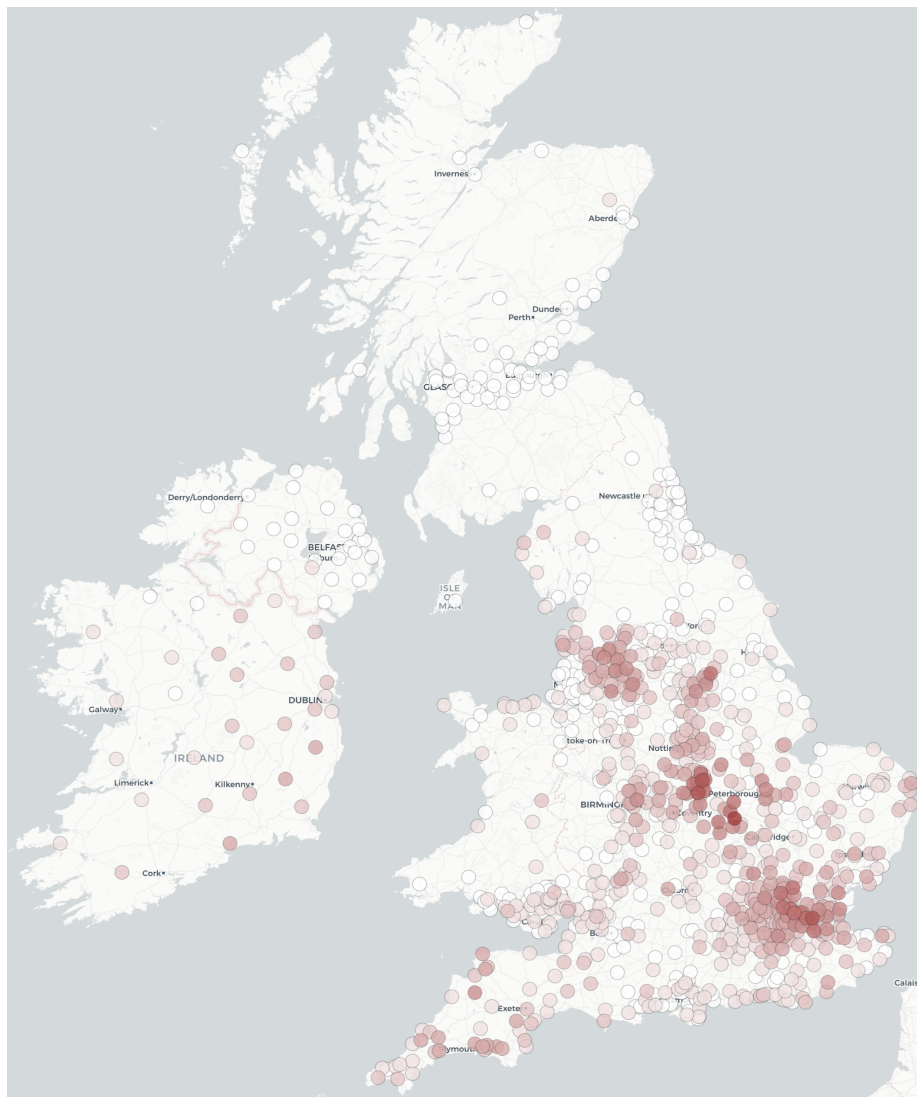
Figure 4.35: The distribution of preposition-determiner dropping on Twitter, 2011-2021

This distribution is similar to that provided by Gopal et al. (2021) (shown in

Figure 2.11, and repeated below in Figure 4.36), which is unsurprising, given that both maps represent PD-DROP distribution on Twitter. There are, however, some important differences. First, the map in Figure 4.35 corresponds to data collected over a longer time period; from 2011-2021, where Gopal et al.'s (2021) data is drawn between 2017-2019. Due to larger time-window, this results in more data, but beyond this, as we saw in the how Twitter data have changed over time, the period in which PD-DROP use was most prolific was between 2011-2015. This difference may explain why the usage rates in the South West and Ireland are more apparent in the present map. This effect is amplified by the fact that Gopal et al.'s (2021) map is based on a corpus search that uses a narrower range of strings, involving the verb *go* and "a small set of frequent, semantically appropriate noun phrases" (Gopal et al., 2021, p.265).[15]
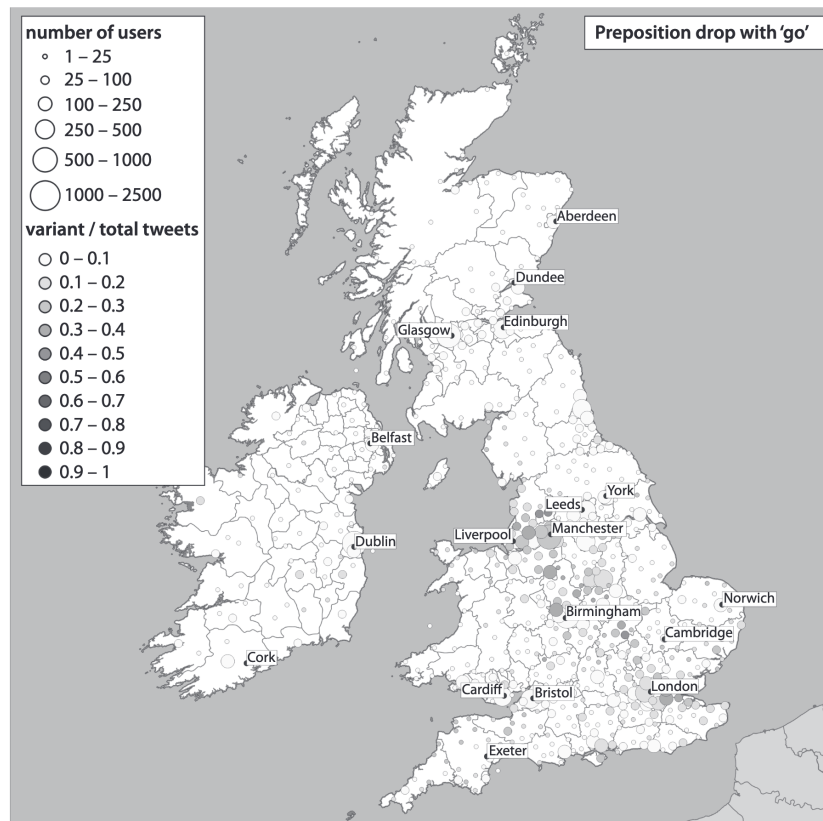
Figure 4.36

[15]Along with the verb *go*, the search included: (to) Amsterdam, go (to) Asda, go (to) (the) chicken shop, go (to) college, go (to) jail, go (to) London, go (to) Manchester, go (to) Nando's, go (to) Paris, go (to) prison, go (to) (the) pub, go (to) school, go (to) Tesco (Gopal et al., 2021, p.265).

Second, the map in Figure 4.35, shows only PD-DROP strings, while Gopal et al.'s (2021) map combines strings which represent P-DROP. As a result, in Figure 4.35, we see a clear absence in Merseyside, which is not visible in Figure 4.36. As we saw in Figure 4.34 in the previous section, this area is dominated by P-DROP, rather than PD-DROP.

Third, there is an important difference in the way in which location data are gathered between the two maps. As described in the background chapter, Gopal et al. (2021) assign location based on an algorithm which scored users proportionately based on 'keywords' and other types of profile metadata they provided, as likely to be from a given location. This is a more direct method than the one underlying the map in 4.35, which is based on Twitter's geo-location algorithm, and includes GPS location data. Broadly, as we saw in the comparison between NWatlas-corp and pDrop-corp (§4.2.2), there is a high level of correlation between the two methods. However, there may be specific consequences for including GPS data. Specifically, the South West region receives substantial numbers of seasonal visitors, who may be recorded in the region, but are not from that region, if, for instance, they are tweeting whilst on vacation. I explore this concern in §4.5.4.

However, the presence of the feature in the South West follows Harris's (1967) account of PD-DROP there. The high rates across the Republic of Ireland are more surprising, given that there are no previous reports of its use there, with a notable absence in Northern Ireland, which, as with [DATALT], patterns instead with Scotland.

Particularly notable is the high concentration of PD-DROP in the East Midlands, which appears to be as pronounced, if not more so, to that which we see in East London (see §4.5.3 for a detailed breakdown of the numbers). This again fits with descriptions in the historical and present-day literature as reported in Evans (1848) who reports PD-DROP use in Leicestershire in the 1840s, and Braber and Robinson (2018) who reports it in present-day Leicester.

Determiner dropping

D-DROP is localised to West and South Yorkshire - around Leeds and Sheffield, with particularly high rates in East Riding, around the city of Hull There are additionally notable rates of use spreading across England, Scotland and Northern Ireland.

The low level of D-DROP use visible in many areas across England is quite scattered, and likely noise in the data, due to *telegraphic writing*, which was expected, as I discussed earlier in the methodology.

The highest rates of use are most notably localised in the area where we find DAR-0 in Barry's (1972) SED map. I return to a discussion of the correlation between the distribution of DAR and [P/D-DROP] in §4.6, but this pattern was expected considering that there is likely a direct link between DAR-0 and D-DROP, as both imply the complete absence of the determiner.
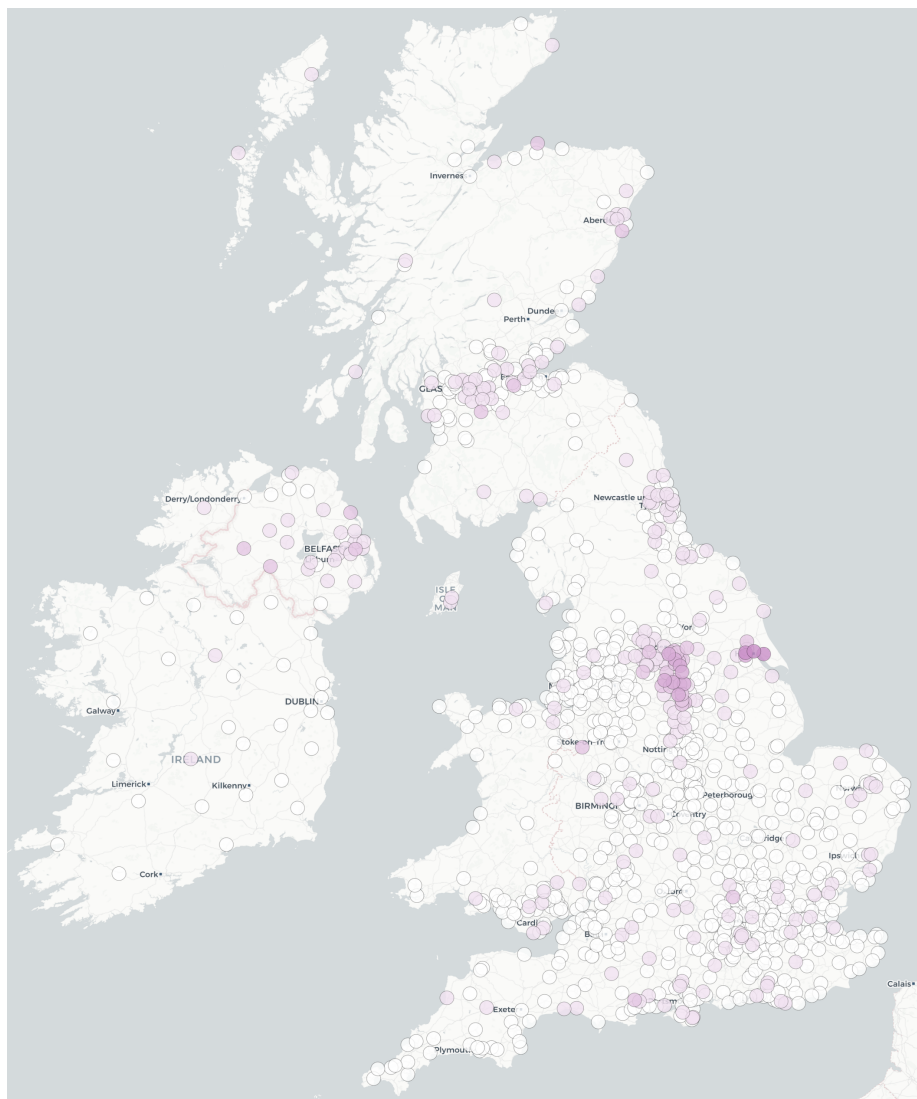
Figure 4.37: The distribution of determiner dropping on Twitter, 2011-2021

It is interesting here to see that Hull exhibits the highest D-DROP rates, with

194

rates around the West Midlands notably more diffuse. Kerswill and Williams (2014), in comparing the effects of levelling between Milton Keynes, Reading and Hull, find that Hull is more linguistically conservative, likely due to its relative geographical isolation.

There does seem to be a consistently higher band of D-DROP use through the North East, the Central Belt of Scotland, and across into Northern Ireland, with a clear divide in Ireland between North and South.[16] The rates are relatively low, and, as we will see, are overwhelmed by standard NO-DROP in the cluster maps. They do, however, follow a consistent theme with other variants that we will encounter; that of a historically rooted inter-dialect continuum between these regions.

K-means clustering

The Elbow plot was again used as a first approximation of the most appropriate number of clusters to choose (see §4.4.1 for a description of this approach).
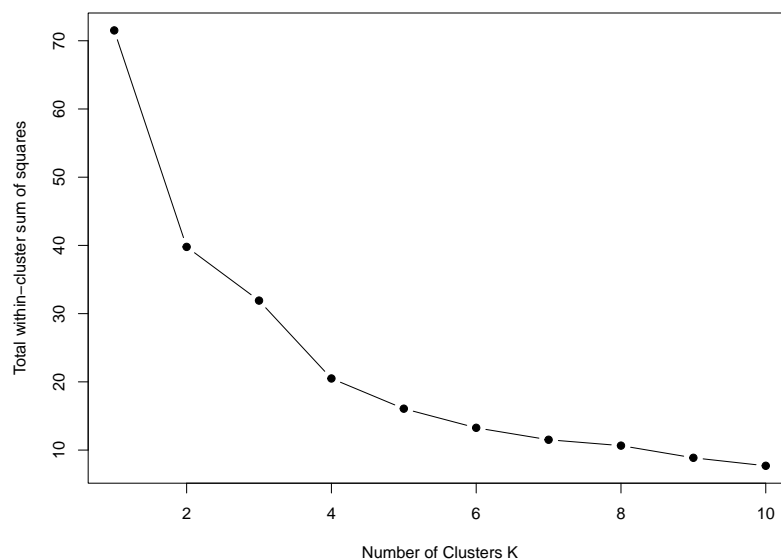
Figure 4.38: Elbow plot to estimate optimal number of clusters: P-DROP

The Elbow plot is not as clearcut as with [DATALT]. It seems that around five or six clusters should be optimal. Running K-means clustering with six

---

[16]Note, with regard to the distribution of D-DROP in Scotland, that the Twitter search did not include *tae*, the alternative Scots spelling of *to*, which may impact the results shown here.

clusters groups the data effectively into two P-DROP-dominant clusters, two PD-DROP-dominant, one D-DROP-dominant, and one NO-DROP (or standard)-dominant.
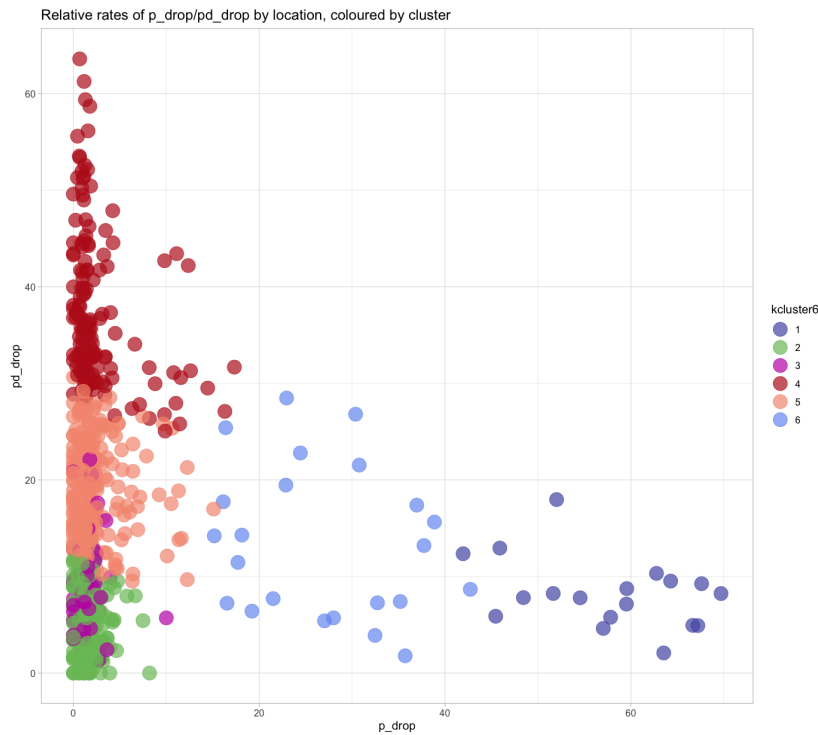
Figure 4.39: Scatter plot for [P/D-DROP] six clusters

This offers a good starting point. However, after some experimentation, it was found that, in fact, using *twelve* clusters, and then grouping clusters into sub-clusters, following the initial six-cluster grouping, offered a more nuanced picture of the data, with better gradation within each main cluster.[17]

Colour coding

Clusters were manually colour coded into sub-clusters as *Red* for PD-DROP, *Blue* for P-DROP, *Magenta* for D-DROP, and *Green* for NO-DROP. Darker shades correspond to higher rates of a given variant. The two clusters with the highest rates of NO-DROP (Clusters 10 and 12) are coloured with the same dark shade of green.

---

[17]In effect, grading clusters into sub-clusters by variant reduces the number of main clusters to six, which fits the number suggested by the elbow plot. An alternative method would be to formally divide the initial clusters by running a hierarchical clustering algorithm, such as the *Ward method* on each initial cluster.

Cluster	P-DROP %	PD-DROP %	D-DROP %	NO-DROP %
1	2.2	9.7	8.2	80.0
2	1.5	8.8	2.7	87.0
3	32.5	9.3	4.6	53.6
4	2.3	6.4	17.5	73.8
5	2.4	49.3	4.7	43.6
6	60.1	6.2	4.3	29.3
7	2.6	33.0	5.3	59.0
8	3.2	23.0	5.3	68.5
9	1.4	12.6	30.4	55.6
10	1.3	1.5	3.3	93.9
11	1.9	16.9	4.2	77.0
12	1.8	1.8	10.2	86.3

Table 4.7: Breakdown of Centres for each cluster: twelve clusters
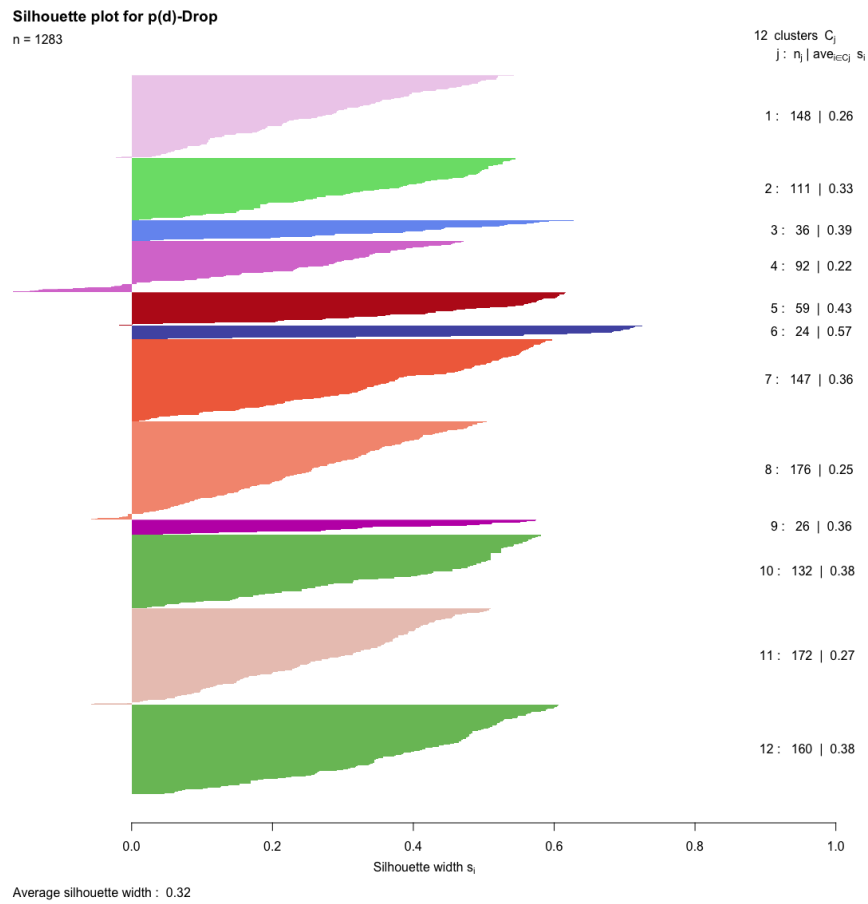
Figure 4.40: Silhouette plot for eight clusters: [P/D-DROP]

The boxplots in Figure 4.41 show the breakdown of cluster distributions for each variant of [P/D-DROP].

197
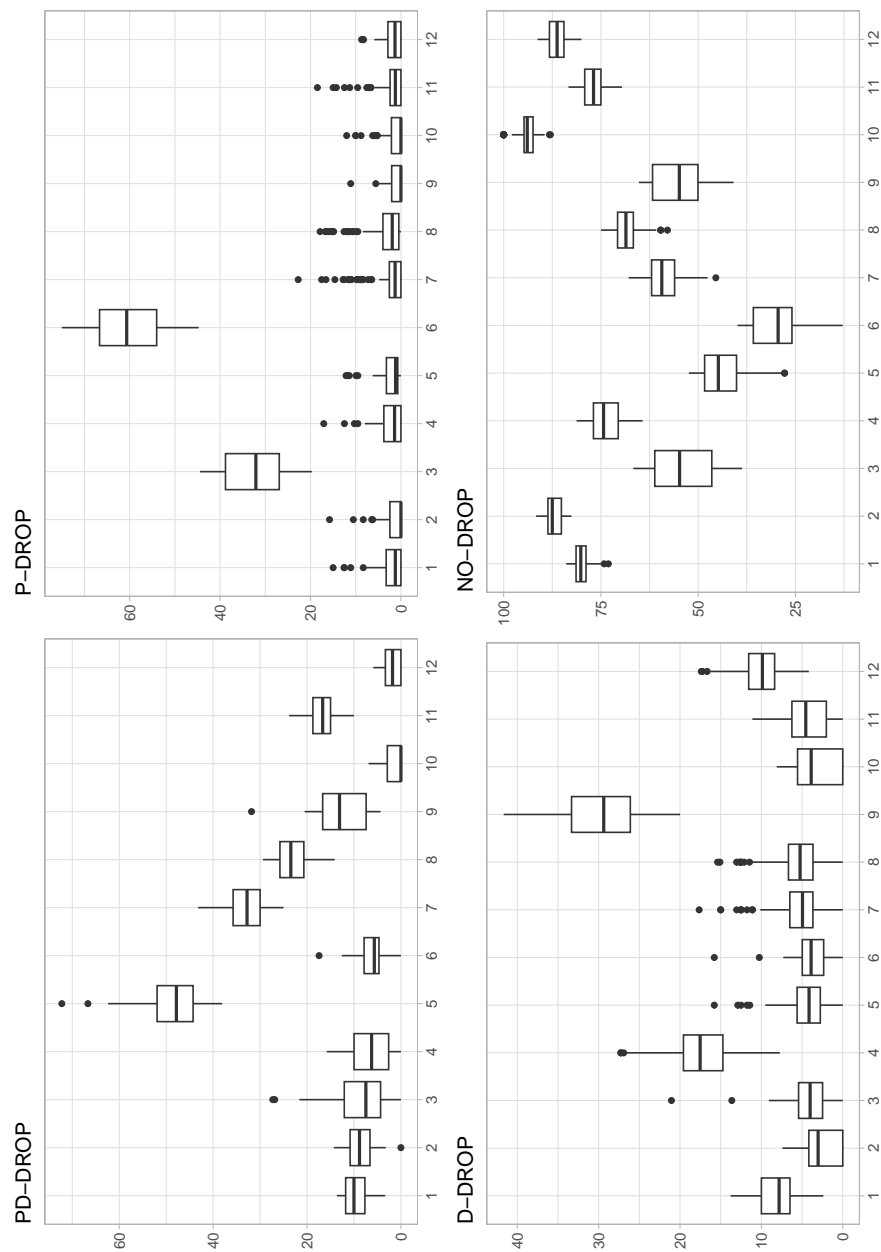
Figure 4.41: Boxplots showing the rates of [P/D-DROP] variants by cluster.

They show that P-DROP is clearly represented in clusters 6 and 3, with ranges of $\approx 55 - 65\%$ and $\approx 25 - 40\%$, respectively. D-DROP is mostly captured by clusters 4 and 9, with ranges of $\approx 15 - 20\%$ and $\approx 25 - 35\%$. NO-DROP is represented across all clusters, which is expected, given its status as the standard, but is most dominant in clusters 1, 2, 10, 11 and 12, with cluster 10 capturing places where NO-DROP is around 90%. PD-DROP is most dominant in cluster 5 representing locations at ranging from 40-50%, but we also see that PD-DROP is widespread with sizeable proportions in clusters 7, 8 and 9. Clusters 10 and 12 are largely absent of P-DROP, however, there are a number of outliers: some places with quite high P-DROP are included in these clusters. This can be seen in the scatterplot in Figure 4.42, below.
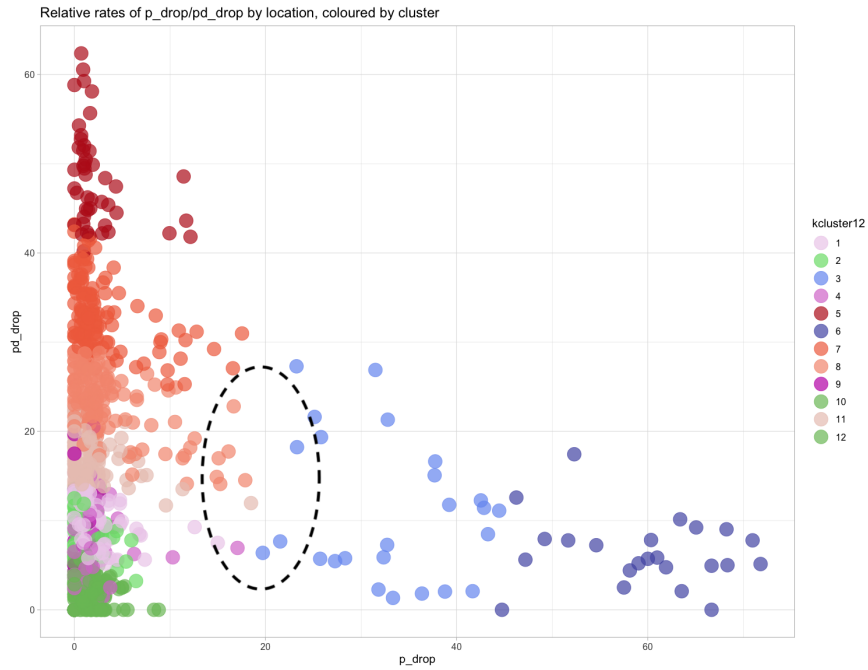
Figure 4.42: Scatter plot for twelve clusters, colour coded following the initial six clusters: P-DROP/PD-DROP. Dotted oval highlights clusters that are near-neighbours.

In Figure 4.42, we see, that most cluster points are contained on the left of the plot, denoting low P-DROP. Blue points correspond to clusters that contain progressively higher rates of P-DROP, and are shown, accordingly, at increasing points on the x-axis. However, most of these points are not bound to the x-axis, that is, they also contain anywhere between $\approx 10 - 20\%$ PD-DROP. Red points, corresponding to high PD-DROP, more tightly coalesce to the y-axis, but there are also a number that stray into higher P-DROP rates. These stray locations

are (unsurprisingly) in the North West, where P-DROP is dominant.

Figure 4.43 shows the same data, but presents the rate of D-DROP on the x-axis, and maintains PD-DROP on the y-axis.
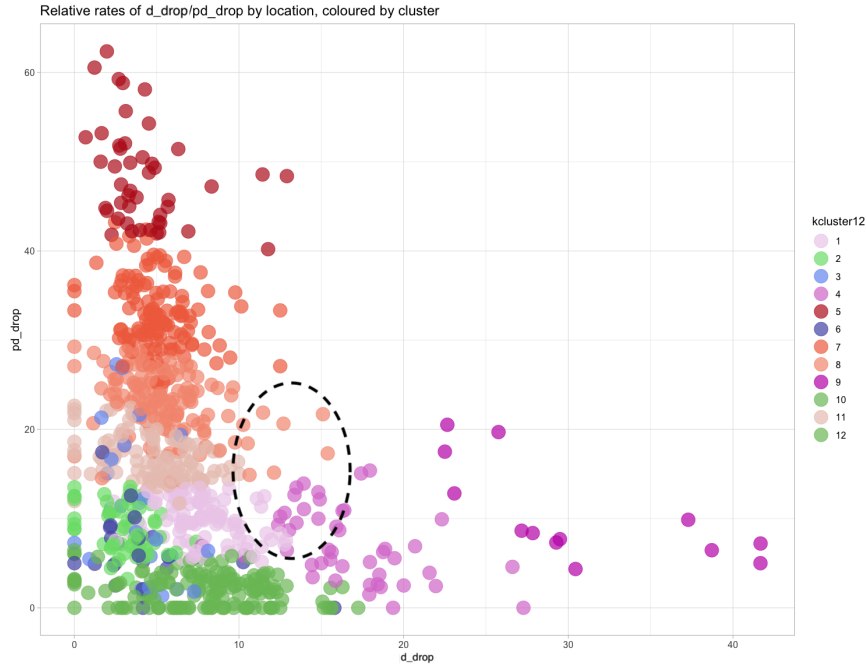
Figure 4.43: Scatter plot for twelve clusters, colour coded following the initial six clusters: D-DROP/PD-DROP. Dotted oval highlights clusters that are near-neighbours.

This allows us to see the spread of D-DROP-dominant clusters, which was obscured in Figure 4.42. We can see from this plot that there are, overall, fewer places that are D-DROP-dominant, and most also contain around 10% PD-DROP. There are also a number of points that are coloured as light magenta, denoting lower rates of D-DROP, that have an equal, sometimes *higher* rate of PD-DROP. They are nevertheless categorised as D-DROP, as it is the rate of D-DROP is the factor that distinguishes them.

Note that PD-DROP is quite high in some places that are coded as P-DROP, and there are edge cases where PD-DROP is slightly *higher* than P-DROP. These clusters are highlighted in Figures 4.42 and 4.43 by the dotted oval. This reflects the initial cluster grouping at *six* clusters: though these locations may have high PD-DROP, they are distinguished by also having high P-DROP: around 20%. The locations that fit into this *mixed* category typically lie on the transition zones between the respective dominant areas.

In Figure 4.44, we see a good separation between clusters: most points are

closely grouped along their respective axes.
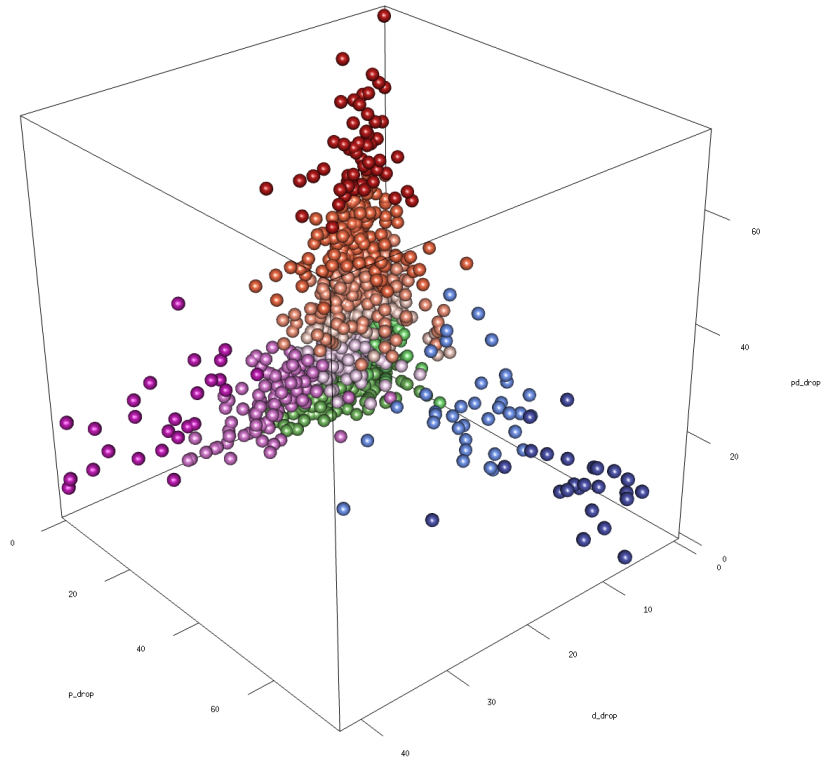
Figure 4.44: 3D Scatter plot for twelve clusters, colour coded following the initial six clusters:
P-DROP/PD-DROP/D-DROP
Interactive version at: `http://bit.ly/3XoiQ8z`

The 3D plot allows us to see the how each cluster colour is distinguished along separate axes. It also shows that, towards the lower end of each axis, there is some degree of overlap between clusters. We do see, again, a few outliers; places, for instance that are coloured blue, corresponding to high-P-DROP, but which are also quite far along the D-DROP axis. Green clusters, which are NO-DROP-dominant (*go to the N*), are bunched around the bottom corner of the plot, as they contain low rates of the three non-standard variants.

Mapping clusters

Here I present all of the cluster data on a map of the UK and the Republic of Ireland.
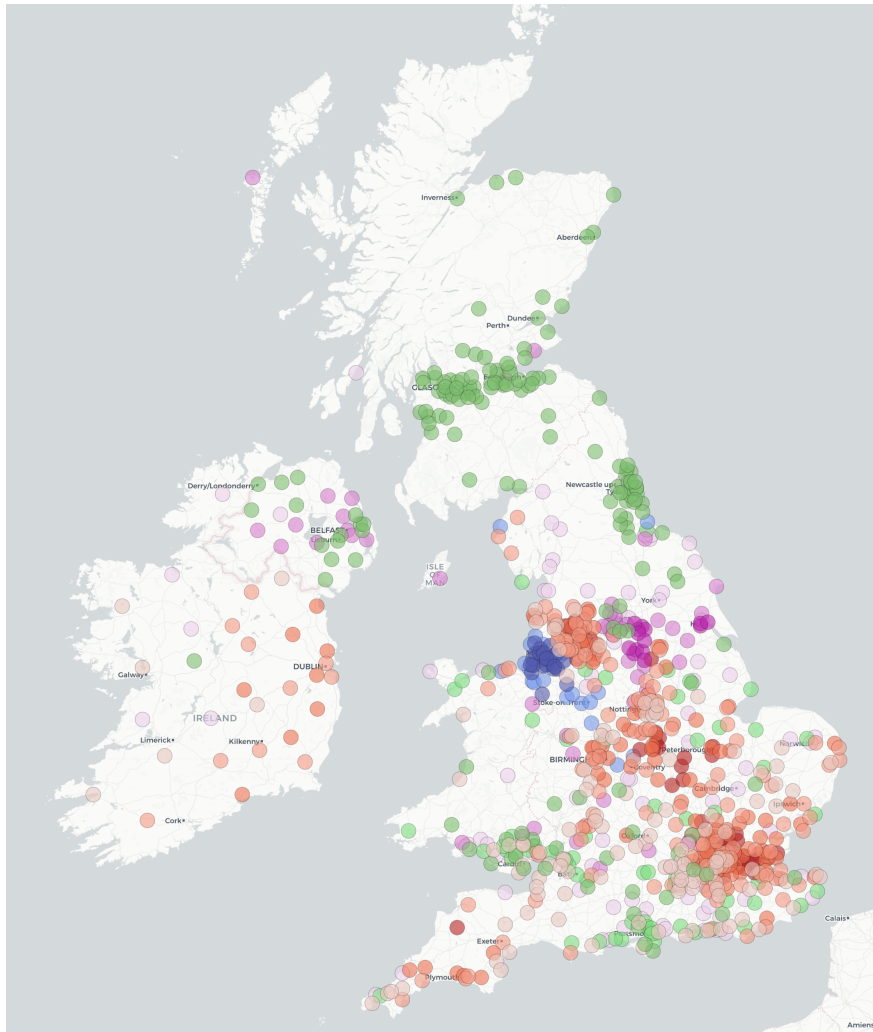
Figure 4.45: Preposition/determiner dropping cluster map. Twelve clusters, based on initial six. Places limited to minimum 30 hits.
Red = PD-DROP (Darkest = ≈ 50%-Lightest = ≈ 17%)
Blue = P-DROP (Darkest = ≈ 60%-Lightest = ≈ 30%)
Magenta = D-DROP (Darkest = ≈ 30%-Lightest = ≈ 12%
Green = NO-DROP (> 80%)
Interactive map available at: `http://bit.ly/4bWQ17O`

4.5.1 Interpreting the cluster data

[P/D-DROP]: Diffusion or multiple innovation

Widespread use of PD-DROP is perhaps the most striking aspect of the data. Rates of above 20% are found throughout England and Ireland, with many places above 30%. Meanwhile, use of PD-DROP is markedly lower, at around 5%, in Scotland and the North East.

202

Such widespread use seems to suggest diffusion and levelling. A recent uptick, akin to what we saw with TH-fronting, may have come about in the context of social media, and other media. It is not well known the effect that social media may have on the diffusion of linguistic features. Kerswill (2003, p.240) suggests that the rapidity of the spread seen with TH-fronting may have been influenced by "spoken media [...] which makes speakers more positively disposed towards the incoming forms they hear from the people they meet". It is possible we are seeing a similar effect here with PD-DROP, driven by social media.

Figure 4.46 presents the cluster map of preposition and determiner-dropping with dotted lines indicating Gopal et al.'s (2021) proposed path of diffusion from SE→NW. We see the clear spatial distribution of three dominant variant types: P-DROP (blue), D-DROP (magenta), and PD-DROP (red). The clustering is regionally differentiated, with P-DROP forming a clear northwestern enclave, particularly concentrated around Manchester, Cheshire, and parts of North Wales. D-DROP, on the other hand, is more restricted, with a cluster centred on West Yorkshire and scattered cases elsewhere. Most notably, PD-DROP dominates the southeastern portion of the map, encompassing Greater London, parts of Essex, and extending into the East Midlands. The red-dominated landscape of this region stands in marked contrast to the northwestern and northern areas.

It is a reasonable starting point to assume that the places with the highest rates of use are those for which the innovation has more established roots. One interpretation of the pattern shown in Figure 4.46 is that the innovation spread North West, from urban centre to urban centre, following a gravity-based model, as predicted by Gopal et al. (2021) on the basis of simulated models of diffusion. Gopal et al. (2021, p.267) describe a "contiguous corridor corresponding to highly densely-populated and well-connected areas".

The pattern that Gopal et al. (2021) describe is evident in the distribution, with a corridor of very high PD-DROP use stretching from London to the North West. However, the picture is complicated by the also-high rates in the South West and Ireland, which do not fit the gravity-based diffusion model that they suggest. Instead, it would appear that a different explanation is warranted, at least for the South West and Ireland.

If PD-DROP had diffused from London to the South West, we would expect to see higher rates in the intervening conurbations such as Bristol. It is also
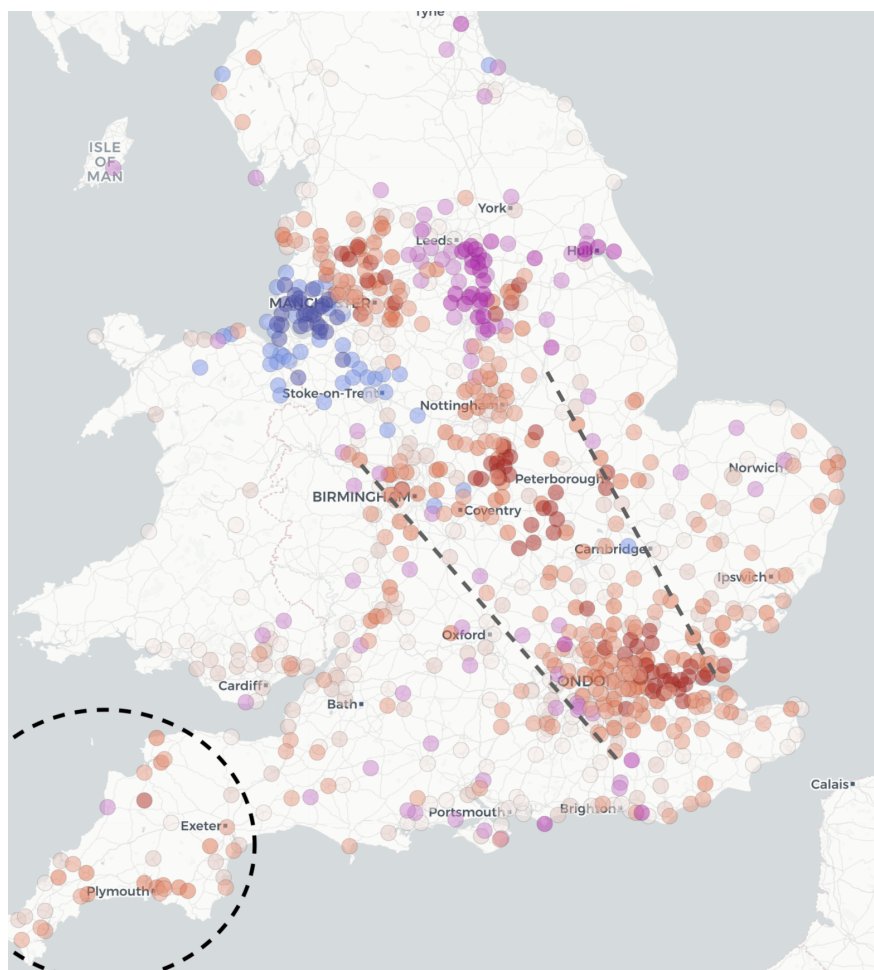
Figure 4.46: Preposition/determiner dropping cluster map.
Clusters removed with over ≈ 85% NO-DROP
South East to North West corridor.

clear from the present data that a SE→NW distribution of PD-DROP does not extend to Liverpool and the surrounding areas, which are instead dominated by P-DROP (blue). Likewise, PD-DROP dominance does not extend to West and South Yorkshire, where there is strong competition from D-DROP (magenta).

Leaving these anomalies to one side for a moment, it is nevertheless still very plausible that a SE→NW diffusion took place for PD-DROP, in addition to its being separately innovated — as tempting as it may be to take one side of the argument, it is possible that both diffusion and independent innovation coexist.

The timeline for this process—innovation to diffusion—is unclear, though

as we have seen, PD-DROP is attested in the North West in the 1880s (Ellis, 1889), earlier in the Midlands (Evans, 1848) and in London in the 1730s, as found in the Old Bailey Corpus (Huber et al., 2016). The literature which has reported PD-DROP as a relatively recent innovation, did not account for these early attestations. Additionally, it is difficult to see how a variant could have spread so rapidly across the country during the 19th century, and the existence of multiple independent sites of innovation seems more likely.

As discussed previously, a popularisation of cross-dialectally extant PD-DROP, in the wake of the development of Multicultural British English in London and Birmingham, accelerated by use on social media, may explain a more recent diffusion 'event' across the 'urban corridor' between London and Manchester.

4.5.2 The North West and Liverpool's historical sphere of influence

P-DROP is likely distinct from PD-DROP, with a separate course of innovation and localised transmission. Much of the spread of Liverpool English to surrounding towns can be explained by the migration of people from Liverpool to those areas during the 20th century, as well as its wider sociolinguistic influence. In Figure 4.47, we see P-DROP extending out from Liverpool to surrounding Cheshire, Lancashire and Staffordshire. In Liverpool's more immediate environs, these patterns are commensurate to the patterns of de-urbanisation, and relocation in the latter half of the 20th century. We see high rates of use, for example, in Ormskirk, Skelmersdale, Runcorn and Warrington, as well as across the River Mersey.

However, I again draw attention to the discrepancy between the distribution of P-DROP and that of the $_p$GTD, shown in Figure 4.23; the $_p$GTD is far more localised to the immediate vicinity of Liverpool. It may be that this is due to factors of identity. Perhaps, P-DROP is more salient than $_p$GTD, and therefore more accessible to speakers who wish to identify themselves with Liverpool. It is certainly true that the division that Knowles (1973) describes between the two competing 'urban fields' of Liverpool and Manchester is clearly represented here, as is his description of Liverpool and Merseyside being "in the North of England but not of it" (p.15).

The distribution also fits the picture described by Watson (2006), where
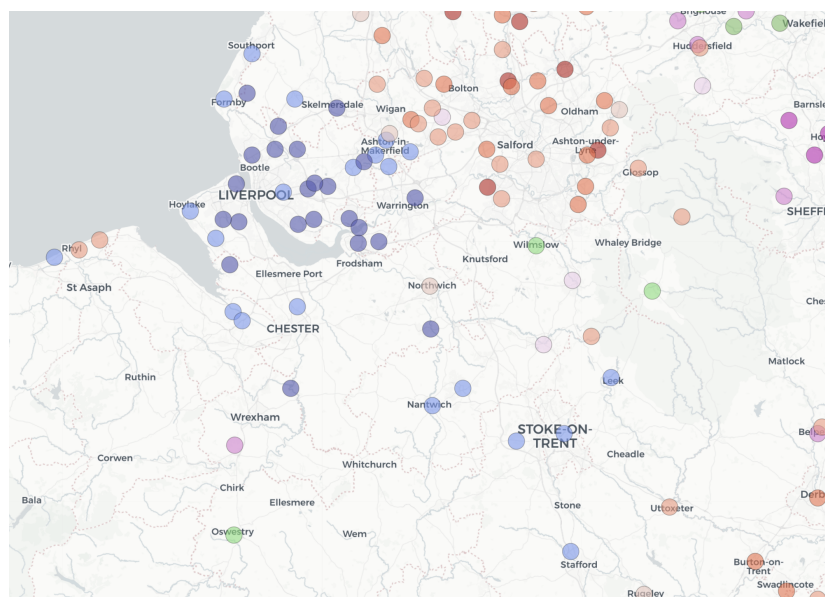
Figure 4.47: Detail of [P/D-DROP] cluster map, showing the North West and Staffordshire. P-DROP use (Blue) extends on a narrow corridor to Stoke-on-Trent

Liverpool seems to be resisting a general trend, in the North, towards regional dialect levelling, as shown by the extension, rather than contraction, of the contexts in which the Liverpool feature, aspiration of /t/ (t→h), may occur.

However, neither recent internal migration or speaker identification with place adequately explain the geographical extent of P-DROP use, which stretches from Dublin to Stoke, Birmingham and up the coast to Workington, as shown in Figure 4.51.

Regarding the further spread of P-DROP, Stoke is of particular interest. Stoke has been previously associated with Liverpool dialect features (Leach, 2018; McMahon & Maguire, 2013). Meta-commentary on the topic is widespread, with comments on discussion forums questioning the link.

Posted March 3, 2008   **People from Stoke...**

Why do they sound like they have scouse accents?
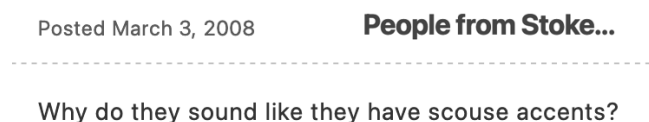
Figure 4.48: Blog comment asking why people from Stoke "sound like they have Scouse accents". Source: `www.liverpoolway.co.uk/index.php?/forums/topic/50208-people-from-stoke/`

The impression that there is a similarity between Stoke and Liverpool is

206

formalised by the two locations' phonetic proximity in McMahon and Maguire's (2013) Neighbournet graph. This used a method to quantitatively measure the phonetic proximity of English varieties, based on data from their *Sounds Comparison Project* (Mcmahon, Heggarty, Mcmahon, & Maguire, 2007).
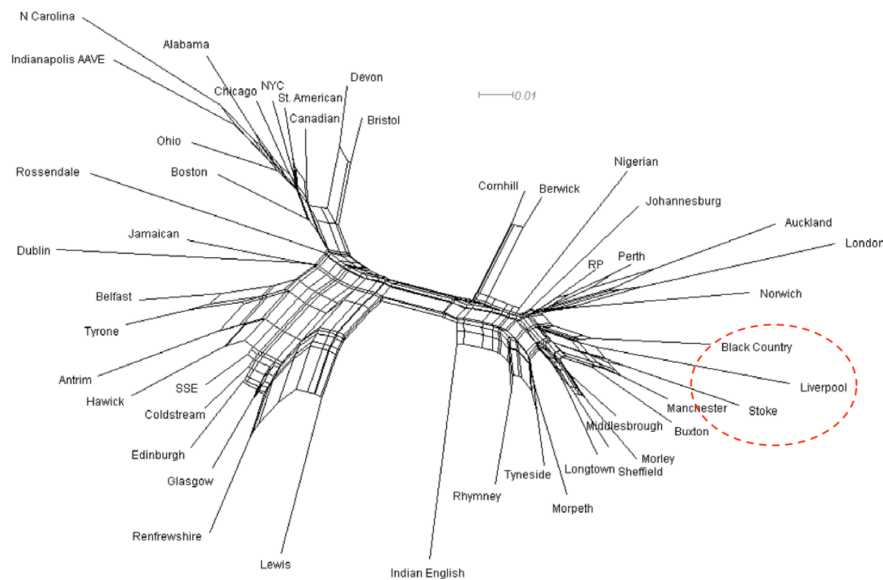
Figure 4.49: Neighbournet showing phonetic proximity of Stoke to Liverpool and Manchester
Source: McMahon and Maguire (2013, p.246)
also: Leach (2018, p.67) (for the highlighting of the proximity of Liverpool and Stoke in the Neighbournet

Here we see that Stoke and Liverpool are immediate neighbours. Interestingly, Maguire comments on this research that "Liverpool and Stoke were more closely related to each other than to most other varieties" (Maguire, 2014, personal communication, reported by Leach (2018, p.68)).[18] In the light of this phonetic connection between Liverpool and Stoke, the additional finding of the presence of Liverpool P-DROP in Stoke is intriguing.

What, then, could be driving this connection? Leach (2018) suggests that there could be an explanation in terms of modern-day contact. People from Stoke frequently travel to bigger cities, such as Manchester, Birmingham and Liverpool for shopping and days out. This, however, does not, in my view, explain such a close connection between the varieties, particularly now with the inclusion of syntactic phenomena.

---

[18]As a side note, it has been reported (though the source is not clear) that the stew from which *Scouser* gets its namesake — *lobscouse* — is also the origin of the label *Lobby*, used to refer to the stew made in Stoke (Williams, 2008).

Here, I suggest that such a linguistic connection is better explained by the strong commercial and industrial interaction that existed between Stoke and Liverpool in the early 1800s. Stoke famously produced pots both for decorative and industrial purposes (Skidmore, 2009). At this time, the Liverpool dialect was forming (Honeybone, 2007), and there was direct trade, following the completion of the Trent and Mersey Canal in 1777, which was lobbied for by Josiah Wedgwood to ensure the safe passage of his pottery to Liverpool, and on to the global markets.

The map in Figure 4.50 shows the route of the Trent and Mersey canal through Staffordshire. Its construction predated railway connections between these places, and would have been the main route for trade.
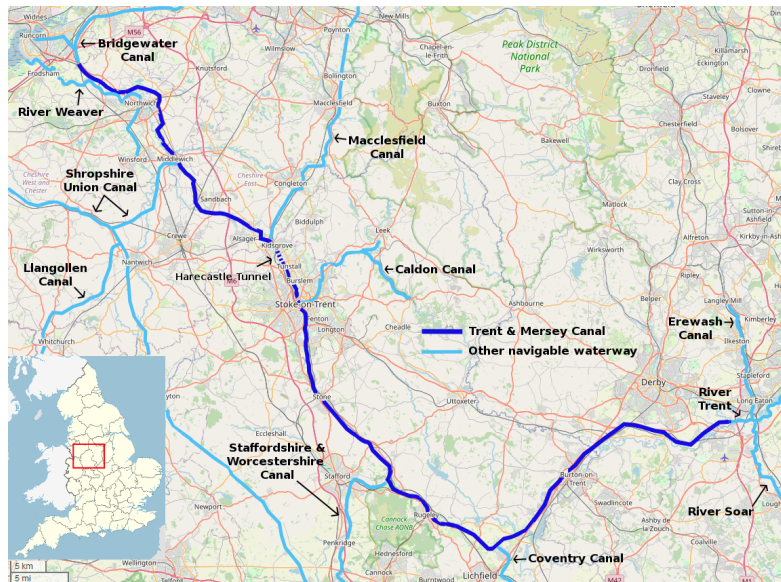
Figure 4.50: Route of the Trent and Mersey canal, approved in 1766, completed in 1777.
Source: https://en.wikipedia.org/wiki/Trent_and_Mersey_Canal

Compare this route to the pattern of P-DROP distribution that we see in Figure 4.51.

In the light of these industrial connections, the distribution pattern of P-DROP, and the close phonetic relationship between Liverpool and Stoke begin to make more sense. Furthermore, as shown in Figures 4.52 and 4.53, this era of industrial connectivity from Liverpool extends to locations along the Lancashire coast, North Wales and across the Irish Sea (Skidmore, 2009).

These facts may explain the otherwise anomalous high rates of P-DROP in Workington, 100 miles north of Liverpool, and obstructed over land by the
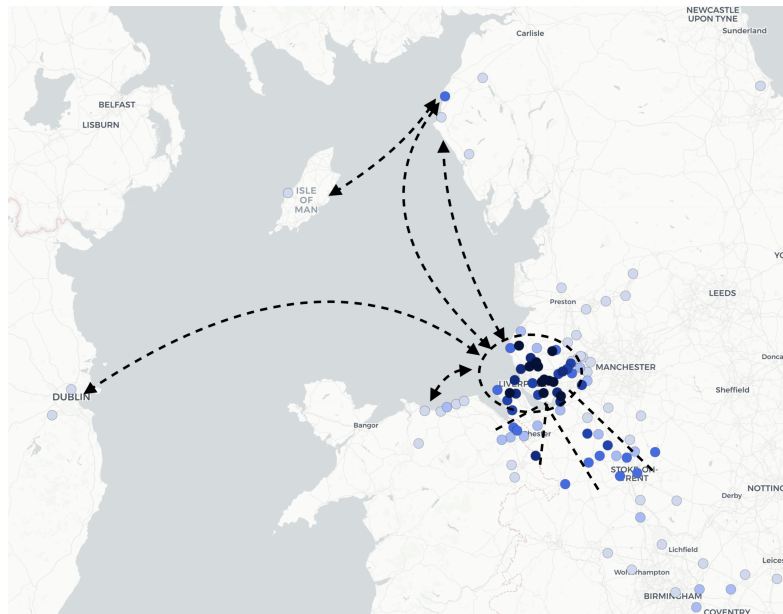
Figure 4.51: Detail from P-DROP single variant map, highlighting North West maritime and canal connections.
Shades represent 10% increments:
Lightest blue > 10%
Darkest Blue > 60%

Figure 4.52: "Pattern of trade in grain and foodstuffs on the North West Coast".
Source: Skidmore (2009, p.83)

mountainous region of the Lake District, as well as in Dublin. Workington, specifically, was a major ship-building port.[19]
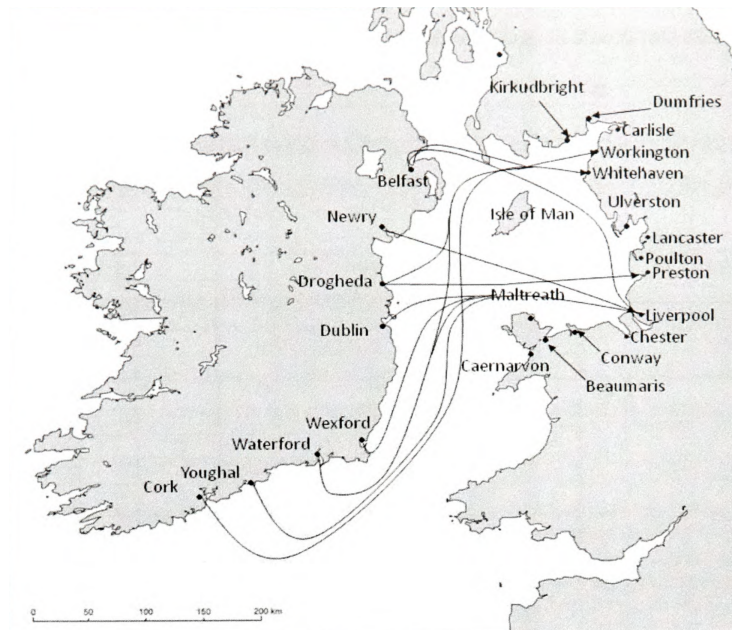
Figure 4.53: "Pattern of trade in grain and foodstuffs from Ireland".
Source: Skidmore (2009, p.83)

Skidmore's (2009) detailed account of the *Maritime economy of North West England in the later Eighteenth Century* shows how the maritime economy during this period was tightly integrated and inter-dependent. That we may be seeing the imprint of these historical trade routes in the language used in modern-day social media is fascinating. Further research into these connections and their linguistic consequences are surely warranted.

For now, we can at least say that, if the distribution of P-DROP to Stoke and Workington is most likely to have occurred during the time when these industrial connections were fully established, then the likely point of innovation of P-DROP would be at around the time that Liverpool English is thought to have first emerged in the early 19th century (Honeybone, 2007). This would substantially put back the likely point of innovation from that which has previously been suggested.[20]

---

[19]The presence of P-DROP in Workington was interrogated, and appears to be genuine: out of 61 instances of [P/D-DROP], 32% are P-DROP, and are produced by separate individuals.

[20]There is, in fact, no real consensus on the point of innovation for P-DROP, though it is generally referred to as a relatively recent phenomenon. Biggs (2018), for example, suggests, based on the fact that older speakers in her sample do not accept various P-DROP phenomena, that the

4.5.3 London and the Midlands

The map presented in Figure 4.54 shows the detail of PD-DROP distribution in in Greater London and the Midlands. The darkest red circles indicate locations where PD-DROP is dominant ($\approx 50\%$), while lighter shades represent slightly lower, but still substantial, rates of use. The geographical concentration of PD-DROP in East London—particularly areas such as Ilford, Barking, Redbridge, and Basildon—is striking. Similarly noteworthy is concentration in and around Leicester, where usage rates are higher than in nearby Birmingham.
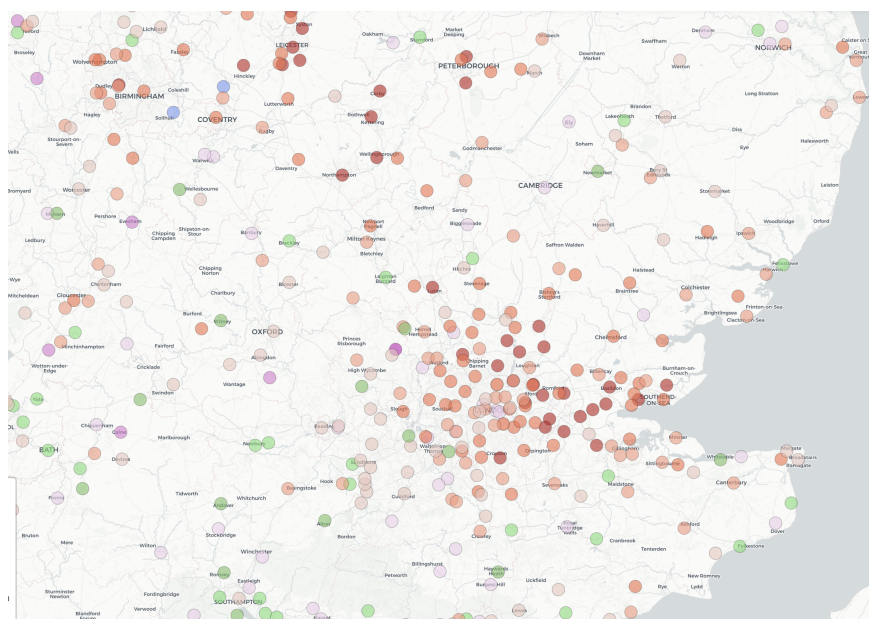
Figure 4.54: Detail of PD-DROP in London and the Midlands. Dark red = PD-DROP-dominant ($\approx 50\%$), Red = very high PD-DROP ($\approx 30\%$)

The pattern extends outward from London along key transport corridors into Kent and Essex, with evidence of a spread westward toward Watford and Croydon, though with less density. This spatial distribution seems to map quite neatly onto the zones most affected by counter-urbanisation, trends discussed in §1.3.2: it was to these areas that many original Cockney speakers relocated, following the decline of industry, rehousing projects and changing economic situation (Cheshire et al., 2011).

Bailey (2018b) suggests that PD-DROP diffused from London → Kent at some

---

grammatical innovation she describes (that of a null functional head that performs some of the functions of the over preposition), is recently innovated. It may be the case (in fact is likely the case) that P-DROP with verbs of motion (as plotted in the Twitter map) preceded such a grammatical innovation.

earlier point. Note here that the rates in Kent are high, but lower than East London and Essex, which again fits the notion that East London is a main hub of innovation in the South East. However, the alternative view, favoured by Bailey (2018b), that PD-DROP was separately innovated in multiple locations, remains quite plausible (Bailey suggests that MLE-PD-DROP and Kent-PD-DROP are independent phenomena). There are a number of factors that support a *multiple innovation hypothesis* (MIH). The main argument put forward by Bailey is on the grounds that we see equivalent phenomena occurring cross linguistically, independent of contact.

Table 4.8 provides the numerical data underlying the map presented in Figure 4.54, listing locations in which PD-DROP exceeds 50% of the total for all variants. Notably, all locations with above 50% PD-DROP use are situated within East London, Essex, Kent, or the East Midlands, reinforcing the idea that PD-DROP is embedded in a broader South East-Midlands dialect area. For instance, in Thurrock, P-DROP accounts for over 62% of all tokens, while in Blaby (Leicestershire), the figure stands at just over 60%.

Location	Region	PDD%	PD%	DD%	NOD%	Count
Thurrock	Greater London	62.37	0.72	1.97	34.95	558
Blaby	Leicestershire	60.55	0.92	1.22	37.31	327
Havering	Greater London	59.26	1.01	2.69	37.04	594
Barking	Greater London	55.67	1.64	3.12	39.57	609
Broxbourne	Greater London	54.29	0.48	4.52	40.71	420
Kettering	Northamptonshire	53.19	0.71	1.65	44.44	423
Gravesham	Kent	52.74	0.68	0.68	45.89	146
Corby	Northamptonshire	52.05	1.03	3.08	43.85	390
Oadby	Leicestershire	51.82	0.45	2.73	45.00	220
Charnwood	Leicestershire	51.47	0.93	2.79	44.81	645
Redbridge	Greater London	51.42	1.58	6.31	40.69	317

Table 4.8: Places where P-DROP is >50%, as percentage of combined counts of all variants, where the combined count is ≥100. All places are in East London, Essex and around Leicester in the East Midlands

Following a recent SE→NW diffusion model for PD-DROP (*go N*), which stimulated increased use in areas where it was already present, it is evident that the places with the highest rates are indeed to be found in London, Essex and the East Midlands. These are also the areas for which we have the oldest historical record.

Here, PD-DROP is often the majority variant, in some places approaching twice the rate of standard NO-DROP. The highest rates are in Essex (Thurrock is over 60%). This would set a likely innovation for PD-DROP in East London,

subsequently being displaced to Essex following Cheshire et al. (2011), who report the traditional working-class varieties of East London as being displaced to the east of the capital following the emergence of MLE. This does seem to suggest a proliferation in PD-DROP use prior to the influence of MLE. The same is true for the East Midlands, where usage rates are higher in Leicester, for which we know PD-DROP has deeper historical roots than the much more recent development of MBE: Blaby, Oadby and Charwood each exhibit PD-DROP rates north of 50%.

Despite these complications, a diffusion from London still seems likely to be playing a role in the more recent, wider adoption of PD-DROP, perhaps serving as a catalyst to accelerate an already active process. However, collapsing P-DROP (*go the N*) and PD-DROP (*go N*), in the way that Gopal et al. (2021) do, oversimplifies the picture: it is clear that P-DROP is structurally distinct from PD-DROP.[21] Given this structural distinction, we should at least consider a separate innovation for P-DROP in the Liverpool area and PD-DROP in the South East. I return to this point in 4.6.

4.5.4 The South West

The rate of PD-DROP in the South West might initially seem surprising, given its peripherality to the main thrust of the apparent diffusion from SE→NW. It may be that some of the variation here can be explained by external visitors: the South West is a popular holiday destination and additionally attracts many university students. However, external visitors alone are unlikely to explain the widespread use across the region. As we saw, Harris (1967) documented PD-DROP use in older male speakers (aged 79 and 84) in South Zeal, Devonshire in the 1960s. These speakers are reported to have spent their entire lives in the village, with little outside contact. Further, as we will see, the presence of PD-DROP in the South West is attested in traditional corpora FRED and BNC amongst older speakers. These facts suggest that it has deeper roots in the region.[22]

Table 4.9 shows remarkably high rates of PD-DROP use across the South West region, with Torridge nearing 50%. The map in Figure 4.55 shows that

---

[21] It should be noted that the majority of search strings used to form Gopal et al.'s (2021) corpus were cases where there would be no determiner in the standard form, such as "going (to) school", however, they did include "going (to) (the) pub", which, based on the frequency counts for pDrop-corp, would comprise a sizeable chunk of the data.

[22] As we saw, PD-DROP is also reported in the South West on eWAVE.

Location	Region	PDD%	PD%	DD%	NOD%	Count
Torridge	Devon	45.00	1.67	3.33	50.00	60
Plymouth	Cornwall	33.58	0.73	3.33	62.36	1501
Teignbridge	Devon	33.10	2.11	4.93	59.86	142
Barnstaple	Devon	24.56	0.00	5.26	70.18	57
Falmouth	Cornwall	17.46	6.35	6.35	69.84	63
Truro	Cornwall	16.92	4.62	1.54	76.92	65
Exmouth	Devon	15.19	1.27	2.53	81.03	79
Newquay	Cornwall	13.43	1.49	7.46	77.61	67

Table 4.9: Places in Devon and Cornwall, combined count is ≥ 50. There are high rates of PD-DROP in the South West.

these higher rates, represented by darker red points, are quite isolated, with a large area of lower rates of use extending beyond the city of Bristol.
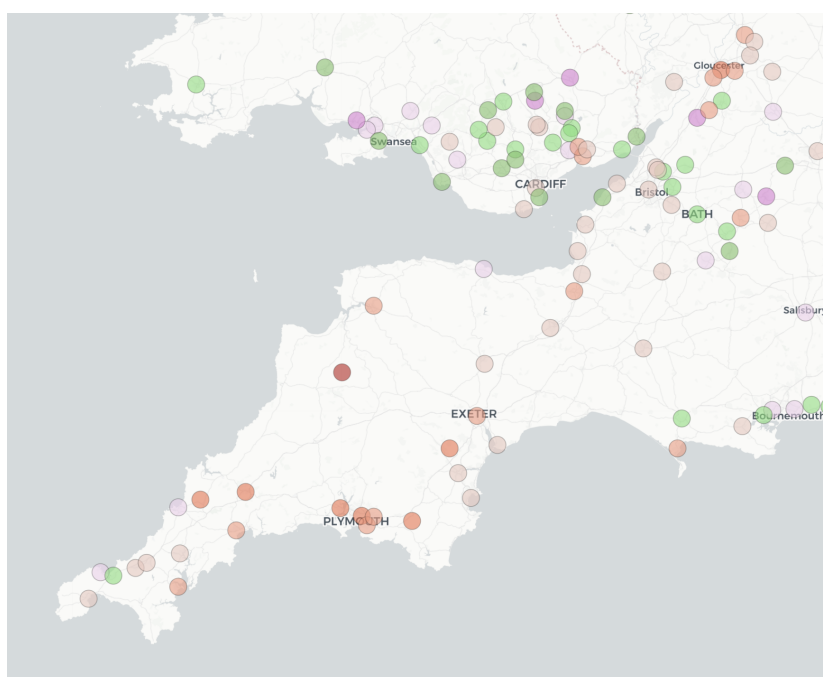
Figure 4.55: Rates of PD-DROP are high in Devon and Cornwall, but there appears to be a gap in the hypothesised spread from London, with substantially lower rates in Bristol.

This pattern suggests that PD-DROP is endemic to the South West region, supported by Harris's (1967) documentation of the feature, in older speakers, as discussed above, and appears to be a distinct phenomenon to the broader transregional dialect zone discussed for the South East and Midlands.

As mentioned, some of what we see in the South West may be explained by seasonal visitors coming from PD-DROP dominant areas in the South East and Midlands. According to Hutton and Murray (2025, p.21), the South West was

214

the most visited region in the UK in 2023 representing 17% of all overnight domestic stays, with a total of 16.4m stays.

To test whether there is a seasonal effect on the relative rates of PD-DROP in the South West, it is worth looking to see if there are notable spikes during the summer and winter months. Figure 4.56 shows relative [P/D-DROP] variant rates at three-month intervals. We can see that PD-DROP use declines steadily over the ten-year period, which mirrors what we see nationwide (recall Figure 4.3). It is also possible to discern spikes during certain times of the year, which may correspond to PD-DROP use by seasonal visitors. However, overall rates are high between 2012 and 2015, following the national pattern. This suggests that there are indeed relatively high rates of endemic use in the South West, outside of the effect of seasonal visitors.
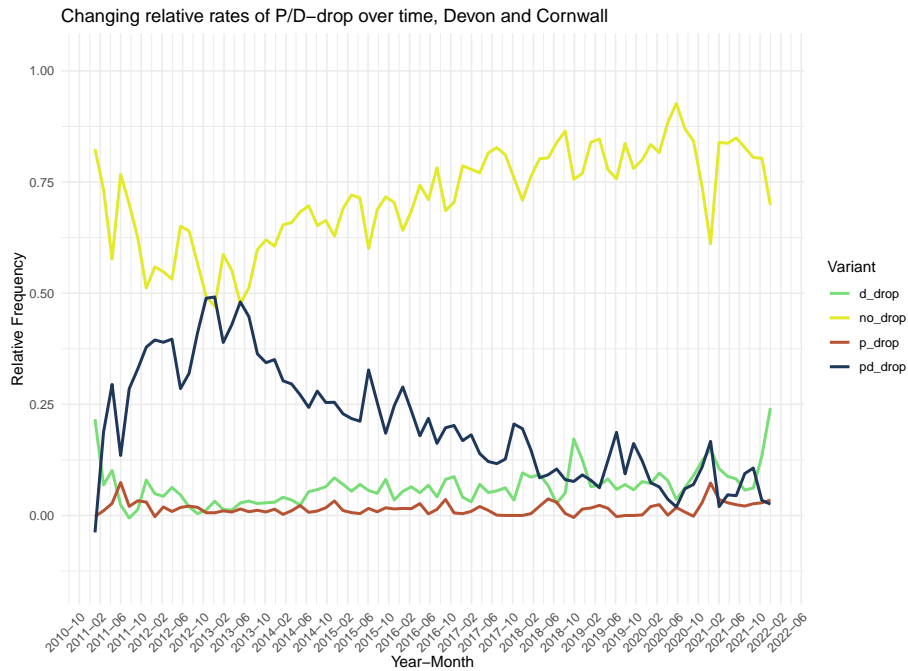
Figure 4.56: Relative rates of [P/D-DROP] over time, Devon and Cornwall
For legibility, dates on the x-axis are displayed at four-month intervals.

4.5.5 Ireland

Following a single-innovation hypothesis originating from the South-East of England, it is again difficult to account for the substantial rate of PD-DROP in the Republic of Ireland (table 4.10 and Figure 4.57).
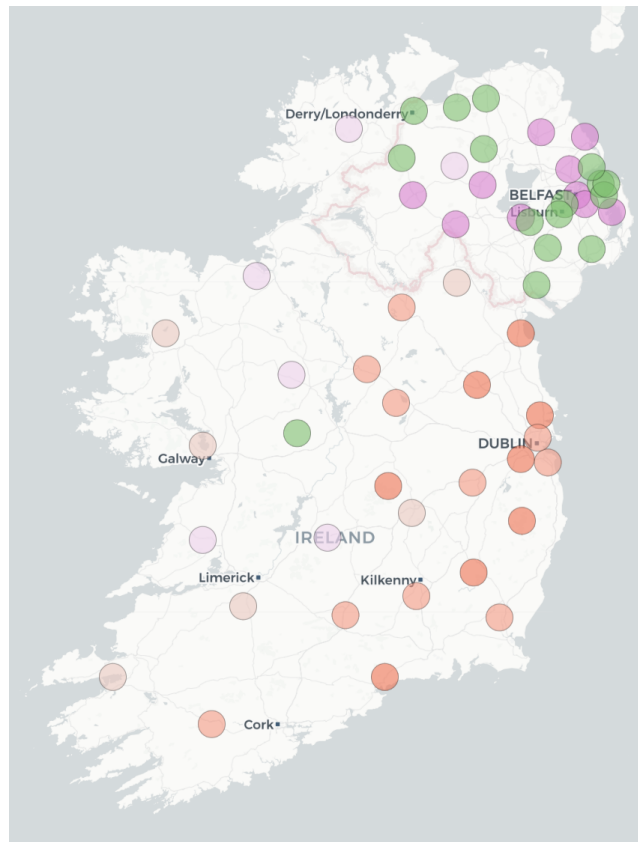
Figure 4.57: Preposition/determiner dropping cluster map, Ireland.
Red = PD-DROP (between ≈ 20%, and ≈ 30%)
Magenta = D-DROP (between ≈ 10% and ≈ 20%)
Green = NO-DROP (≈ 90%)

The distribution in the Republic of Ireland is not limited to urban centres on the east coast, as might be expected if it were recently adopted via diffusion from England, but is found in high rates across the eastern half of the country. This leaves either independent innovation, an earlier spread, at the time when Ireland was undergoing language shift to English, or, as mentioned, the effect of social media on the popularisation of the feature may itself be playing a role.

High PD-DROP rates in Ireland are surprising: it is not apparent on eWAVE, and appears limited in Gopal et al. (2021). Lack of attestation in Gopal et al.'s (2021) data may be attributed to the time frame of their dataset (2017-2019): as reported in §4.3, the nature of Twitter data changed markedly after 2015, resulting in a dramatic decrease in the rate of PD-DROP. In the present data, which charts 2011 to 2021, PD-DROP appears to be well-established.

The establishment of English in Ireland took place over some centuries.

216

Location	PDD%	PD%	DD%	NOD%	Count
Carlow, Ireland	38.4	1.0	4.5	56.1	198
Waterford, Ireland	30.5	0.9	7.0	61.6	583
Wicklow, Ireland	30.2	1.1	5.8	62.9	550
Offaly, Ireland	29.6	1.2	6.2	63.0	162
Meath, Ireland	28.0	3.8	5.0	63.2	636
Fingal, Ireland	26.0	9.5	6.2	58.3	1304
Wexford, Ireland	23.7	1.2	3.9	71.2	413
Kilkenny, Ireland	23.3	0.9	6.5	69.4	232
Cork, Ireland	22.8	1.4	5.6	70.3	2575
Kildare, Ireland	22.6	2.1	6.1	69.2	904

Table 4.10: Top ten locations in Ireland for P-DROP as percentage of count totals.

However, a final *shift* to English in Ireland was, by all accounts, sudden, with a particularly rapid increase in the early 20th Century (Chiosáin, 2006). It is not implausible that PD-DROP diffused to Ireland during this period, and has since been transmitted inter-generationally.[23]

It is notable that while Dublin is categorised in the PD-DROP cluster group (red), it is in fact an edge case: on previous runs, the cluster algorithm occasionally marked Dublin as P-DROP. The presence of P-DROP in, and around Dublin can be seen on the single-variant map in Figure 4.34. This likely reflects historical —and present-day—contact/migration with Liverpool. A further possibility is that Irish PD-DROP is actually connected with South-West English PD-DROP. It is established that the first incursions of English into Ireland was from South Western varieties. This, of course, is speculation, but it may be worth further investigation; recall the eWAVE attestations of the feature in Newfoundland, an area also largely settled from the South West.

4.5.6 Scotland and Northern Ireland

A detailed examination of the distribution in Scotland and Northern Ireland is beyond the scope of the present work. It is however informative that PD-DROP is almost entirely absent from Newcastle and north, in Scotland, and in Northern Ireland. In terms of a SE→NE diffusion narrative, we would have to say that such a process has not yet reached the North East.

It is also interesting that the distribution in Scotland mirrors what we found with [DATALT], showing a continuity between the North East, Scotland and Northern Ireland, reflecting their shared linguistic heritage, stemming, at its

---

[23]The fact that we also see parity between the Republic of Ireland and southern England with the pronominal ditransitive is suggestive, in this regard.
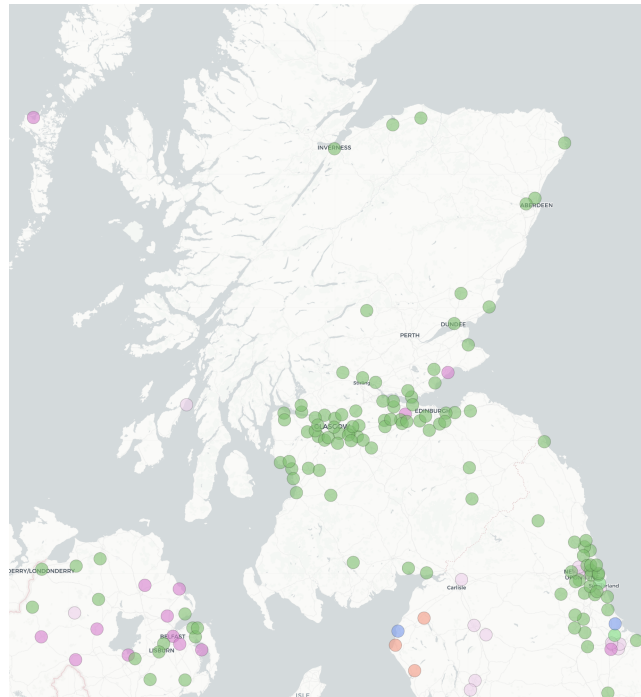
Figure 4.58: [P/D-DROP] cluster map, Scotland and North East England. The map is dominated by green points (NO-DROP-dominant), with a few places as light magenta (moderate D-DROP). Green = 90-10, NO-DROP-D-DROP
Light magenta = 80-20, NO-DROP-D-DROP

root from historical spread of Northumbrian Old English, creating a dialectal base that laid the foundation for both Northern English and Early Scots (Aitken 1985; Lass 1992).

The distribution that we see with both [DATALT] and [P/D-DROP] do seem to reflect the notion that the relationship between Scots and Northern English can be described as a dialect continuum, rather than a set of sharply bounded varieties (Wales, 2006). Meanwhile, more recent sociolinguistic work, paints a picture of a 'translocal' dialect network characterised by ongoing patterns of contact and mutual influence (Milroy & Milroy, 1985) rather than a set of historically connected dialects, which again, seems to be reflected in the distribution of [P/D-DROP] and [DATALT].

The continuity additionally extends into Northern Ireland. The fact that the pattern of distribution of both [DATALT] and [P/D-DROP] extends to Northern Ireland, and particularly that it does not extend into the Republic of Ireland gives some indication of the persistence of the features over time; the Plantation of Ulster, which brought tens of thousands of settlers from the Scottish

Lowlands into north-eastern Ireland, dates back to the 17th century. I discussed this in greater depth in relation to the distribution of ₚGTD and D-DROP. It is interesting to see that the same pattern is again visible here.

4.5.7 Summary

The superlocalisation of London English to the surrounding counties, likely brought with it the spread of PD-DROP to those areas in the period after the second world war.

The possibility that the PD-DROP (*go N*) found in the Midlands (Birmingham and Leicester) and Manchester, originally diffused from London via a SE-NW corridor, seems unlikely, given that the existence of the feature in the Midlands predates the period of counter-urbanisation. During the 19th century, the direction of travel was in into, rather than out of London, making it more likely that PD-DROP entered from the Midlands, into the dialect mix in the capital during this period.

However, a more recent re-kindling of the feature may have occurred, along the lines presented by Gopal et al. (2021). If this were the case, there would appear to be two lines of resistance against PD-DROP spread. One in West Yorkshire, including the city of Leeds, and the other in the North West, including Stoke-on-Trent. Here, we see a dominance of D-DROP (*go to N*) and P-DROP (*go the N*), respectively (magenta and blue in Figure 4.46).

Such a distribution does not, however, rule out the potential for the existence of multiple sites of independent innovation, and does not account for the presence of localised pressure towards P-DROP and D-DROP from pre-existing linguistic phenomena such as DAR/preposition reduction. The gravity-based model is additionally not a good fit for the high *relative* rates of PD-DROP outside of the SE-NW corridor, such as in the South West of England and Ireland.[24]

The distribution of P-DROP is particularly interesting: clearly dominant in Liverpool, but extending further into the interlying region towards Manchester. There also seems to be a corridor of P-DROP-dominant locations which extends South East from Liverpool, towards Birmingham, including Stoke-on-

---

[24]I should highlight an important distinction in the presentation of the data in this section. The focus is on the relative rates of each variant. The total number of occurrences is not shown, rather there is a cut-off of a minimum of 30 total matches for a location to be represented on the map. This is in contrast to the presentation in Gopal et al. (2021), who present overall rates as the size of a given point, and the relative rate as the shade-intensity. Focusing on relative rates over total rates was a methodological decision: above a certain number of matches, the critical question is how frequently a given variant occurs relative to how frequently it could have occurred.

Trent. This distribution seemed to align with historical trade routes, discussed in §4.5.2.

4.6 [P/D-DROP] and DAR

We now turn to an alternative possibility: that at least P-DROP and D-DROP —and to some extent, the Midlands and Northern variant of PD-DROP —emerged out of the development or reanalysis of various manifestations of DAR. This hypothesis is first motivated by a striking overlap between the spatial distribution of each [P/D-DROP] variant with the DAR areas marked out in the SED, as shown in Figure 4.59. If this is the case, it may be that underlying all of these variations was a grammatical innovation which loosened the requirement for prepositions in specific contexts, and this has since being playing out at different rates across different varieties. Alternatively, it may be that prepositions, in certain contexts, never had a particularly strong grip in the first place.

The extent to which Barry's (1972) isoglosses, based on SED data, align with the distribution of [P/D-DROP], in the North, is striking. The SED isoglosses trace, almost perfectly, the divisions between P-DROP in Merseyside, corresponding to DAR-θ, PD-DROP around Greater Manchester, corresponding to DAR-ʔ, and D-DROP in Yorkshire, to DAR-X. The alignment between DAR-θ SED and the P-DROP area is particularly notable, including Liverpool and its environs, with a corridor extending into Staffordshire towards Stoke-on-Trent.

In §4.5.2, I argued that the spread of P-DROP towards Stoke-on-Trent could be attributed to contact along the 18th and 19th century trade corridor between Stoke and Liverpool, with the likely innovation occurring in Liverpool. On the assumption that DAR-θ predates P-DROP, given a hypothesised transformation from DAR-θ→P-DROP, the SED and Twitter comparison suggests that, along with processes of diffusion and transmission of an innovation, areas along the diffusion path additionally possessed DAR-θ as an input.

A transformation from DAR-θ to P-DROP involves an additional step; the loss of the preposition, leaving VERB-θ-NOUN, and a subsequent reanalysis of [θ] to *the*. As discussed in the background to this topic, there need not be a direct transformational link between DAR and P-DROP, it may be that the presence of DAR-θ results in an increased likelihood for P-DROP to emerge via reanalysis.

However, it seems that Biggs's (2018) approach to the P-DROP found in Liverpool, where an inherently null functional element inherits some, but not all,
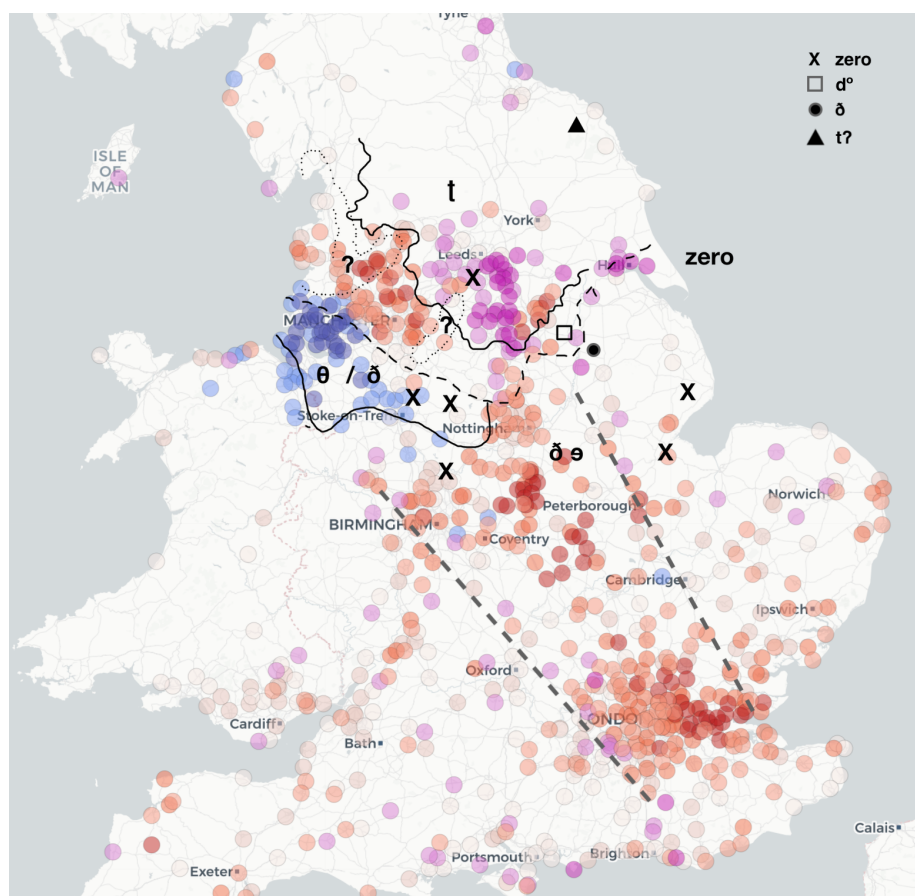
Figure 4.59: Preposition/determiner cluster map overlaid onto Barry's (1972) SED DAR map.
Blue = P-DROP dominant
Magenta = D-DROP dominant
Red = PD-DROP dominant

properties of overt prepositions offers a clearer potential structural continuity between DAR-θ and P-DROP. I explore this in the discussion chapter.

The correspondence between DAR-ʔ and PD-DROP in the Greater Manchester region is also noteworthy. We know that Manchester drew predominantly from local, rural populations during its population boom in the 19th century (Knowles, 1973). In is conceivable, in the context of the resulting dialect mix in the newly expanding city that this resulted in the transformation from DAR-ʔ→PD-DROP at that time. This would fit also fit the historical record for the early attestations of PD-DROP in Ellis (1889) in Cheshire in the 1880s. Such a transformation is, of course, speculative, but seems more likely than diffusion from the South East in the 19th century.

The division between PD-DROP and D-DROP, closely follows the isogloss

between DAR-ʔ and DAR-X. This line may be explained in part by topography. The Pennine hills run also follow this division, which would have formed something of a natural barrier before the establishment of modern transport infrastructure (discussed further in §4.8).

Finally, we see that the pocket of high-PD-DROP use in the East Midlands, around Leicester, sits within another DAR area. Here, the correspondence is not so clear, but we at least can say that some forms of DAR would have served as inputs to the urban varieties being established during the 19th century. The relative proximity to London, however, renders influence from the capital more tenable, particularly in the latter part of the 20th century.

4.6.1 [P/D-DROP] and DAR in traditional corpora

In this section I present data from FRED and the spoken portions of the BNC XML edition (a slightly revised version of the original 1994 release). Here I present indicative examples, however, I do not present a quantitative analysis. This is due to the fact that there are simply too few results, when broken down by location, to do so meaningfully. Additionally, location information for BNC is very broad, making direct comparisons difficult. The examples presented here represent most (in some cases *all*) of the results found in FRED. For BNC, there are additional data, presented in the Appendix.[25] It should also be noted that, while the geographic area covered by BNC and FRED is broad, it is not necessarily evenly balanced between regions. London, for example, is under-represented relative to its population size, while Kent is comparatively over-represented.

The data presented here are nevertheless informative as a historical record. This is particularly the case with the FRED data. The fact, for instance, that there are multiple attestations of PD-DROP (*go N*) in the South West by older speakers born around the turn of the 19th-20th century, is good evidence that the variant is both older and more widespread than has been suggested. This enriches the present-day Twitter data, helping us to make stronger claims about the distributions we see there.

---

[25]BNC 2014 may present the opportunity for a quantitative analysis, but is less useful from the standpoint of offering a historical record. I leave this to future work.

DAR-t in FRED

There are 39 instances of DAR-t when it follows the preposition *to* in FRED, following :

(41) a. I went to t' funeral (Prescot, Lancashire)

  b. See, I go to t' nephew on Sunday (Prescot, Lancashire)

  c. we used to run up to t' pawn shop (Prescot, Lancashire)

  d. their fields went right down to t' river (Ambleside, Westmorland)

  e. Of course she was going to t' cobblers (Middlesbrough, Yorkshire)

The majority of these examples are in the expected areas: Lancashire and Yorkshire. Examples such as these are likely predecessors to D-DROP, following assimilation of the preposition *to* with the reduced determiner *t'*.

(42) to t' → to

Examples of P-DROP in FRED

There were no recorded instances of DAR-θ in FRED, however, there are a number of examples of P-DROP. It is notable that attestations of P-DROP are geographically widespread. There are several in Lancashire, which is in the North West, but identified as broadly DAR-ʔ in the SED map (Figure 2.16). The presence of P-DROP in London and Kent is perhaps more puzzling, given the present-day localisation to the North West.

(43) Kent

  a. someone had to go the pub with a stone jar and get a gallon (Lydd, Kent, b.1895)

  b. go in there to tend them . You went the gypsies' caravan (Whitstable, Kent, b.1886)

  c. used to have to go the road and ask them to hold the train for an hour (Lydd, Kent, b.1895)

The presence of P-DROP in Kent may be explained by the substantial localised influence of Northern dialects in pockets of the South East resulting from the migration of coal miners as documented by Hornsby (2018). Kent's population exploded following the opening of the first deep coal mines in the

223

area, and tens of thousands of migrant workers were recruited to work from around the country, particularly from the Midlands and North. Hornsby reports some instances of DAR in his study, along with other features usually associated with northern varieties. He suggests the mass migration resulted in new dialect formation and the creation of a local *koiné*.

P-DROP in Lancashire is also quite common, which broadly fits with the present-day distribution.

(44) Lancashire

    a.   want to go back to farm service, so <u>I went the Criterion</u> and I served mi time as a cook (Barrow, Lancashire, b.1892)

    b.   to go in on the Strand, we used <u>to go the dance</u> (Barrow, Lancashire, b.1892)

    c.   And then we used to <u>go the backside of Ramson</u> (Preston, Lancashire, b.1908)

    d.   We used to have our fortunes, we used <u>to go the van</u>, they didn't need to come round (Crompton, Lancashire, b.1930)

There are two examples from the North East, in County Durham. This is interesting as there does appear to be a P-DROP foothold in Hartlepool in the Twitter data. It is not clear why this is the case, but it is perhaps a detail that is worth looking into in future work. Hartlepool itself was, of course, a major shipping port, which would have been in interaction with other major industrial ports in the 19th C. I return to the role of maritime trade and the distribution of P-DROP in §4.5.2.

(45) North East

    a.   do what you like after that. I used <u>to go the boozer</u> and have a couple of pints (Birtley, County Durham, b.)

    b.   and <u>she went the Empire</u>, <u>I went the Corn Exchange</u> but that's beside the point (age 46)

There is one example in Nottingham. This is not so surprising given the present distribution which extends from Liverpool through the Midlands.

(46) after doing the clipping wagons on <u>I went the other end</u> (Nottingham, b.1897)

D-DROP in FRED

Examples of D-DROP (*go to N*) are common in FRED, and found predominantly in Lancashire and the Midlands amongst older speakers born around the turn of the century.

(47)   a.   you could do a lot with that six pence. We used <u>to go to pictures</u>, used to <u>get in pictures</u> for three pence (Nottingham, b.1902)

       b.   I used to get sixpence every night off her for <u>to come to pictures</u> (Choppington, Northumberland, b.1890)

       c.   we used <u>to go to baths</u>, you see. In Cable Street? (Preston, Lancashire, b.1904)

       d.   course and then when we used <u>to go to pictures in afternoon</u> Saturday afternoons (Nottingham, b.1902)

PD-DROP

PD-DROP is perhaps unsurprisingly found in the South East:

(48)   South East

       a.   Oh no , no . Let me see now , I was four when <u>I went Church of England School</u> (Prescott, Lancashire b.1906)

       b.   I'm not! He said, <u>I'll go prison</u>! (Poplar, London, b.1924)

       c.   You had to work , <u>go night, night shift</u> for it (Faversham, Kent, b.1890)

Particularly revealing is that PD-DROP is also found quite frequently in the South West.

(49)   South West

       a.   this one had a ticket <u>to go pictures</u> and one had a ticket to go somewhere else (Cornwall, b.1904)

       b.   but if 't was any special service, <u>her'd come church</u> first (Totnes, Devon, b.1892)

       c.   after we let them go off, <u>they'd go stores</u> (Sunnyside, Sommerset, b.1894)

       d.   Well us used to be shoved off there Saturday afternoons and <u>go pictures</u> (Totnes, Devon, b.1910)

e. And I was in, <u>I went Bournemouth</u> for a holiday during the war (Buckfast, Devon, b.1913)

Given that current thinking on London PD-DROP is that it was quite recently innovated, perhaps, some *forms* as recently as the emergence of MLE (Bailey, 2018b), it is again informative that the examples in FRED are both geographically dispersed and attested in older speakers. There are several examples of PD-DROP in the South-West used alongside traditional dialect features such as the first-person plural nominative *us*.

P-DROP and PD-DROP in the BNC

Metadata are patchy in the BNC, sometimes the exact age of the participant is provided, other times it is a broad range. Location data are also limited, categorised as broadly South, Midlands and North. Some individual speakers are labeled according to their 'dialect'. It is also sometimes possible to narrow down the location to a degree from the context. Even so, the data do illustrate the point that P-DROP and PD-DROP are widespread in England, and have been for some time.

Recall the breakdown of BNC texts, reproduced in Table 4.11 below. The corpus is not regionally balanced, with the South representing 44% of the corpus, by word count.

	texts	w-units	%	s-units	%
Unknown	35	448458	4.30	27496	2.64
South	311	4687877	45.03	457726	44.09
Midlands	213	2492236	23.94	240306	23.14
North	349	2781280	26.71	312552	30.10

Table 4.11: Distribution of textual material by region
Source: `http://www.natcorp.ox.ac.uk/docs/URG.xml?ID=BNCdes`

The region identified as 'North' includes the North West (Manchester and Liverpool), up to the North East (Durham and Northumberland). The Midlands includes Lincolnshire across to Staffordshire and Shropshire.

The search criteria used were only for the verb *go* (past, present and continuous tenses) followed by (the) NOUN:

```
pdDrop: ("go"|"went"|"going") [pos="NN0"|pos="NN1"]
pDrop:  ("go"|"went"|"going") "the" [pos="NN0"|pos="NN1"]
```

All searches were carried out via on the BNC XML Edition (2007), via the *CQPweb* interface (`https://cqpweb.lancs.ac.uk/bncxmlweb/`), which is a slightly revised version of the original 1994 corpus. Results of the search query were manually inspected to retrieve only true examples of each structure.

As the corpus is not regionally balanced, attestations of P-DROP and PD-DROP are compared to the rate of NO-DROP in each region.

P-DROP

There were 22 examples of P-DROP in the BNC XML edition, spoken corpus (accessed via CQPweb). Again, as a benchmark, counts are compared to the rate of NO-DROP in each region.

Region	P-DROP	NO-DROP	% P-DROP
South	5	398	1.3%
Midlands	8	259	3.1%
North	9	249	3.6%

Table 4.12: Rate of P-DROP by region in the spoken portion of BNC XML edition (QCPweb)

(50) South: 5 examples

    a. Poor Keith had to go to this , he had to go the station every day on his bike (age N/A, recorded 1985-1993)

    b. The ones that smile, incidentally, usually go the bankruptcy court (age N/A, recorded 1985-1993)

    c. We went the bank the other day (SW, age 50, recorded 1992)

    d. Then that's going the bank (SW, age 20, recorded 1992)

    e. cos we asked you to go the cinema all the time (age 13, recorded 1985-1993)

Examples identified as Midlands are mostly located in Wales and the sub-region identified as North West Midlands (Shropshire). This fits with the distribution found in the Twitter data.

(51) Midlands: 8 examples (5 shown here)

    a. Yeah, but well you've got ta go the shop and tell her (Wales, age 25-34, recorded 1992)

    b. she'll have one. we were going we were going the house (Wales, age 28, recorded 1992)

c. could've have asked him. When did you go the shop? (NW Midlands, age 8, recorded 1992)

d. Well I know with two ladies th their daughters wanted to, them to go the house and they wouldn't. (NWM, age 53, recorded 1992)

e. They said, no we want to go the party. (NW Midlands, age 53, recorded 1992)

For the North, more precise location metadata are lacking for many examples. There are four examples for speakers who are identified as have a 'North East' dialect, though three of these are from the same speaker. It is nevertheless interesting to see that, again, there does seem to be a foothold for P-DROP in the North East.

(52) North: 9 examples (5 shown here)

a. if we wanted to go the match we had to pay (NE, age 53, recorded 1991)

b. how come he's good enough to go the Sea Cadets but he's not good enough to go college? (Central North, age 35-44, recorded 1992)

c. and she went the Empire, I went the Corn Exchange but that's beside the point (also in FRED) (NE, age 46, recorded 1992)

d. No never mind and you're going the doctor's (NE, age 46, recorded 1992)

e. I mean when we went the estate agents and solicitors (NE, age 46, recorded 1992)

PD-DROP

There was a total of 33 attestations of PD-DROP in the BNC XML spoken corpus (accessed via CQPweb). The results are presented in the table below.

Region	PD-DROP	NO-DROP	% PD-DROP
South	7	398	1.8%
Midlands	6	259	2.3%
North	20	249	8%

Table 4.13: PD-DROP attestations in BNC (QCPweb)

It is very notable that there are substantially higher rates of PD-DROP the 'North' region. Numbers are relatively low, however, and the regions covered

are broad. I present these data as indicative rather than comprehensive. Nevertheless, they are antithetical to the idea that PD-DROP is originally southern.

I present five examples from each region here, the full list of examples is available in the Appendix.

(53) **South**: 7 examples (5 shown here)

    a.   admit their part in the crime but they won't have to <u>go court</u>.

    b.   No, no! [unclear] every time <u>you go loo</u> (East Anglian, age 31, recorded 1991)

    c.   over her money situation? Well I, <u>I went guarantor</u> for that. (age 25-34, recorded 1991)

    d.   How much is it when you want <u>to go cinema</u>? (age 35-44, recorded 1991)

    e.   it on a Saturday they , they won't do their, <u>go football</u> or something (age 60＋, recorded 1991)

(54) **Midlands** 6 examples (5 shown here)

    a.   . that er eighty thirty five S: mm , oh <u>I'll go bed</u> then while you see (age 82, recorded 1994)

    b.   Go on what? <u>I went hospital</u> with Lynn (age 15-24, recorded 1992)

    c.   I'm going to Espania tomorrow, cos <u>I'm going holiday</u> to El Spania (age 15-24, recorded 1992)

    d.   Sure you don't want <u>to go badminton</u>, mm, ah? (age 0-14, recorded 1992)

    e.   I said to mum I think off when <u>we went bingo</u> and I was smoking (age 45-59, recorded 1992)

(55) **North** 20 examples (5 shown here)

    a.   we'd go down the station, <u>went pub</u> (North England, age 68, recorded 1994)

    b.   <u>Should go pub</u>. S: I can't afford it. S: Should save up (North England, age 37, recorded 1994)

    c.   got ta <u>go dentist</u> next week (North, age N/A, recorded 1992)

    d.   two in the morning before you go out and two before you <u>go bed</u> at night (North recorded 1985-1993).

    e.   So where is that <u>I'll go Doctor</u> [gap:name] for my hair (North, recorded 1985-1993)

Example (3a) is from an exchange between two older speakers in their late 60s, recorded in 1994, reference to old currency denominations allows us to get a sense of the historical period. We can also locate the speakers to Yorkshire.

PS1F1: we stay, when we we're in Yorkshire all

KCSPS001: Oh

PS1F1: the apprentices in, <u>in mill</u> where we were, we were all bound together

KCSPS001: Aye

PS1F1: we'd go down the station, <u>went pub</u>, there's er, I think were two and seven pence I'm not right sure

The first underlined phrase, *in mill*, is an example of DAR-Ø, with the determiner completely absent. This fits the SED distribution for DAR-Ø in parts of Yorkshire. It shows that speaker *PS1F1* is a DAR-Ø user. This is helpful when we look at their next utterance, where we find the phrase *went pub*, an example of PD-DROP. Seeing PD-DROP in the context of Yorkshire dialect, including DAR-Ø, frames it well outside of a hypothetical London-specific point of origin.

The next dialogue is more recent, discussing present life in 1994. It is coded vaguely as North England, but the use of *owt* is associated with Yorkshire and other Northern varieties. Again, the use of PD-DROP in this dialectal context sets it apart from the South-East.

PS1FC: Well you've got a right interesting life, you've got.

KCXPSUNK: I know.

PS1FC: Cars, home and done owt.

KCXPSUNK: Mm.

PS1FC: Is that it? Full stop?

KCXPSUNK: Just about.

PS1FC: : <u>Should go pub</u>.

KCXPSUNK: : I can't afford it.

4.6.2 [P/D-DROP] and DAR in the Twitter corpora

D-DROP and DAR-t/DAR-Ø

The clearest link between DAR and [P/D-DROP] is with the D-DROP variant ('going to pub'), which aligns spatially with DAR-t and DAR-Ø as charted in

the SED (see again Figure 4.59). Phonologically, of the [P/D-DROP] variants, D-DROP is closest to its DAR counterpart.

(56) a. going to t' pub (DAR-t)
      b. going to Ø pub (DAR-Ø)

Here, D-DROP shows a direct correspondence to DAR-Ø (56b) and given that they are surface-identical, linking the two is straightforward. It additionally seems relatively unproblematic that D-DROP is at least an orthographic presentation of DAR-t, or else the result of processes of assimilation between the preposition *to* and the reduced article *t'* in actual speech practice. Given that the preposition in DAR is also frequently reduced, rendering *t' t' NOUN*, assimilation to *to* or *t'* appears likely. It is notable that a sizeable portion of DAR examples on Twitter are self-conscious parodies or metacommentary on regional practice.

(57) DAR-t in parody and metacommentary

      a. Busy af day, taught all morning then had <u>to go t'Yorkshire</u> this afternoon and now water polo. I'm frigged!

      b. see also: 'tek yer coat off inside or you won't feel't benefit' and 'eee well al <u>go t't foot of our stairs</u>' aka my grandma's most used sayings

      c. @LexxClarke true, I don't want to get airs and graces I might have to take t'coal out of t'bath!

      d. @spikeygo In Leigh is is definitely <u>going t'asda</u> (Leigh)

      e. Waiting for a Nandy "I spoke to a lady in my constituency who said (strong northern accent) "Eye up, I don't think it's reet tha them rich folk wi automothingies get to go t'ospital and not pay for parking. A kick in teeth for folk like me" (normal accent) which I listened to."

Parody and metacommentary are indicative of the social salience of DAR, and the particular social associations it indexes. Hollmann and Siewierska (2011) suggest that DAR is to some extent performed as an *act of identity*, following Le Page and Tabouret-Keller (1985). These associations are informative as they may provide social motivation for a shift both to, but in the case of P-DROP, *away from* DAR, particularly in the context of urban speakers who wish

231

to differentiate themselves from perceived rural speech.[26]

However, DAR-t is also used without apparent parody:

(58) DAR-t without parody

    a.   thought she was <u>going t nandos for a valentines meal</u> (Knowsley)

    b.   Have you got any signed copies for sale? Chloe <u>going t'Egypt</u> on Sunday would make a great holiday read for her. (St. Helens)

    c.   Are we sure it was wandering? It may have been going t'shops or for its covid jab. ☺ (Manchester)

    d.   hope you have fun are you going t'game? (St. Helens)

    e.   shocking when you can't pop t'local without an earful from the wife (Warrington)

I report these examples as *without parody*, however, this is a qualitative judgement. They are still clearly self-conscious representations of speech practice, a phenomenon frequently reported in social media exchanges. There is a sense that users are self-transcribing how they would speak (Eisenstein, 2013).

Ditransitives

DAR is also found with ditransitives such as:

(59) my mum <u>sent me t shop</u> to get some food, i come back with a christmas tree! haha she told me to put it up why shes out,but how do i do lights

The use of DAR with ditransitive *send* in this way bears a notable resemblance to the P-DROP phenomena reported in the North West:

(60) me nan sent me the shop

This parallel is made more apparent in the next section, with attestations of DAR-θ/DAR-ð.

P-DROP and DAR-ð/DAR-θ

DAR-ð/DAR-θ, orthographically realised as DAR-th, presents an issue that some examples may simply be typos - with the final *e* missing. There are some examples, however, where *th* is repeated consistently within the same tweet. I

---

[26]Recall the discussion on *wools* and *plastic scousers* on page 38.

present here multiple tweets from single authors in order to show the consistency with which they employ this kind of self-stylised representation of speech.

(61) author_ID: 778171800 (Chorley)

    a. Its not just about pretty passin football.Its about courage, <u>taking th game to th opposition</u> & scrappin4your lives.

    b. i disagree. Th reason Gfh cant make money@ #lufc is cos kb took th money out of th club & is still paid as part of th hidden deal.

    c. why didntU explain th stadium sketch to th fans?Becoming v hard to trust anything u say.Always implied but never confirmed #lufc

(62) author_ID: 22832667 (St. Helens)

    a. oh ye true lets <u>go th whole hog</u> n <u>go Thailand</u> or hong kong.guna say Sydney then but ive been. so u fancy South Africa then.?

    b. ha u funny.<u>i luv th vanessa</u> show shes fab.n ben had bit banter wiv him on ere.hes cool guy,he folows me.ur my fav couple tho x

    c. oh what a mad day had a man <u>at th bus stop</u> ask me what r u doin up north. i said i llive here. ha think he thought i ws someone else.WEIRD

(63) author_ID: 2781489293 (Runcorn)

    a. I concede he must hv been consulted before <u>th bid went in</u>. But he could hv been testing <u>th water himself</u>. If hes genuinely wanted at Goodison there is no better place..

    b. Absolutely.What we spent 12 months was only because we knew 90 mill was coming in for Lukaku. Less than 5 weeks <u>t the END of th transfer window</u>. We are going to end up with no one of any substance nd stuck with Koemans misfits..

    c. EFC turn down bid from Leipzig for Lookman. Thank God for that. He cut it <u>in th Bundesliga</u> on loan...which is precisely why he went. Hes ready now.

P-DROP: th'GOAL

Most revealing, in terms of a connection between DAR and P-DROP, are examples where the *th* is used in the [P/D-DROP] paradigm, with the verb *go/come*

233

and *th'*GOAL, shown in (64).

(64) Examples of P-DROP: th'GOAL

    a. cheeky git, but thanks pal. Back in England in 2 days will have to go th'pub soon (Wigan)

    b. I need go th'opticians or get a labrador. (Wigan)

    c. it was horrific! Nothing in tho n ceebs going th'asda (Wigan)

    d. Just nipping th'Asda in my new coat (Warrington)

    e. Disaster averted! 17YO came home & I nipped th'offie (St. Helens)

    f. M + S, Sainsburys and Tesco only allow 1 person in per "family" except in very special circumstances. Sedate andvsatisfying shopping. Another reason I dont go th'asda

Evidence of grammaticalisation

I now explore the possibility that P-DROP may have emerged from DAR-θ/ð, via a process of erosion of the preposition and subsequent reanalysis of *th'* to *the*, as in:

(65) [t' th'] → [th'] → [the]

We might expect *the* to have appropriated some of the grammatical function of the preposition. If this were the case, then we might expect to find examples where reanalysed *the* is able to function as a preposition. This could explain the phenomena in the North West where speakers use *the* where we would ordinarily find *to*, as in:

(66) a. I went the Tesco before and bought a tub of ice-cream, I'm now eating fish with salad cream on it

    b. when ye go the tesco what meal deals r ye chattin?

    c. it gave him a new lease of life for a bit, he even went the Tesco alone

    d. Think I'll just nip the Asda instead 😂

It is notable that this phenomenon is frequently the subject of metacommentary:

(67) a. Ermmmmm scooooooose me, I live in Whiston, originally from

Kenny and I'm no Wool. My postcode is still an L and I still <u>go THE asda</u> 😂😂😂

b. Scousers 'going the Asda' is definitely a thing.

> Replying to @northnorth_west
>
> Lebanon is another country that often gets a definite article. I might use it for, say, The Scotland from now on...
> "You having a weekend away? Lovely, where you going?"
> "The Wales"

Figure 4.60: Tweet excerpt commenting on the phenomenon: *going the Wales*

It is also possible to have *the* preceding a place name in where we would ordinarily find *to*, as in:

(68) a. The @CurveFashionFes is coming the Liverpool this Autumn
b. If I'm ever dumb enough to <u>go the pool</u> ('pool' here is a common shortening of 'Liverpool')
c. Wish u were <u>coming the Manchester uk</u> 🇬🇧 again or Liverpool
d. she's tryna get ya to <u>go the Amsterdam</u> tomorrow

Hall (2019) finds evidence, in London PD-DROP, for the frequently-discussed claim that proper nouns should be analysed as "complex definite descriptions, which enter the derivation as NPs"(p.14), for examples such as "going Margate". If such nouns already possess a DP, it is curious that we should find them fronted by *the*.

Irish influence

Interestingly, there may be a role here for the influence of Irish. Irish combines the definite article with the preposition. Irish influence on the development of Liverpool English, directly or indirectly may be playing a role in the use of the structures shown above.

(69) Beidh mé ag dul   go dtí   an   Fhrainc sa tsamhradh
     I   will be going to [the] France in   the summer.

(70) Itheann daoine pasta san Iodail
     People eat   pasta in [the] Italy.

Filppula (1999) outlines a number of cases where Irish substratum influence on Irish English brought about the 'overuse of articles'. Of particular interest here is variation found in the use of articles with "the names of geographical areas and localities, public institutions, buildings and monuments" (Filppula, 1999, p.60).

(71) From Filppula (1999, p.60):

    a. Well, it's this side of Roundwood, *the Sally Gap*.

    b. you would get people that'd give you a good deal of the lowdown of *the County Wicklow*.

    c. Until such time as *the Belvedere College* moved in.

This connection is speculative, however, but certainly worth considering in future work.

PD-DROP and DAR-ʔ

A factor motivating the loss of the preposition in DAR would also explain the picture in the area around Manchester which is mapped in the SED to DAR-ʔ, and which corresponds to PD-DROP (*going pub*) in pDrop-corp. Here, where the article was already reduced to a glottal-stop in traditional regional speech, losing the preposition (via assimilation) results in VERB-ʔ-NOUN which is very close to PD-DROP, and likely identical in its written form. If this were the case — that PD-DROP in Manchester actually involves a trace, previously glottalised, preposition-determiner — we might expect to see some structural distinction between Manchester PD-DROP and London PD-DROP where both the preposition and determiner appear to be structurally absent (Hall, 2019) (though recall, Bailey's (2018b) analysis of Kent PD-DROP *does* involve silent preposition/determiner rather than full incorporation to the verb).

Initial evidence that PD-DROP is descended from regional dialect can also be found in dialect writing, where the author imitates features that they associate with traditional dialect such as the respelling of *gotta→gorra* (a), *fut* (b), and the dropping of the determiner in the PP adjunct *in morning*, in (c)

(72)   a. Gorra... As in, I've <u>gorra nip shop</u> dya wonnote?

      b. In an attempt to feel young and fit in with the cool young folkers I have observed over the last few weeks, I'm now <u>going fut nip shop</u>

wearing shorts.[27]

    c.   Reported to police <u>going station in morning</u> but I have it on CCTV

Use of *nip* with PD-DROP is interesting, as it contravenes what is thought to be permissible with PD-DROP as described in the South-East and appears to be limited to the Midlands and North West, though it is not possible to verify this without a full, unrestricted, corpus for the UK. It is worth checking if it is found in unrestricted Twitter corpora that cover the South-East.

Variation by GOAL NOUN

If PD-DROP in the North is related to DAR, we might expect that it would be available with a wider range of GOAL NOUNS than that found in the South East. The evidence here is mixed. There is robust quantifiable evidence that P-DROP and D-DROP are available with a range of nouns that are found rarely in the South East. With PD-DROP, however, the overall trend in the North and Midlands appears to mirror the South East across the range of GOAL nouns tested.

This said, there are examples in the data for PD-DROP in the Midlands and North which are not permitted in the South East, and which, as a result, suggest a distinct underlying structure; namely the remnants of functional elements which we might expect to see if they were recently elided following the completion of reductive processes.

4.7 [P/D-DROP] attested in North West, but disallowed in South East

With the influence of DAR on PD-DROP in the Midlands and North in mind, we now turn to look in greater detail at examples of PD-DROP which have been found to be illicit for speakers in the South East (Hall, 2019), but that are attested in NWatlas-corp.[28]

The following examples show that, at least for some cases, surface PD-DROP is not of the sort found in the South East. However, such examples do not entail that South Eastern PD-DROP is not present in the North. The point is that there

---

[27]The use of *fut* here, is informative, an perhaps a reduction of *for to*, though this is not clear.

[28]Examples are attested with enough frequency to indicate that they are licit, however, the specific nature of each phenomenon means that there are too few examples for a quantitative analysis.

is a surface-identical structure in the North, likely descended from DAR and which likely pre-existed the influence of southern PD-DROP.

PD-DROP in the South East is reported to exhibit a 'radical absence' of preposition and determiner (Hall, 2019) which has various consequences for the grammar. These are outlined in the following sections.

PD-DROP with relative clauses

Due to the absence of the determiner, South East PD-DROP lacks a definite reading, rendering examples with a relative clause which describes a specific noun such as (73) as ungrammatical (Hall, 2019). Such instances are not common in NWatlas-corp, but there are some attestations:

(73) a. Need <u>to go bank which is legit about 5 minutes away from my house</u> but I'm so lazy (Coventry, pDrop-corp)

    b. Taxi drivers are dicks trying to get me to leave my phone with him while <u>I went bank which was next to the car</u> (Tameside, Greater Manchester, pDrop-corp)

    c. Just ordered me & Twig matching Christmas bow & face mask, y'know cos we can <u>go places that we can wear them</u> at Christmas? (Liverpool)

The examples suggest that there is room, at least, for the presence of a null functional head which licenses a specific NOUN which may be referenced in a relative clause. These data run counter to what has been found in London PD-DROP where such constructions are illicit.

P-DROP and PD-DROP with verbs other that *come* and *go*

It is reported (Bailey, 2018b; Hall, 2019), that PD-DROP is usually only permissible with *come* and *go* for most speakers, with some other verbs permitted where they are semantically weak. Myler (2013) finds P-DROP to be licit in Ormskirk with greater range of verbs than PD-DROP in the South East, including verbs *nip, pop, jog and drive*. Biggs (2018) finds an even wider range of verbs available with P-DROP in Liverpool, including *manner of motion* verbs such as *amble, cartwheel*.

Accordingly, we do find P-DROP in NWatlas-corp with a greater range of verbs than has been reported with South East PD-DROP.

(74)   a.   Am I able to <u>pop the shop</u> tomorrow for pjs or do I have to click and collect? Is there x large or xx large available?

       b.   Can't be bothered cooking, or <u>walking the chippy</u> orrrr waiting for my brother to go for me..DILEMMA

       c.   youre not <u>nipping the bank</u> to extend your overdraft. Life is good.

       d.   <u>Driving the shop</u> as I'm too lazy to make anything

       e.   <u>Walking the bus stop</u>, I hope the bus doesn't take long!

       f.   oh fab, will <u>take a trip the drs</u>, thank you xx (Garston)

We have already seen that PD-DROP is possible with *nip*. Along with *nip* and *pop*, we also find *walk* and *drive*, in NWatlas-corp.

(75)   a.   @jonnycheetham5 shit that looks wel nice think i might have <u>nip shop</u> for a few haha

       b.   Wish you could just <u>nip shop</u> and buy a toned body. Sod exercise

       c.   <u>Walking shops</u> with mel and aim x

       d.   Why dose it rain when I start <u>walking shop</u>, so cold

Ditransitive *take*

We have seen that P-DROP is frequent with ditransitive *take*, where it appears to follow a similar schema to ditransitive *bring*. It is also found with some frequency in north-western PD-DROP, which is not reported for SE-PD-DROP.

(76)   a.   Was going <u>take him emergency vets</u> this afternoon

       b.   brilliant any chance phil will <u>take us station</u>

       c.   want Sunday cos I won't <u>take them town</u> on a fri

       d.   May have to beg my dad <u>to take me bookies</u>

       e.   my old boy <u>took me races</u> and encouraged gambling

This is, again, further evidence of north-western PD-DROP being structurally distinct from that found in the South East.

Pluralised GOAL nouns

As seen in the last two examples above ((76d) and (76e)), and those in (77) and (78) below, pluralised nouns are common in the North West, both with P-DROP and PD-DROP. This runs counter to previous reports for both structures. A critical aspect of the 'radical absence' of preposition and determiner (Hall,

2019) in London PD-DROP is that it predicts the GOAL noun to be necessarily singular, disallowing the structures in (78).

Whilst unreported, pluralised GOAL nouns are not necessarily expected with P-DROP given Biggs's (2018) analysis that P-DROP involves a covert prepositional element and obligatory determiner.

(77) a. Get onto the Peterborough chairman Phil, if ppg comes in it's gonna go the courts (Birkenhead)
　　 b. a lot of people don't have the money to go the games (Liverpool)
　　 c. I really hope i get to go the parks one day!
　　 d. If people are stupid enough to still go the pubs it's their own fault
　　 e. Kudos to the lunatics that keep going the aways

If North West PD-DROP also retains some silent structure, we should expect such examples to be attested, which we do, and with some frequency.[29]

(78) PD-DROP with pluralised GOAL nouns. London judgements reported in Hall (2019)
　　 a. people who go food banks (*London, attested NW)
　　 b. its orange wednesday you have to go cinemas (*London, attested NW)
　　 c. Ppl shouldn't be going pubs & restaurants (*London, attested NW)
　　 d. all these [people] going parks and beaches (*London, attested NW)

It is interesting that a number of cases involve discussion related to the pandemic. This does give them an unusual framing, and it is conceivable that the use of the structure was stretched in these circumstances. This said, the fact that it is felicitous to include pluralised GOALs in the structure, regardless of the context, still weighs against a structural interpretation that involves pseudo-incorporation, contra Hall's (2019) analysis of the phenomena in London.

It is worth noting that there was one example of PD-DROP with pluralised GOAL nouns found in pDrop-corp, in the London borough of Enfield (79).

(79) Hate going doctors and hospitals 😩 (Enfield)

---

[29]There are many examples with *cinemas* as the GOAL noun. I have included one example (78b). However, I suspect that this actually refers to a familiar place, rather than being a true pluralised form.

Whilst this is only one example, it nevertheless indicates that plural GOAL nouns may be licit in the South East, for some speakers, contrary to Hall's (2019) findings.

Adjectival modification

London PD-DROP, *obligatory adjacency*, which follows if the noun is *pseudo-incorporated*, disallows intervening material between the verb and the NP, such as adjectival modification of the GOAL noun, but again, these are attested in the North West (80).

(80) PD-DROP with modified GOAL nouns. London judgements reported in Hall (2019)

    a.   I wanna go new places kyle. I want to see the world (*London, attested NW)

    b.   yes we can go new pub for lunch the glass horse (*London, attested NW)

    c.   going to be mental if all 4 coaches go same pub at stop off (*London, attested NW)

    d.   Mums just rang me to say we're going big tesco (*London, attested NW)

    e.   Gutted cant go big reunion tomorrow now at least X factor will be on (*London, attested NW)

    f.   even if that means going different shops

Whilst adjectival modification like this is not very frequent, (there are 8 attestations total), there are nevertheless enough examples to indicate that it is productive. Again, without an open Twitter dataset for the South East, it is not possible to verify, at this stage, if there may be such examples there too, though the current understanding is that it is ruled out there (Bailey, 2018b; Hall, 2019).

***Straight/Right* modification**

Due to the hypothesised absence of prepositional structure in London and Kent, PD-DROP is disallowed with *straight* modification (SM) as in (81) (Bailey, 2018b; Hall, 2019). This is perhaps the strongest evidence for the existence of a NULL functional head in both North West P-DROP and PD-DROP.

(81)   They went straight pub (*London/Kent)

Biggs (2018) finds that Liverpool SM with P-DROP is accepted. Accordingly, SM is well-attested in the NWatlas-corp data (82).[30] The availability of SM is one of the facts that points to the existence of the functional element x in Biggs' analysis which is able to select for the GOAL noun, and license case.

There were 27 cases of SM with P-DROP in NWatlas-corp.

(82)   a.   presentations on Monday we will go straight the library and start robs xx

       b.   Can't wait to finish Tuesday then just go straight the elephant for a few pints with work

       c.   When you go straight the airport from a night out..😂😫

       d.   Remember watching this goal and going straight the field next day to recreate the strike

       e.   literally dont know what a healthy level of angst is to introduce? i go straight the major character death and its like. calm down.

It is notable here that the structure is also permissible with a complex abstract noun like 'the major character death', and not just nouns denoting familiar or anaphoric places.

There is one apparent example of *right*-modification (RM) with P-DROP:

(83)   another with a massive future...he will go right the top too! Only 19....frightening! (Liverpool)

This is unexpected, given the unacceptability of RM for Biggs' respondents. However, with only one example, it is hard to draw firm conclusions.

SM distinguishes Northwest P-DROP from Southeast PD-DROP (Biggs, 2018; Hall, 2019), and from the variety described by Myler in Ormskirk. Revealingly, in the Northwest, SM (but no RM) is also possible with PD-DROP and is as frequent as with P-DROP in NWatlas-corp, with 31 occurrences. A number of examples are presented in (84).

---

[30]Interestingly, Myler (2013) finds SM to be illicit in Ormskirk with P-DROP. However, he notes that this is based on only two native speaker judgements.

As we will see, the survey presented in the following chapter finds this structure to be well accepted for most respondents across the region, including Ormskirk. It is possible that this reflects an ongoing change towards the variant identified by Biggs (2018) in Liverpool.

(84) a. <u>I'll go straight bookies</u> mint odds them

     b. Finishing work and <u>going straight A&E</u>

     c. Finishing work at 7 and <u>going straight rugby</u> is the biggest ball ache ever

     d. pack your case we're <u>going straight Skegness</u> after ☼⛵🏝

     e. Can't keep finishing work and <u>coming straight town</u>, mess

This is further indication that the Northwest variety of PD-DROP is distinct from that found in the South East, and does involve some silent material that can be modified by an adverbial like *straight*. We have seen that *straight*-modification is seen as a robust measure of (silent) PP-hood, and these data provide resultantly strong evidence that there is a NULL PP present in Northern PD-DROP.

These data additionally complicate the interpretation that PD-DROP in the North West diffused from London. The fact that Northern PD-DROP is structurally distinct would suggest that it is independently innovated, rather than diffused.

Prepositions other than *to* or *at*

Preposition drop is typically only found with prepositions *to* or *at*. Biggs (2018) finds this to be the case for her Liverpool survey respondents. Dropping of *at* is supported in NWatlas-corp:

(85) a. ok babe <u>I'm staying his</u> all weekend

     b. <u>staying my nans</u> tonight, another dinner today for me yay.

     c. <u>Dan stayed his friends</u> last night

However, there are attested examples in NWatlas-corp which appear to show dropping of other prepositions such as *for, on* and *in*. The availability of the structure with a greater range of prepositions is compatible with the DAR-derived analysis. As we have seen, DAR is available across a wide range of prepositions (Hollmann & Siewierska, 2011), not just *to* and *at*.

(86) a. Labour, <u>they arrived the scene</u> with insincerity & lies. This is not breaking news [ON/AT]

     b. <u>Picture of Jesus arrived the post</u>. Looks like a 1970s singer songwriter [IN]

     c. Best things about being home is <u>going hill walks</u> and then off to

> ibrox at night [ON]
>
> d. Ok <u>waiting the last people</u> and we are ready to start this [FOR]
>
> e. me patiently <u>waiting the london shows</u> [FOR]
>
> f. <u>I'm still patiently waiting the letter</u> for permission to buy consumables from Roche [FOR]

It is not certain, however, from the examples with *waiting*, if we are truly looking at P-DROP, or if we are seeing a re/misspelling of *awaiting*.

P-DROP with *Put it*

Where *put* is a ditransitive and the THEME is pronoun *it*, it is possible to have the THEME-GOAL order. If this were analysed as P-DROP, the preposition *in* would be elided.

(87) a. They didn't. <u>I put it the correct bin</u>.

b. There <u>putting it the right places</u> but no bins to put it in

c. I can't I'm not allowed <u>to leave it the house</u> until tommorow

d. Sellotape the blade up and <u>put it the recycle bin</u>, same as a scalpel blade.

e. And the best thing about <u>putting it the oven</u>...IT DOESNT SHRINK

The availability of P-DROP with prepositions other than *to* or *at* runs counter to Biggs's (2018) finding in Liverpool. The data make sense, however, if we consider P-DROP as a descendent of DAR-θ, which, is frequent across a wider range of prepositions (Hollmann & Siewierska, 2011) (see Figure 2.17 on page 94). Additionally, at least the dropping of *for* and *in* are attested with PD-DROP in NWatlas-corp.

(88) a. I'm currently <u>waiting update</u> from my Cardiologist

b. Nothing on clearing so <u>she has to wait year!</u>

c. No power and <u>waiting phone calls</u> not good.

d. <u>Can't wait tomorrow!</u> ✐

e. temptation of <u>staying bed</u> and not going to my 9am is strong

f. stuck between wanting to go town and menace, and <u>staying bed</u>

These facts again offer some degree of differentiation between NW-PD-DROP and SE-PD-DROP, and suggest that, at least to some extent, NW-PD-

DROP is independently innovated. However, the evidence remains stronger, and clearer for P-DROP/D-DROP. Bailey (2018a) does, in fact report PD-DROP with a broader range of prepositions in Kent:

(89) From Bailey (2018a)

    a. it smelt coffee

    b. they go holiday

    c. So we went coffee and then came here

The observation that PD-DROP is permitted only with familiar or institutional places is challenged with data from NWatlas-corp, which finds numerous examples where the preposition is dropped from the infinitival form of a following verb:

(90) a. <u>cant wait finish!</u>

    b. <u>Cant wait start drivinggg</u>

    c. <u>can't wait release all my frustration</u> out in the gym

However, again, Bailey (2018a) does report that PD-DROP is occasionally also found outside of cases with locational GOALs in Kent.

P-DROP with abstract nouns

P-DROP is also licit with abstract nouns such as *conclusion/realisation*:

(91) a. <u>Come the conclusion</u> that I'm going to have to pick that book up because it seems very interesting

    b. When I start writing about myself I quite quickly <u>come the realisation</u> I'm the biggest nobhead I know.

    c. "Mr BMW" in Germany would call Merckel to ensure no deal happens.. 🥴 I've <u>come the same conclusion.</u>

Examples of P-DROP with abstract nouns reveals the versatility of the phenomenon, and shows that it is available across registers. As noted earlier, this is supported by the fact that while PD-DROP saw a dramatic decline in usage as the nature and use case for Twitter evolved, P-DROP remained relatively stable.

4.8 Comparing [P/D-DROP] and [DATALT] distribution

A note on topography

We can make some observations on the cluster patterns for [DATALT] and [P/D-DROP] in the Northern cross-section. The boundaries to the west of the Pennines are more closely aligned than those to the east. Liverpool is characterised by high-$_p$GTD and high-P-DROP. Additionally, there is a clear boundary between Liverpool and Manchester for both [DATALT] and [P/D-DROP] (TGD →GTD and P-DROP →PD-DROP, respectively). However, the [DATALT] boundary is notably more restricted immediately to Liverpool, whereas the [P/D-DROP] boundary extends further into the interlying region between the two main conurbations. This difference, it turns out is crucial to the distribution of available grammars in the region, as we will see.
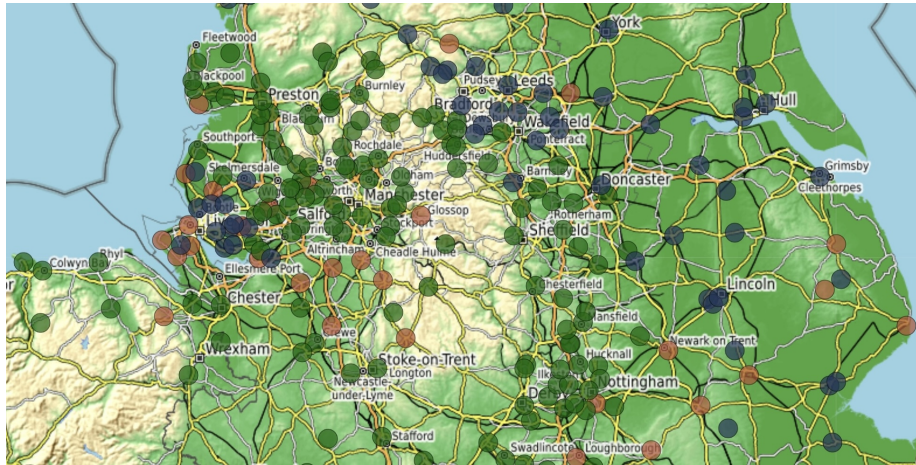
Figure 4.61: Pronominal ditransitive distribution on topographical map.

The eastern boundary of [DATALT] lies a few miles further east than [P/D-DROP], there is nevertheless a clear edge between $_p$TGD and $_p$GTD which lies south and west of Leeds-Bradford and east of Halifax and Huddersfield. This boundary extends south, marking a narrow transition zone between places such as Wakefield ($_p$GTD) ←→ Barnsley ($_p$TGD) and Sheffield($_p$TGD) ←→ Doncaster($_p$GTD).

The boundary between PD-DROP (red) and D-DROP (magenta) clearly follows the edges of the Pennines, while $_p$TGD is distributed across both sides. The alignment of PD-DROP and D-DROP to this topographical boundary is interesting. It likely speaks to the deep historical roots of D-DROP in the area

246
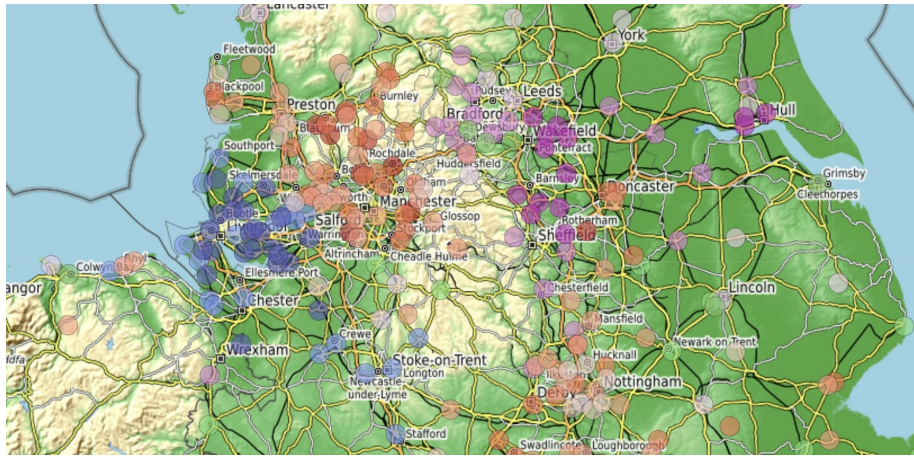
Figure 4.62: Preposition/determiner drop distribution on topographical map.

to the east (West and South Yorkshire), perhaps a relic from a time when the semi-mountainous region would have formed a greater impedance to contact and resultant diffusion. This finding is reminiscent of previous work, such as Britain's (2001), which showed how diffusion had been impeded by an area of once-swampland (the Fens) before it was reclaimed, resulting in quite different varieties being thrown into contact. It is interesting to consider here, what the contact situation now is between varieties which favour P-DROP and those which favour D-DROP, and potential linguistic consequences.[31]

The relative extent of ₚGTD around Liverpool is notably smaller than P-DROP. P-DROP use extends across the North West region up to the edges of Manchester's sphere of influence, but ₚGTD is more closely bound to Liverpool's immediate vicinity. This is interesting; it may be that P-DROP is more available for diffusion that ₚGTD, which fits with the notion that pronominal ditransitives seem to exhibit a remarkable resistance to spread over long stretches of history.

4.8.1 Probabilistic connection between P-DROP and TGDs

Here I return to the question of a correlative link between P-DROP and TGD. The main intersection between the two phenomena is where preposition dropping is implicated, for some speakers, in TGDs. Haddican (2010) found that for at least some of his Manchester respondents, TGDs fit the criteria for being underlyingly PDAT with a dropped preposition.

---

[31] I should note here that the clusters showing D-DROP dominance are essentially abstractions - it is not the case that there is no PD-DROP there, it is just that D-DROP is marked as dominant.

247

Haddican makes the observation that "subjects accept theme-goal ditransitives with full DP inanimate goals, only if they also accept null/deleted *to* sentences (and not vice-versa)" (Haddican, 2010, p.12). Biggs (2018) extended this finding for her Liverpool sample. Crucially, it is likely not the case that a TGD-as-PDAT grammar exists only in one location, more that this grammar will correlate with the availability of other structures in predictable ways. Specifically, the availability of both $_p$TGD (*gave it him*) and ditransitive P-DROP (*take them the match*) will predict the availability of TGD-fullDP (*gave the letter the bank*). However, given that the correlated structures *are* geographically variant, it follows that TGD-as-PDAT will be more likely to be located in locations where we find the following conditions:[32]

1. $_p$TGD is frequent

2. P-DROP is frequent

3. ditransitive P-DROP is frequent

Liverpool itself is in fact not the best candidate, despite Biggs's (2018) findings. As noted, TGD rates are low in Liverpool, which is an issue for an analysis which claims that TGD involves preposition dropping. Of course, low rates do not mean that it is absent, however.

We have seen that ditransitive P-DROP with two full DP objects is frequent with the verbs *take* and *bring* in NWatlas-corp, and tends to occur in Liverpool and surrounding areas, following the distribution of P-DROP with *go* and *come*.

Of particular interest is the area between Liverpool and Manchester which exhibits both high $_p$TGD and high P-DROP (indicated by the dashed blue oval in Figures 4.63 and 4.64). Biggs (2018) makes the point that her description of a grammar which fits the diagnostics for a P-DROP-TGD may not be strictly located in Liverpool, rather that "the availability of such a construction will correspond to systematic and productive variation in other aspects of that individual's grammar" (Biggs, 2018, footnote 5, p.4).

It follows, then, that the area in which we are most likely to find speakers whose grammar fits this profile in this interlying region. However, with *give*, situations where both objects are full DPs are very infrequent, and as such, it

---

[32]Note here that I am not assuming a deep syntactic connection between these structures, though this may be the case, for now I am drawing attention to the probabilistic tendency of structures to co-occur in a given location. Whilst this may suggest a deeper structural connection, I leave syntactic analysis to future work.
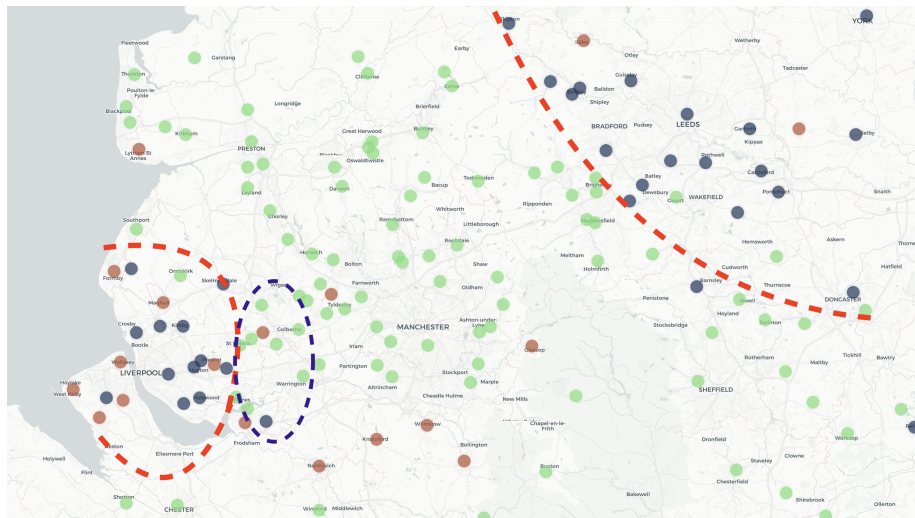
Figure 4.63: Detail from [DATALT] cluster map with boundaries (dashed red lines). TGD-P-DROP intersection indicated with dashed blue oval
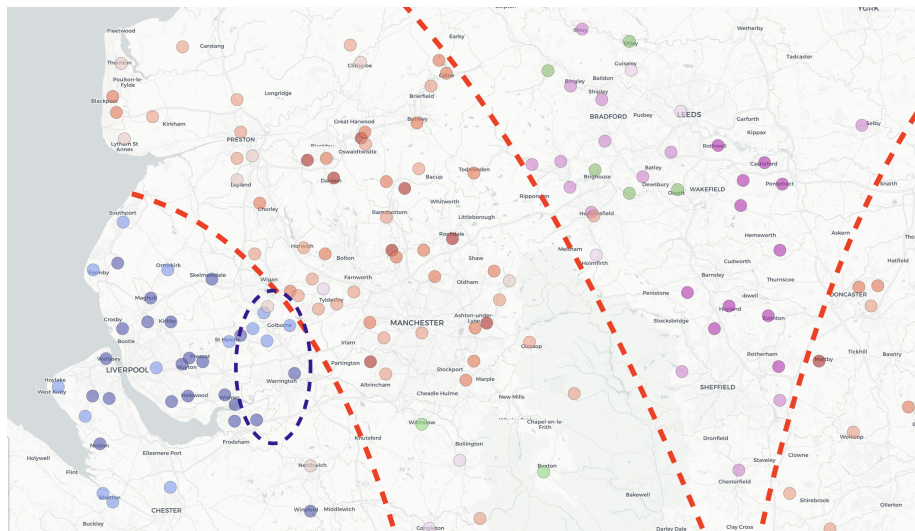
Figure 4.64: Detail from [P/D-DROP] cluster map with boundaries (dashed red lines). TGD-P-DROP intersection indicated with dashed blue oval

is difficult to draw firm conclusions on the corpus data alone. I return to this point in the next chapter.

4.9 Summary

[DatAlt]

The distributions of the two variables under investigation show quite different characteristics. The story of [DatAlt] is that of a gradual shift. ₚTGD, previously a dominant feature of British English, declined in use over the course of the 19th century and was overtaken by ₚPDAT and ₚGTD in the south of England. ₚTGD use has remained high in the Midlands and North West, but the area in which it dominates has shrunk since the 1950s, as shown by comparisons between the current spread and that shown in the SED.

It was also shown that ₚGTD is dominant across a broad inter-dialectal continuum, from the East and North East of England, through Scotland and Northern Ireland. Such a continuity suggests that ₚGTD dates back at least as far as the original Ulster Plantations, in the 17th century, which brought Scots into Northern Ireland. An even earlier date—during the period of Scandinavian settlement in the 9th and 10th centuries—, whilst speculative, does fit the distribution patterns that we see, and to which a coherent narrative may be applied.

In terms of the present-day structural properties of [DatAlt], the idea that TGD is related to GTD, rather than being PDAT with an elided preposition is supported by the corpus evidence. Verbs which were previously thought to favour PDAT were shown to do so across geographically distinct varieties, and crucially, for the most part, did not show the corresponding increase in TGD that might be expected if TGD were underlyingly PDAT. Following Biggs's (2018) suggestion that Liverpool English is an exception to this rule, we saw that there does exist a geographical area where both P-DROP and TGD are prevalent, but this area lies to the east of Liverpool, rather than within Liverpool itself.

[P/D-DROP]

I have argued that P-DROP and D-DROP likely emerged from DAR. The story of PD-DROP is considerably more complicated. The point of its innovation is not certain, and it may be that it there is a general propensity for preposition omission across Englishes. This said, the Twitter data show particular strongholds in the Midlands and East London, with what appears to be an independent instantiation in the South West. Historical evidence shows that the earliest

attested occurrence of PD-DROP is in the Old Bailey Corpus in 1737, with the feature attested in the East Midlands around the 1850s, and 1880s in the North West. However, whilst it is tempting to assert that this suggests an initial innovation in London, it is difficult to justify such a rapid diffusion in the 19th century as discussed in 2.5.1. Additionally, the Midlands and North West offer additional mechanisms by which PD-DROP may have emerged via internally driven change in the context of DAR, a potential mechanism for which I outlined in §4.6. The geographical overlap between instantiations of [P/D-DROP] and corresponding instantiations of DAR is certainly striking.

We saw that the relative rates of PD-DROP in the BNC XML edition (1994), are higher in the North than in the Midlands and South. This finding is curious, considering what we see in the Twitter corpus. Overall counts are low, however, and it may be simply that the particular locations sampled in the BNC disproportionately favour those with higher rates of PD-DROP.

We also saw evidence that the PD-DROP found in the NWatlas-corp was structurally distinct to that reported in the South East. Here, it would be advantageous to look at equivalent Twitter corpora for the structural status of PD-DROP in the Midlands, South East and South West which might help to complete the picture, but was unfortunately unavailable for the current project.[33] It may be, for example, that further data would reveal that South East PD-DROP is available across a greater range of contexts than has been reported in recent literature, I leave this to future research.

Bailey's (2018b) observation that there may be two types of PD-DROP in the South East appears, at first glance, to be borne out. The data show that there is a core PD-DROP area in East London, which appears to have deep roots, that likely then spread to Essex, Kent, and surrounding areas. The role of MLE, with multiple input varieties known to also exhibit PD-DROP, has likely played a role in recent history. However, its influence seems uncertain, given that the spread to Essex and Kent, where its usage frequency is most pronounced, occurred prior to MLE's development.

Regarding the present-day widespread use of PD-DROP, we do see evidence of —the likely rapid—diffusion of PD-DROP in recent history. This trend, perhaps partly fuelled by social media, has created an environment, to some extent independent of geography, where speakers have greater license to use

---

[33]Such an analysis is only possible with an open and unrestricted corpus of tweets. Such corpora do exist for the UK as a whole, albeit for shorter time windows.

PD-DROP, where it was already historically present.

Regarding P-DROP, we saw that use is not limited to Liverpool and the North West, but also extends into North Staffordshire, and around the Lancashire coast. Looking at historical trade routes, it was possible to date the probable innovation of P-DROP to around the time Liverpool English is thought to have formed. As such, I suggested that the formation of P-DROP likely occurred around the start of the 19th century, emerging from a reanalysis of the regional variant of DAR (DAR-θ). We also saw that there is the additional potential influence from the Irish language on some aspects of P-DROP.

Structural acceptance in the North West of England

> "[A] renewal of geographical linguistics [which] combines the methodology of sociolinguistics [...] with the aspirations of generative syntacticians to show that not only usage data but also carefully controlled judgment data vary with geography."
>
> (Cornips & Gregersen, 2016, p.511)

5.1 Introduction

We have already seen that the nature of P-DROP is variable in the North West and that PD-DROP in the North West appears to be distinct to that reported in the South East by Bailey (2018b) and Hall (2019).[1] We have also seen that the THEME-GOAL order is widely used in the North West where both objects are pronouns (*it me*), but that Liverpool is an island that bucks this trend, instead

---

[1]As discussed, it is not clear whether the distinction between North West PD-DROP and South East PD-DROP is simply due the larger sample available in the North West. Future work using social media and/or survey data in the South East may shed light on this.

strongly favouring the opposite order (_me it_).

In this chapter, the focus remains on the North West region with the introduction of a mass-participation acceptability judgement survey. Here, both of the variables documented in the Twitter corpora — [DATALT] and [P/D-DROP]— show a sudden shift in the relative frequency of variants over the space of only a few miles, indicating the presence of two distinct linguistic boundaries. Linguistic border regions are of particular interest as they offer the possibility of investigating micro-variation between speakers.

In addition, as discussed, several recent papers have investigated variation in [DATALT] and [P/D-DROP] in Manchester (Haddican, 2010; Haddican & Holmberg, 2012), Liverpool (Biggs, 2018) and Ormskirk, a town 12 miles north of Liverpool (Myler, 2013). The focus on the North West also offers the chance to extend recent work on the nature of [P/D-DROP] outside of London following Hall's (2019) recent work on PD-DROP in the capital.

5.2 Goals of the survey

The main goals of the survey are to:

1. Supplement the dialect atlas and cross validate results of the Twitter corpus by testing if judgements match what we would expect given the distribution of surface strings found there.

2. Extend the work in the recent syntax literature (Biggs, 2018; Haddican, 2010; Haddican & Holmberg, 2012; Myler, 2013) by providing substantially more data points to allow for a more fine-grained analysis of the distribution of phenomena.

3. Demonstrate a proof of concept for a methodology involving sixth-form colleges in this kind of study.

4. Create a _socially informed_ dataset such that results may be differentiated along traditional sociolinguistic criteria (age, social class, gender, identification to place). _This last goal was unfortunately made unworkable due to the pandemic, which meant that a sufficient range of participants was not gathered._

5.2.1 Research questions addressed in this chapter

First, the survey addresses:

RQ (8) To what extent do speaker judgements correlate with corpus frequencies for a given location?

This is the point of entry for the current survey. Mean acceptance ratings of sentences that are attested in the area in which the respondent resides are likely to show some correlation to corpus frequencies. From the point of view of the respondent, while they may not have a given structure in their own grammar, they will at least have a degree of exposure to it in their linguistic environment.

In such cases, we should expect to see some degree of correlation between the survey responses for a given location and the frequency of use of a given structure in the Twitter corpus. Further to this initial goal, the survey aims to more deeply probe questions relating to the underlying nature of the structures available in the North West, and to test recent findings reported in the literature.

Second, we assess the extent to which the area predicted in the Twitter data to be the most likely host for a grammar that supports full-DP GOAL-THEME ditransitives with *transfer* (see Figures 4.63 and 4.64) is found to do so:

RQ (9) To what extent is there evidence for $_{\text{DPDP}}$TGD in the location predicted from the Twitter data? Is there a grammar in the North West for which TGDs are underlyingly PDAT-P-DROP and is this grammar geographically restricted?

Third, we look at further evidence that PD-DROP in the North West is structurally distinct from that found in the South East:

RQ (10) Is PD-DROP in the North West akin to that found in the South East?

Finally, we look at the relative distribution of THEME-passives versus GOAL-passives, following Haddican and Holmberg's (2012) work in Manchester.

5.3 Background to the Survey

Dialect surveys

The manner in which dialect surveys are conducted and distributed within a population — the demographic characteristics of survey participants; where

a participant grew up, where they have spent time as an adult, their particular career path, whether they attended university, which particular university and even which subject they take at university (and countless other variables) — are all potentially important factors to take into account. However, these methodological considerations are, at times, overlooked, particularly when the research being conducted has its root in formal syntactic theory, where such questions have traditionally not been the central to the agenda. Often due to time and/or funding constraints, linguistic surveys are distributed amongst the personal contacts of the researcher, or perhaps amongst undergraduate students, and usually undergraduate *linguistics students*. Of course, this may not be a problem — it depends on the particular questions being investigated, and how the questions are framed.

In the case of formal syntax, which probes categorically grammatical or ungrammatical structures, the background of the participant may not be critical. More likely, if the participant is a colleague of the researcher — another university lecturer —, their background can be assumed to be such that they will have full command of the standard form of the language, in addition to a regional variety. Formal syntactic models have, until recently, therefore tended to model the standard language.

Contemporary methodology employed in dialect atlas projects has used online tools like Amazon Mechanical Turk (AMT) (cf. Wood, Zanuttini, Horn, & Zentz, 2020). This involves sending small payments to participants sourced from a large pool through the AMT system. It is possible to target participants from certain areas. More detailed biographical information can be gathered in the survey itself, typically asking where they grew up, their line of work, age, gender and so on. This method can be very effective, especially for smaller projects, where results are needed quickly.

The Survey of British Dialect Grammar (SBDG)

An alternative to using university networks to distribute a survey, is to go into schools. The obvious advantage in working with schools is that, unlike universities, the majority of students will, usually, have grown up in — and their language use will be more reflective of — the local area.[2]

---

[2]There are exceptions to this, of course. Particularly in the inner city, school children may have moved around considerably more. Overall, it is more likely that school children will be local than university students, but it should of course be considered on a case by case basis.

A notable example of this approach is the Survey of British Dialect Grammar (SBDG) (Cheshire, 1986), which was conducted mainly in the 1970s and early 1980s. The main objective was to document and analyse morphosyntactic variation across traditional dialects of English spoken in the British Isles. Such a large-scale project would ordinary require significant resources, however, the decision was taken to work with schools, in part as a way to avoid the considerable costs associated with a larger scale distribution using traditional methods. This was particularly the case in the 1980s, before modern recording and transcribing equipment and before the ability to conduct interview online. However, rather than it being a limitation, the investigators incorporated working with schools into the research agenda.

The aim of the project was twofold: first, to get a handle on the grammatical variation that existed in British English at the time; the features, their 'nature' (structure and status in terms of LVC) and geographical spread. Second, the project functioned as a way to push back against the dominance of 'standard English' in education at that time, and raise the awareness—and status—of the non-standard varieties that the majority of children used.

Importantly, in relation to the present work, this study demonstrates that working with schools can be a fruitful method for gathering linguistic data.

Syntactic Atlas of the Dutch Dialects (SAND)

The Syntactic Atlas of the Dutch Dialects (SAND) (Barbiers & Bennis, 2007), led by Sjef Baribiers at the Meertens Institute in the Netherlands, was conducted as a collaboration between several instituions in the Netherlands and Belgium. The project was a foundational as a study of geographic variation in syntax. At the time of its inception, in the early 2000s, there were relatively few systematic studies of geographic microvariation in the syntax of a single language. The project broke considerable new ground in that it brought typologists and generative linguistics together on the same project. Whilst this was initially borne out of necessity—the Dutch contingent were generative linguistics and the Flemish were typologists—there was an expectation that a descriptive level could be reached that would satisfy both camps. However, as Barbiers and Bennis (2007, p.56) report, this proved to be very challenging, with 'fierce debates' between the groups. Such a collaboration, appears to have been a fruitful exercise, leading to the conclusion that "data cannot be taken

as theory-neutral objects that receive an interpretation within a particular theoretical framework. The data themselves are theoretical objects." (Barbiers & Bennis, 2007, p.56).

These difficulties were also encountered in the current project, which is at its core, a corpus-dialectological approach to linguistic variation, but which sought to engage with, and draw from, theories borne out of formal syntactic theory.

Scots Syntax Atlas (SCOSYA)

The Scots Syntax Atlas (SCOSYA) (Smith et al., 2019), was similar to the SANDS project in that it involved a collaboration between different types of linguists. In this case between generative syntacticians and sociolinguists. The project aimed to update and extend previous atlases of Scots dialects by using a combination of sociolinguistic interviews and surveys.

The project gathered both syntactic and phonological data, with a team of research assistants who visited each place in person. The primary output of the project was the creation of an interactive atlas of Scots, based on both a grammaticality judgement survey across 122 locations, and sociolinguistic interviews. As such, the project serves as a rich resource for future research.

As SCOSYA demonstrates, this approach is particularly valuable when investigating remote communities, such as those found in the Scottish Highlands, and northern coastal areas. The project was a major contribution to the field, and a substantial undertaking.

This work is particularly relevant to the present study in that it combines the analysis of language use with acceptance data, and demonstrates how acceptance data can be used effectively for the analysis of sociolinguistic variation (Jamieson et al., 2024). Additionally it demonstrates clearly that sociolinguistic and formal syntax approaches may be fruitfully combined. In particular, the project's focus on both the syntax-morphology interface and the role of dialect leveling in microparametric syntax offers an effective template for this kind of study.

The present work borrows from SCOSYA in taking a combined approach to geographic microvariation in regional English. In the future, a more elaborate study of the North West would do well to additionally integrate sociolinguistic interviews alongside the corpus data. This would allow for a richer considera-

tion of social factors influencing the variability that we see in the data.

A key finding of the project was that syntactic judgements ratings and usage data often diverge: speakers may reject features that they are found to use in casual conversation. This demonstrates that speakers are often unaware that they use a given feature and highlights the importance of triangulating methods.

5.4 Comparing usage and judgement data

The question of the extent to which usage and judgement data are comparable has surfaced several times in the recent literature (cf. Bermel & Knittl, 2012b; Bresnan, Cueni, Nikitina, & Baayen, 2007; Gerasimova & Lyutikova, 2020). The question is not straightforward. On the one hand, we know that frequency in a corpus holds very little bearing on grammaticality, indeed, this is a maxim of theoretical syntax. On the other hand, we also expect that, in the context of competing variants, high frequency variants are likely to be rated more favourably by participants than those with low frequency. We saw this in the close correlation between the acceptance rates of $_P$TGD in the Our Dialects Project (MacKenzie et al., 2022) and usage rates on Twitter.

The debate is often framed between the extent to which grammatical knowledge is gradient and probabilistic or whether it is categorical. The reality is probably a combination of the two and depends critically on the sentence being judged. There is also no real sense in which a gradient survey response necessarily equates to a gradience in linguistic knowledge; a gradient response may indicate a gradient sociolinguistic evaluation of a categorically grammatical sentence.

There are, therefore, some crucial factors to consider when assessing the extent to which the acceptance of a given sentence string will show a relationship with its frequency in a language corpus for a given location.

1. We consider the task of sentence judgement not to be an open and shut case of speaker intuition or what is parsable by their internal linguistic system, but that such a task is in itself a kind of linguistic performance, unavoidably coloured by individual perception (Schutze, 1996; Schutze & Sprouse, 2014).

2. Where a survey uses gradient measures of acceptance, such as that used in the current project, responents are given the freedom to report a degraded

259

acceptance of a given structure, rather than outright rejection. Degraded acceptance may be taken as indicative of a structure's presence in the dialect grammar, with other, possibly sociolinguistic, factors reducing perceived acceptance.

3. If the task is a choice between multiple ways of expressing the same proposition, and each of these ways is attested in corpora of language used locally to the respondent then they will likely have a clear sense that a structure is used by people around them, even if they may not themselves use it. Here, it follows that we are much more likely to see a measured gradience in acceptance ratings that, in turn, reflect the frequency of occurrence in those corpora.

5.5 Survey rationale and predictions

Whilst both Biggs and Haddican focus on Liverpool and Manchester English respectively, neither study claims that each place strictly hosts a given grammar. Haddican's (2010) results indicated a degree of inter-speaker variation within his Manchester sample, some of whom seem to align with the alternative P-DROP analysis of THEME-GOAL ditransitives elaborated by Biggs in Liverpool.

Meanwhile, Biggs makes a critical clarification to the claim she is making with regard to 'Liverpool English':

> "The claim of this paper is not that certain linguistic forms will only occur in precise geographical regions. Rather, the claim is that the availability of such a construction will correspond to systematic and productive variation in other aspects of that individual's grammar (Biggs, 2018, p.4, fn5)."

The suggestion, then, is that there exists a location non-specific grammar in the region for which TGDs and TPASS are underlyingly P-DROP. The proposal that there are distinct grammars underlying identical surface strings maps to MacKenzie's (2019) concept of *covert representational variability*. The idea here, is that child learners, when presented with language data, arrive at surface forms via different mechanisms, within the bounds of what is permitted by the broader grammar of the particular variety. This idea has been discussed in relation to Korean pronouns, where learners arrive at one of two types of a surface-identical pronoun (Kim & Han, 2016). Given that, in the Korean case,

the choice between the options is thought to be essentially random, we would correspondingly not anticipate there to be tractable spatial variation in the distribution of one or the other grammar.

In the case of TGD-as-P-DROP, however, given that the conditions which predict the occurrence of the grammar are geographically locatable, the suggestion put forward here is that the source of the grammar should be so too.

5.6 Survey methodology

What the survey measures

The principal method of investigation is a large-scale sentence acceptability survey distributed through sixth-form colleges in the region. The methodology for the survey follows that of the recent literature that focuses on syntax variation in the North West, primarily Haddican (2010); Haddican and Holmberg (2012). This involves using a sliding scale for sentence acceptance between 0 and 100 rather than a 5-point Likert scale, or the **, *, ?,?? scale traditionally used in syntactic theory. The methodology aims for transparent reporting of the actual survey data presented as numeric values, rather than by the reported responses recorded by the investigator. Additionally, the survey here has, as its focus, variation between individual speakers as well as the cumulative responses of speakers by area. Gradient acceptance ratings were chosen both for parity with Haddican's (2010) method and for the reason that they map more easily to the corpus usage frequency data.

Survey distribution

In order to increase the chance that survey respondents were local to the area, sixth-form colleges were chosen over universities. Other possible recruitment sites that were considered were libraries, local community centres, online discussion forums such as Reddit or more locally oriented universities that tend to recruit students from the local area. However, sixth-form colleges offer a bridge between high school and university and students are almost exclusively between the ages 16-18 and, as a result, students are more likely to have grown up in the local area. This assumption was validated in the results: the postcodes that students provided in response to the question of where they *grew up* were, in the vast majority of cases, local to the institution.

Recruiting schools

Institutions were selected from the .gov database (`https://get-information -schools.service.gov.uk/`) which offered English Language A-level. A spreadsheet was then compiled with the relevant contacts for each school. Schools were first sent an email with a cover letter, and a *press release* (Figure 5.1), providing background information on the project, and a brief outline of what the survey would entail.

Dialect variation in the North West of England:

An opportunity for sixth form students to participate in university research.

Part of ESRC-funded project: Investigating dialect borders; *language, identity and mind*.
Jon Stevenson, Paul Kerswill and Ann Taylor. University of York, UK.

UNIVERSITY
of York

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

The project so far
The first part of the project used Twitter messages to map language use across the UK (see map below).

This innovative research has revealed distinct geographical patterns in how we use language.

Some areas group together to form larger regions while distinct borders are revealed where the pattern abruptly changes over just a few miles.

Getting students involved
We are now looking to gain a more detailed understanding of these border areas by conducting a language survey.

This is where we are looking for the participation of **English Language A-Level students** to take and then help distribute the survey.

The first border area we are looking at is in the north-west between Liverpool and Manchester.

Linguistic atlas
A principal outcome of this research will be an interactive language atlas combining the Twitter and survey data.

The final atlas will be made freely available to participating schools.

We believe the atlas will make a valuable teaching resource for future classes particularly for topics relating to language and identity and social media.

Opportunity for students
- Insight into real research from a leading Higher Education department
- Hands-on experience with a language data-investigation

Invitation to join
Your sixth form has been selected as one of a number offering English Language A-Level in the region and we would like to invite you to take part.

Participation involves two components:

(1) In-class survey
The survey takes just 20 minutes to complete in class, where conditions can be controlled. It will be completely anonymous.

(2) Distribution
Students then simply ask friends and family at home and in the neighbourhood to take the same survey.

When will this take place?
The survey will be available to access from 1st November 2019 to 1st March 2020. Ideally the In-class session should take place by January.

Registration
To register your interest and confirm participation, email by

Monday 21st October

Jon Stevenson: **js1472@york.ac.uk**

Map of Twitter messages showing preference for structures like:

"gave it to me" (red)
"gave me it" (blue)
"gave it me" (green)

Interactive version:

https://tinyurl.com/y3qdalk8

Please use freely.

Figure 5.1: Press release sent out to schools.

Participation in the survey was presented as an *opportunity to participate in*

university research. This was not merely a selling point, but a genuine proposal. A-Level English language already offers a *data investigation* component, which is essentially a mini linguistic fieldwork project. The idea was that students would have follow-up opportunities to engage further with the work. Unfortunately, however, the timing was such (March 2020) that the pandemic cut short the exercise (see information box below). Despite this setback, there is considerable scope for the development of this concept: working with sixth form students in the gathering of linguistic data in a way that is of real mutual benefit. The response from teachers was enthusiastic, with one school having the story published in a local newspaper.[3]

Schools were then contacted twice by phone to follow up on the email as a way to gauge interest. Finally, a doodle poll was sent to the main contacts (usually the department head), asking to indicate whether they would like to participate. Out of around 40 contacted, 15 agreed to participate, with 7 actually taking part. The lower number was likely due to COVID. Even with 7 participating institutions, there was good coverage of the North West region.

Ethical considerations

Participation in the survey was entirely voluntary. This was made clear on the cover page to the survey, which students had to check off that they had read before be able to continue to the next page. No personally identifiable data were gathered, and schools were given access to the survey prior to running such that they were able to verify that it did not contravene their ethics protocols.

Because the nature of the questions involved grammaticality judgements, rather than, for instance, recordings or social media texts, there was no way that student anonymity could be compromised, and no room for students to expose potentially compromising information. Any future project that might include such data would, of course, be an entirely different undertaking.

5.6.1 Sample population

The results of the survey have considerably less range than was anticipated due to the disruption of the onset of the COVID-19 pandemic. The original idea was that students should first take the survey in class, then, second, distribute

---

[3]I return to this in *Future directions* in the next chapter.

it amongst older family and community members. For every student, there was anticipated to be an additional three or four survey participants, with a greater and more representative age-range. However, only one institution was able to complete the distribution of the survey.

Whilst the results from this one institution demonstrate the viability of this approach — and its applicability to future research — the results overall do not provide enough data to allow for meaningful subgroup analyses at the level of age and social class. The vast majority of participants are in the 16-18 age-range, and responses in each location are not numerous enough to differentiate along the lines of social class. However, the data provided still offer useful insights into the structures under investigation, their geographical variability and still serve as a useful cross-validation of the Twitter data.

There was a total of 342 completed responses, more or less evenly split across the region. The vast majority of respondents are in the 16-18 age group. As mentioned, the initial concept involved students taking the survey in class and then distributing it amongst family and friends, including older speakers, but that this was rendered impossible for most of the participants due to the pandemic. One Sixth Form was able to complete the wider community section, however, demonstrating the workability of the concept.

Demographic data were gathered: as well as age and location (the first part of the postcode, for the place they grew up), participants were asked to provide the work profile and location of their parent/guardian. Due to overall numbers being lower than expected, it was not possible to use this demographic data quantitatively, when broken down by location. For now, I present only the overall results by location.

It may be possible to find trends in the data between acceptance ratings when looking at the region as a whole, or perhaps grouping into two large groups: East⟵⟶West, for example. Another possibility would be to gather more data by conducting a follow-up survey. This could be achieved by re-approaching schools, and/or releasing the survey to a wider online audience, perhaps via online communities, or via other on-the-ground community centres (libraries, youth centres, interest groups etc.). For now, I leave this to future work.

5.6.2 Initial survey questions

The start of the survey included phonological and lexical questions relating to phenomena known to be variant in the region. This is relevant in terms of the methodology, as it provided a gentler introduction to the survey, before the test sentences. With lexical items, for instance, pictures of objects were presented and respondents asked to select the name of the object from a list (e.g. Barm/Roll/Bap etc.) for a *bread roll*. Phonological questions involved reporting whether two words rhymed, following other recent work (Leemann et al., 2018; MacKenzie et al., 2022), such as *book-spook* and *singer-finger*. The results of this part of the survey will be presented in separate work.

5.6.3 Test sentences

A set of 62 test sentences were used, combined with 50 filler sentences, making a total of 112. Sentences were presented on three consecutive pages, in a pseudo-randomised order: they were first randomised, then manually adjusted such that there was a sufficient gap between test sentences.

Figure 5.2: Extract of test sentences as presented to participants

Sentences were judged by moving a toggle along a sliding bar between 0% and 100%, as shown in Figure 5.2 above. This mirrors the method used in Haddican (2010), which was completed on paper and had participants shade along

a similarly gradient scale. It is distinct from the methods used in Biggs (2018); Myler (2013), which followed more standard, categorical sentence ratings.

The test sentences followed those used by Haddican (2010) and Biggs (2018), where the goal was principally to gauge regional variability in the acceptability of:

1. Preposition dropping with intransitive and ditransitive verbs of motion (*going the pub* and *took them both the zoo*)

2. Preposition-determiner drop is accepted (*went pub*)

3. Straight-modification with P-DROP (*went straight the pub*)[4]

4. Ditransitives with *transfer of possession* verbs, such as *give* with THEME-GOAL/ GOAL-THEME orders

    (a) With a variety of other verb classes *refuse/envy*
    (b) With full-DP complements

The nature of the test sentences was such that they included a range of standardly accepted constructions, such as:

(1)   a.   Cute hamster! .... Yes, my dad gave her to me
        b.   Yes, his brother introduced me to her

These were broadly accepted across the board, and as such, to some extent served to calibrate participant responses. The full list of test sentences is provided in the Appendix.[5] Table 5.1 below shows only those results most pertinent to the discussion presented in the coming pages. Where additional sentences are involved, these are indicated in the relevant sections.

Sentence	Verb	Type	Cat.	Str.	THEME	GOAL
Are you <u>heading t'pub</u> later?	*head*	trans.	motion	DAR		
He won't be coming - he's <u>flying Germany</u> tomorrow	*fly*	trans.	motion	P-DROP		

---

[4]Given the finding in NWatlas-corp that *straight* modification is also present in the North West with PD-DROP (*went straight pub*), it would be interesting to test speaker acceptance of this too. Unfortunately, speaker use of this construction was not found until after the survey had been conducted, and did not present as a structure that should be tested.

[5]All of the survey results are available to view at `http://nwdialectatlas.uk`

I think he said he would take them both the zoo after work	*take*	ditr.	motion	P-DROP	P	DP
No questions asked, they went straight the pub after that	*go*	trans.	motion	P-DROP		
She's staying John's place tonight	*stay*	trans.	locat.	P-DROP		
They were saying that the letter was sent Mark.	*send*	ditr.	give	TPASS	DP	DP
I'm going shop first thing tomoz do you want anything	*go*	trans.	motion	PD-DROP		
After the game, he shouted the results the crowd	*shout*	ditr.	comm.	TGD	DP	DP
I didn't know that he gave the letter the bank on friday	give	ditr.	*give*	TGD	DP	DP
I should tell you, we gave it the girl already	*give*	ditr.	give	TGD	P	P
It's a scanner/Printer thing. Someone gave it me	*give*	ditr.	give	TPASS	DP	DP
I'm not certain but I think that the voucher was sent him	*send*	ditr.	give	TPASS	DP	DP
They mentioned that it was given her	give	ditr.	*give*	TPASS	P	P

Table 5.1: List of test sentences discussed in this chapter (full list of test sentences are presented in the Appendix).

In order to make sentences feel more natural, they were presented with surrounding context. There is a balance to be struck here: adding additional context may make sentences more natural, but they may also introduce confounding factors, where there are other parts of the sentence that may influence the sentence rating. Additionally, adding context makes it more difficult to compare structures as minimal pairs in the way that is usually expected in psycholinguistic experiments. The decision was taken to forgo some testing precision in favour of making sentences more palatable to (young) respondents who were presented with quite a long list of sentences to judge. If there were 100 sentences of minimal pairs, extreme judgement fatigue would arguably compromise the validity of the results.

The effect of context was mitigated by underlining the parts of the sentences that were to be judged by respondents. Respondents were instructed to judge the acceptability of the underlined constructions in context. The result appeared effective. Respondents completed the survey without apparent issues, and feedback was positive for the vast majority.

5.6.4 Data processing

Normalising responses: z-score transformation

The pre-processing for the survey responses uses a short R script that follows the standard procedure for calculating z-scores described in Schutze and Sprouse (2014):

> "For a given participant P, calculate the mean and standard deviation of all of P's judgments. Next, subtract each of P's judgments from the mean. Finally, divide each of these differences by P's standard deviation."

A potential issue with this approach is that raw values at the extreme ends of the scale (0 or 100), are converted to a different z-score for each participant. This means that for a respondent A, who generally gives sentences a high rating, 100 might produce a z-score of 1.5, whilst another respondent B, who tends to give low ratings for sentences, 100 might produce a z-score of 2. On the one hand, this is the point of z-score transformation: in some sense we want to say that a raw score of 100 has less 'weight' for respondent A than respondent B. On the other hand, when a sentence is rated at 100, there is an intuitive sense that it should correspond to an equal position of grammatical acceptance between participants. In this view, 100 simply means 'categorically acceptable', and in the case of A and B above, to rate one as 1.5 and the other as 2 could be viewed as a misrepresentation of the results.

This concern is not, however, a serious one, if our purpose in performing a z-transformation of the scores is to look for correlations between sentence types, as the z-score will be constant for each participant. To take the above example — if both participant A and B rate a pair of sentences (S1 and S2) and we wish to compare the two ratings — it does not matter if participant A's score of 100 translates to a different z-score to participant B's score of 100. What matters is that the z-scores for S1 and S2 are equivalent for each participant. So,

Category	z-score range
Strong accept	> 1
Weak accept	0-1
Weak reject	-1-0
Strong reject	< -1

Table 5.2: Breakdown of rating categories and corresponding z-scores.

following the above example, if both participants score both S1 and S2 at 100, then this would translate to a z-score of 1.5 for both sentences for participant A and likewise a z-score of 2 for both sentences for participant B.

Where it is more problematic is when we try to apply a label to a given z-score. For the purpose of mapping geographical variation in sentence ratings, there is a temptation to want to say that a z-score of more than 2 corresponds to a stronger level of acceptance than a z-score of 1.5. This is not always the case, however, as we have seen above.

Categorising z-scores

In order to effectively display results on a map, z-scores were categorised as either *strong accept, weak accept, weak reject* or *strong reject*. The criteria for the categorisation protocol are set out in Table 5.2.

In some sense it is arbitrary where exactly the line is drawn, but these categories make logical sense. However, they are also idealisations of the data. For this reason, along with the maps are presented boxplots which chart the distribution of the actual z-scores, show the range of responses and the median overall ratings by location.

5.7 Survey results and discussion

We now turn to look at the survey results. We will see that the results confirm a number of predictions: Liverpool and environs do pattern differently to neighbouring Manchester across a range of diagnostics and the distributions largely support those found in the Twitter data. At the same time, the results only partially support the data presented in the recent literature (Biggs, 2018), and as a result, point to different conclusions.

Mapping the results

Whilst the maps group results into four categories based on z-score, boxplots offer a more fine-grained view. A small box denotes agreement between respondents whilst a larger box denotes a divergence. This is particularly useful in the context of linguistic borders. Notice how in Figure 5.6, Manchester and surrounding places both rate the pronominal TGD highly (around a z-score of 1), and are converging on this response. Liverpool is overall low, with diverging responses, while its surrounding towns have a relatively high median, but with large numbers of respondents giving a less favourable score.

Rather than mapping results by institution, results were mapped according to postcode area. This to some extent mitigates any particular bias in responses from a given classroom setting. Out of the 7 participating schools, there were 8 postcode areas represented.

Acceptance of pronominal ditransitives

I begin by addressing the acceptance of pronominal ditransitives in GOAL-THEME and THEME-GOAL orders (_me it_ and _it me_). This also most directly answers the question of the extent to which the survey data corroborate the Twitter frequency data.

RQ (8) To what extent do speaker judgements correlate with corpus frequencies for a given location?

GOAL-THEME

Starting with the GOAL-THEME order (_me it_), we see that it is widely accepted across the region, even in places that are THEME-GOAL-dominant in the corpus data. This is indicative of its normative status. It also gives some context to the code-switching phenomena that we saw in the Twitter data. There are many speakers who essentially have access to both orders, but, in usage, favour $_p$TGD. We do see a slightly degraded acceptance for some speakers in St.Helens and Culcheth.
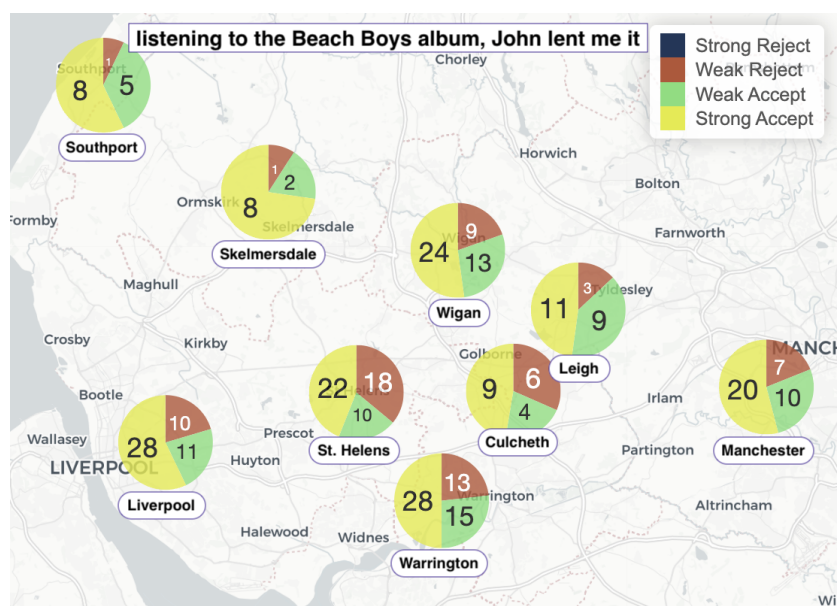
Figure 5.3: Geographical distribution of GTD with two pronominal objects "I'm not sure why I'm listening to the Beach Boys album, <u>John lent me it</u>"
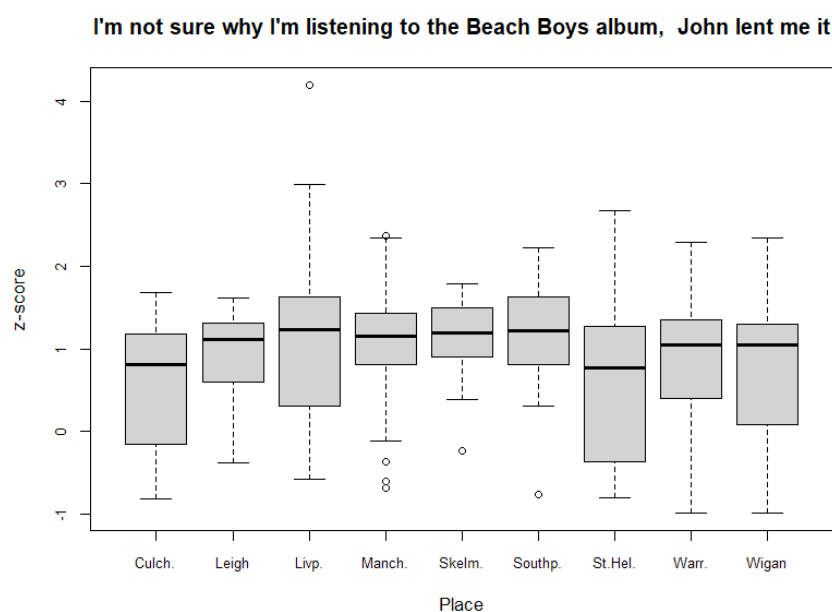
Figure 5.4: Geographical distribution of GTD with two pronominal objects "I'm not sure why I'm listening to the Beach Boys album, <u>John lent me it</u>"

Acceptance ratings for $_p$TGD (_ *it me*), on the other hand, align closely with the Twitter distribution: ratings are high across the board, outside of Liverpool, though notably less-so in the more geographically proximate locations of St.Helens and Warrington. Within Liverpool, ratings are measurably lower, with only around 25% of respondents categorised as *strong accept*.
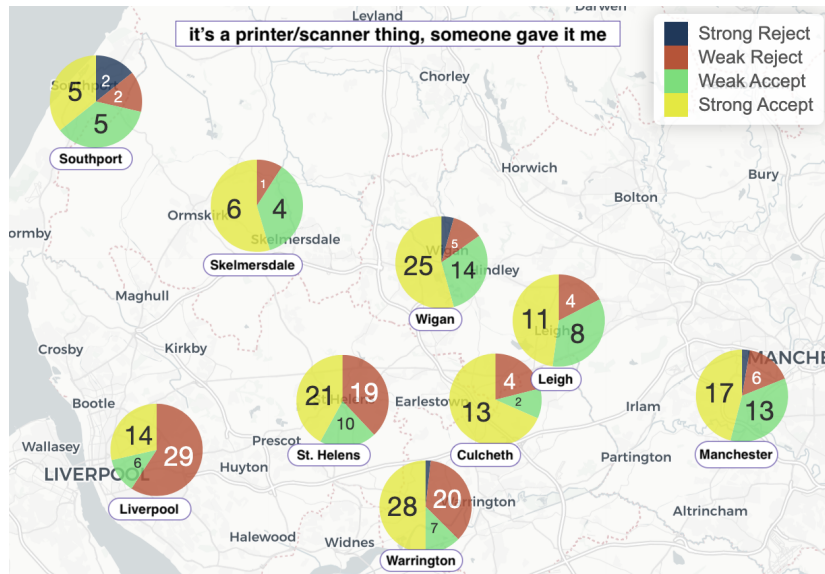
Figure 5.5: Geographical distribution of TGD with two pronominal objects
"It's a scanner/Printer thing. Someone gave it me"

The median z-score, shown in Figure 5.6, is around 1, indicating broad acceptance. Again, this is not surprising, given the preferential use of $_p$TGD in the region, found in the Twitter data. Liverpool, meanwhile, has a median z-score of less than 0, though the range is quite large, a significant proportion of respondent also accept $_p$TGDs. The median z-score across the other locations is around 1, which corresponds to the *strong accept* category. It is interesting that $_p$TGD is as degraded as it is for Liverpool respondents, even if this follows the usage data. There may be an element of sociolinguistic judgement in these ratings: not so much that it is ungrammatical, but perceived as *not-Liverpool*.

These results are comparable to those found in the *Our Dialects Survey* (MacKenzie et al., 2014), Table 5.3.

Postcode area	ₚTGD (n)	Total (n)	% ₚTGD
Liverpool (L)	130	482	27%
Manchester (M)	665	1113	60%
Warrington (WA)	247	504	49%

Table 5.3: Proportion of people who report using ₚTGD themselves from the *Our Dialects Survey* (MacKenzie et al., 2014)
Data supplied by personal correspondence from George Bailey.
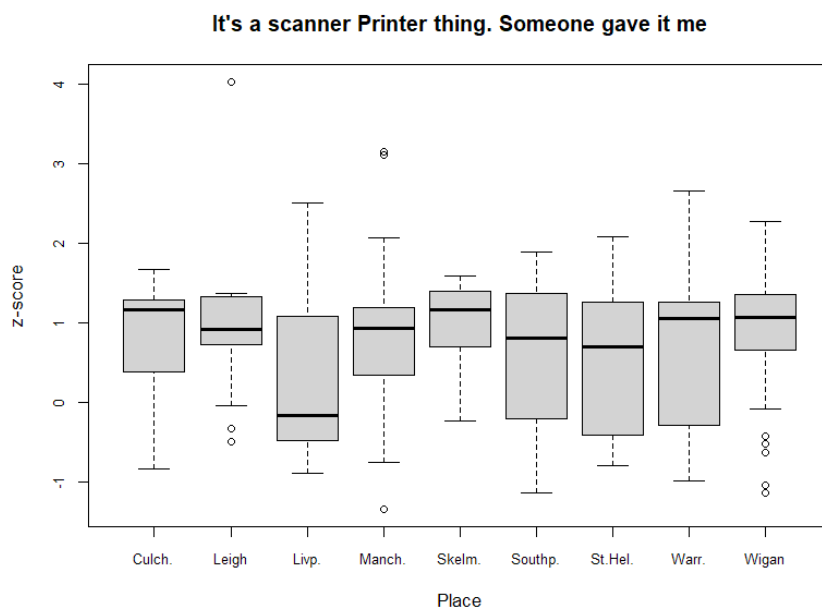
Figure 5.6: Boxplot of z-scores for TGD with two pronominal objects
"It's a scanner/Printer thing. Someone gave it me"

Where the THEME is pronominal and GOAL is a DP, as in the test sentence *I should tell you, we gave it the girl already*, acceptance is predictably lower than where both objects are pronouns. This follows the cline in acceptability ratings reported by Haddican (2010) (see Figure 2.1, page 55). Overall acceptance is still quite high however, and follows the geographical distribution for ₚTGDs.
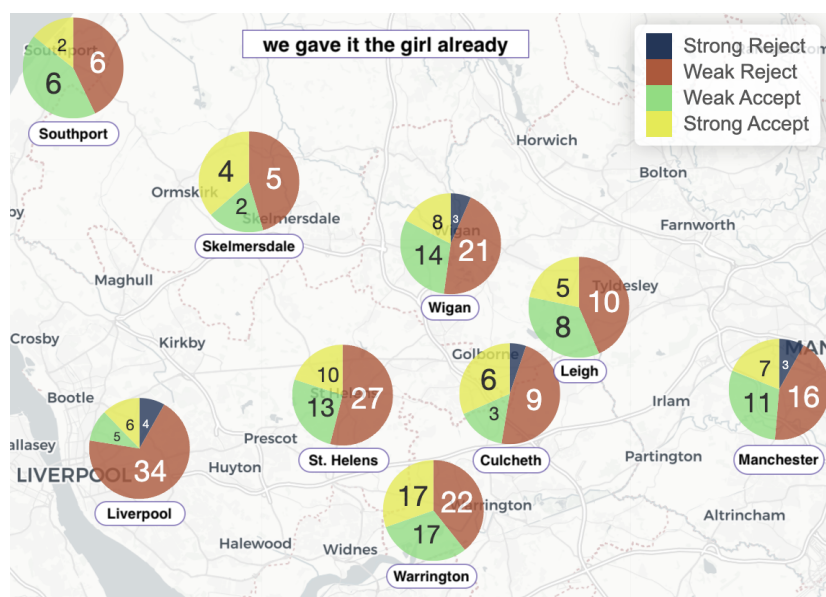
Figure 5.7: Geographical distribution of TGD with pronoun 'it' and DP recipient.
"I should tell you, we gave it the girl already"

Again, Liverpool shows the lowest approval ratings. This is perhaps the first indication that the availability of P-DROP does not have a bearing on the acceptance of TGDs with full-DP objects. Rather, the strongest predictor of the acceptance of the sentence with pronoun-DP, appears to be acceptance of pronoun-pronoun.

I take this as evidence in support of the claim that TGDs do not, for most speakers at least, involve preposition drop, at least not the *kind* of preposition drop that is available in Liverpool with *motion* verbs. This is in-line with the wider evidence from the Twitter data regarding the distribution of different ditransitive verbs: that, for example, latinate verbs such as *explain* are near-categorically ₚPDAT, and we see scant examples with ₚTGD or ₚGTD.

I should tell you, we gave it the girl already
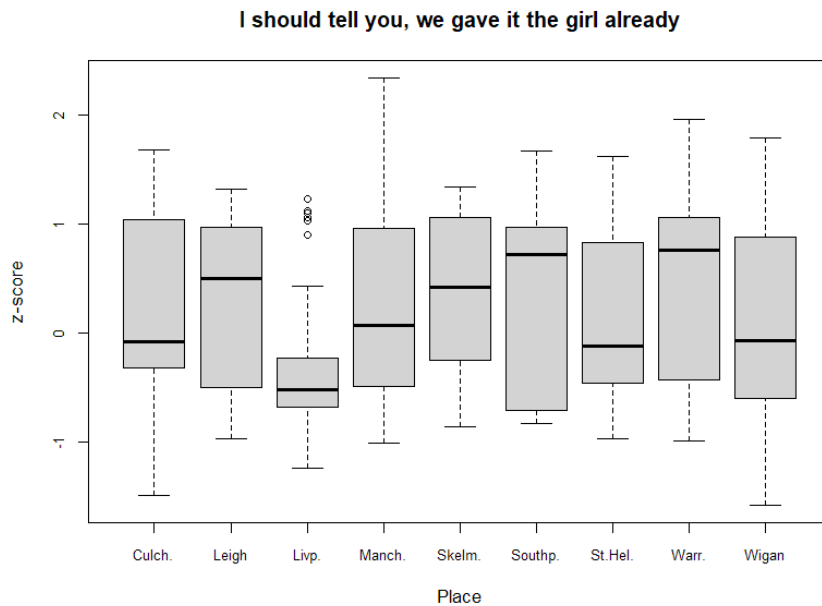
Figure 5.8: Boxplot of z-scores for TGD with pronoun 'it' and DP recipient.
"I should tell you, we gave it the girl already"

Ditransitive *take*

We have already seen from the Twitter data that P-DROP is, on the other hand, frequent with motion verbs in the North West. We saw this with intransitive verbs *go, come* and with ditransitive verbs *take, bring* and *send*. There is accordingly widespread acceptance of the ditransitive form of P-DROP with full DP objects, across the North West, though notably less-so in Manchester and Wigan. This fits the expected pattern for P-DROP with the verb *take*.
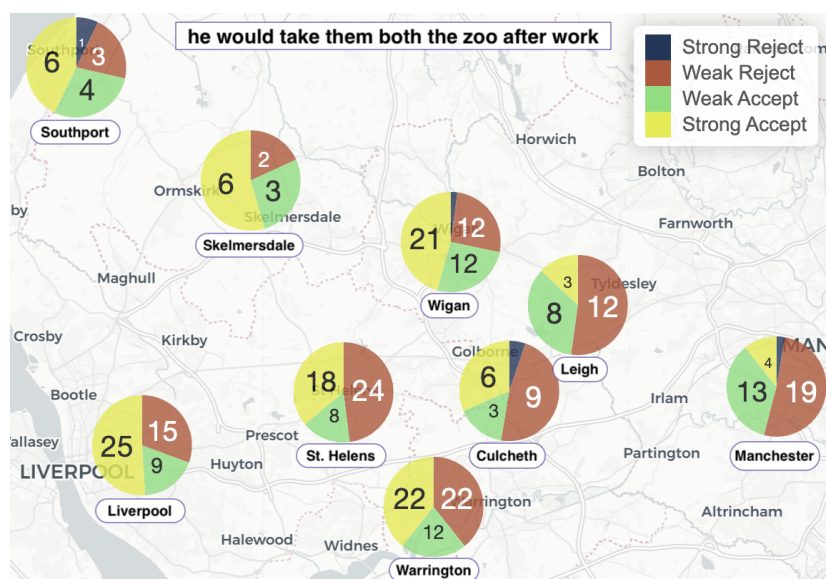
Figure 5.9: Map of acceptance scores for P-DROP with full-DP objects:
"I think he said he would <u>take them both the zoo</u> after work"

Acceptance for ditransitive *take* with the preposition absent in *he took them both the zoo,* is strongly accepted for about half of respondents outside of Manchester. It is notably degraded in Culcheth and Leigh, which neighbour Manchester, indicating a gradual decline in acceptance the further away from Liverpool (and closer to Manchester) we get.

Again, if the P-DROP we find across the region was available with TGDs, we would expect to see a correspondingly higher rate of acceptance for sentences like *gave it the girl already* in places such as Liverpool (Figures 5.7 and 5.8). Instead, it is notable that TGD-pDP is degraded for Liverpool speakers. If it were preposition drop, we might expect that it would be better accepted, given the high frequency and acceptance of P-DROP amongst Liverpool speakers. However, it seems that, as reported, the restriction on permissible verbs with [P/D-DROP] to those of *motion towards a directional goal*, holds.

I think he said he would take them both the zoo after work
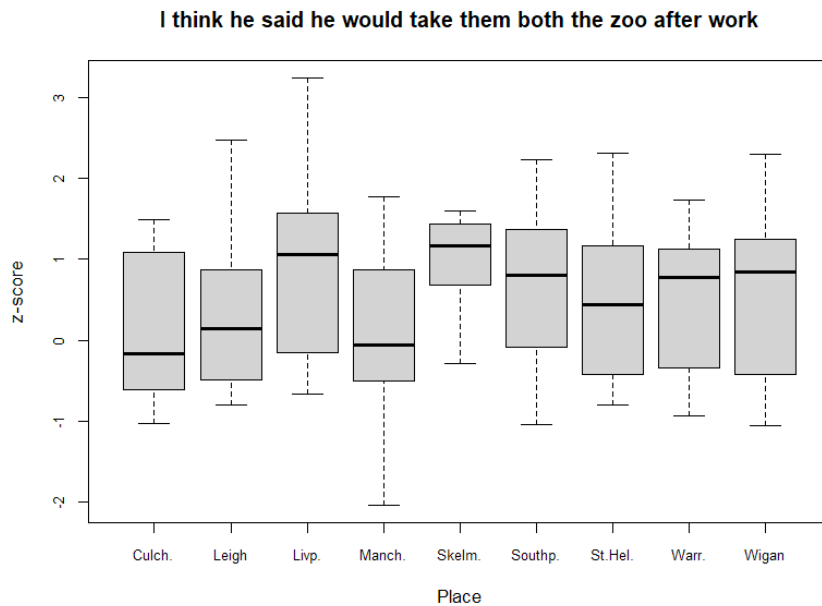
Figure 5.10: Boxplot of z-scores for P-DROP-*take* two DP objects
"I think he said he would take them both the zoo after work"

Give, TGD with two full DP objects

Here, the results start to get particularly interesting. We have seen that:

1. Pronominal TGD is widely used and accepted in Manchester and the region as a whole, but not in Liverpool.

2. TGDs where the THEME is pronominal and the GOAL is a DP are accepted, albeit to a slightly lesser degree, following the same geographical distribution

3. P-DROP with ditransitive *take* is available with two full DP objects and is broadly accepted everywhere apart from Manchester, with a sloping off of acceptance as we approach its sphere of influence.

This leaves the question:

What, then, for TGDs with the verb *give* and two full DP objects, which involves *transfer of possession*, not *motion*?

Here, I address the research question:

278

RQ (9) To what extent is there evidence for DPDP TGD in the location predicted from the Twitter data? Is there a grammar in the North West for which TGDs are underlyingly PDAT-P-DROP and is this grammar geographically restricted?

Bearing in mind that Biggs (2018) reports this structure as accepted in her Liverpool sample, it is perhaps surprising that we do not see broad acceptance there. Figure 5.11 shows instead that it is only weakly accepted by about 20% of respondents, with around 70% of respondents categorised as *weak reject*. Again, it seems—from these data at least—that the availability of P-DROP with verbs of *motion* does not predict the availability of TGDs with full-DPs with verbs involving *transfer of possession*.
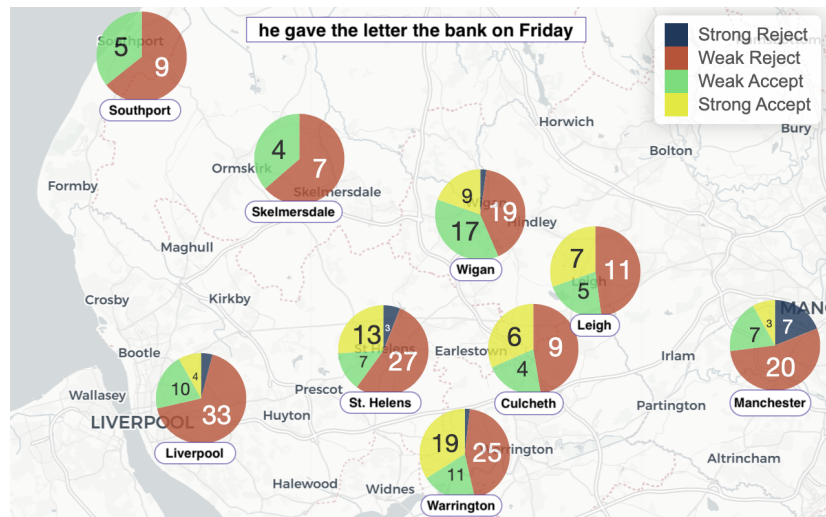
Figure 5.11: Geographical distribution of TGD with two DP objects
"I didn't know that <u>he gave the letter the bank on Friday</u>"

However, the structure is well accepted for many speakers in the interlying region - Culcheth, Leigh and Warrington. We begin to see a pattern emerging: locations that accept P-DROP-fullDP and pTGD/TGD-pDP also accept TGD-fullDP. It is this exact location that was identified in the geographical distribution of the Twitter data as being the likely location for a grammar that permits TGDs will two full DP complements (see Figures 4.63 and 4.64).

Biggs (2018) argues for the existence of an innovated null functional head (x) which permits preposition dropping across both *verbs of motion* and *transfer of possession* verbs, such as *give,* and underlies the pronominal form (_*it me*) for

those speakers. What these data appear to show is that if this grammar exists, it is located primarily in the region where all the conditions are adequately met; not so much in Liverpool, but in the region between Liverpool and Manchester. A possibility regarding Biggs's (2018) results, is that her informants were of the smaller set in Liverpool who do follow the inter-regional trend, or that they were otherwise influenced by it. We do see that there is *some* acceptance for the structure in Liverpool, after all.
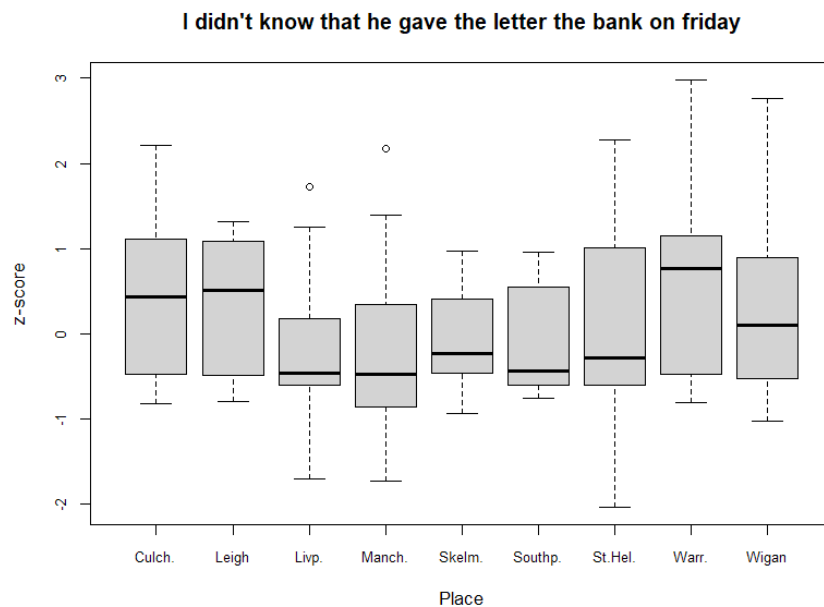
Figure 5.12: Boxplot of z-scores for TGD with two DP objects
"I didn't know that he gave the letter the bank on Friday"

These data suggest that learners who have abundant access to P-DROP in verbs of motion, by proximity to that structure's main host (Liverpool), and who also have access to the pronominal form of the TGD are able to port preposition-dropping to cases where TGD-*give* has two full DP objects, following a conventionalised interpretation of the data provided to them in their linguistic environment.

This suggestion is reminiscent of Hawkins's (2004) *Performance—Grammar Correspondence Hypothesis*, which states that:

> Grammars have conventionalized syntactic structures in propor-
> tion to their degree of preference in performance, as evidenced by

patterns of selection in corpora and by ease of processing in psy-
cholinguistic experiments. (Hawkins, 2004, p.3)

Note, that, full-DP TGD-*give* is *not* found in the corpus data. If it is used, it
seems that it is used infrequently. The connection to the *Performance-Grammar
Hypothesis*, is that it is the structures which appear to feed acceptance of full-DP
TGD-*give* that are frequent.

If we consider that $_{\text{DPDP}}$TGD is underlyingly a prepositional dative, there is
some reason to believe that it shares its underlying semantics which Goldberg,
Fauconnier, Lakoff, and Sweetser (1995) consider to be a "metaphorical exten-
sion" of the CAUSED-MOTION construction (p.89) where they label PDAT as a
TRANSFER-CAUSED-MOTION construction. We can see from this how, for some
speakers — who already have TGDs — the availability of P-DROP with *motion*
verbs may be transferred to PDAT, with a dropped preposition.

Another possibility is that acceptance of *transfer* TGDs with full DPs is driven
by a phonological process related to DAR. I return to respondent acceptance of
DAR shortly.

PD-DROP distribution

We saw that the distribution of PD-DROP (*go N*) is most strongly attested in
Manchester and locations in its immediate vicinity. Here, again the survey
data serve, in part, to corroborate the Twitter findings (RQ (8)).

We can see from the acceptance of the following two structures: *I'm going
shop* and *he's flying Germany*, that the geographic bounds are broader for ac-
ceptance ratings than they are for usage. Bearing in mind that respondents
are mostly in the 16-18 age range, this may be an indication of the ongoing
diffusion of the structure.
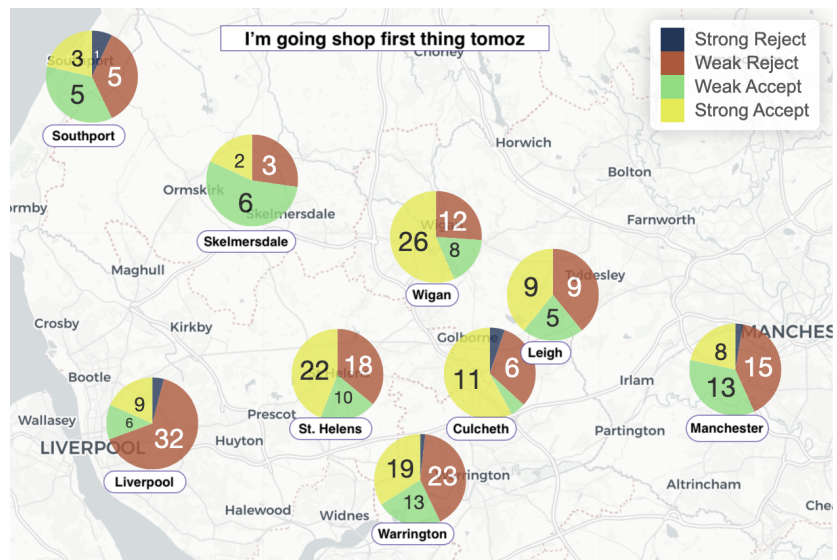
Figure 5.13: Geographical distribution of PD-DROP-*go*
"<u>I'm going shop</u> first thing tomoz do you want anything"

Liverpool is the only location that shows a marked overall dispreference for PD-DROP, which fits with what we expect if we consider that P-DROP is the dominant variant here.
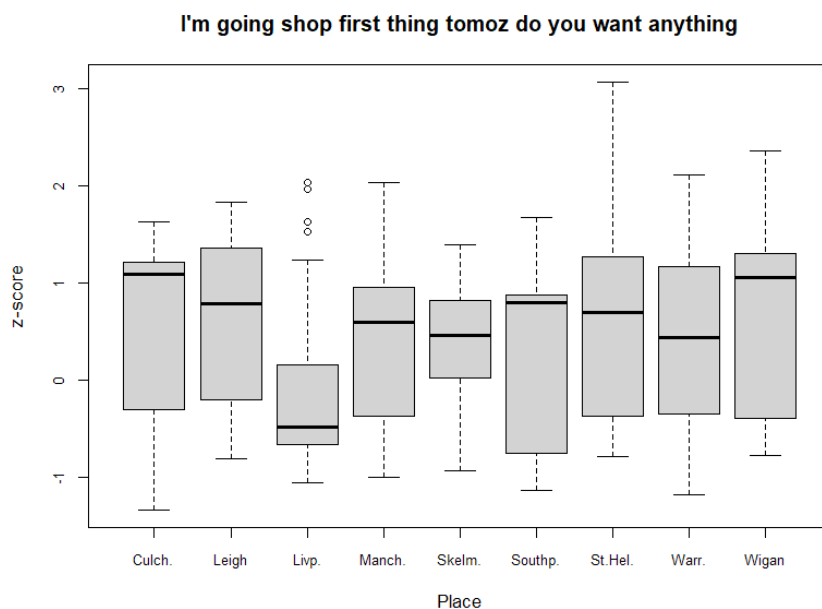
Figure 5.14: Boxplot of z-scores for PD-DROP-*go*
"I'm going shop first thing tomoz do you want anything"

Interestingly, the sentence *he's flying Germany tomorrow*, which does not have an overt determiner in the standard structure, is also disfavoured in Liverpool. This is, on first impression, counter-intuitive: given that P-DROP is available in Liverpool, we might expect that this sentence would be more widely accepted there.
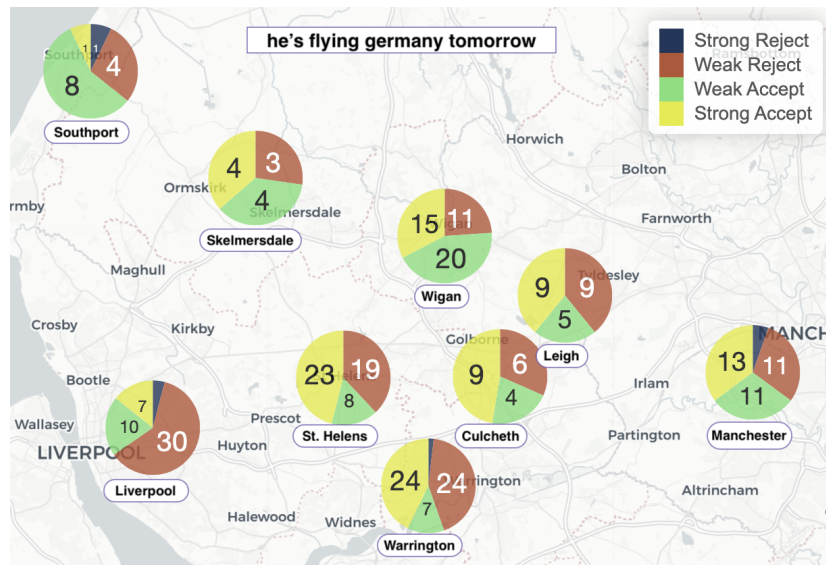
Figure 5.15: Geographical distribution of PD-DROP-*flying*
"He won't be coming - <u>he's flying Germany</u> tomorrow"

Such structures were found to be accepted in Liverpool (Biggs, 2018) and Ormskirk (Myler, 2013). It is not clear why it should be degraded in the current survey data for Liverpool specifically. However, one possibility is that is reflects a particular distinction where the GOAL noun is a *country*. We saw in the Twitter data that there is an (apparently recent) trend to 'insert' a determiner in such contexts, as in:

(2)   Going the Germany tomorrow.

Here, I suggested that the determiner has appropriated some of the function of the preposition in the process of reducing and assimilating with the preposition. Another possibility is a restriction on the verb *fly*. We have seen that, for some P-DROP varieties, such as in London, there is a narrow range of possible verbs that are permissible with P-DROP (Bailey, 2018b; Hall, 2019).
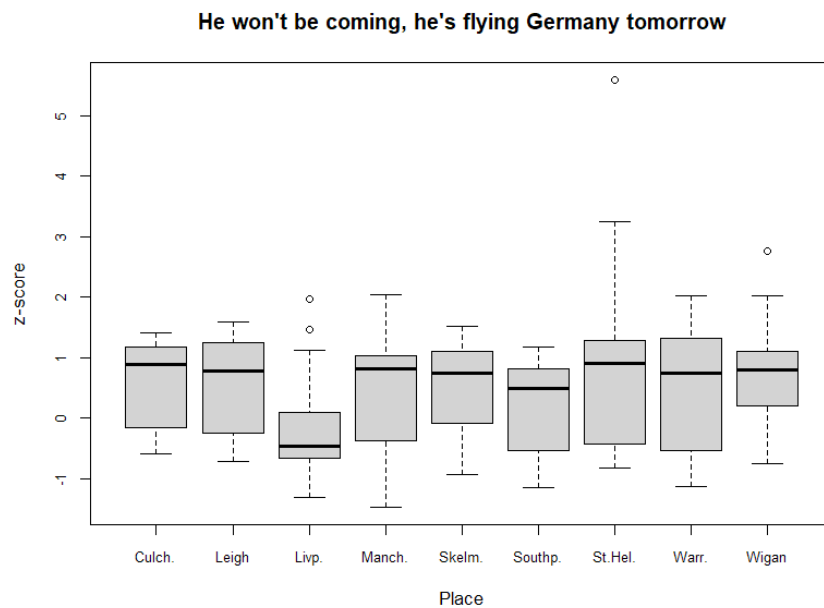
Figure 5.16: Boxplot showing z-scores for PD-DROP-*flying*
"He won't be coming - he's flying Germany tomorrow"

Definite article reduction

For *definite article reduction* (DAR), the sentence "are you heading t'pub later" is well accepted for many speakers outside of Liverpool, including in Manchester.
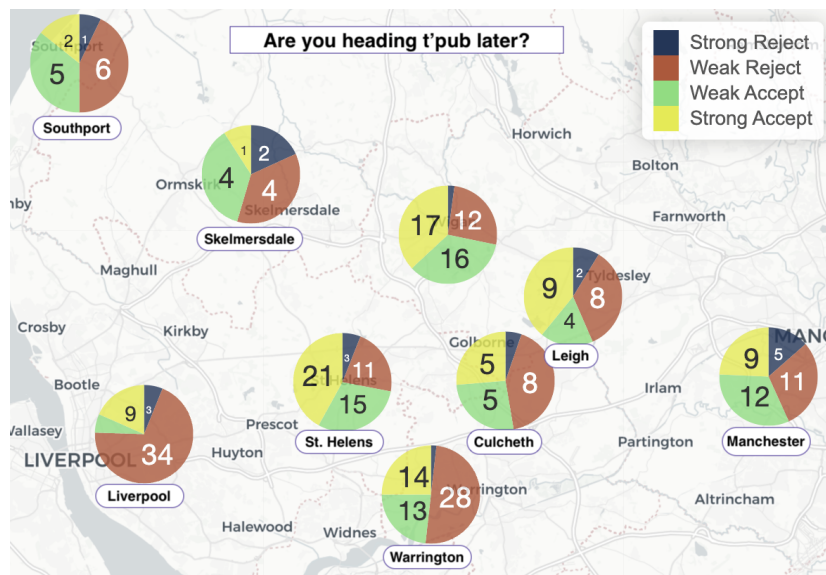
Figure 5.17: Definite article reduction in NWS

284

The form of DAR tested was DAR-THEME, where the article is reduced to [t], and the preposition is omitted. However, the test sentence *Are you heading t'pub later?* could also be interpreted as DAR-Ø/ʔ, where the abbreviated *t'* corresponds to a reduced preposition, and the determiner is absent or reduced to a glottal. Thus, respondents may be interpreting the structure either way.

Given the SED distribution, and what has been discussed in relation to DAR-θ distribution in the North West (that the P-DROP area identified on Twitter shows a close correspondence with DAR-θ), it would be good to see acceptance ratings of DAR-θ as well as DAR-ʔ. Unfortunately, DAR-θ was not tested.

Are you heading t'pub later
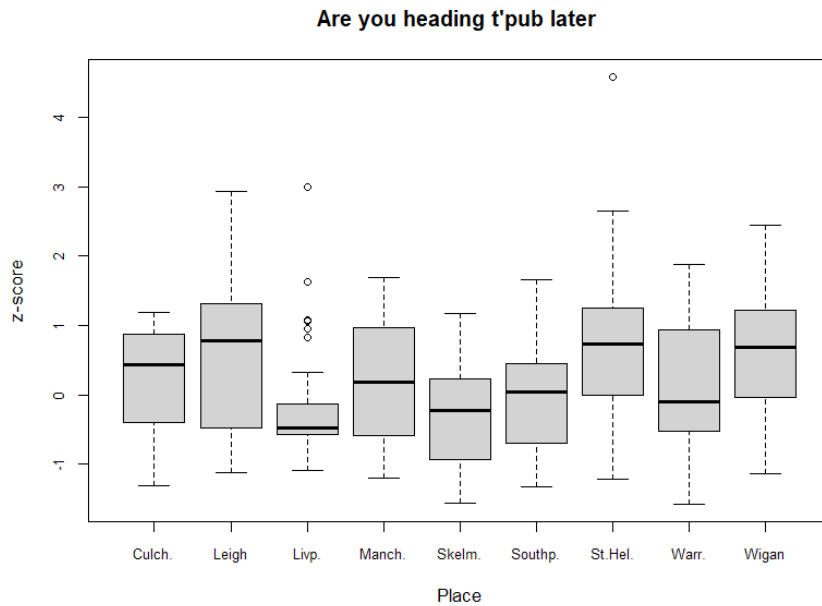
Figure 5.18: Definite article reduction in NWS

However, the distribution in acceptance of DAR-ʔ, shown here, is informative. Recall the SED map (presented in Figure 2.16, reproduced below). The area where DAR-ʔ is accepted in the survey corresponds to the area in the SED that is marked as [ʔ].
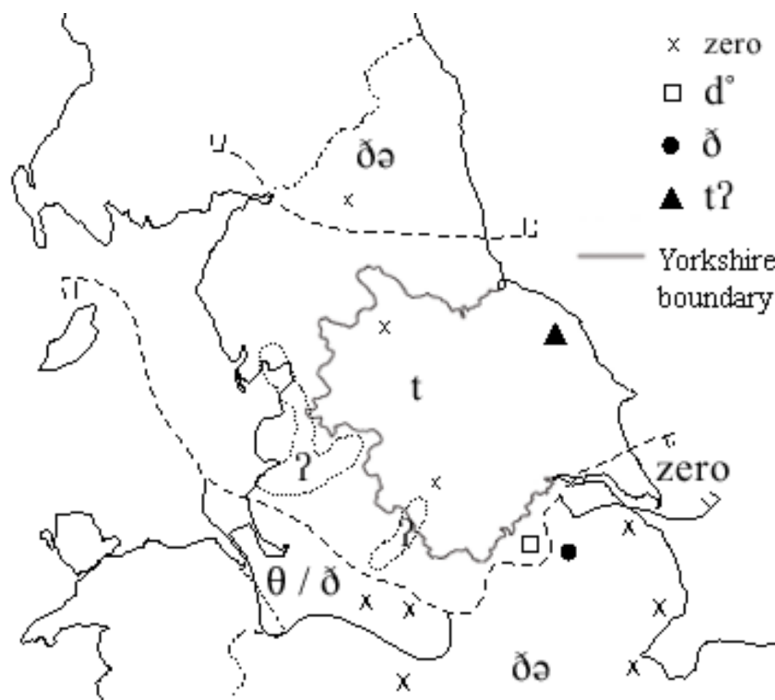
Figure 5.19: DAR map (shown again here).
Source: Barry (1972)

This is, again, predominantly the *inter-lying* region, between Liverpool and Manchester. What we are seeing here, then, is that speakers in this interlying region have access to the full range of grammatical options. If P-DROP is a direct descendant of DAR-θ in Liverpool, which, for the most part, no longer accepts DAR-θ, the *inter-lying* region, appears to have maintained both.

We return again to what is driving the availability of P-DROP and PD-DROP phenomena and to what may explain the acceptance of TGD-*give* with two DP objects (*gave the letter the bank*). The argument presented thus far aligns with Biggs's (2018) suggestion that it is the availability of both P-DROP and ₚTGDs that has resulted in an innovation which permits TGDs with two DPs. However, none of the research conducted in the North West has thus far taken the presence of DAR into account.

If we reconsider the sentence *gave the letter the bank* in the light of an interaction with DAR, the connection is not initially obvious. After all, there is no apparent reduction in the determiner. However, given the proposed process for a reanalysis of DAR-θ in Liverpool (θ→ *the*), a possible link emerges.

Consider the same sentence rendered as DAR-θ/DAR-ʔ:

286

(3)   a.   he gave th'letter (t')th'bank
       b.   he gave ʔ letter (t')ʔ'bank

The argument follows that learners in the interlying region have a number of options available to them that would result in acceptance of $_{DPDP}$TGD. They have P-DROP, which is likely a development from DAR *and* they have access to a still-active form of DAR, in addition to pronominal TGD.

P-DROP with *manner of communication* verb: *shout*

Contrary to Biggs's (2018) finding, preposition drop does not appear (for most speakers) to extend to ditransitives with *manner of communication* verb *shout*, which standardly does not occur in the GOAL-THEME order.

(4)   ??he shouted the crowd the results

These particular results do, however, align with Myler (2013), who found Ormskirk P-DROP to only be licit with *motion* verbs (*come, go, head*).
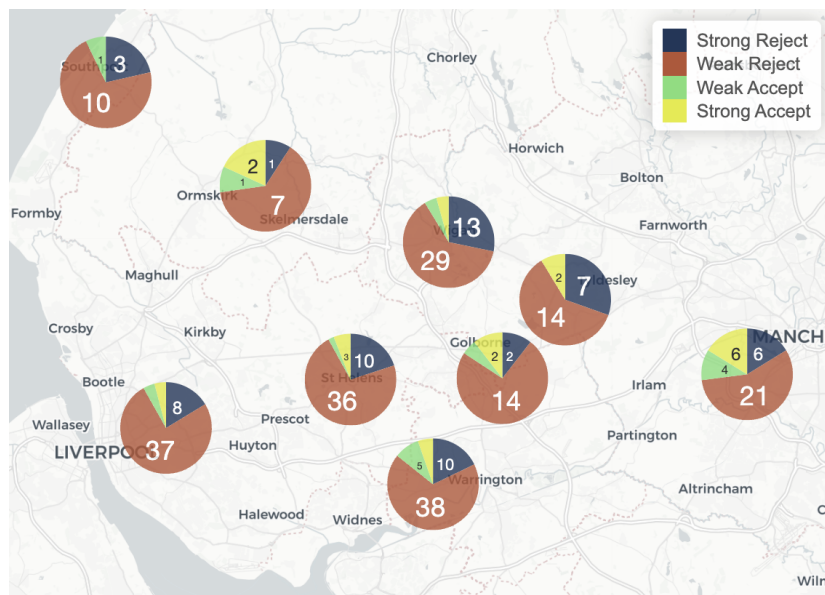
Figure 5.20: "After the game he shouted the results the crowd"

As a sanity-check, respondents did strongly accept the sentence *with* the preposition, across the board (the box-plot in Figure 5.21).

287
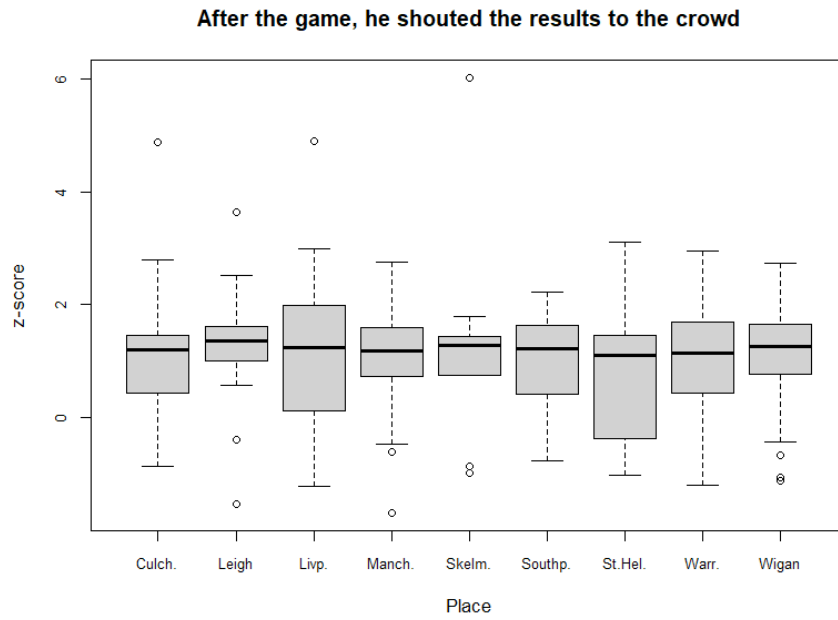
Figure 5.21: *Shout* without the preposition omitted.
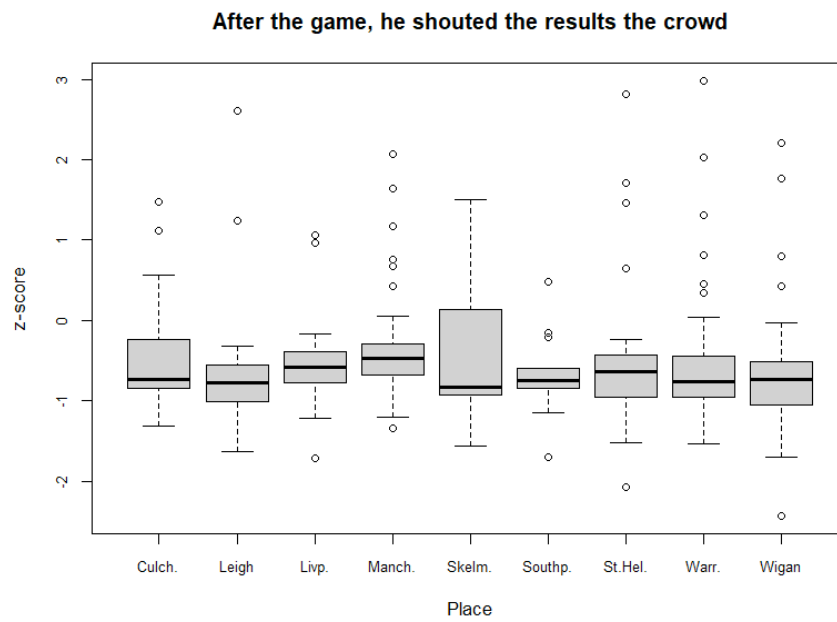
Figure 5.22: *Shout* with the preposition omitted.

In Figure 5.22, whilst overall ratings are low, we do still see quite a number of outliers; respondents who were apparently OK with dropping the preposition here. Acceptance, however, is not as strong, or widespread as with *give*, and does not pattern spatially with the P-DROP available with *motion* verbs *go* or *take*, in the way that it does with *give*. These results suggest that P-DROP is not simply phonological: if it were, there would be no reason why it should not be available with verbs like *shout*. Non-acceptance with *shout* fits with the diagnostic that P-DROP is preferentially available with verbs that carry a *motion* reading, one analogous to it, such as with *give* carrying a potential MOTION-CAUSED-POSSESSION reading.

Summary of grammars

Grammar	ₚTGD	ₚₒₚTGD-*give*	ₒₚₒₚTGD-*give*	P-DROP-*go*	PD-DROP	DAR
1. Inter-reg	✓	✓	✓	✓	✓	✓
2. Manch.	✓	x	x	x	✓	✓
3. Livp.	x	x	x	✓	x	?

Table 5.4: Summary of grammars in the North West.

Evidence of silent/NULL prepositional material

If northern PD-DROP is distinct from southern PD-DROP, in that it does not involve the radical absence of preposition and determiner, then we expect it to be possible to get a definite reading of the GOAL noun. Recall that there was substantive evidence for a distinction along these lines in the Twitter data.

Figure 5.23 shows that there is a notable degree of acceptance where the GOAL is modified such that it refers to a specific place:

(5)   can we go [park with the fountain]
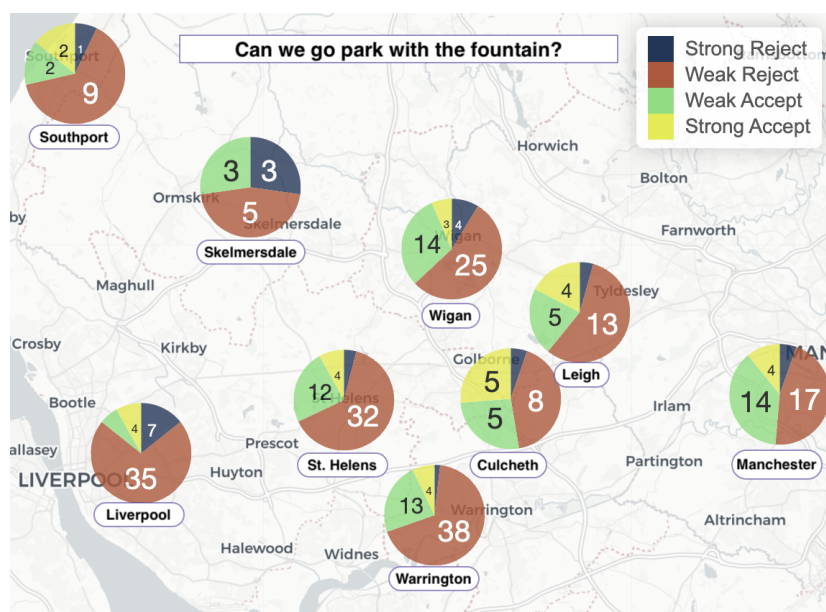
Figure 5.23: "can we go park with the fountain?"

These results add further support to the claim that North West PD-DROP does not follow the same constraints as that found in the South.

Straight-modification

Straight-modification (SM) with P-DROP is well accepted for most speakers. It is accepted by fewer respondents in Manchester, which is expected for P-DROP in general.
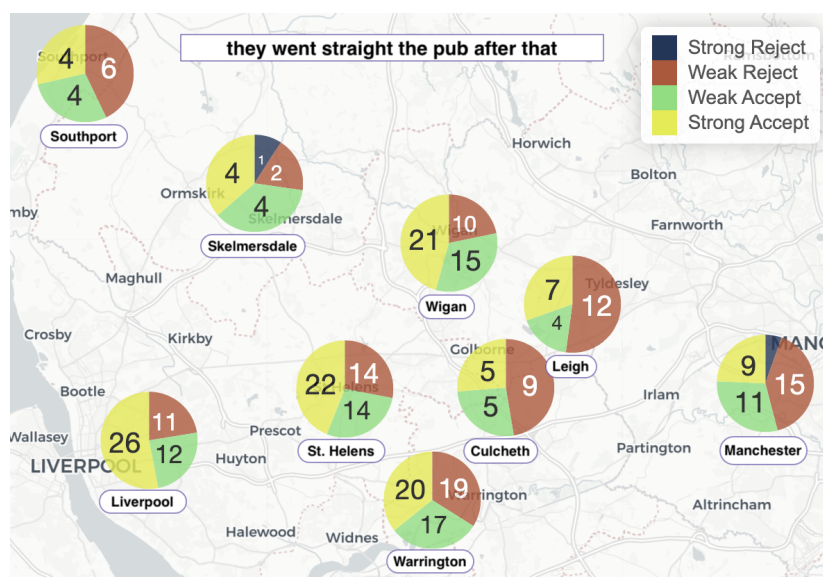
Figure 5.24: *Straight*-modification: Responses to "no questions asked, <u>they went straight the pub</u> <u>after that</u>"

While Myler (2013) reports that *straight*-modification is not possible in Ormskirk, this was based only on the judgements of two native speakers.
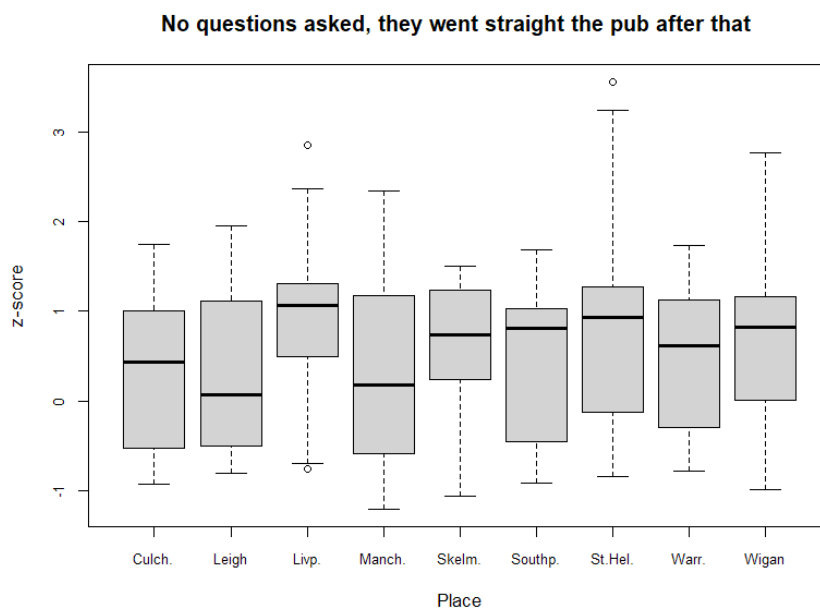
Figure 5.25: Boxplot showing z-scores for *straight*-modification: "no questions asked, <u>they went straight the pub</u> after that"

The widespread acceptance of SM with P-DROP across the region is strong evidence for the existence of silent PP structure, rather than incorporation to the verb, as suggested by Myler (2013) for P-DROP and Hall (2019) for London PD-DROP. Rather the data suggest an alignment with Biggs's (2018) proposal that there is a null functional head of some description for P-DROP.

Stative construction: preposition *at*

Stative P-DROP is not well accepted in Liverpool, but widely accepted inter-regionally. This is unexpected, given that Biggs (2018) reports it as licit in her Liverpool sample.
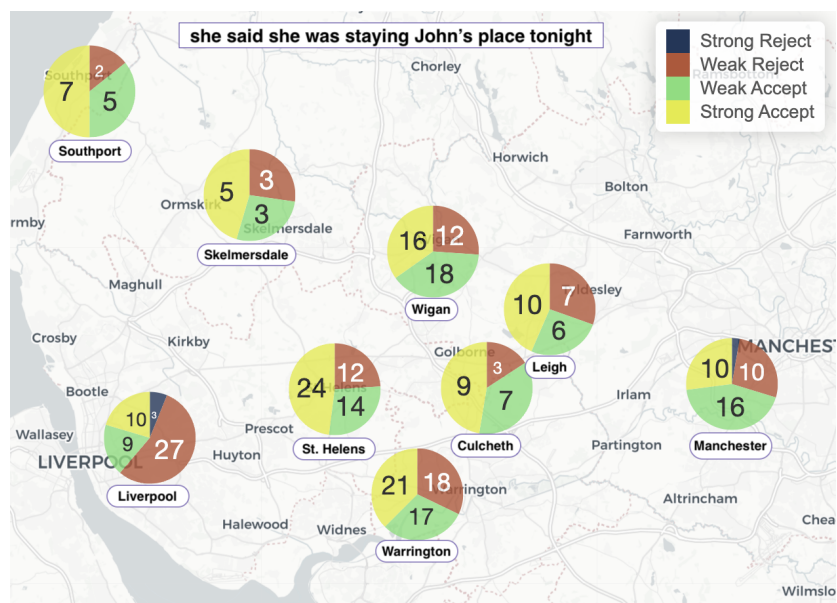
Figure 5.26: Geographical distribution of stative *at*-dropping:
"She said <u>she's staying John's place</u> tonight"

This is somewhat puzzling, although these results do fit with the proposal that the grammar Biggs (2018) was describing for Liverpool exists more in the inter-regional area than in Liverpool itself.

Quantifier scope

Outside of the standard acceptance rating questions, the survey also probed the semantic interpretation of the sentence in (6). The point of this was again to test the kinship of $_P$TGD with GTD.

(6)   If it didn't hurt so much you know I would <u>give it you all</u>

The idea was that if ₚTGD is a true DOC, that is, equivalent to ₚGTD with a reversed order, then (6) *all* would scope over the THEME rather than the GOAL, as it does standardly with the ₚGTD carrying the meaning: "to give all of something to someone" (7)

(7)   If it didn't hurt so much you know I would <u>give you it all</u>

If, on the other hand, ₚTGD were underlyingly ₚPDAT, then we would expect (6) to carry a different reading, where *all* scopes over the GOAL rather than the THEME, as in (8)

(8)   If it didn't hurt so much you know I would <u>give it to you all</u>

If a speaker has access to both P-DROP and TGD-asGTD, then we might expect (6) to be ambiguous between the two interpretations. If a speaker has neither P-DROP or TGD-asGTD, then we expect them to find the sentence to be ungrammatical.

   Participants were asked to select from one of four options on the meaning of (6):

1. everything is given to someone
2. everyone is given something
3. either of the above
4. neither of the above

   The results, plotted in 5.27 show a clear preference for the *all-to-one* reading for the majority of respondents with a remarkable consistency between places. These results again suggest that ₚTGD is better understood as underlyingly ₚGTD for most speakers.
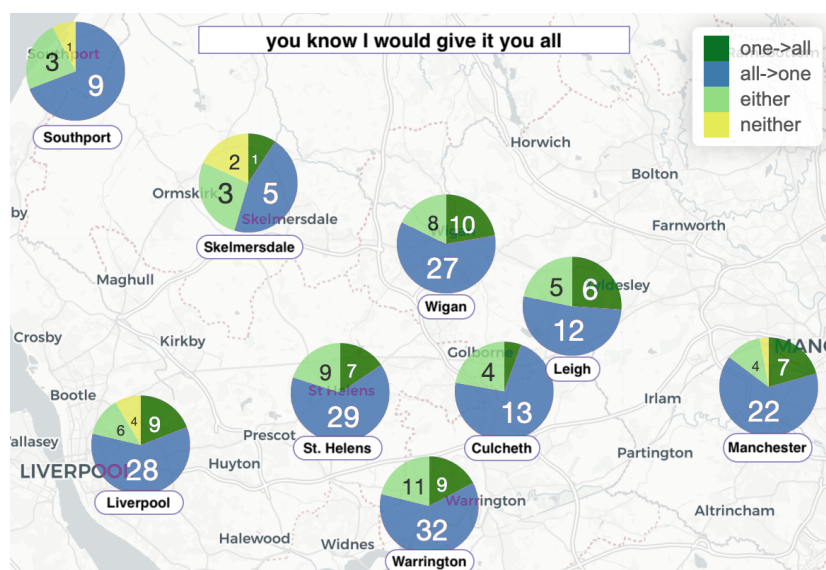
Figure 5.27: Quantifier scope: "If it didn't hurt so much you know I would give it you all"

A sizeable proportion of respondents favour the *one-to-all* or *either* readings, indicating that there do appear to be some speakers for whom ₚTGD is, or *can be* arrived at via preposition drop. This finding corresponds to Haddican (2010), who finds inter-speaker variability, and Biggs (2018), who finds that TGD-as-PDAT in Liverpool, with the qualification, that she may not be describing a grammar located exclusively in Liverpool, but one that is found in the North-West, coexisting with the grammar found in Manchester.

Theme passives

Finally, we look at the distribution of TGDs versus GTDs in active and passive contexts. Recall Haddican and Holmberg's (2012) finding that there was a correlation between acceptance of THEME-passives and acceptance of TGDs in their Manchester sample.

(9)   Active TGDs

    a. 'It's a scanner/Printer thing. <u>Someone gave it me</u>'

    b. 'I should tell you, <u>we gave it the girl</u> already'

    c. 'I didn't know that <u>he gave the letter the bank on friday</u>'

(10)   Active GTDs

    a. '...I'm listening to the Beach Boys album, <u>John lent me it</u>'

b. 'I should have told you, <u>he lent the guy it</u> after all'

c. 'Apparently <u>he sent the office the proposal</u> yesterday'

(11) Theme passives

a. 'They mentioned that <u>it was given her</u>'

b. 'I'm not certain but I think that <u>the voucher was sent him</u>'

(12) Goal passives

a. 'It was clear that <u>she was given it</u>'

b. 'There's no doubt that <u>he was sent the book</u>'

Next, we take those respondents with an active contrast above zero, which corresponds to a preference for TGDs over GTDs. Following (Haddican & Holmberg, 2012), the prediction here is—if TPASS (11) is derived from TGD (9)—that speakers who prefer TGD should show a higher rating for TPASS. That is, we should find a correlation between THEME-GOAL orders in passive and active contexts. This was found to be the case in Haddican and Holmberg (2012) in Manchester. The aim in the present section, then is to assess whether the prediction holds out across the wider region.

Figure 5.28 shows that the prediction is largely borne out. Here, 'Active contrast' represents the mean z-score for TGD (9) minus the mean z-score for (10), by participant.

Contrast in z–scores between TGD and GTD
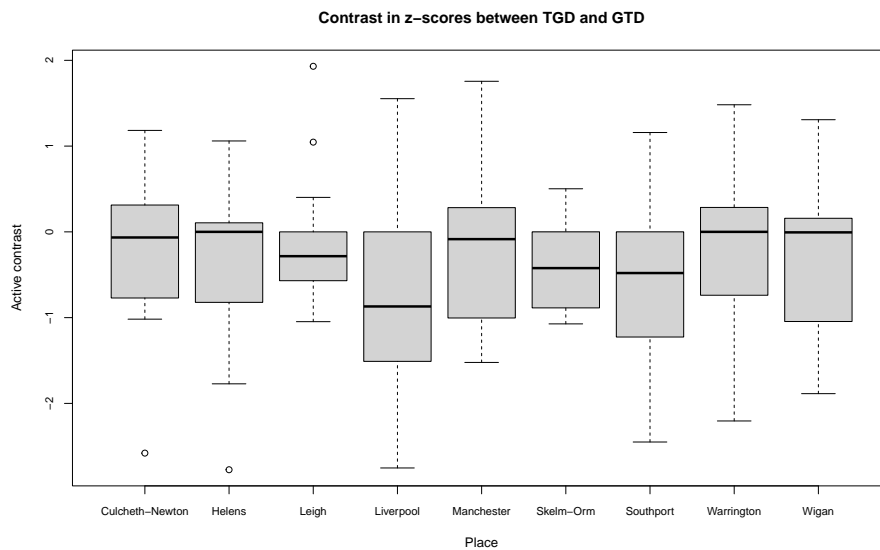
Figure 5.28: Boxplot showing active contrasts by place. Zero represents parity in mean acceptance between TGD ("gave it me" and "gave it the guy") and GTD ("gave me it" and "gave the guy it").

Contrast in z–scores between Tpass and Gpass, between TGD and GTD preferers.
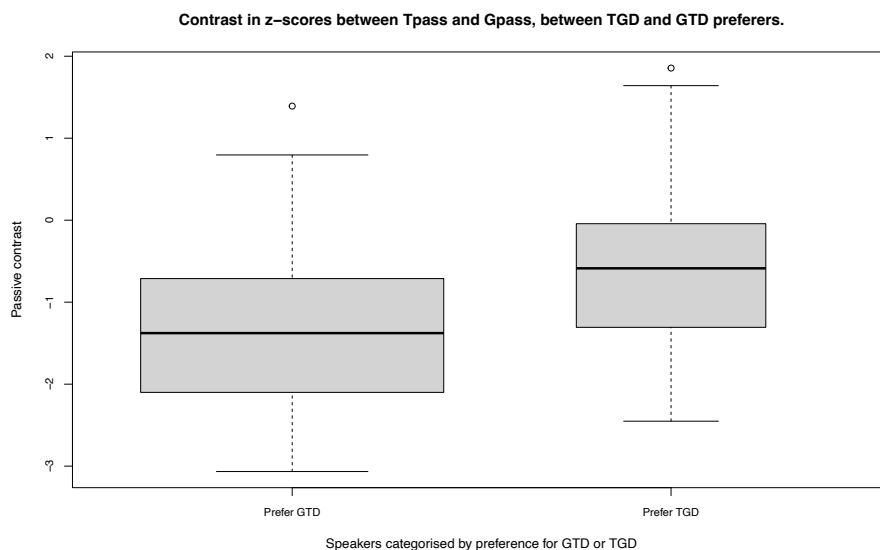
Figure 5.29: Boxplot showing passive contrasts by speakers with preference for GTD or TGD. Box width corresponds to number of respondents in category. < 0 represents a preference for GPASS; > 0 represents preference for TPASS; zero represents parity.

We can see from Figure 5.30, that there are very few points in the top right of the plot which correspond to a high TPASS and low TGD rating (grammar 4), fitting our prediction.

Correlation between active and passive contrasts, all places
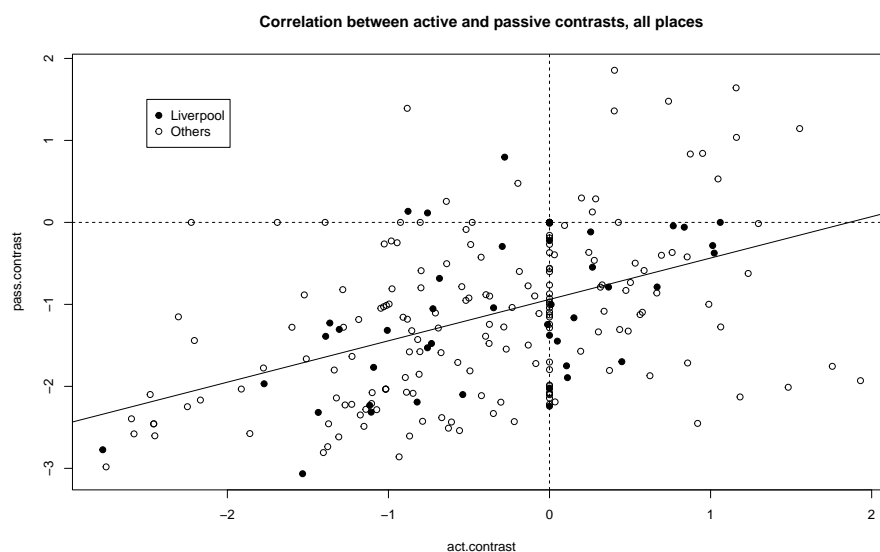
Figure 5.30: Correlation between active and passive contrasts, across all places in the North West, following Haddican and Holmberg (2012). Filled points correspond to Liverpool speakers. Zero represents parity in acceptance between types.

DP-Theme passives

A key observation in Biggs (2018) is the availability—in Liverpool—of theme-passives and active TGDs where one or both objects are full DPs. This runs counter to Haddican's (2010) finding that such constructions are ungrammatical for most speakers in his Manchester sample. As Haddican (2010) uses the unavailability of fullDP TGD and TPASS to argue against pronominal TGD involving preposition drop, so Biggs (2018) takes their availability in Liverpool as an argument in its favour.

As mentioned, the fact that Haddican's sample includes some speakers who do accept fullDPs in these constructions, and Biggs does not supply individual speaker responses, leaves open the possibility that both Manchester and Liverpool have some speakers with each acceptance pattern – and by extension – with each grammar. If this were the case, we would expect to see a comparable proportion of each response type in each place and, further, given that the underlying mechanism is likely opaque to learners, we would not expect to see an association of a given type with a given place as we do with the overtly variant structures under investigation. And this is, indeed, what the data appear to show (figure 5.31).

Given Biggs's (2018) finding that TPASS with full noun objects is acceptable

to her Liverpool informants, it is surprising that it is so heavily degraded in the Liverpool group in NWS (Figures 5.31 and 5.32).
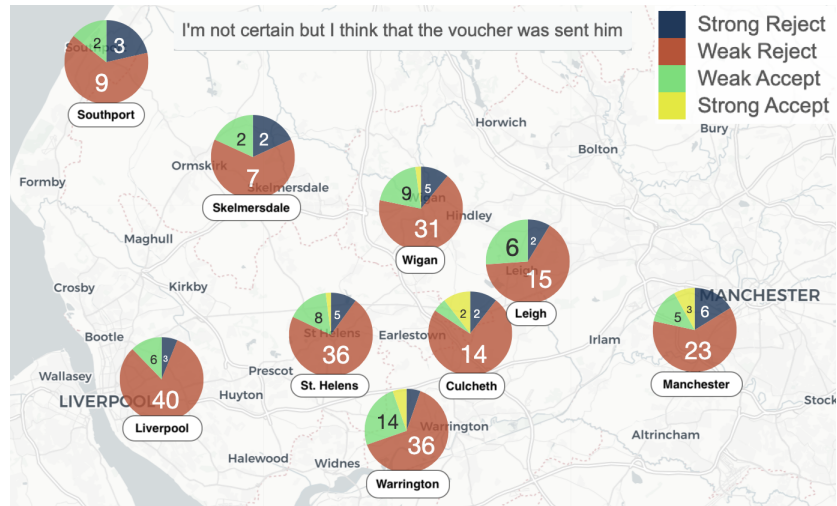
Figure 5.31: Geographical distribution of theme passives with full DP
"I'm not certain but I think <u>the voucher was sent him</u>."

Whilst TPASS with full DPs is not well accepted on average across the region, which corresponds to Haddican and Holmberg's (2012) findings in Manchester, we do see evidence that a subset of speakers accepts the construction.
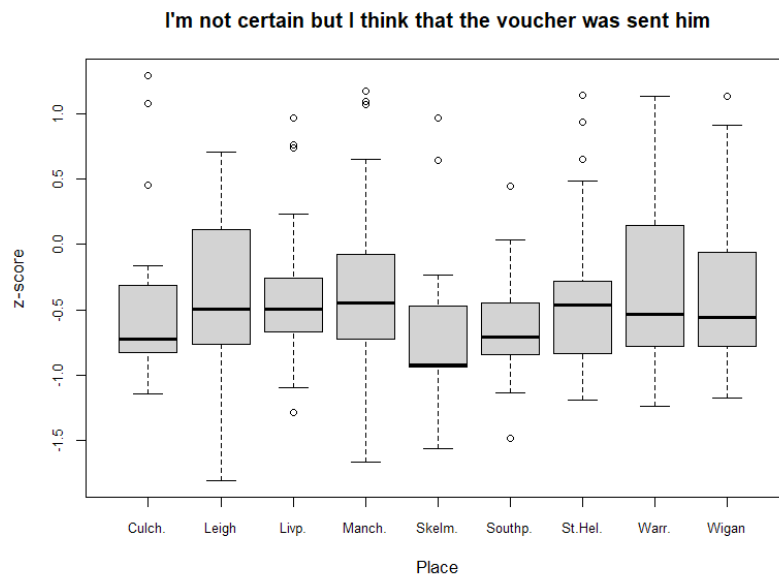
Figure 5.32: Boxplot of theme passives with full DP
"I'm not certain but I think <u>the voucher was sent him</u>."

Corpus evidence for TPASS is also low; there were scant examples in the Twitter datasets. It may be that TPASS is associated with older, regional speech patterns.

Again, the boxplot in Figure 5.32 shows a median z-score of around -0.5, with little variation between places. It is notable however, that Leigh and Warrington, particularly exhibit more variation in acceptance, with a greater range of responses. There are additionally quite a number of *outliers*, represented by the small circles above the boxes, which indicate respondents who gave high acceptance ratings; many with a z-score of above 0.5.

These results are consistent with Haddican and Holmberg's (2012) proposal of the existence of multiple grammars in the region, one of which permits TPASS.

5.8 Chapter summary

The data presented in this chapter confirmed:

1. That the distribution of pronominal THEME-GOAL ditransitives and P-DROP align with the usage frequencies found in the Twitter corpus.

2. That the availability of $_{\text{DPDP}}$TGD does occur in the predicted inter-regional area where the conditions for its occurrence are fully met.

3. That P-DROP is only available where *motion* to GOAL is implied for most speakers, and by analogy in the case of *give* in the interlying area.

4. North-West PD-DROP appears to have different constraints to that found in the South East, which aligns with the patterns found in the Twitter data.

5. That the availability of TPASS is dependent on the availability of TGD in active contexts, following Haddican and Holmberg (2012), and that this holds across the region.

5.9 Conclusion

Overall, the results presented in this chapter demonstrate the potential for working with Sixth Form students to gather these kinds of data. It is a shame that it was not able to fully realise this potential due to the incompletion of the community fieldwork aspect of the project, as a result of the Covid-19 lockdown. However, the successful completion of the fieldwork component by one

school demonstrated the workability of the approach. There is considerable potential here for further collaboration in this way, that could offer substantial research opportunities; combining important school *Outreach* work with the gathering of useful data.

The survey also demonstrates a way that large-scale social media data may be supplemented with data gathered through other means, and that the resulting data may be integrated effectively. This was one of the driving motivations for the initial conception of the survey; considering how Twitter data may be corroborated outside of comparisons to previous corpora.

Responses were predictable from the usage data, both in terms of where we would expect a known variant, such as $_\text{P}$TGD to be accepted and not accepted, and in terms of where we expected the wider grammatical conditions for acceptance to be met. There were some surprises, such as the degraded acceptability of *flying Germany* in Liverpool, and THEME-passives, which had been reported as accepted in previous literature.

The results were shown to have a good level of consistency: respondents gave predicable ratings, for example, where sentences were standard (without any dropped prepositions), responses were well accepted across the board, indicating that they were paying attention, and not suffering judgement fatigue.

Finally, the results presented in this chapter are a subset of the overall results. There are additional data present in the survey which may shed additional light on some of the findings. Specifically, it would be interesting to see the extent to which there are correlations in individual responses between structures and phonetic data (such as *book-spook*).

CHAPTER 6

Discussion and conclusion

6.1 Introduction

This thesis examined the geographic distribution of the variants of *ditransitives,*
preposition dropping, their structural variability, and the extent to which there
are localised connections between them. In the case of preposition dropping,
a case was made for a historical connection with *definite article reduction.*

Here, I most directly addressed the question:

RQ (2) What is the comparative geographical spread of variants of (a) the pronom-
inal ditransitive (pDit) and (b) preposition/determiner dropping (P-Drop)?

As a recap, ditransitives exhibit variable word order in the UK and Ireland
where both objects are pronouns. The extent of this variability was demon-
strated with the mapping of variants found on Twitter, across a wide range of
verbs. The extent of the variability was shown to be delimited by verb type,
following previous assertions, such as latinate verbs (e.g. *explain, deliver*) and
denial of possession verbs (e.g. *refuse, deny*). Further, the distribution of verb
by type showed that, for the most part, THEME-GOAL pronominal ditransitives
pattern like those with the GOAL-THEME order, suggesting that they are derived
not by preposition-drop, but by a reversal of objects.

Variants of preposition and determiner dropping with *intransitive verbs of motion towards a directional goal* (*go*, *come*) were also shown to exhibit tractable geographic variability across both the UK and Ireland. Here, the distribution was compared to that found by previous work (Gopal et al., 2021) and to the distribution of *definite article reduction* (DAR) reported in the Survey of English Dialects. The possible connection between DAR and [P/D-DROP] variants was further explored through a structural analysis of texts from traditional corpora and tweets.

By gathering two types of location data, which could be compared in the North West, and looking at the change in time over the ten-year period, I also addressed the question of data reliability:

RQ (1) How reliable are geocoded Twitter data? To what extent do they correlate with Twitter data located based on user-reported place of origin? How have Twitter data changed over time?

Here, I showed that Twitter geolocation data do function, for the most part, as an adequate proxy for user place of origin. This said, user place-of-origin remains preferable. The method used for getting place of origin data; finding instances where a user reports having *grown up* in a place, and then finding other texts that they produced, is, in fact, more scalable to alternative platforms than relying on geolocation data. This is important in the light of the fact that, since the takeover of Twitter by Elon Musk, Twitter data, which was the only real source for geolocated social media, are now prohibitively expensive to access. For instance, platforms such as *Reddit*, *Tiktok* and *YouTube* do not offer geographic data (at least not with any granularity), but it would, in theory be possible to locate users based on this kind of self-reported location data. The method used in NWatlas-corp showed that instances of users reporting where they *grew up* or where they are *from* are surprisingly common.

6.2 Overview of findings

6.2.1 Geohistorical and structural distributions

Regarding the present-day distribution of the phenomena, I explored some of the potential geohistorical underpinnings:

RQ (3) What historical factors may explain the geographic distribution uncov-

ered in (2)?

Geohistorical distribution of [DATALT]

Cluster analysis of the wider distribution of variants of the pronominal ditransitive revealed distinct regional groups which likely have substantial time-depth. The distribution of the $_p$GTD was shown to align with that of place names which have Scandinavian suffixes, lending support to Gast's (2007) proposal that the variant may be connected with Old Norse contact. This claim remains speculative, and perhaps will remain so, given a lack of geo-locatable historical corpora at the kind of scale that would be required. On the other hand, it is difficult to find a better explanation for such a close geographical alignment.

$_p$TGD distribution revealed that there is a contiguous region of use which closely aligns with the region identified as $_p$TGD-dominant in the *Survey of English Dialects* (SED) (Orton & Dieth, 1962), as mapped by Kirk et al. (2014). A sharp boundary was identified on the eastern edge of this region, where $_p$TGD gives way to $_p$GTD. The focus of the present study was on the North West, but there is substantial scope for further study on this eastern boundary. In particular, there is a rapid shift in relative use between the cities of Sheffield and Doncaster, which are only 15 miles apart, presenting a promising site for further study. The same is true across the north-eastern boundary, for instance, between Halifax and Bradford.

Structural distribution of [DATALT]

Regarding the structural distribution of [DATALT], I presented data which broadly supported Haddican's (2010) proposal that the $_p$TGD is akin to $_p$GTD rather than $_p$PDAT with a dropped preposition. This addressed:

RQ (4) What quantitative evidence can be brought to bear on the TGD-as-GTD question?

Here I showed that latinate verbs (*explain, deliver*) favour, as has been reported (Levin, 1993), almost exclusively $_p$PDAT regardless of geographic location. Importantly, there were very few instances of the $_p$TGD with these verbs even in locations where $_p$TGD is otherwise dominant. This is not expected if $_p$TGDs are underlyingly $_p$PDAT. There were some indications of inter-speaker variability here, however, and evidence that $_p$TGD is at least more likely to occur with

latinate verbs than ₚGTD.

Geohistorical distribution of [P/D-Drop]

The exploration of [P/D-Drop] distribution led into the question of the extent to which it is compatible with the distribution of definite article reduction, and if there could be a transformational connection:

RQ (5) To what extent do [P/D-Drop] variants align with DAR distribution?

A key finding in this thesis was that there does seem to be a connection between [P/D-Drop] and *definite article reduction*. The data show that there is a marked geographical correspondence between DAR and P-Drop/D-Drop. A connection between the two phenomena is additionally supported by corpus evidence, and a mechanism for deriving P-Drop/D-Drop from DAR is relatively straightforward.

This discussion led directly to the question of whether PD-Drop had undergone a recent diffusion from the South East towards the North West:

RQ (6) Was PD-Drop independently innovated in multiple locations, or did it spread from one point of innovation in the South East of England?

The situation for PD-Drop is not so clearcut, but evidence was presented which show that PD-Drop is older than has previously been suggested, and its presence in the South-West amongst older speakers adds another dimension to the proposal of a (recent) diffusion from the South East. It seems that a development towards PD-Drop is ripe in English, occurring to various degrees in multiple locations, but that (East) London has more recently acted as a particularly dominant engine for growth.

Viewing [P/D-Drop] as variable with four possible variants made it possible to identify distinct geohistorical regions. Variation in [P/D-Drop], has only recently come to light. Whilst determiner reduction and absence has been well documented in the DAR literature, preposition dropping has received less attention until more recently. P-Drop and PD-Drop are however not new; the frequency of attestations in traditional corpora are comparable to that found for [DatAlt]. Recent interest in [P/D-Drop] may have resulted from a surge in PD-Drop usage that has taken place in London, likely spearheaded through the emergence of what has been dubbed *Multicultural London English*.

It is likely this *surge* in use that has since undergone a process of diffusion from London. In the North, the ground for PD-DROP appears to have already been laid. A London surge did not introduce a new feature, rather it licensed the use of an extant feature, or set of features. The linguistic conditions in the North were such that they may not have *required* London influence.

Regarding P-DROP, specifically I showed how the distribution pattern is best explained by contact phenomena that would have been most intensive during the early-to-mid 19th Century. There is considerable scope for further investigation here, but the results do indicate a likely point of innovation for P-DROP which is further back than has previously been indicated.

6.2.2 $\varkappa$ and DAT-θ: connection?

I mentioned in the results that it seems that Biggs's (2018) approach to the P-DROP found in Liverpool, where an inherently null functional element inherits some, but not all, properties of overt prepositions, offers a clearer potential structural continuity between DAR-θ and P-DROP. I elaborate on this now.

The first point is that both P-DROP and PD-DROP exhibit *straight*-modification. That is, they are both possible with the adverbial *straight*, which must be modifying some kind of functional head, which is doing the work of an overt preposition.

(1)   a.   When you <u>go straight the airport</u> from a night out..
          b.   <u>I'll go straight bookies</u> mint odds them

I take this as evidence both of a null functional head *and*, that this head is historically linked to DAR which is, itself, variably available between speakers.

In the case of (1a), the historical root is DAR-θ/DAR-ð. We saw evidence of this in the striking geographical overlap between occurrences of tweets containing structures like (1a), and DAR-θ/DAR-ð. I also presented evidence from of still existent DAR-θ in north-west tweets, fulfilling the same structural role.

(2)   I need <u>go th'opticians</u> or get a labrador.

Crucially, in such examples, the preposition is absent. Here, I suggested that the preposition has assimilated to the reduced determiner.

Further, I suggested that there has been a process of reanalysis of this kind of structure which likely occurred, during the time that Liverpool English was

itself forming, where the reduced determiner-preposition element was reinterpreted as *the*:

(3)   t' th' → th' → *the*

This, I suggested, may explain the phenomenon we see now, in North West English, where *the* is used where there is used in lieu of a preposition, where we would not usually find a determiner:

(4)   I'm going the Wales tomorrow

A connection with DAR also explains the availability of P-DROP with a broader range of prepositions than that found in Biggs's (2018) study. Biggs found P-DROP to only be available with prepositions *at* and *to*. I showed examples of its availability with *on, in* and *for*.

(5)   a.   a picture of Jesus arrived the post
       b.   I put it the correct bin
       c.   Ok waiting the last people
       d.   they arrive the scene

These data are important, we know that DAR is available across a wider range of prepositions than just *to* or *at.*

Regarding (1b), we see a complete absence of overt preposition and determiner, but the availability of *straight,* is again strong evidence that the structure is both distinct from south eastern PD-DROP, *and* that it, too, is historically connected to DAR. This connection, I argued, is at odds with the idea that northern PD-DROP arrived *only* via diffusion from the South-East.

6.2.3 Survey

Here I first addressed the question of a continuity between survey and Twitter results:

RQ (8) To what extent do speaker judgements correlate with corpus frequencies for a given location?

Here, results were mixed: $_p$TGD and P-DROP phenomena did, indeed correlate with corpus frequencies. However, $_p$GTD, which is frequently used only in

Liverpool (in the North West), was actually broadly accepted across the region. This could be a result of the ₚGTD's relatively recent rise to prominence as a canonical variant (Gerwin, 2014). It may also indicate the beginnings of a shift towards the Liverpool pattern (following the broader reach of P-DROP), and perhaps the early stages of diffusion.

The second question addressed in the survey was whether we would find evidence for the ₚₚₚₚTGD in the predicted area, based on that location meeting the principal conditions for it to be present:

RQ (9) To what extent is there evidence for ₚₚₚₚTGD in the location predicted from the Twitter data? Is there a grammar in the North West for which TGDs are underlyingly PDAT-P-DROP and is this grammar geographically restricted?

The finding that the predicted area did indeed show a marked acceptance for ₚₚₚₚTGD, while areas outside did not, was a particularly interesting outcome. However, this area was also shown to exhibit a uniquely-wide acceptance of DAR, which may also play a role in the acceptance for ₚₚₚₚTGD. In any case, this finding prompts additional research into this particular subregion, which was shown to be particularly active in terms of socio(linguistic) variability.

The final question asked in the survey built on the findings presented in the Twitter corpora:

RQ (10)   Is PD-DROP in the North West akin to that found in the South East?

I showed that PD-DROP in the North West was attested under conditions that we would not expect for South-Eastern PD-DROP, such as pluralised GOAL nouns, *adjectival modification* and *straight*-modification. These data suggested that northern PD-DROP is, in fact, not akin to South Eastern PD-DROP. This finding additionally cast some doubt on a diffusion model from SE→NW.

The survey data supported this finding, showing that *definiteness* is possible with NW PD-DROP for a sizeable proportion of speakers (as in *can we go park with the fountain*).

This said, some doubts remain. It is not completely clear that the constraints on SE PD-DROP are as strict as have been reported (Bailey, 2018a); there are indications that the tight conditions reported for South East PD-DROP are, in fact, considerably broader than has been suggested. Bailey (2018a), shows

data in a conference presentation which suggest that PD-DROP is available, for instance with a wider array of prepositions, and GOALs. Additionally, without an unrestricted Twitter corpus for the South East, it remains an open question if we might find examples there which do align with those presented here for the North West. This is, again a potentially useful line of enquiry in a currently-developing research topic.

6.3   Strengths and Limitations

Search then download / Download then search methods

One of the key benefits of the data gathered for the UK and Ireland was the fact that it was able to cover an entire ten-year period across all nations by searching for list of possible strings for each variant of [DATALT] [P/D-DROP]. This allowed for detail that has not been seen before and for the revealing of the changing nature of Twitter data over time. However, this method was also a limitation: relying on set strings, however exhaustive, made it impossible to search for patterns in the data that were not included in the initial search.

Having a third, *unrestricted* corpus for the North West, mitigated this limitation for that region, allowing for a closer examination of phenomena there.

Amount of data available

The amount of data available was also both a key strength, and in some sense, a limitation. Whilst it revealed a wide range of unexpected facts, it was not possible, in one project, to cover all of these facts.

Part of the motivation for the project was to create an atlas for the wider UK and Republic of Ireland, to find specific areas that may warrant further study and then demonstrate such a further study in one of those areas. Several possible areas were highlighted, and the North West was selected as a case study. However, as discussed, it is problematic to view one region in isolation, particularly when the claim is that a given structure follows distinct constraints to that reported in another region.

As a result, the project has perhaps raised more questions than it has answered. This is, however, arguably a good result in that it prompts the need for further research.

Key limitations of the survey

The survey is not exhaustive, lacking the full spectrum of minimal pairs and other comprehensive elements that might be expected or preferred. It attempts to achieve two objectives: firstly, to conduct a traditional dialect survey; and secondly, to investigate indicative structural phenomena. Strict minimal pairs were intentionally limited to make the survey more engaging for 16-18 year-olds in a classroom setting. Despite these limitations, it is still possible to glean some insights and to identify connections within the data. While the findings may be speculative, they provide enough information to suggest further fruitful lines of enquiry.

One area of interest is the potential for showing systematic patterns in individual users. For example, if an *individual* accepts one linguistic structure, does this predict their acceptance of certain other structures? Additionally, the data may reveal whether speakers can be grouped into clusters and whether these clusters exhibit geographical patterns.

6.4 Future directions

1. **Possibility of psycholinguistic study**. Where there may be distinct underlying structure to identical surface strings between neighbouring dialect regions (c.f. Object-shift-TGD vs Prep-drop-TGD in Manchester/Liverpool), to what extent is this difference detectable by psycholinguistic experimentation?

2. **Revised survey**. In future work, it would be beneficial to develop a shorter survey, including more refined set of verbs based on Twitter atlas. It would additionally be advantageous to include a greater range of local dialect terms, as this is what some participants mentioned, in the survey feedback, to be lacking.

3. **A more geographically expansive survey**. An obvious next step is to run a survey across a wider stretch of the UK, to explore the constraints on the phenomena discussed in this project, and others. The successful execution of the North West survey demonstrated the viability of the method of working with sixth form students. This could be readily scaled across broader regions.

4. **The inclusion of census data**. Comparing geolocated Twitter messages

with census data was shown in Baxter and Stevenson (2024) to be effective in supplying additional biographic data, where such data are lacking in the Twitter data. This could be achieved in the UK as well, and the additional census data may shed light on some of the patterns that we see.

5. **Further investigation into particular border regions**. The map data revealed a number of distinct isoglosses, such as in South and West Yorkshire, the investigation of which was beyond the scope of the current project. An aspiration of the present work is that these data may be used to identify such areas of interest, and to stimulate future work.

6. **Deeper analysis of survey results**: The data produced by the survey could be analysed more deeply. There is substantial room for future work here. Specifically, the analysis of the within speaker correlation between structures, could be particularly fruitful.

6.5 Conclusion

The thesis presented in this document focused on two syntactic phenomena: preposition dropping and ditransitives, and one phonological variable: DAR. Two principal data sources were used: Twitter data (Chapter 3) and a mass-participation survey (Chapter 5). Twitter data revealed substantial geographic variation for both phenomena with distinct border regions where the relative frequency of each variant changed dramatically over the space of a few miles. Interestingly, *where* the borders fell is distinct between the two syntactic variables.

One of the aims of the Twitter chapter was to create an atlas of use for the UK and Ireland as a whole. The idea here was to get an overview that could be used to then focus in on a particular region that would offer further scope for study for the survey. A number of potential candidates fit these criteria. The Scottish-English border, a border between the South-East and Midlands, between West and East Yorkshire are all possibilities and offer potential for future study. The final area chosen for the survey is the North-West region: Manchester, Liverpool and the area in between. As with any research project, but particularly here, where the initial dataset is so broad, there are more questions raised than there are answered. There were multiple potential border regions and sites identified that warrant further investigation. These offer ex-

citing new opportunities for research, both in terms of academic understanding and broader outreach.

References

Adger, D. (2019). *Language unlimited: The science behind our most creative power*. Oxford: Oxford University Press.

Allen, C. L. (2006). Case Syncretism and Word Order Change. In A. van Kemenade & B. Los (Eds.), *The Handbook of the History of English* (pp. 201–223). Oxford: Blackwell.

Anthony, L., & Hardaker. (2022). *FireAnt (Version 2.2.1)*. Tokyo, Japan: Waseda University.

Bailey, L. R. (2018a). *Let's Go Sussex - Null Prepositions in Southeast English.* [Paper presentation]. University of Sussex.

Bailey, L. R. (2018b). Some characteristics of Southeast English preposition dropping. *IBERIA: An International Journal of Theoretical Linguistics*, *10*, 24.

Barbiers, S. (2005). Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics*. In K. P. Corrigan & L. E. A. Cornips (Eds.), *Syntax and variation: Reconciling the biological and the social* (pp. 233–264). Amsterdam; Philadelphia: John Benjamins.

Barbiers, S., & Bennis, H. (2007). The Syntactic Atlas of the Dutch Dialects. *Nordlyd*, *34*(1). 10.7557/12.89

Barrie, C., & Ho, J. C.-t. (2021). *academictwitteR: Access the Twitter Academic Research Product Track V2 API Endpoint.*

Barry, M. (1972). The Morphemic Distribution of the Definite Article in Contemporary Regional English. In M. F. Wakelin (Ed.), *Patterns in the folk speech of the British Isles* (pp. 164–181). London: Athlone Press.

Baxter, K., & Stevenson, J. (2024). Ain't + infinitive verb in Black-

/African American English. In S. Wagner & U. Stange-Hundsdörfer (Eds.), *(Dia)lects in the 21st century: Selected papers from Methods in Dialectology XVII,* (pp. 31–55). Berlin: Language Science Press.

Beal, J. C. (2004). *English in modern times*. London: Arnold.

Bech, K., & Walkden, G. (2016). English is (still) a West Germanic language. *Nordic Journal of Linguistics*, *39*(01), 65–100. 10.1017/S0332586515000219

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. 10.21105/joss.00774

Bermel, N., & Knittl, L. (2012a). Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory*, *8*(2), 241–275. 10.1515/cllt-2012-0010

Bermel, N., & Knittl, L. (2012b). Morphosyntactic variation and syntactic constructions in Czech nominal declension: Corpus frequency and native-speaker judgments. *Russian Linguistics*, *36*(1), 91–119. 10.1007/s11185-011-9083-x

Biber, D. (1991). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., Gray, B., & Staples, S. (2016). Contrasting the grammatical complexities of conversation and academic writing: Implications for EAP writing development and teaching. *Language in Focus*, *2*(1), 1–18. 10.1515/lifijsal-2016-0001

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English* (Vol. 2). Harlow: Pearson.

Biggs, A. (2018). Locating variation in the dative alternation. In C. Lignos, L. MacKenzie, & M. Tamminga (Eds.), *The locus of linguistic variation* (pp. 1–32). Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Braber, N., & Robinson, J. (2018). *East Midlands English* (No. Volume 15). Boston Berlin: De Gruyter Mouton.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J., & Nikitina, T. (2009). The gradience of the dative alternation. In

L. Uyechi & L.-H. Wee (Eds.), *Reality Exploration and Discovery: Pattern Interaction in Language & Life* (pp. 161–184). Stanford: CSLI Publications.

Britain, D. (2001). Welcome to East Anglia! two major dialect boundaries in the Fens. In J. Fisiak & P. Trudgill (Eds.), *East Anglian English* (pp. 217–242). Cambridge: D.S. Brewer. 10.2307/j.ctvc16qkh.15

Britain, D. (2010). Supralocal regional dialect levelling. In D. Watt & C. Llamas (Eds.), *Language and identities* (pp. 193–204). Edinburgh University Press.

Britain, D., Kolly, M.-J., & Leemann, A. (2018). Using impact to make an impact? In D. McIntyre & H. Price (Eds.), *Applying Linguistics: Language and the Impact Agenda* (1st ed., pp. 83–97). Milton Park, Abingdon, Oxon; New York, NY: Routledge. 10.4324/9781351055185

Cardoso, A. (2015). *Dialectology, phonology, diachrony: Liverpool English realisations of PRICE and MOUTH* (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh.

Champion, T. (2005). Population movement within the UK. In R. Chappell (Ed.), *Focus on People and Migration* (pp. 91–113). London: Palgrave Macmillan UK. 10.1007/978-1-349-75096-2_6

Cheshire, J. (1986). Survey of British dialect grammar. *Journal of Multilingual and Multicultural Development*, *7*(5), 430–430. 10.1080/01434632.1986.9994259

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics*, *15*, 151–196.

Chiosáin, M. N. (2006). Language shift in early twentieth-century Ireland. *Proceedings of the Harvard Celtic Colloquium*, *26/27*, 370–384.

Cornips, L., & Corrigan, K. P. (2005). *Syntax and variation : Reconciling the biological and the social*. Amsterdam ; Philadelphia: John Benjamins.

Cornips, L., & Gregersen, F. (2016). The impact of Labov's contribution to general linguistic theory. *Journal of Sociolinguistics*, *20*(4), 498–524. 10.1111/josl.12197

Cornips, L., & Poletto, C. (2005). On standardising syntactic elicitation techniques (part 1). *Lingua*, *115*, 939–957. 10.1016/j.lingua.2003.11.004

Davis, M. (2014). *"I'm going shop": Preposition dropping in British youth dialects*. https://wordjazz.blog/2014/02/06/im-going-shop-preposition-dropping -in-british-youth-dialects/ [Accessed: 30/08/2023].

De Cuypere, L. (2014). The Old English to-dative construction. *English Language*

and Linguistics, 19, 1–26. 10.1017/s1360674314000276

De Cuypere, L. (2015). A multivariate analysis of the Old English ACC + DAT double object alternation. *Corpus Linguistics and Linguistic Theory*, *11*(2), 225–254. 10.1515/cllt-2014-0011

De Cuypere, L., De Coster, E., & Baten, K. (2014). The acquisition of the English dative alternation by Russian Foreign Language Learners. *Phrasis (Gent): Studies in Language and Literature*, 187–212.

Deumert, A., & Lexander, K. V. (2013). Texting Africa: Writing as performance. *Journal of Sociolinguistics*, *17*, 522–546.

Dorantes, A., Sierra, G., Donohue Pérez, T. Y., Bel-Enguix, G., & Jasso Rosales, M. (2018). Sociolinguistic corpus of WhatsApp chats in Spanish among college students. In L.-W. Ku & C.-T. Li (Eds.), *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media* (pp. 1–6). Melbourne, Australia: Association for Computational Linguistics. 10.18653/v1/W18-3501

Drummond, R. (2018). *Researching urban youth language and identity*. Cham, Switzerland: Palgrave Macmillan.

Edwards, V., Trudgill, P., & Weltens, B. (1984). *The grammar of English dialect: A survey of research: A report to the ESRC Education and Human Development Committee*. London : Merstham, Redhill, Surrey: Economic and Social Research Council ; Available from the School Govt. Pub. Co.

Edwards, V., & Weltens, B. (1985). Research on non-standard dialects of British English: Progress and prospects. In W. Viereck (Ed.), *Focus on: England and Wales. Varieties of English Around the World* (pp. 97–139). Amsterdam and Philadelphia: John Benjamins.

Eisenstein, J. (2013). Phonological factors in social media writing. In *Proceedings of the NAACL/HLT 2013 Workshop on Language Analysis in Social Media (LASM 2013)* (pp. 11–19). Atlanta, Georgia: Association for Computational Linguistics.

Eisenstein, J. (2018). Identifying regional dialects in on-line social media. In C. Boberg, D. Watt, & J. A. Nerbonne (Eds.), *The handbook of dialectology* (pp. 368–383). Hoboken, NJ: Wiley Blackwell.

Ellis, A. J. (1889). *The existing phonology of English dialects, compared with that of West Saxon speech*. New York: Greenwood Press.

Emonds, J. (1972). Evidence that indirect object movement is a structure-preserving rule. *Foundations of Language, 8*(4), 546–561.

Emonds, J., & Faarlund, J. T. (2014). *English: The language of the Vikings* (No. 3). Olomouc: Palacký University.

Evans, A. B. (1848). *Leicestershire words, phrases and proverbs. Collected by Arthur Benoni evans, D. D., head master of market Bosworth Free Grammar School.* London: William Pickering.

Evans, A. B. (1881). *Leicester words, phrases and proverbs. Collected by the late Arthur Benoni Evans, D. D., head master of Market Bosworth Free Grammar School. Edited with additions and an introduction by Sebastian Evans.* London: Trübner and Co.

Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive Written Discourse as an emergent register. *Written Communication, 8*(1), 8–34. 10.1177/0741088391008001002

Filppula, M. (1999). *The grammar of Irish English: Language in Hibernian style* (No. 5). London; New York: Routledge.

Fox, S. (2010). Ethnicity, religion and practices: Adolescents in the East End of London. In C. Llamas & D. Watt (Eds.), *Language and identities* (pp. 144–156). Edinburgh University Press.

Fox, S., Khan, A., & Torgersen, E. (2011). The emergence and diffusion of Multicultural English. In F. Kern & M. Selting (Eds.), *Ethnic styles of speaking in European metropolitan areas* (pp. 19–44). Amsterdam: Benjamins.

Gast, V. (2007). I gave it him – on the motivation of the 'alternative double object construction' in varieties of British English. *Functions of Language (special issue:Ditransitivity), 14*(1), 31–56.

Gehrke, B., & Lekakou, M. (2013). How to miss your preposition. *Studies in Greek linguistics, 33*, 92–106.

Gerasimova, A., & Lyutikova, E. (2020). Intralingual variation in acceptability judgments and production: Three case studies in Russian Grammar. *Frontiers in Psychology, 11*.

Gerwin, J. (2013). Give it me!: Pronominal ditransitives in English dialects. *English Language and Linguistics, 17*, 445–463.

Gerwin, J. (2014). *Ditransitives in British English dialects.* Berlin, Boston: De Gruyter.

Goldberg, A. E., Fauconnier, G., Lakoff, G., & Sweetser, E. (1995). *Constructions: A construction grammar approach to argument structure* (Vol. 25).

Gopal, D., Blaxter, T., Willis, D., & Leemann, A. (2021). Testing models of diffusion of morphosyntactic innovations in Twitter data. In A. Ziegler,

S. Edler, & G. Oberdorfer (Eds.), *Urban Matters: Current approaches in variationist sociolinguistics* (Vol. 27, pp. 253–278). Amsterdam: John Benjamins Publishing Company. 10.1075/silv.27

Grieve, J. (2016). *Regional variation in written American english*. Cambridge: Cambridge University Press. 10.1017/CBO9781139506137

Grieve, J., Montgomery, C., Nini, A., Murakami, A., & Guo, D. (2019). Mapping lexical dialect variation in British English using Twitter. *Frontiers in Artificial Intelligence, 2*. 10.3389/frai.2019.00011

Grondelaers, S., Van Hout, R., Van Halteren, H., & Veerbeek, E. (2023). Why do we say *them* when we know it should be *they*? Twitter as a resource for investigating nonstandard syntactic variation in The Netherlands. *Language Variation and Change, 35*(2), 223–245. 10.1017/S0954394523000121

Haddican, B. (2010). Theme–goal ditransitives and theme passivisation in British English dialects. *Lingua, 120*, 2424–2443.

Haddican, B., & Holmberg, A. (2012, October). Object movement symmetries in British English dialects: Experimental evidence for a mixed case/locality approach. *The Journal of Comparative Germanic Linguistics, 15*, 189–212. 10.1007/s10828-012-9051-x

Hall, D. (2019). P-D-Drop and pseudo-incorporation in London English. In *Proceedings of NELS* (Vol. 49, pp. 1–16).

Harley, H., & Jung, H. K. (2015). In support of the P $_{HAVE}$ analysis of the double object construction. *Linguistic Inquiry, 46*(4), 703–730. 10.1162/LING_a_00198

Harris, M. B. (1967). *The phonology and grammar op the dialect of South Zeal, Devonshire.* (Unpublished doctoral dissertation). University of London, London.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford ; New York: Oxford University Press.

Hawksey, M. (2018). *TAGS: https://tags.hawksey.info/*.

Haynes, K. E., & Fotheringham, A. S. (1984). *Gravity and spatial interaction models*. Beverley Hills: Sage.

Hollmann, W. B., & Siewierska, A. (2011, January). The status of frequency, schemas, and identity in Cognitive Sociolinguistics: A case study on definite article reduction. *Cognitive Linguistics, 22*(1). 10.1515/cogl.2011.002

Holmberg, A. (2016). Norse against Old English: 20-0. *Language Dynamics and*

Change, 6(1), 21–23. 10.1163/22105832-00601004

Honeybone, P. (2007). New-dialect formation in nineteenth century Liverpool: A brief history of Scouse. In C. Grey, K. Watson, & A. Grant (Eds.), *The Mersey Sound: Liverpool's Language, People and Places* (Vol. 2007, pp. 106–140). Liverpool.

Hornsby, D. (2018). 3. A new dialect for a new village: Evidence for koinéization in East Kent. In L. Wright (Ed.), *Southern English Varieties Then and Now* (pp. 74–109). Berlin ; Boston: De Gruyter Mouton. 10.1515/9783110577549-004

Huber, M., Nissel, M., & Puga, K. (2016). Old Bailey Corpus 2.0. hdl:11858/00-246C-0000-0023-8CFB-2

Hutton, G., & Murray, A. (2025). *Tourism: Statistics and policy. Research briefing.* House of Commons Library.

Ilbury, C., Grieve, J., & Hall, D. (2024). Using social media to infer the diffusion of an urban contact dialect: A case study of Multicultural London English. *Journal of Sociolinguistics*, josl.12653. 10.1111/josl.12653

Jamieson, E., Smith, J., Adger, D., Heycock, C., & Thoms, G. (2024). 'When intuitions (don't) fail': Combining syntax and sociolinguistics in the analysis of Scots. *English Language & Linguistics*, 1–34. 10.1017/S1360674323000679

Johnston, P. (1997). Older Scots phonology and its regional variation. In C. Jones (Ed.), *The Edinburgh History of the Scots Language* (pp. 47–111). Edinburgh University Press. 10.1515/9781474410977-005

Jones, T. (2015). Toward a description of African American Vernacular English dialect regions using "Black Twitter". *American Speech*, *90*, 403–440. 10.1215/00031283-3442117

Kerswill, P. (2003). Dialect levelling and geographical diffusion in British English. In D. Britain & J. Cheshire (Eds.), *Social dialectology: In honour of Peter Trudgill* (pp. 223–243). Amsterdam: Benjamins.

Kerswill, P., Cheshire, J., Fox, S., & Torgersen, E. (2012). English as a contact language: The role of children and adolescents. In D. Schreier & M. Hundt (Eds.), *English as a contact language* (pp. 258–282). Cambridge: Cambridge University Press.

Kerswill, P., & Williams, A. (2000). Creating a new town koine: Children and language change in Milton Keynes. *Language in society*, *29,* 65–115. kerswill

Kerswill, P., & Williams, A. (2014). Dialect levelling: Change and continuity in Milton Keynes, Reading and Hull. In P. Foulkes & G. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 141–162). London: Routledge.

Khan, A. (2006). *A sociolinguistic study of Birmingham english: Language variation and change in a multi-ethnic British community.* (Unpublished doctoral dissertation). Lancaster University.

Kim, K.-m., & Han, C.-h. (2016). Inter-speaker variation in Korean pronouns. In *Inter-speaker variation in Korean pronouns* (pp. 347–372). De Gruyter Mouton. 10.1515/9781614517016-013

Kirk, J. M., Sanderson, S., & Widdowson, J. D. A. (Eds.). (2014). *Studies in linguistic geography: The dialects of English in Britain and Ireland* (First issued 1985 ed.). London New York: Routledge, Taylor & Francis Group.

Knowles, G. O. (1973). *Scouse: The urban dialect of Liverpool* (Unpublished doctoral dissertation). University of Leeds, Leeds.

Kortmann, B., Lunkenheimer, K., & Ehret, K. (Eds.). (2020). *The Electronic World Atlas of Varieties of English (eWAVE). Version 3.0.*

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, *1*(1), 97–120.

Labov, W. (1996). When intuitions fail. In I. L. McNair, K. Singer, L. Dolbrin, & M. Aucon (Eds.), *Papers from the parasession on theory and data in linguistics* (Vol. 32, pp. 77–106). Chicago Linguistic Society.

Labov, W. (2007). Transmission and diffusion. *Language, 83*(2), 344–387. 10.1353/lan.2007.0082

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., & Nagy, N. (2011, September). Properties of the sociolinguistic monitor: The sociolinguistic monitor. *Journal of Sociolinguistics, 15*(4), 431–463. 10.1111/j.1467-9841.2011.00504.x

Leach, H. (2018). *Sociophonetic variation in Stoke-on-Trent's pottery industry* (Unpublished doctoral dissertation). University of Sheffield.

Leemann, A., Kolly, M.-J., & Britain, D. (2018). The English Dialects App: The creation of a crowdsourced dialect corpus. *Ampersand, 5*, 1–17. 10.1016/j.amper.2017.11.001

Le Page, R., & Tabouret-Keller, A. (1985). *Acts of identity*. Cambridge: Cambridge University Press.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation.*

Chicago: University of Chicago Press.

Lodge, K. (2010). Th'interpretation of t'definite article in t'North of England. *English Language & Linguistics*, *14*(1), 111–127. 10.1017/S1360674309990384

MacKenzie, L. (2019). Perturbing the community grammar: Individual differences and community-level constraints on sociolinguistic variation. *Glossa: a journal of general linguistics*, *4*(1), 28. 10.5334/gjgl.622

MacKenzie, L., Bailey, G., & Turton, D. (2014). *Our dialects: Mapping variation in English in the UK.*

MacKenzie, L., Bailey, G., & Turton, D. (2022). Towards an updated dialect atlas of British English. *Journal of Linguistic Geography*, *10*(1), 46–66. 10.1017/jlg.2022.2

Maechler, M., Rousseeuw, Peter, Struyf, Anja, Hubert, Mia, & Hornik, Kurt. (2025). *Cluster: Cluster analysis basics and extensions. R package version 2.1.8.1.* https://cran.r-project.org/web/packages/cluster/index.html.

Mcmahon, A., Heggarty, P., Mcmahon, R., & Maguire, W. (2007). The sound patterns of Englishes: Representing phonetic similarity. *English Language & Linguistics*, *11*(1), 113–142. 10.1017/S1360674306002139

McMahon, A., & Maguire, W. (2013). Computing linguistic distances between varieties. In M. G. Krug (Ed.), *Research methods in language variation and change* (pp. 421–432). New York: Cambridge University Press.

Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of linguistics*, *21*, 339–384.

Mitchell, H. (1799). *Scotticisms, vulgar Anglicisms, and grammatical improprieties corrected, with reasons for the corrections.* Glasgow, Printed by Falconer & Willison for the author.

Morgan-Lopez, A. A., Kim, A. E., Chew, R. F., & Ruddle, P. (2017). Predicting age groups of Twitter users based on language and metadata features. *PLOS ONE*, *12*(8), e0183537. 10.1371/journal.pone.0183537

Myler, N. (2013, October). On coming the pub in the North West of England: Accusative unaccusatives, dependent case, and preposition incorporation. *The Journal of Comparative Germanic Linguistics*, *16*, 189–207. 10.1007/s10828-013-9055-1

Ojanen, A.-L. (1985). Use and non-use of prepositions in spatial expressions in the dialect of Cambridgeshire. In W. Viereck (Ed.), *Focus on England and Wales* (pp. 179–212). Amsterdam Philadelphia: J. Benjamins.

Orton, H., & Dieth, E. (1962). *Survey of English dialects*. Leeds: E.J. Arnold, University of Leeds.

Orton, H., Sanderson, S., & Widdowson, J. (Eds.). (1978). *The Linguistic atlas of England*. London: Croom Helm.

Page, R. E. (2012). *Stories and social media: Identities and interaction* (No. 3). New York: Routledge.

Parish, W. D., Shaw, W. F., & Masters, J. W. (1888). *A dictionary of the Kentish dialect and provincialisms in use in the County of Kent*. Lewes Farncombe.

Paul, K., Wittenberg, E., & Wiese, H. (2008). "Ich mag so Wasserpfeifeladen" : The interaction of grammar and information structure in Kiezdeutsch. In *Vortrag auf der 5. Internationalen Jugendsprachekonferenz, Kopenhagen, 27.-29.3.2008.*

Pavalanathan, U., & Eisenstein, J. (2015). Audience-Modulated Variation in Online Social Media. *American Speech, 90*(2), 187–213. 10.1215/00031283-3130324

Pielco, & Haccer. (2017). *TWINT - Twitter Intelligence Tool*. TWINT Project.

Pooley, C., & Turnbull, J. (2005). *Migration and mobility in Britain since the eighteenth century*. Routledge. 10.4324/9780203976548

Quist, P. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism, 12*, 43–61.

R Core Team. (2013). *R a language and environment for statistical computing: Reference index*. Vienna: R Foundation for Statistical Computing.

Ramisch, H. (1997). Dialectological and cartographical features of the Computer Developed Linguistic Atlas of England (CLAE). In *Issues and methods in dialectology* (pp. 224–233). Bangor [Wales]: Dept. of Linguistics, University of Wales Bangor.

Robinson, P. (1997). *Mapping of Ulster-Scots*. https://www.ulster-scotsacademy.com/research/gregg/mapping-ulster-scots.php [Accessed: 12/10/24].

Rupp, L. (2007). The (socio-)linguistic cycle of Definite Article Reduction. *Folia Linguistica Historica, 28-1-2*(41), 215–249. 10.1515/flih.2007.215

Rupp, L., & Tagliamonte, S. A. (2019). "They used to follow Ø river": The Zero Article in York English. *Journal of English Linguistics, 47*(4), 279–300. 10.1177/0075424219865933

Schutze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.

Schutze, C. T., & Sprouse, J. (2014). Judgement data. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). Cambridge: Cambridge University Press.

Shortis, T. (2015). *Orthographic practices in SMS text messaging as a case signifying diachronic change in linguistic and semiotic resources* (Unpublished doctoral dissertation). UCL (University College London), London.

Shortis, T. (2016). Texting and other messaging: Written system in digitally mediated vernaculars. In V. Cook & D. Ryan (Eds.), *The Routledge Handbook of the English Writing System* (pp. 487–516). Abingdon: Routledge.

Siewierska, A., & Hollmann, W. (2007). Ditransitive clauses in English with special reference to Lancashire dialect. In M. Hannay & G. J. Steen (Eds.), *Structural-functional studies in English Grammar: In honour of Lachlan Mackenzie* (Vol. 83, pp. 83–102). Amsterdam; Philadelphia: John Benjamins Publishing.

Skidmore, P. F. (2009). *The maritime economy of North West England in the later Eighteenth Century* (Unpublished doctoral dissertation). University of Greenwich, London.

Smith, J., Adger, D., Aitken, B., Heycock, C., Jamieson, E., & Thoms, G. (2019). *The Scots Syntax Atlas* (Tech. Rep.). University of Glasgow.

Smith, J., & Holmes-Elliott, S. (2022). Mapping Syntax and the Sociolinguistic Monitor. In T. Christensen & T. J. Jensen (Eds.), *Explanations in sociosyntactic variation* (pp. 58–89). Cambridge, United Kingdom ; New York, NY: Cambridge University Press.

Stevenson, J. (2016). *Dialect in digitally mediated written interaction: A survey of the geohistorical distribution of the ditransitive in British English using Twitter* (Masters by Research). University of York.

Stevenson, J. (2021). Preposition drop and ditransitives in North West British English: Twitter and survey data [Paper presentation]. In *Morphosyntactic variation and change in the 21st Century*. University of Cambridge.

Strelluf, C. (2019). Positive-anymore, American regional dialects, and polarity-licensing in tweets. *American Speech*, *94*(3), 313–351. 10.1215/00031283-7587883

Strelluf, C. (2020). needs + PAST PARTICIPLE in regional Englishes on Twitter. *World Englishes*, *39*(1), 119–134. 10.1111/weng.12451

Stroud, C. (2004). Rinkeby Swedish and semilingualism in language ideological debates: A Bourdieuean perspective. *Journal of Sociolinguistics*, *8*(2),

196–214. 10.1111/j.1467-9841.2004.00258.x

Tagg, C. (2009). *A corpus linguistics study of SMS text messaging* (Unpublished doctoral dissertation). University of Birmingham.

Tagliamonte, S. A. (2014). A comparative sociolinguistic analysis of the dative alternation. In R. T. Cacoullos, N. Dion, & A. Lapierre (Eds.), *Linguistic variation: Confronting fact and theory* (pp. 297–318). New York: Routledge.

Tagliamonte, S. A., Durham, M., & Smith, J. (2014). Grammaticalization at an early stage: Future be going to in conservative British dialects. *English Language and Linguistics*, *18*, 75–108. 10.1017/s1360674313000282

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, *1*(2), 179–195. 10.1017/S0047404500000488

Trudgill, P. (1974). Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Lang. Soc*(2), 215–246.

Trudgill, P. (1986). *Dialects in contact: Peter Trudgill* (No. 10). Oxford: Blackwell.

Upton, C. (2006). Modern Regional English in the British Isles. In L. Mugglestone (Ed.), *The Oxford History of English* (pp. 379–414). Oxford: Oxford University Press.

Upton, C., Parry, D., & Widdowson, J. D. A. (1994). *Survey of English Dialects: The Dictionary and Grammar*. Psychology Press.

Vasko, A.-L. (2010). *Cambridgeshire Dialect Grammar* (Vol. 4). Helsinki: VARIENG.

Wales, K. (2006). *Northern English: A cultural and social history*. Cambridge, UK ; New York.

Watson, K. (2006). Phonological resistance and innovation in the North-West of England. *English Today*, *22*(02), 55. 10.1017/S0266078406002100

Watts, E. L. (2005). *Mobility-induced dialect contact: A sociolinguistic investigation of speech variation in Wilmslow, Cheshire* (Unpublished doctoral dissertation). University of Essex, Essex.

West, H. F. (2015). *Accent variation and attitude on the Merseyside/Lancashire border: A sociophonetic study of Southport and Ormskirk* (Unpublished doctoral dissertation). University of York.

Wiese, H. (2009, May). Grammatical innovation in multiethnic urban Europe: New linguistic practices among adolescents. *Lingua, 119*(5), 782–806.

10.1016/j.lingua.2008.11.002

Williams, S. (2008). *Staffordshire foods: Lobby - a traditional Potters' dish.* www.bbc.co.uk/stoke/content/articles/2008/07/25/lobby_feature.shtml [Accessed: 12/04/2024].

Willis, D. (2020). Using social-media data to investigate morphosyntactic variation and dialect syntax in a lesser-used language: Two case studies from Welsh. *Glossa: a journal of general linguistics*, *5*(1), 103. 1-33. 10.5334/gjgl.1073

Wood, J., Zanuttini, R., Horn, L., & Zentz, J. (2020). Dative Country. *American Speech*, *95*(1), 3–45. 10.1215/00031283-7587901

Woolford, E. (1993). Symmetric and asymmetric passives. *Natural Language & Linguistic Theory*, *11*(4), 679–728. 10.1007/BF00993017

Wright, J. (1898). *The English Dialect Dictionary, being the complete vocabulary of all dialect words still in use, or known to have been in use during the last two hundred years*. London; New York: H. Frowde; G. P. Putnam's Sons.

Yáñez-Bouza, N. (2016a). Early and Late Modern English grammars as evidence in English historical linguistics. In M. Kyto & P. Pahta (Eds.), *The Cambridge Handbook of English Historical Linguistics* (pp. 164–180). Cambridge: Cambridge University Press. 10.1017/CBO9781139600231.011

Yáñez-Bouza, N. (2016b). "May Depend on Me Sending It You": Double Objects in Early Grammars. *Journal of English Linguistics*, *44*, 138–161. 10.1177/0075424216630793

Yáñez-Bouza, N., & Denison, D. (2015). Which comes first in the double object construction? *English Language and Linguistics*, *19*, 247–268. 10.1017/s136067431500012x

Zehentner, E. (2016). *On competition and cooperation in Middle English ditransitives* (Doctoral Thesis).

Zehentner, E. (2018). Ditransitives in Middle English: On semantic specialisation and the rise of the dative alternation. *English Language and Linguistics*, *22*(01), 149–175. 10.1017/S1360674316000447

Zehentner, E. (2019). *Competition in language change: The rise of the English dative alternation* (Vol. 103). Berlin: de Gruyter Mouton.

Code switching: users who switch between variants

Here I briefly present interesting data that show code-switching between variants. These facts offer considerable scope for further investigation which were beyond that of the current project. However, I nevertheless draw attention to them here.

It is perhaps expected that users will commonly switch between normative and non-standard structures, such as between ᵨPDAT and ᵨGTD, depending on context and register.

Switching between two non-standard structures should be expected only in locations where a speaker has access to both alternatives. We might anticipate that non-standard switching would occur more frequently in border areas, where a speaker is more mobile, or if they are accommodating to another user.

Switching variants between tweets

For both [DATALT] and [P/D-DROP], whilst it is a minority of users who switch between non-standard structures *between*, rather than *within* tweets, it still represents something of a frequent occurrence, as shown in Table A.1.

Impressionistically, a significant proportion of users who exhibit switching

Corpus	Variants	Users who switch	Total unique users
pDit-corp	TGD $\Longleftrightarrow$ GTD	1161	104,104
pDrop-corp	PD-DROP $\Longleftrightarrow$ P-DROP	2074	193,326

Table A.1: Total counts for structure switching within the same tweet between TGD $\Longleftrightarrow$ GTD and PD-DROP $\Longleftrightarrow$ P-DROP.

appear to have moved between locations, or are located in border areas. It certainly looks to be a potentially interesting avenue of inquiry, but for now, I leave further investigation into the phenomenon of individual switching to future research.

Multiple variants in the same tweet

Running the *kwic* concordance function from the *quanteda* package (Benoit et al., 2018) in R extracts examples of strings into separate rows, even if they belong to the same tweet. This makes it possible to track instances of variant switching *within* a given tweet. Entries which had the same unique tweet *id* were first grouped, then, filtered such that only groups with more than one variant were returned.

```
multi<-d %>%
group_by(id) %>%
filter(n_distinct(variant) > 1) %>%
ungroup()
```

The facts presented below are important because they lie at the intersection of grammar and usage, two areas of linguistics that are often separated in the pursuit of what we might call a theoretical idealisation of categoricity in syntactic theory.

[P/D-DROP]

For pDrop-corp, out of 390,657 unique tweets, 399 contained multiple *distinct* variants.

There were 13 combinations of P-DROP and PD-DROP.

(1)   a.   I don't know why people wouldn't <u>go the gym</u>, its actually amazing, do you <u>go gym</u>?

        b.   Don't <u>go pub</u>/ open pub/close pub. Every stupid ***** is <u>going the beach!</u>

c. Who's gonna <u>go the pub</u> if they can't have a drink that's what you <u>go pub</u> for

d. <u>come cinema</u> instead, and <u>come the zoo</u> with me on a day I can actually afford to <u>go the zoo</u> :D

e. hmm contemplating <u>going the gym</u> tonight!!! or shal i just <u>go pub</u> with my girl ...

There were 78 containing both NO-DROP and P-DROP:

(2) a. Haha it's when you tell someone you're <u>going the pub</u> for tea and they ask you why you would <u>go to the pub</u> and drink tea and not beer ⊗

b. She's <u>gone the theatre capital of the world</u> and <u>went to the theatre</u>? What a fool.

c. Someone <u>come the cinema</u> with me to see the new Purge Movie?! Literally cannot remember the last time <u>I went to the cinema</u> 🍿 📽

d. standard day for me and Beth, <u>go to the gym</u>, go out for food and now we're in the pub in our gym stuff drinking pints of lager, what was the ***** point in <u>going the gym</u> 😂

e. When you <u>go to your birds families</u> for a roast and <u>go the pub</u> but the girls <u>go to the cinema</u>.

150 tweets which contained both NO-DROP and PD-DROP.

(3) a. I need to stop <u>going to the gym</u> on the days I'm working even if it means me only <u>going gym</u> twice a week.. all this neck & shoulder pain is not the one 🔒🔒

b. I've been <u>going to the gym</u> regularly for a few weeks now, and I've never really felt like I've progressed or worked hard enough, but more and more people have told me I look healthier and asked if I have been <u>going gym</u>.

c. Really don't know whether to watch the football or <u>go to the gym</u> because I know it'll be absolutely dead tonight! I could watch it THEN <u>go gym</u>, but then will people do the same and it'll just be busy again? See my predicament...

For pDit-corp, out of 155,646 unique tweets, 185 contained multiple distinct variants.

There were 22 tweets containing both TGD and GTD.

(4)   a.   Like ACTUALLY <u>lend me it</u>; I will promise to <u>give it them back</u> when my loan comes in January! :(

    b.   haha, pop it in your account and tell him he hadn't <u>gave it you</u> yet he just said can he <u>give you it</u> mwahaha

    c.   I didn't <u>give you it</u> because I had too, <u>I gave it you</u> because I thought I could trust you

    d.   you <u>sent it me</u> on the 9th, i <u>sent you it</u> on the 11th, i love being right ;)

    e.   currently sat for 20 mins in subway they better <u>feed it me</u> and <u>give me it</u> for free had a belly full i have

Combinations of PDAT and TGD numbered 59.

(5)   a.   Sorry should have read, <u>my son sent it me</u> via a an ex racer Richie Ellis who <u>sent it to him</u>

    b.   Picking up book 4 my sous chef, his girlfriend has got for his Christmas present. Should I <u>post it to her</u> or <u>give it him</u>?

    c.   Asked for melatonin today...will see how we get on. Soo reluctantly to <u>give it me</u>. For gods sake U <u>offered it to me</u> last visit

    d.   Oh, <u>give it her now</u>, whilst she still wants it, in case someone else <u>gives it to her</u> or spoils it. But I am a soft touch.

    e.   Need 50k anyone want to <u>give it to me</u> and <u>I will give it them back</u> in about 6 hours waiting for coins

The remaining 104 were combinations of GTD and PDAT.

(6)   a.   Maybe a fan <u>sent you it</u> anonymously because they think you're awesome? ☐ At least that's why I'd <u>send it to you</u>.

    b.   I think I <u>sent it to you</u> before but that's where I stayed and wrote about. If it wasn't you let me know and I can <u>send you it</u>.

Again, I report the phenomenon of code-switching here in passing, and

leave a full treatment to future investigation.

Full search code

```r
librarian::shelf(tidyverse,purrr,data.table,jsonlite,dplyr,
    academictwitteR,devtools,reshape2,foreach,doParallel,
    tictoc,utils,quiet = TRUE)


#devtools::install_github("cjbarrie/academictwitteR", build
    _vignettes = TRUE)
#install.packages("academictwitteR")


numCores <- detectCores()-1


# load bearer token
bearer_token <- readChar("bearer_token_JS.txt", 114)


# set name
name<-'ditVerbsUK'


# set start and end dates
start <- "2012-01-01T00:00:00Z"
 end <- "2022-01-01T00:00:00Z"
```

```r
# set max tweets and bin size
max_tweets <- 10000000
bin_size <- 50000

# import places master list
pm<-readRDS('places_master.rds')

# import grep
grep<-read.csv('verbList.csv')

# Loop verbs
for (s in 1:nrow(grep)) {
  # define search string for outer loop
  tw_search<-tolower(unique(as.character((as.vector(grep[s,
      2:7]))))))
  tw_search<-tw_search[nchar(tw_search)>0]
  verb<-as.character((as.vector(grep[s,1])))
  q<-paste0(tw_search,collapse = ' OR ')
  start_tweets<-start
  end_tweets<-end

  # build query:
  query <- paste0('(',q,')', ' has:geo -is:retweet (place_
      country:GB OR place_country:IE)')

  # set path
  path <- paste0("data/jsons/",name,"/",verb)

  # create inner loop defining max_tweets per bin
  for (b in 1:(max_tweets/bin_size)) {

    # (re)set max_tweets_bin
    max_tweets_bin <- bin_size

    # set current bin path from type and verb
    path_bin <- paste0(path,'/bin_',b)
```

```r
 # if enddate file does not exist, start process. If it
     does, skip to next loop.
if (file.exists(paste0(path,"/complete"))){break}

if (!file.exists(paste0(path_bin,"/enddate"))){
  # create bin directory
 dir.create(paste0(path_bin),recursive = TRUE,
     showWarnings = F)

 # if past the first bin, set end date from previous bin
 if (b>1) {
   end_tweets<-readLines(paste0(path,"/bin_",b-1,"/
       enddate"))
 }

 # GET TWEETS
 cat('\n === Fetching bin',b,'for',verb,"(verb",s,'of',
     nrow(grep),") === \n")

 # check if data collection for bin has already been
     started, if true, set new end date.
 if (!length(list.files(path_bin,pattern = "^data"))<2){
     metadata <- readLines(file.path(path_bin,"query"))
     files <- list.files(path_bin, pattern = "^data",
         full.names = TRUE,include.dirs = TRUE,recursive
          = TRUE)
     datatmp<-data.table()
     cat('\n === Retracing bin end date to restart
         process === \n')
     registerDoParallel(cores=numCores)
     datatmp<-foreach (i=seq_along(files),.combine=dplyr
         ::bind_rows) %dopar% {
       jsonlite::fromJSON(files[[i]],simplifyVector =
           TRUE,flatten = TRUE)
     }
     stopImplicitCluster()
     end_tweets<-min(datatmp$created_at)
     max_tweets_bin<-max_tweets_bin-nrow(datatmp)
```

```r
        if (max_tweets_bin>499){cat('\n === Resuming bin
            from last point:',end_tweets,'=== \n')}
        else {cat("\n === Bin already full === \n")}
    } else {if (b>1) {cat('\n === New end date carried
        from last bin is:',end_tweets,'=== \n')}}


if (max_tweets_bin>499){
get_all_tweets(query,
                    start_tweets,
                    end_tweets,
                    bearer_token,
                    file = NULL,
                    data_path = path_bin,
                    export_query = TRUE,
                    bind_tweets = FALSE,
                    verbose = T,
                    n = max_tweets_bin,
                    page_n = 500)
}



# extract data from jsons and set new end date for next
    bin

# get file names from bin
data_files <- list.files(path_bin, pattern = "^data",
    full.names = TRUE,include.dirs = TRUE,recursive =
    TRUE)
user_files <- list.files(path_bin, pattern = "^users",
    full.names = TRUE,include.dirs = FALSE,recursive =
    TRUE)

## GET DATA
if (!length(data_files)<2){
registerDoParallel(cores=numCores)
data <- data.table()
data <- foreach (i=seq_along(data_files),.combine=dplyr
    ::bind_rows,.multicombine = TRUE) %dopar% {
```

```r
    df<-jsonlite::fromJSON(data_files[[i]],simplifyVector
        = TRUE,flatten = TRUE)
  }
  stopImplicitCluster()
  # get new end date and write enddate file
  enddate<-min(data$created_at)
  write(enddate,paste0(path_bin,'/enddate'))
  } else {
  x<-'complete'
  write(x,paste0(path,'/complete'))
  cat('\n Verb complete, sleeping for 60 seconds before
      starting next verb \n')
  Sys.sleep(60)
  {break}
  }
  }
}
  }
```
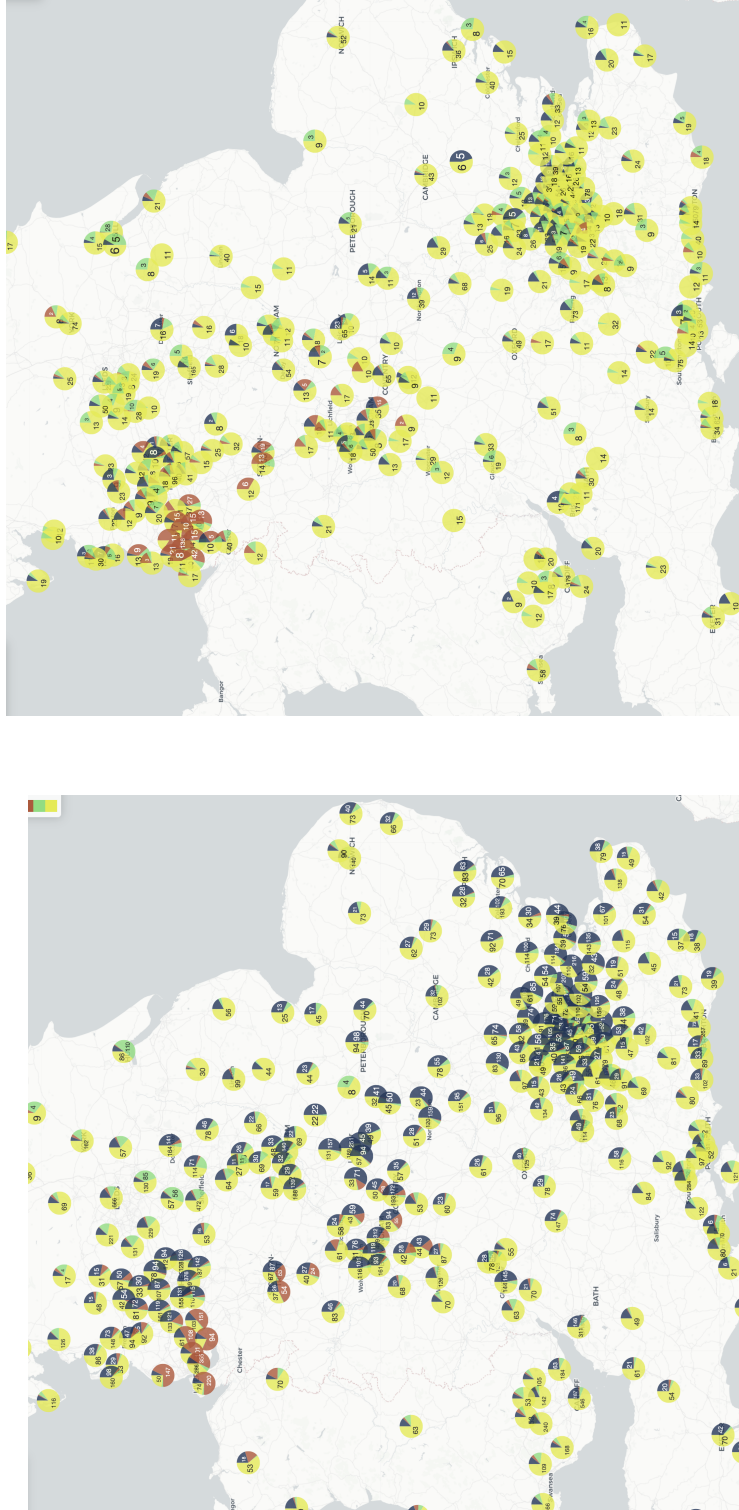
(a) [P/D-DROP] 2013

(b) [P/D-DROP] 2020

Figure B.1: The difference [P/D-DROP] variation between 2013 and 2020.
Yellow: NO-DROP; Blue: PD-DROP; Red: P-DROP; Green: D-DROP

335

Examples from BNC XML edition (spoken corpus)

C.1 PD-DROP

(1) **South**: 7 examples

    a. admit their part in the crime but they won't have to <u>go court</u>. A caution does n't mean a criminal record but it does

    b. No, no! [unclear] every time <u>you go loo</u> (age 45-59, 1991)

    c. over her money situation? Well I, <u>I went guarantor</u> for that. (age 25-34, 1991)

    d. How much is it when you want <u>to go cinema</u>? (age 35-44, 1991)

    e. it on a Saturday they , they won't do their, <u>go football</u> or something (age 60+, 1991)

    f. You're on PS0GG: but <u>I'm going race track</u> as well so I 'm on the race (age 15-24, 1992)

    g. He wanted to know if <u>I was going park</u> (0-14, 1992)

(2) **Midlands** 6 examples

    a. . that er eighty thirty five S: mm , oh <u>I'll go bed</u> then while you see (age 82, 1994)

b. Go on what? <u>I went hospital</u> with Lynn (age 15-24, 1992)

c. I'm going to Espania tomorrow, cos <u>I'm going holiday</u> to El Spania (age 15-24, 1992)

d. Sure you don't want <u>to go badminton</u>, mm, ah? (age 0-14, 1992)

e. I said to mum I think off when <u>we went bingo</u> and I was smoking (age 45-59, 1992)

f. on the corner here is this road, where you <u>go park road</u> there (age 45-59, 1992)

(3) **North** 20 examples

a. we'd go down the station, <u>went pub</u> (North England, age 68, 1994)

b. <u>Should go pub</u>. S: I can't afford it. S: Should save up (North England, age 37, 1994)

c. got ta <u>go dentist</u> next week (North, age N/A, 1992)

d. two in the morning before you go out and two before you <u>go bed</u> at night (North 1985-1993).

e. So where is that <u>I'll go Doctor</u> [gap:name] for my hair (North, 1985-1993)

f. and hair and that here S: No , no when you <u>went football</u> I did it (North England, age 68, 1994))

g. I can do this I can <u>go radio four</u> (age N/A, 1985-1993)

h. Do n't get settled cos I have <u>to go toilet</u> (age 45-59, 1992)

i. D'ya want a cup of Dawn? What time do <u>you go hospital</u>? (age 45-59, 1992)

j. No, no when <u>you went football</u> I did it, how's it doing? (age 60+, 1992)

k. I says erm yo it'd be cheaper <u>to go bed and breakfast</u> ! (age 60+, 1992)

l. Candleford and there <u>they went breakfast</u> was bread and dripping (age 45-59, 1992)

m. I've been in with the you know, <u>to go bingo</u>. I'll just hold it like that, eh? (age 25-34, 1992)

n. it's not Barton, it begins with B. He used <u>to go Station</u> (age 35-44, 1992)

o. then turn left, <u>go Cock Inn</u>, erm that er, turn that into Dane there (age 60+, 1992)

    p.   Sea Cadets but he's not good enough <u>to go college</u>? (age 35-44, 1992)

    q.   no we went to Aldy and we went and <u>we went Co-Op</u> and where else? (age 15-24, 1992)

    r.   mind ya fillings next week, yeah, got ta <u>go dentist</u> next week (age 15-24, 1992)

    s.   I think I'm gonna <u>go drama class</u> (age 35-44, 1992)

C.2 P-DROP

(4)   South: 5 examples

    a.   Poor Keith had to go to this , he had to <u>go the station</u> every day on his bike (age N/A, 1985-1993)

    b.   The ones that smile, incidentally, usually <u>go the bankruptcy court</u> (age N/A, 1985-1993)

    c.   We <u>went the bank</u> the other day (SW, age 50, 1992)

    d.   Then that's <u>going the bank</u> (SW, age 20, 1992)

    e.   cos we asked you <u>to go the cinema</u> all the time (age 13, 1985-1993)

(5)   Midlands: 8 examples

    a.   You passed up the chance <u>to go the pit</u> then? (age N/A, 1985-1993)

    b.   It was from the credit that the, <u>I went the education secretary</u> (age N/A, 1985-1993)

    c.   Yes Daisy? Can <u>I go the toilet</u>? (age 9, 1985-1993)

    d.   Yeah, but well you've got ta <u>go the shop</u> and tell her (Wales, age 25-34, 1992)

    e.   she'll have one. we were going <u>we were going the house</u> (Wales, age 28, 1992)

    f.   could've have asked him. When did you <u>go the shop</u>? (NW Midlands, age 8, 1992)

    g.   Well I know with two ladies th their daughters wanted to, them <u>to go the house</u> and they wouldn't. (NWM, age 53, 1992)

    h.   They said, no we want <u>to go the party</u>. (NW Midlands, age 53, 1992)

(6)   North: 9 examples

    a.   Don't you just do n't you <u>go the Registry Office</u>? (age 25, 1985-

1993)

b. if we wanted to go the match we had to pay (NE, age 53, 1991)

c. Yeah, cos you wan na go the garage don't you? (age 50, 1992)

d. We went the shop and she Anthony a pound, she give Sam a pound (age 50, 1992)

e. how come he's good enough to go the Sea Cadets but he's not good enough to go college? (Central North, age 35-44, 1992)

f. and she went the Empire, I went the Corn Exchange but that's beside the point (also in FRED) (NE, age 46, 1992)

g. No never mind and you're going the doctor's (NE, age 46, 1992)

h. I mean when we went the estate agents and solicitors (NE, age 46, 1992)

i. Alright Matt, Matt will you go the chippie? (Scottish, age 35, 1992)

Pronominal ditransitives by verb

All clusters

verb	ᴘTGD	ᴘGTD	ᴘPDAT	total
explain	0.6	0.1	99.3	6102
introduce	0.0	0.8	99.2	491
deliver	0.8	0.6	98.6	1154
address	1.5	0.0	98.5	131
describe	1.1	0.5	98.4	440
donate	1.0	0.7	98.3	299
mention	1.4	0.3	98.3	2512
leave	3.7	1.7	94.6	4025
hand	2.6	4.2	93.2	4209
read	5.5	1.5	93.0	1302
post	3.4	4.2	92.5	2970
kick	6.8	3.4	89.8	59
forward	2.9	9.6	87.4	1090
credit	11.3	1.6	87.1	124
pass	7.4	6.9	85.7	1243
sell	6.1	10.7	83.2	4879
bring	6.7	11.7	81.6	5649
award	4.8	16.2	79.0	105
fax	7.0	14.0	78.9	57
owe	7.6	17.8	74.6	1282
email	9.4	18.5	72.1	4286
loan	6.1	24.9	69.1	181
give	13.8	20.4	65.8	59867
feed	9.4	27.4	63.3	1037
send	13.6	23.5	62.9	37764
mail	8.1	30.6	61.4	991
offer	15.4	25.2	59.4	1333
lend	20.1	32.7	47.1	2420
deny	14.0	43.4	42.7	143
pm	8.5	55.5	36.0	317
show	10.2	56.4	33.4	6905
promise	17.8	51.7	30.5	118
message	12.9	59.3	27.8	1669
pay	25.4	49.8	24.9	209
text	15.5	67.1	17.4	3661
ask	6.4	84.0	9.6	156

Table D.1: Relative frequencies of [DATALT] variants by verb, ordered by ᴘPDAT

Cluster 1

verb	$_p$TGD	$_p$GTD	$_p$PDAT	total
deliver	0.0	0.0	100.0	148
explain	0.6	0.0	99.4	963
describe	1.1	0.0	98.9	93
introduce	0.0	1.4	98.6	73
donate	0.0	1.8	98.2	55
mention	1.5	0.4	98.0	454
leave	4.6	4.0	91.4	581
read	4.3	4.8	90.9	209
post	2.1	10.1	87.8	466
hand	1.2	12.1	86.7	646
forward	1.8	18.9	79.3	164
pass	5.5	17.0	77.5	200
bring	1.8	24.8	73.4	997
sell	2.9	27.7	69.5	802
email	5.5	37.7	56.9	751
owe	2.7	43.6	53.7	188
offer	4.6	48.2	47.2	218
give	5.0	49.4	45.6	11186
feed	2.9	52.6	44.6	175
send	4.6	51.6	43.8	6998
mail	1.5	54.9	43.6	204
lend	6.1	67.5	26.3	539
pm	4.3	71.0	24.6	69
show	3.3	80.2	16.5	1503
message	1.6	86.3	12.1	315
pay	13.7	76.5	9.8	51
text	2.6	90.5	6.9	823

Table D.2: Cluster 1 ($_p$GTD-dominant): relative frequencies of [DATALT] variants by verb, ordered by $_p$PDAT

Cluster 2

verb	$_p$TGD	$_p$GTD	$_p$PDAT	total
describe	0.0	0.0	100.0	65
introduce	0.0	0.0	100.0	65
explain	0.6	0.0	99.4	893
mention	2.0	0.2	97.8	460
deliver	2.3	0.0	97.7	175
donate	3.6	0.0	96.4	55
leave	5.9	1.5	92.6	608
hand	8.9	1.7	89.4	576
read	11.1	1.0	87.9	198
post	10.0	3.3	86.7	460
forward	7.6	7.1	85.3	170
sell	19.7	3.6	76.7	857
pass	20.1	3.5	76.4	254
bring	23.0	6.9	70.1	1021
owe	24.7	9.4	66.0	235
email	29.2	12.9	57.9	770
mail	34.4	9.2	56.4	163
feed	28.2	18.8	53.0	202
give	45.1	8.4	46.6	10705
send	47.3	11.7	41.0	6726
offer	44.0	16.4	39.6	250
lend	56.8	16.8	26.4	511
show	40.3	36.6	23.1	1093
message	42.9	42.1	15.0	366
text	50.3	39.5	10.2	765

Table D.3: Cluster 2 ($_p$TGD-dominant): relative frequencies of [DATALT] variants by verb, ordered by $_p$PDAT

Cluster 3

verb	ₚTGD	ₚGTD	ₚPDAT	total
explain	0.6	0.1	99.4	3311
donate	0.7	0.0	99.3	143
deliver	0.7	0.5	98.9	615
introduce	0.0	1.2	98.8	249
describe	0.9	0.4	98.7	223
address	1.7	0.0	98.3	60
mention	1.4	0.4	98.2	1050
hand	1.5	1.9	96.6	2049
post	1.4	2.3	96.3	1395
leave	3.1	1.3	95.6	1899
read	4.8	0.5	94.4	564
pass	3.1	4.4	92.5	522
bring	2.2	7.2	90.6	2705
forward	1.3	8.2	90.4	594
sell	2.5	7.1	90.4	2067
credit	12.5	0.0	87.5	56
give	5.9	10.2	83.9	27215
owe	3.0	13.5	83.5	563
award	0.0	17.0	83.0	53
loan	1.4	16.9	81.7	71
email	5.6	13.0	81.4	2207
send	4.5	15.2	80.3	15497
mail	3.4	18.1	78.4	408
offer	7.6	19.2	73.1	536
feed	6.6	20.4	73.0	485
lend	9.7	22.1	68.2	972
deny	13.5	35.1	51.4	74
promise	3.8	45.3	50.9	53
pay	8.2	41.0	50.8	61
pm	1.5	48.1	50.4	131
show	4.1	48.3	47.7	2739
message	2.4	54.3	43.3	637
text	5.2	65.4	29.4	1260
ask	12.3	76.5	11.1	81

Table D.4: Cluster 3 (ₚPDAT-dominant): relative frequencies of [DATALT] variants by verb, ordered by ₚPDAT

APPENDIX E

Survey test sentences

---

sentence	verb	type_verb	verb_category	case	type_structure	description	theme	goal
After the game, he shouted the results the crowd	shout	ditransitive	communication		tgd	manner of communicat	dp	dp
After the game, he shouted the results to the crowd	shout	ditransitive	communication		pdat	manner of communicat	dp	dp
And after that shops we went	go	transitive	motion	allative	pd-drop	discourse shift		dp
And after that the shops we went	go	transitive	motion	allative	p-drop	discourse shift		dp
And then she asked me, who was the laptop given by Sophie?	give	ditransitive	give		passive		dp	dp
Apparently he sent the office the proposal yesterday	send	ditransitive	give		gtd		dp	dp
Are you heading t'pub later?	head	transitive	motion	allative	DAR	traditional dialect		
Can we go park with the fountain?	go	transitive	motion	allative	pd-drop	modified np (with pp)	p	p
Cary whispered the news me in class	whisper	ditransitive	communication		tgd		dp	p
Cute cat! Yes, my dad gave her me	give	ditransitive	give		tgd		p	p
Cute hamster! .... Yes, my dad gave her to me	give	ditransitive	give		pdat		p	p
Gary refused his bike me	refuse	ditransitive	refuse		tgd		dp	p
Great dog! .... Thanks, my brother gave me him	give	ditransitive	give		gtd		p	p
He began the station and went by foot, no idea where he is now.	begin	transitive	locative	locative	p-drop			
He has been refusing us his attention	refuse	ditransitive	refuse		gtd		dp	p
He's my only son, I will give him it all	give	ditransitive	give		gtd		p	p
He won't be coming - he's flying Germany tomorrow	fly	transitive	motion	allative	p-drop	no determiner		
I can't say I envy it him	envy	ditransitive	refuse		tgd		p	p
I didn't know that he gave the letter the bank on friday	give	ditransitive	give		tgd		dp	dp
I like the way he meandered his way the office, not a care in the world	meander	transitive	motion	allative	p-drop	manner of motion		
I'm going shop first thing tomoz do you want anything	go	transitive	motion	allative	pd-drop	unmodified		
I'm not certain but I think that the voucher was sent him	send	ditransitive	give		tpass		dp	p
I'm not sure, I think they followed river back	follow	transitive	motion	perlative	DAR	traditional dialect		
I'm not sure why I'm listening to the Beach Boys album, John lent me it	lend	ditransitive	give		gtd		p	p
I'm on my way, didn't Mark lend the drill them already?	lend	ditransitive	give		tgd		dp	p
I'm pretty sure he went with them the bar but he never called....	go	transitive	motion	allative	p-drop	intervening pp		
I'm working the library today, don't try and call!	work	transitive	locative	locative	p-drop	ambiguous		
I offered them the ticket yesterday	offer	ditransitive	give		gtd		dp	p
I should have told you, he lent the guy it afterall	lend	ditransitive	give		gtd		p	dp
I should tell you, we gave it the girl already	give	ditransitive	give		tgd		p	dp
I think he said he would take them both the zoo after work	take	ditransitive	motion	allative	p-drop	intervening pp	p	dp
I thought she said it's Chester they're moving?	move	transitive	motion	allative	p-drop	discourse shift		dp
I was just wondering, who was the card given John by?	give	ditransitive	give		tpass		dp	dp
If he guessed my bank balance I was going to give it him all	give	ditransitive	give		tgd		p	p
If he wins, they are going to give it all to him	give	ditransitive	give		tgd		p	p
In any case, who was the car given her by?	give	ditransitive	give		pdat		dp	p
Isn't it the cinema they're going, not the pub?	go	transitive	motion	allative	tpass	discourse shift		
It's a scanner/Printer thing. Someone gave it me	give	ditransitive	give		tgd		p	p

Example	Verb	Transitivity	Class	Alternation	Form	Note		
It's been bugging me, who was the book given to?	give	ditransitive	give		passive		dp	p
It was clear that she was given it	give	ditransitive	give		passive		p	p
Its for the best that they donated them it	donate	ditransitive	donate		gtd		p	p
Joe plodded the pub after work to watch the game	plod	transitive	motion	allative	p-drop	manner of motion		
Last night we went the pub with the big sports screen	go	transitive	motion	allative	p-drop	modified np (with pp)		
Maggie's just taking her time, ambling the shop after work	ambling	transitive	motion	allative	p-drop	manner of motion		
My dad was asking, who was the ticket given to John by?	give	ditransitive	give		passive		dp	dp
My friend wanted to know, who was the ticket given to her by?	give	ditransitive	give		passive		dp	p
No questions asked, they went straight the pub after that.	go	transitive	motion	allative	p-drop			
She denied the ice cream the child	deny	ditransitive	refuse		tgd		dp	dp
She donated the Alder Hey fund her loose change.	donate	ditransitive	donate		gtd		dp	dp
She's staying John's place tonight	stay	transitive	locative	locative	p-drop	no determiner		
She's waiting outside, she said she left everything the table for you	leave	transitive	locative	locative	p-drop			
She told me she envies you it	envy	ditransitive	refuse		gtd		p	p
She was so happy she virtually cartwheeled shop	cartwheel	transitive	motion	allative	pd-drop	manner of motion		
So, Shelley stashed the beers the freezer a few hours before	stash	transitive	locative	locative	p-drop			
That guy was asking, who was the book given?	give	ditransitive	give		passive		dp	p
The code was whispered Mary before Sally knew what was happening.	whisper	ditransitive	communication		tgd	nominalised	dp	dp
The cupboards are bare - an errand the shops is called for!	errand	transitive	motion	allative	p-drop			
The game was on so to the pub we came	come	transitive	motion	allative	no-drop	discourse shift		
The guys headed directly the office, so they said...	head	transitive	motion	allative	p-drop	modified verb		
There's no doubt that he was sent the book.	send	ditransitive	give		passive		dp	p
They mentioned that it was given her	give	ditransitive	give		tpass		p	p
They said they were coming with us cinema but didn't show up in the end	come	transitive	motion	allative	pd-drop	intervening pp		
They were saying that the letter was sent Mark.	send	ditransitive	give		passive		dp	dp
Very generous that he contributed them his money	contribute	ditransitive	donate		gtd		dp	p
We were talking about it, who was the cat given to by Karen?	give	ditransitive	give		passive		dp	p
Well, it was in good faith that I donated it them...	donate	ditransitive	donate		tgd		p	p
When they die they will give it all him	give	ditransitive	give		tgd		p	p
Yes, his brother introduced me her	introduce	ditransitive	donate		tgd		p	p
Yes, his brother introduced me to her	introduce	ditransitive	donate		pdat		p	p

tgd with pronominal theme and full DP goal

(1) a. Could have <u>give it the rest of us</u>

b. If you can't tell, <u>give it the attackers</u> as an advantage

c. Don't <u>give it the press</u> lol

d. have saved there money and <u>give it the food banks</u> imo

e. That's what I would do for Great British Bake Off, Stick a pie in a barm and <u>give it the judge's</u>

f. Mate I've took you of my xbox <u>I've give it the kids</u>

g. Too late <u>sent it Royal Mail</u>

h. That's where he'll be <u>putting it the big boss</u>

i. i <u>took it the bar</u> and got a free vodka (Birkenhead)

j. <u>give it the aintree KFC</u> mate (Birkenhead)

k. Can't even <u>take it the game</u> (Birkenhead)

l. If they go to the bother of dumping on a road, why not just <u>take it the tip</u>? (Birkenhead)

m. My dad dropped one of our gerbils when I was a kid, <u>took it the vet</u> and it's back leg got put in a cast

n. I had one of these when I was a kid but wouldn't <u>take it the match</u>

(Liverpool)

o. I double dare u <u>to take it the match</u> on sat (Liverpool)

p. It looks like I have fixed it myself instead of <u>taking it the shop</u> (Liverpool)

q. But if it's ****, <u>take it the dump.</u> (Liverpool)

r. <u>I've just took it the garage</u> (Liverpool)

s. I had it and <u>my Nan took it the shop</u> and I was gutted (Liverpool)

t. Rather <u>take it the tip</u> or weigh it in but that's not legal for me, so tip it is (Liverpool)

u. <u>send it the papers.</u> That's as good as finding Jesus in your toast. (Liverpool)

v. imagine if they did <u>send it the customs</u> (Liverpool)

w. My lad has put this over my ticket and told me <u>to take it the game</u> with me to stop my coat from stinking! (Liverpool)

x. Take to the corner, <u>take it the corner</u> (Manchester)

y. So my mum bought my Uncle a birthday cake and <u>brought it the pub</u> (Ormskirk)

z. ill pass you it as soon as im done, I only got it yesterday I need it for this presentation im <u>taking it the office today</u> (Runcorn)

p-Drop-*take* with full DP objects

(1)   Animate THEME

  a. he'd lost his marbles and was trying <u>to send the dogs the shops</u> with his shopping lists!! (Liverpool)

  b. Ok so I'm defo <u>taking the boys the park</u> tomorrow early

  c. Enjoyed the choirs and singing, but after that <u>we took the kids the community centre</u> then headed home. (Birkenhead)

  d. He's more Home Brew Hobby, had his kitchen redone divorced. Fox is more <u>taking the kids the zoo</u> on his access day divorced. (Birkenhead)

  e. When <u>taking the dogs the park</u> seemed like such a good idea (Liverpool)

  f. Hate <u>taking the cats the vet.</u> (Liverpool)

  g. Can't wait <u>to take the kids the match</u> when there old enough (Liverpool)

  h. my mum decided to <u>take the dog the park</u> apparently (Liverpool)

  i. Can't believe me mums just sent me <u>to take the dog the groomers</u> and her appointment was yesterday (Liverpool)

  j. As of Friday we can't even do a simple thing like <u>take the kids the playground.</u> (Liverpool)

k. Actually can't wait <u>to take the baby the zoo</u> tomorrow, big kid (Liverpool)

l. We're taking advantage of ferrying my Mother-in-law to Southport hospital <u>to take the dogs the beach</u> on Monday. (Runcorn)

m. what are we to do with this man in charge he is <u>taking the club the depths of despair</u> (Runcorn)

n. Really hope I only have to do half a day tomorrow, I want <u>to take the dog the beach</u> for a walk (St. Helens)

o. could never <u>take a bird the cinema</u> us, be well embarrassing (Warrington)

p. Bored of all this spice girl and liverpool talk so <u>brought the kids the pics</u>, treated meself to a mag (Warrington)

q. Can see some blue sky, <u>take the kids the park</u> for a kick about (Warrington)

r. got <u>to take the mrs the emergency docs</u> so I won't be on tonight (Wigan)