# 2   Methodology

In Chapter 1, example (15), we introduced the interaction between Luke and his caregiver, Molly, providing an initial 'taste' of the data to come. In this chapter, we provide some essential background information on the community these data come from, including the social, cultural and historical context and the nature of the linguistics that have arisen in that context. We provide a short summary of the adult data to be used as a baseline for comparison, but the bulk of the chapter concentrates on a detailed description of the caregiver/child corpus, including qualitative and quantitative analyses of the type of data with which we are working. The final part of the chapter details our analytical approach: how the methodologies of variationist sociolinguistics are employed to answer the research questions set out in Chapter 1.

The Research Site

The research site centres on the small fishing town of Buckie (population c. 9,000) and surrounding villages on the north-east coast of Scotland, sixty miles northwest of Aberdeen (Figure 2.1). The dialect spoken in this community is very distinctive (e.g. Smith 2000, 2001, 2004), and its distinctiveness can be explained in part by the sociocultural history in which the variety arose. To understand the linguistics of present-day Buckie and its surrounding villages, we need to look to the Buckie of the past.

Buckie in the Past

The community of Buckie has its roots as a fishing settlement of the eighteenth century, and the population grew steadily from that time onwards. A fishing boom in the nineteenth century meant that the coastal settlements along the Moray Firth grew dramatically. Despite this growth, Buckie was a close-knit community, where social ties were dense and multiplex (Milroy 1980). In his study of north-east fisher families in Victorian Scotland, Blaikie (2002: 22) points out that while other 'Scottish communities were sufficiently open to allow considerable migratory traffic in marriage partners and incoming

24
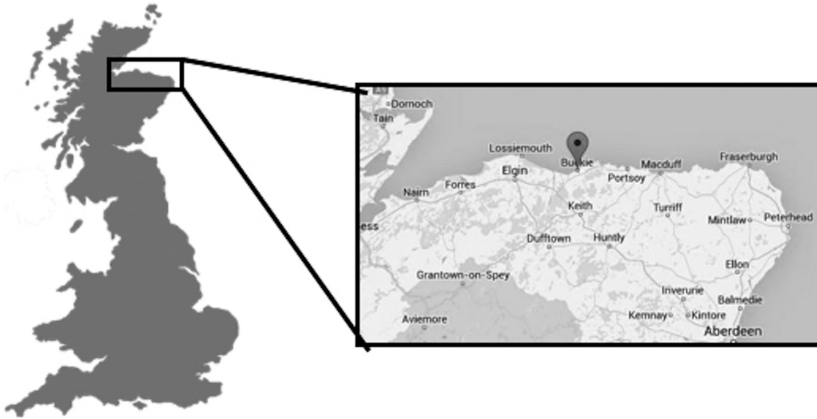
2.1 The research site: Buckie, Scotland (Google 2016)

families', this was not the case in the fishing villages of the north-east coast. As a result, 'the frequency of particular surnames within fishing communities indicates pronounced genetic concentration and reflects high levels of marital isonymy (marriage between individuals with the same surname)'. This is best demonstrated by the village of Portessie, just outside Buckie, where 41 per cent of the population were Smiths and the remaining 59 per cent shared just four other surnames (Blaikie 2002: 23). Such statistics demonstrate that in these small fishing communities on the north-east coast, 'their discrete social world maintained a hermetic family culture that brooked no incursion from the surrounding hinterland' (Blaikie, 2002: 29). Moreover, in these close-knit fishing communities, there were few class-based divisions of the type we think of in the present day. The New Statistical Account of Scotland for Rathven Parish (of which Buckie was a part), states that 'the fishermen, as individuals, are nearly placed upon a footing of equality. Their pursuits are similar; hence their language and dealings are almost the same' (Gordon 1845: 256).

Buckie in the Present Day

It is no surprise to find that Buckie has changed significantly over the last hundred years, embracing many of the innovations that have arisen over the subsequent decades. However, unlike many other rural areas in more recent history, the population of Buckie has remained stable; Figure 2.2 charts census data outputs from 1910 onwards.
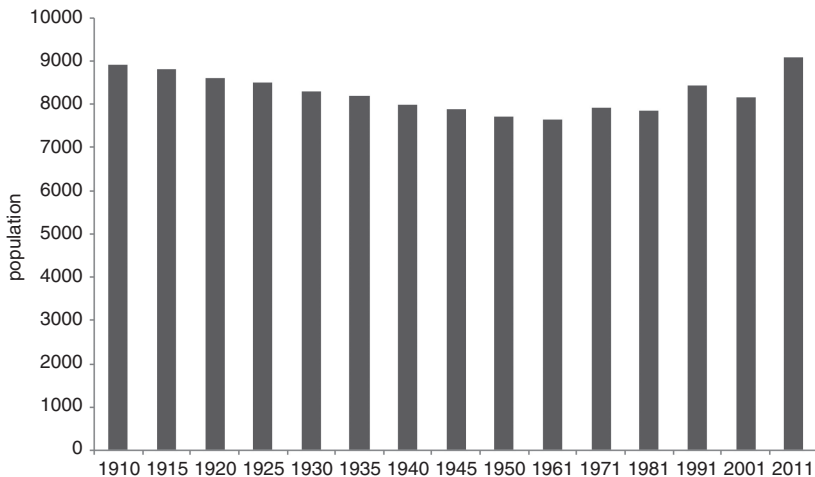
2.2 Population statistics for Buckie, 1910–2011 (adapted from University of Portsmouth et al. 2016)

The stability of the population may come from economic independence: first, in the form of the fishing industry and then, with the decline of fishing, the rise of employment in the oil industry from the 1970s onwards. This employment has meant that working-age adults, and particularly the young, do not need to leave the area to find employment (although this situation might change, given the dramatic decline in North Sea oil in the past few years). Further, this stability may be indicative of social stability more generally. The 2011 Census showed that Buckie has a total population of just over 9,000, with 99.3 per cent 'white' and 90.6 per cent 'white Scottish'. Of the population, 3.1 per cent were born outside the UK, and the majority of the non-Scottish population are from England (5.0 per cent). These statistics provide a snapshot of the demographics of Buckie: a small community with a tightly circumscribed ethnic base and a continuing sense of close-knit networks. It has certainly moved on from its Victorian past, but it retains a level of sociocultural cohesiveness that may be quite unusual in the twenty-first century. This is perhaps best demonstrated by the language used in the community.

The Linguistics of Buckie

The majority of speakers in Scotland speak Scots, which originally derived from Old Northumbrian English. In the present day, the term 'Scots' encompasses a wide spectrum of varieties, with so-called Broad Scots at one end and

Scottish Standard English – a variety generally described has being 'standard English with a Scottish accent' – at the other. The variety used in Buckie, and in this north-east corner more generally, is known as Doric and is at the broad end of the linguistic spectrum of Scots. Despite the fact that 'predictions of its demise continue apace, the "Doric" defiantly persists as an integral part of the region's identity and self-image' (McClure 2003: 15). This statement is very much borne out in practice: step out in the streets of Buckie and you will hear a very distinctive variety of Scots, as demonstrated in (19) from Poppy, a seventy-one-year-old speaker in the community, who is telling our fieldworker, Rose, about New Year's Eve (in Scots, 'Hogmanay').

(19)   ROSE: In that day, your father wouldna've gien to a pub, would he have?
        POPPY: No, your father di— Now Hogmanay the wifies got a wee sherry in a wee sherry glass. And the mannies got a whisky. That was your Hogmanay.
        ROSE: That was it.
        POPPY: That— there was nothin' in atween.
        ROSE: No.
        POPPY: A sherry or a whisky, that was fit— That was it and you used to hae oot a wee spread. Maybe a bit o' shortbread or a sponge that your mother had made. You used to sit and take the bells in. Ken this, fan I think aboot it, abody was happy.
        ROSE: Aye.
        POPPY: We'd all be whooching it up and obviously you n— you tried the sherry, like, fan you was a bairn.
        ROSE: Aye.
        POPPY: If you were caught, God help you.
        ROSE: Oh me, Poppy!

This short extract is full of forms from older stages of Scots and spans lexical, phonological and morphosyntactic variation. The extract includes lexical items such as *ken* for *know, wee* for *small* and *wifie* for *woman*, the negative clitic *na* for *n't* (as in *your father wouldna've gien to a pub*) and the prefix *a-* for *be-* (in *there was nothin' in atween*). To add to historical retention, there is innovation: the isolated nature of the north-east led to the development of a number of features that are not found elsewhere in Scotland, as exemplified in the use of *fan* for *when* (in *fan I think aboot it*) and its cohort of other *wh*-question words (including *fit* for *what* and *fa* for *who*). Thus retention and innovation have gone hand in hand in providing a highly differentiated variety at the lexical, phonetic and grammatical levels of language use.

The linguistics of this community hold a mirror to the larger context of development over the past hundred years, and Buckie's past is imprinted in its linguistic present. At the same time, Buckie is not completely immune to change. While many new so-called 'diffused forms', such as th-fronting and /l/ vocalisation (e.g. Kerswill 2003) remain absent in the dialect, despite

widespread use elsewhere, other supralocal forms, such as the *be like* quotative in (20) from thirteen-year-old Griff, and glottal replacement, as in example (21) from twelve-year-old Gillian, have spread rapidly in this community:

(20)   GRIFF: I asked Mum 'Fit did you do without a mobile phone?' and she *was like* 'Well I play my— eh I played Space Invaders' or something. I *was like* 'Mam, get with the times'.

(21)   GILLIAN: Well I ken how to do i[ʔ] on the iPhone bu[ʔ] I wouldna ken how to do it— like tell you fae tha[ʔ] . . . but there's usually like a li[ʔ]le bu[ʔ]on in your se[ʔ]ings and then it's like your Facebook contacts but then you can like change i[ʔ].

Against the backdrop of high mobility, urban growth and loose networks, this rural area remains largely monoethnic, with a stable population and very close-knit ties. In many ways, this community is quite unrepresentative of twenty-first century life. However, we argue that such a community provides an excellent test site for examining the mechanisms involved in the acquisition of structured variation at the earliest stages exactly because of its circumscribed make-up. As demonstrated in Chapter 1, we are still very much at the early stages in establishing how and when children acquire variation. Buckie thus provides an appropriately controlled research site for examining the processes underlying sociolinguistic development in these early stages.

We turn now to the data used in the present study, beginning with the key corpus of caregiver/child speech.

The Caregiver/Child Corpus

The review of previous research shows that in order to fully understand the acquisition of variation, we needed data not only from children; 'equal attention must be given to the detailed template' of caregiver speech (Labov 2001: 416). In addition, the recorded speech needed to replicate as far as possible the natural vernacular interactions that take place on a daily basis between caregiver and child. Thus, the first hurdle was to be able to find caregiver/child pairs who would be willing to take part in the study, and the second was to find a way to collect data that would model as closely as possible the everyday interactions that take place in these early years between these pairs.

Recruiting Participants

The first task of recruitment was greatly facilitated by the fact that the first author, Jennifer, is from Buckie. Most of her family still live there, as did her mother, Moira, who agreed to be fieldworker in the crucial phase of data collection. Moira's involvement meant direct access to this tight-knit

community, which circumvented the need for 'community gatekeepers' or 'brokers' (e.g. Schilling-Estes 2007). Moreover, Moira was an extremely well-known figure in the community, which meant that the potential difficulty of, for example, 'cold-calling' was circumvented. Using the criteria for selection noted below, she identified potential participants and contacted them initially by telephone. She then made a follow-up face-to-face visit to each caregiver to talk them through what the recordings would involve. This approach proved to be invaluable in persuading caregivers to take part, with 91 per cent of those approached agreeing to participate in this study.

We used judgement sampling (e.g. Chambers 2003: 44), in this recruitment phase, as we had a checklist of predefined criteria for the participants. In order to control the sample as much as possible, we started with a highly elaborated list of criteria for inclusion in the study, including, for example, no older siblings. However, as with most studies of this type, we had to remove a number of constraints in order to recruit enough participants.

We focused in the end on four main criteria:

(1) both parents must have been born and raised in the area
(2) no referrals for speech or language therapy
(3) the mother was the main caregiver (i.e. there was no substantial time with extended family, childminders, etc.)
(4) no child in formal, state-run nursery education.

These criteria were imposed because they have implications for the ambient language to which the child is exposed and how this might impact on language production. For example, previous studies have found that parental input impacts a child's ability to acquire a particular dialect (e.g. Payne 1976; Kerswill & Williams 2002), while the more formal language used in institutional settings such as formal education may also have an impact (e.g. Romaine 1984; Patterson 1992; Chevrot et al. 2000).

We also wanted to capture children's vernacular use at a crucial stage of language development, and 'the 3- to 4-year old level is a critical period for the acquisition of dialectal norms of the speech community' (Roberts & Labov 1995: 110). In addition, while most studies of language acquisition start where language starts (i.e. in the first months, when the child moves from babbling to the one-word phase), we needed data that would allow us to investigate a range of variables, from individual phonemes to use of whole clauses. For this reason, we targeted children between the ages of 2;6 and 4;6, with the final sample ranging from 2;10 to 4;2. Data collection ran from December 2003 to June 2004.

Table 2.1 shows the caregivers and children in this sample.

The Recordings

Studies of first language acquisition are often longitudinal, consisting of a number of recordings over a set period of time. In this study, we chose to

2.1 *Caregiver/child corpus*

Child's name	Caregiver's name	Child's sex	Child's age at time of recording
Elizabeth	Martha	F	2;10
Jade	Avril	F	2;10
Ricky	Sheila	M	2;10
Emma	Danielle	F	2;10
Ellen	Joyce	F	2;10
Jennifer	Karen	F	2;10
Max	Alice	M	2;11
Charlie	Amy	M	2;11
Isobel	Mary	F	3;0
Heather	Tania	F	3;1
Kevin	Joanne	M	3;1
Kerry	Paula	F	3;2
Stephen	Donna	M	3;2
Jake	Liz	M	3;2
Becky	Amanda	F	3;2
Billy	Sarah	M	3;2
Lyle	Fran	M	3;3
Annie	Kimberly	F	3;3
Gareth	Carol	M	3;3
Lucy	Lesley	F	3;4
Ellie	Gail	F	3;5
Luke	Molly	M	3;5
Kieran	Ruth	M	3;6
Gus	Mandy	M	3;9
Oliver	Judy	M	3;9
Marie	Eileen	F	3;10
Izzie	Denise	F	3;11
Ella	Louise	F	3;11
Dan	Suzie	M	4;2

gather cross-sectional data – recordings of each speaker pair across one time period. In doing so, our aim was to gain a 'snapshot' of caregiver/child interaction across a range of ages. Further, in much of the previous research of preschooler language, the researcher has been present during the recordings. Crucial to our endeavour was obtaining data that replicated as far as possible the everyday interactions of caregiver and child. Having the researcher present, even though in this case they are from the community, would highly compromise this endeavour, as issues of the Observer's Paradox (Labov 1972) are even more acute where preschool children are concerned; the presence of an 'unknown' observer in close proximity may result in a child being 'speechless'. This was clearly demonstrated in cases where the first author visited some participants to check on progress: despite her attempts to engage in

conversation, the lips of the child were clamped tightly shut, only to be opened to utter *I'm nae speakin' to you*, or some other equally withering response. This meant that we needed to come up with a data collection technique that would allow us to capture more felicitously the everyday speech patterns between parent and child. To do this, we provided the caregivers and children with lightweight recording equipment (Sony MZ-R700 MiniDiscs and Sony ECM-T145 lapel microphones) and asked them to record themselves in a range of situations over a one-week period. Each caregiver and child had a separate recorder and lapel mic, which meant that they were not constrained to sit in one place for an hour's recording or to be in close proximity to each other at all times. The lightweight recorders were placed in the child's trouser or coat pocket, and the lapel mic was placed under clothes to ensure that it stayed in place.

We asked that the recorded interactions total ten hours. This may seem like quite a lot, given that sociolinguistic research typically aims for between one and two hours of recorded speech during a sociolinguistic interview (Labov 1984), but simply put, kids talk less. Roberts (2002: 336), for example, estimated that eight to fourteen hours of young children's speech is required to collect data comparable to that found in one to two hours of adult speech. We were not prescriptive as to what activities the caregivers and their children should engage in, and we stressed that they should not do anything 'special' for the recordings, but merely carry on their everyday lives. This, of course, is partly a fiction – in everyday life it is unusual to be hooked up to a recorder – but what we were aiming for was the closest possible approximation to caregiver/child interaction and the language that arises from this.

The Challenges of Data Collection with Young Children

As with all data collection, a number of problems arose, exacerbated by the fact that this procedure had not been tried and tested to any extent in previous research. The first problem was time: although we aimed to have each participant complete the recordings within one week, in reality most pairs took longer, with an average of two weeks' collection time. The majority of pairs completed all ten hours of recording, but this was not without its issues, as we will discuss.

The next problem was that some children simply did not want to wear the recording equipment, as shown in (22) and (23):

(22)   Billy (3;2) and Sarah
        SARAH: Oh has your microphone come off?
        BILLY: Aye. I bite it.
        SARAH: Dona bite it.
        BILLY: Why?
        SARAH: Nae bite it.
        BILLY: Why?

(23)   Charlie (2;11) and Amy
    AMY: Put on your microphone. I've got my one on.
    CHARLIE: No, I don't want to.
    AMY: Aye, but listen.
    CHARLIE: I don't want mine.
    AMY: Aye, but we're going to go over to gran's and get your jelly, remember?

A range of creative tactics was employed by the caregivers to attempt to persuade the children to wear the equipment, from general cajoling to more overt bribes, as shown in example (24):

(24)   Annie (3;3) and Kimberly
    KIMBERLY: We'll take our special recorder to Elgin with us.
    ANNIE: No.
    KIMBERLY: How? Pop it in your baggy.
    ANNIE: No, it's annoying me.
    KIMBERLY: It doesna annoy you in your bag though, does it? Nae at all. Keep it on. Remember what we said about Santa? Remember this is the way that Santa hears you, isn't it? He can listen to what you're saying and say, 'I think Annie maybe has been a good girl 'cause I can hear her speaking'. And then Santa can hear athing that you're wanting for Christmas, can't he?

Once the recording equipment was in place, another problem arose. We had suggested that the recorders be carried in a trousers or coat pocket, but we had not foreseen the weight of the recorder as an issue (25–26):

(25)   Annie (3;3) and Kimberly
    KIMBERLY: I'll get your bag to put your recorder in so it doesna annoy you. 'Cause that's that trousers that made it fall down the last time, wasn't it?
    ANNIE: Aye.

(26)   Kieran (3;6) and Ruth
    KIERAN: I don't want it on.
    RUTH: Leave it on. Just for half an hour.
    KIERAN: My trousers is falling down.
    RUTH: Well I'll sort your trousers out. Here, wait a minute.
    KIERAN: Here, take that off.
    RUTH: No, you have to keep it on. Kieran, half an hour.
    KIERAN: For half an hour?
    RUTH: Aye. Right, now go play.

The microphones were also dropped, pulled and even eaten (27):

(27)   Max (2;11) and Alice
    ALICE: Right, just leave that. Try not to eat the microphone. Can't be hungry, you've just had your tea.
    MAX: But I'm still hungry.

It has to be said however that many children positively loved the novelty of being recorded, as demonstrated in (28–30):

(28)   Oliver (3;9) and Judy
        OLIVER: Mam, look at me, I'm walking and running.
        JUDY: You're walking and running? You jogging? Jogging with your celebrity mic pack on, are you?
        OLIVER: Aye.
        JUDY: And we're gonna shout, 'I'm a celebrity, get me out of here!' Will we?
        OLIVER: Aye.
        JUDY: Right, hang on to it just like that.
        OLIVER: 'Celebrity, get me out here!' Come on, mam, we're a celebrities.

(29)   Kevin (3;1) and Joanne
        KEVIN: Is this my microphone, mam?
        JOANNE: That's your microphone.
        KEVIN: [shouts into microphone] Mam! Mam! Mam!

(30)   Annie (3;3) and Kimberly
        ANNIE: I'm making it all bonnie.
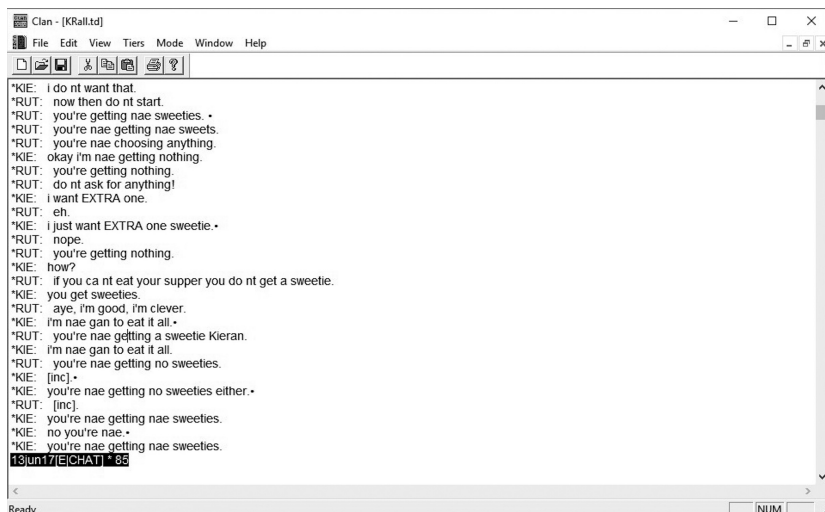   KIMBERLY: You're making your microphone bonnie?

A more serious problem arose with regard to the quality of the data. Our request that the participants carry on as normally as possible resulted in some of the recordings not being of high enough quality for fine-grained linguistic analysis. For example, a walk along the cliffs in this north-east corner of Scotland in high winds meant that the voices were all but obscured by the inclement weather. Some of the bath-time recordings were similarly compromised, with the splish-splash of water drowning out conversation. In the end, for each caregiver/child pair, we selected five hours of speech in which the quality was good enough from the ten hours available in order that all speakers had the same amount of data in terms of length of recording.

In sum, caregivers and children go for walks in the wind, take baths, eat baked beans. This is the language of the home, not of the laboratory, and this methodology decisively captures the diverse linguistic contexts of use that arise naturally between caregiver and child in everyday interaction. Our next task was to transcribe the data.

Transcription

To process the data, we created time-aligned transcripts of each sound file using CLAN (MacWhinney 1991), a computer programme designed to facilitate the transcription of child language data (Figure 2.3).

A strict transcription protocol was followed for all recordings, which laid out the conventions for representation of the speech data. We did not attempt to represent at this first pass every nuance of variation found in the data, as this

```
Clan - [KRall.td]                                          —    □    ×
File  Edit  View  Tiers  Mode  Window  Help                      _ ⊡ ×

*KIE:   i do nt want that.
*RUT:   now then do nt start.
*RUT:   you're getting nae sweeties. •
*RUT:   you're nae getting nae sweets.
*RUT:   you're nae choosing anything.
*KIE:   okay i'm nae getting nothing.
*RUT:   you're getting nothing.
*RUT:   do nt ask for anything!
*KIE:   i want EXTRA one.
*RUT:   eh.
*KIE:   i just want EXTRA one sweetie.•
*RUT:   nope.
*RUT:   you're getting nothing.
*KIE:   how?
*RUT:   if you ca nt eat your supper you do nt get a sweetie.
*KIE:   you get sweeties.
*RUT:   aye, i'm good, i'm clever.
*KIE:   i'm nae gan to eat it all.•
*RUT:   you're nae getting a sweetie Kieran.
*KIE:   i'm nae gan to eat it all.
*RUT:   you're nae getting no sweeties.
*KIE:   [inc].•
*KIE:   you're nae getting no sweeties either.•
*RUT:   [inc].
*KIE:   you're nae getting nae sweeties.
*KIE:   no you're nae.•
*KIE:   you're nae getting nae sweeties.
13jun17[E|CHAT] * 85

Ready                                                    NUM
```

2.3 Screenshot of transcription in CLAN

would effectively mean carrying out multiple analyses at the same time as attempting the already onerous task of transcription. Instead the transcription protocol represented faithfully the local lexical and morphosyntactic differences found in each pair's interactions (e.g. in Figure 2.3, *how* for *why*), but used standard English orthography for more fine-grained phonological differences. For example, the lexical item *sweetie* in the transcription in Figure 2.3 appears as both *swee[t]ie* and *swee[ʔ]ie* in the actual recording. The same is true of *-ing* forms; many of these appeared with the alveolar realisation, but all were represented with *-ing* in orthography. To avoid full analysis in this initial stage, we did not annotate such variable phonological realisations. The utterances were broken down into finite clauses, and these 'first pass' transcripts were further annotated in subsequent analyses where necessary.

These protocols were followed diligently in the transcription phase, but we adapted them to aid the reader in interpreting the interactions from here onwards, with key glosses provided in the section that follows.

Key Linguistic Differences in the Examples

Negation appears with the clitic *na*:
don't   *dona* pronounced [dɪnʌ]
haven't   *havena*

can't	*canna*
won't	*wi(ll)na*
isn't	*isna*
aren't	*arena*
couldn't	*couldna*

The non-cliticised negative marker is *nae*, pronounced [ne]:

He's *not* here   He's *nae* here.

Prefix *be-* can also be realised as *a-* in preposition:

between	atween
before	afore
beside	aside

Every- can alternate with a- in quantifiers:

everybody	abody
everything	athing
everywhere	awey (where wey is where)

There is an alternation between [ʍ] and [f] in wh- question forms, resulting in the following paradigm (which we analyse more fully in Chapter 5):

what	fit
where	far
when	fan
who	fa
why	fitna wye (literally 'in what way')

There are also many 'unshifted' vowels, such as *oot* for *out, aboot* for *about, stane* for *stone* and *hame* for *home*. These are represented in the standard form, unless they are the focus of analysis (see Chapter 5).

As with all other varieties, there are lexical items specific to Scots and/or Buckie. These are too numerous to mention here, but notable examples are *ken* for *know, mind* for *remember, wir* for *our, wifie* for *woman* and *ain* for *one*. (Some of these lexical alternations are examined in more detail in Chapter 4.) Diminutives are used very frequently in the dialect, including *sweeties* (as in Figure 2.3), *boaties, housies* and *glovies*, among many others. Where understanding of the data is compromised, we include a gloss within the interactions in square brackets. Speech that is incomprehensible is marked with [inc].

The fully transcribed data totalled c. 700,000 words and revealed a number of themes with regard to the types of interaction between caregiver and child, as we now detail.

Given that we asked the caregivers to simply carry on their everyday routine, it was not surprising to find that a lot of interactions took place around food, including preparing or eating breakfast, lunch and dinner, as in (31) and (32):

(31)   Ellen (2;10) and Joyce and muck
        JOYCE: What do you want for your dinner?
        ELLEN: Muck [dirt].
        JOYCE: What do you want to eat for your dinner?
        ELLEN: Muck.
        JOYCE: You don't eat muck. Would you like soup?
        ELLEN: Muck.

(32)   Dan (4;2) and Suzie and pancakes
        SUZIE: Come back here and eat that or you willna be getting nae pancake out of Baxters.
        DAN: No!
        SUZIE: Aye, you ca— dona hae to eat that bit, you can just eat the— that bit.
        DAN: Only this bit?
        SUZIE: Aye. Oh! What a mess. Your hands is all sticky now, you'll hae to wash them. Look at the mess, it's over the table. You're a right messy loon [boy]!
        DAN: Um-hum-hum.
        SUZIE: And far you gan now? That doesna look finished to me. And dona bother asking for sweeties if you canna eat your dinner.

Quite a lot of data revolved around the simple function of getting dressed (33):

(33)   Oliver (3;9) and Judy and Spiderman
        JUDY: Uh-huh. And do you want boxers or pants?
        OLIVER: No mam, I'm wanting pants.
        JUDY: Pants the day? Nae boxer shorts?
        OLIVER: No I'm just wanting, um and I'm wanting my Spiderman jumper.
        JUDY: Your Spiderman t-shirt?
        OLIVER: Aye.
        JUDY: Well, I'll have a look see fit we can do.

Some of the data were more set activity-based, often around games, artwork or reading as in (34–36):

(34)   Ella (3;11) and Louise and I spy
        ELLA: I spy with my little eye something the colour is pink.
        LOUISE: Ella's jacket.
        ELLA: I spy with my little eye something the colour of red.

LOUISE: The Tweenies book. I spy with my little eye something the colour silver.

ELLA: Wir car!

(35)   Emma (2;10) and Danielle and counting fingers

DANIELLE: Right, you got to spread out your hand again if you're gan to draw around your handprint. How many fingers have you got? Let me hear you counting.

EMMA: One, two, three, four.

DANIELLE: Four hands— four fingers. Wow, so you have!

(36)   Lucy (3;4) and Lesley and Scooby Doo

LESLEY: [reading] A few days later the gang came back for more donuts. But the shop looked different. And Dora was packing everything up. She was closing Doodle's Donuts. 'I have a big problem', she told the gang. 'Donuts have been disappearing' [stops reading]. Fit's that picture of there? Fit's that?

LUCY: A beastie [insect].

LESLEY: And fit's that?

LUCY: A ghost. Mam, fit's that?

LESLEY: Scooby Doo's van.

As to be expected in child data, there was much discussion centred around make-believe (37). Interestingly, this was often done in a very 'knowing' way, as in (38) and (39):

(37)   Jade (2;10) and Avril and a cup of tea

JADE: I'm mading you a cup of tea, Mam.

AVRIL: Okay.

JADE: Here's a cup of tea there, Mam. Here, Mam, here's a cup of tea.

(38)   Lucy (3;4) and Lesley and Kika

LUCY: You can do it on Kika.

LESLEY: Fa's [Who's] Kika?

LUCY: Remember my imaginary friend?

LESLEY: I canna see her.

LUCY: Look. She's there.

LESLEY: Right, okay then. Fit kind of hair has Kika got? Has she got long hair or short hair?

LUCY: She's got girl hair. Perfume hair.

LESLEY: Perfume hair. Is it long? Has she got bobbles in it?

LUCY: No it's erm, purple hair.

LESLEY: Purple. I've never seen a little girl with purple hair afore.

LUCY: That's nae her hair colour.

LESLEY: No, purple's that colour, isn't it?

LUCY: Aye, so that's her hair up to there. And her car's a bus.

LESLEY: Her car's a bus?

LUCY: Aye.

LESLEY: A bit like J. C.'s ain [one]?
LUCY: No, a top ain.
LESLEY: Oh a double decker ain.
LUCY: Aye.
LESLEY: Like fit we was on in Portugal? How come she's got such a big car like?
LUCY: It's nae got a roof up to the bitty on the—
LESLEY: Oh so it's a double-decker. And it's nae got a roof on the top.
LUCY: No.
LESLEY: Like an open air ain?
LUCY: And ken fit?
LESLEY: Fit?
LUCY: Kika goes up the top. She keeps coming upstairs with me. And I na like it when she comes up the top with me.
LESLEY: Do you nae?
LUCY: No. It's funny. Erm she broke her bus on two days.
LESLEY: Did she?
LUCY: And she broke her bus.
LESLEY: How did she break it? Did she get a flat wheel?
LUCY: Aye.
LESLEY: Did she get a new tyre put onto it?
LUCY: Aye.
LESLEY: Well that was a bit silly. How did she manage that?
LUCY: She— they just kicked it. The children. It went down in the road. And she couldna drive home to the bobbies [policemen].
LESLEY: [laughing] Did you dream this, did you?
LUCY: Aye . . . no. Kika dreamed this.
LESLEY: Oh was Kika telling you like?
LUCY: Aye.
LESLEY: You've a right good imagination, haven't you?

(39)   Luke (3;5) and Molly and doggies
LUKE: I'm gan to put the paws on. Do you need a paw? Do you need another paw? Look, I'm a dog woof woof. Do you need a paw?
MOLLY: Good dog. Clever dog. Well done.
LUKE: Are all the pieces on?
MOLLY: Are all the pieces of what on?
LUKE: Of the dog.
MOLLY: Aye.
LUKE: Right I'm nae a doggy anymore.
MOLLY: You're nae a doggy anymore?
LUKE: No.
MOLLY: How nae?
LUKE: Cos I'm nae. I'm a doggy. I'm a doggy. I'm a doggy. Do you need a paw? Do you? Do you?
MOLLY: Paw. Good doggy. Clever doggy, well done.

Most of the interactions were between the caregiver and child. However, there were also 'third parties' involved, especially fathers, grandparents and, most often, siblings (40). Common themes were sibling rivalry (41) and sibling blame (42):

(40)   Ellie (3;5) and Gail and her baby brother Tommy and Barbie plates
GAIL: You're doing really well. Tommy's really eyeing up that chicken dippers, you ken.
ELLIE: [addressing Tom] Look, do you see Barbie?
GAIL: [addressing Tom] See fan she eats all her dinner, Tommy, she sees Barbie at the bottom o' her plate.
ELLIE: We- we'll buy a Barbie plate for him?
GAIL: Oh I na think he'll like Barbie.
ELLIE: Nae like Barbie?
GAIL: Boys dona like Barbie. No, I think he'll maybe like Bob.
ELLIE: Well we've got a bumble bee ain for him!

(41)   Lucy (3;4) and her older sister Beatrice and Bob the Butcher
LUCY: I'm finished. Are you finished, Mam?
LESLEY: I am, aye.
LUCY: Come on then, Mam.
LESLEY: Are you gan to get your frock on and we'll go up to Bob's and get this mince?
LUCY: No, but we're watching the wee— the video.
LESLEY: Well can we nae do it when we come back down?
LUCY: Okay then.
BEATRICE: I'm finished.
LESLEY: Get your frock on then.
LUCY: Dona ken far it's at.
LESLEY: Far did you put it?
LUCY: I na ken.
BEATRICE: It's down there.
LESLEY: I see it. You're sneaky, you was hiding it fae me.
LUCY: I dona ken far it's at now.
LESLEY: There it goes.
LUCY: I dona ken far it's at.
LESLEY: Are you wanting me to go myself?
BEATRICE: I'll come with you.
LUCY: I can see it. [addressing Billie] You dona ken far Bob's is at.
BEATRICE: Aye I do.
LUCY: No you do not.
BEATRICE: Aye I do so.
LUCY: No you do not.
BEATRICE: Aye I do so. It's across there.
LUCY: I've been to Bob's afore, haven't I mam? I've been to Bob's afore, haven't I mam?
LESLEY: Aye.

(42)   Gareth (3;3) and Carol and Humpty Dumpty
  CAROL: Who's that?
  GARETH: Humpty Dumpty.
  CAROL: Humpty Dumpty. Are you gan to do Humpty Dumpty for mammy
    then?
  GARETH: Yeah. Humpty Dumpty.
  CAROL: —ty sat on a . . .
  GARETH: Wall.
  CAROL: Humpty Dumpty had a big f—
  GARETH: Fall.
  CAROL: Fall. All the King's . . .
  GARETH: All the King's.
  CAROL: ho— horses
  GARETH: horses.
  CAROL: . . . and all the King's men couldn't put Humpty together again.
  GARETH: Together again.
  CAROL: Uh-huh.
  GARETH: I bet he slipped.
  CAROL: [gasps] Who did that?
  GARETH: Nathan.
  CAROL: No way!
  GARETH: Yes, it was.

A proportion of the data revolved around politeness routines, with overt
teaching from the caregivers, as in (43):

(43)   Lucy (3;4) and Lesley and burping
  LESLEY: What do you say? Pa—. What do you say? What do you say? Pa—.
    What do you say? Pardon me. Don't you?
  LUCY: No.
  LESLEY: You div [do] so.
  LUCY: You say 'I'm not allowed burping'.
  LESLEY: You are allowed burping, but you're supposed to say 'pardon me' or
    'excuse me'.
  LUCY: No.
  LESLEY: Aren't you?
  LUCY: Excuse me.
  LESLEY: Aye, that's better.

Not all interactions took place in the house: the mobility of the recording
equipment meant that data could also be collected on, for example, walks
(44) and going to the shops (45).

(44)   Kerry (3;2) and birdies
  DAD: See all them birdies, Kerry?
  KERRY: I dona see any bears, or any birdies. I dona hear them either. I can't
    see them.

(45) Gus (3;9) and Mandy and the fish shop
GUS: Mam, what's that? Mam, what's that?
MANDY: That's kippers.
GUS: What's that? That?
MANDY: That's the same fish.
GUS: Mam, what's that big ones? Mam, what's that big ones?
MANDY: Mmm?
GUS: See fishies in there?
MANDY: Kippers are tasty.
GUS: Mam, see inside there. Inside.
MANDY: It's fish. Look!
GUS: Mam . . .. Mam . . .. Mam!
MANDY: Uh-huh.
GUS: Mam, what's them? Mam, what's them? What's them?
MANDY: Scallops.
GUS: Mam, I'm hungry again.

Not surprisingly with such large amounts of data, there was quite a bit of scolding (46)!

(46) Gareth (3;3) and Carol and being cheeky
CAROL: You're winding me up the day. Stop it!
GARETH: You're winding *me* up.
CAROL: You're being cheeky as usual.
GARETH: *You* are being cheeky.

And, of course, there were many occasions when nature called (47):

(47) Max (2;11) and Alice and you know what
MAX: I'm gan to poop. Need a poop.
ALICE: Hmm?
MAX: Need a poop. Need to do a poop. I need to do a poopy now.
ALICE: Take off your trousers.
MAX: But I've got braces.
ALICE: Well I'll help you.

Finally, in line with previous studies in first language acquisition (e.g. Brown & Hanlon 1970), we should point out that there were very few examples of correction of non-target forms (so-called negative evidence), as in (48):

(48) Gareth (3;3) and Carol and broke-blaming
GARETH: Why you broked it?
CAROL: I never broked it.
GARETH: What if Nathan broked it?
CAROL: Broke it.
GARETH: What if Nathan did broke it?
CAROL: We'd have to clap his lug, wouldn't we?

GARETH: Yeah and give him a hot backside.
CAROL: And a hot backside, aye.

Most of these went uncommented upon, as in (49):

(49)   Isobel (3;0) and Mary and mading things
        ISOBEL: What are you mading?
        MARY: I'm making a bed.

As this demonstrates, we have an extremely rich array of data from these young children and their caregivers, providing an excellent resource for tracking the emergence of structured variation at these early stages.

The Community Corpus

To provide the most controlled comparison between community norms on the one hand and caregiver and child norms on the other, ideally we would have speech data from the caregivers talking to their children and also talking to other adults. Although we do not have such data, we have access to a large database of speech from other adult speakers in the community, which we use as a baseline for comparison of general community norms.

The community data we employ here was collected in 2013 as part of a larger project on bidialectalism within and across speakers (Smith 2013–2016). In this project, we recorded speakers talking to a community insider and then a community outsider. For these purposes, we use only the data from interviews with the community insider, as these represent the most vernacular form of speech as accessed through the 'sociolinguistic interview' (Labov 1984). We should note that these data were collected ten years after the caregiver/ child data, which may compromise issues of comparison, especially in cases where a variable is in the process of change. To provide the best comparison possible, we use the middle-aged speakers' data only: eight men and eight women, aged thirty-five to fifty-five. We selected these speakers because when we factor in time of data collection (caregiver/child corpus in 2003–4, community corpus in 2013), the caregivers in these data – in their late twenties and early thirties at the time of recording in 2004 – would now be close in age to this middle-aged generation in 2013. This community corpus totals 135,000 words.

As with the caregiver/child data, these data represent highly vernacular use. The fieldworker for this project, Rose, knew all the participants well, and the resulting recordings are relaxed and spontaneous, as demonstrated in example (50) from Sandy, a middle-aged man, and in (51) from Lana, a middle-aged woman:

(50)   SANDY: There's a quinie that won America's Got Talent a few year ago and
                 I forget her name but she has just— ken she was fit, aboot ten or

something but she'd this voice that just came oot o' her and you're
like 'Wow, far's that coming fae, you ken fit I mean?'

(51)   LANA: Far I went there to get my trench coat fan they come in aboot. Got my
trench coat oot o' Time Machine and your leather jackets was oot o'
Sheila's shop, mind, in Buckie? Fit was it ca'ed again? Was it just
ca'ed Sheila's? Mind she did a' the leather jackets with the— 'cause
I've still got hit, I've still got my leather jacket and my nevica and my
trench coat. You nae mind they were a' different colours?

The Analyses

As indicated in Chapter 1, the language we first learn is our caregivers' verna-
cular (Labov 2001: 307). The fundamental question arising from this is *when* and
how do children acquire the variable norms of community and caregiver?
In order to answer that question, our analyses centre on a range of linguistic
variables taken from different levels of the grammar: lexical (52), lexical-
phonological (53), phonetic (54) and morphosyntactic (55):

(52)   Billy (3;2) and Sarah
        SARAH: Fit did he make the girls do?
        BILLY: *Cry.*
        SARAH: *Cry. Crying.* See fit's that girl got for her tears?
        BILLY: Aye she's *greetin'.*
        SARAH: She's *greetin'.* Why is she *greetin'*?
        BILLY: I dona ken.
        SARAH: 'Cause Georgie Porgie came out to give her a kiss.

(53)   Lucy (3;4) and Lesley
        LUCY: I could phone the bobbies.
        LESLEY: Oh.
        LUCY: Kick the bobbies, kick the washing machine, so they can get me
                 [uː]t.
        LESLEY: Aye.
        LUCY: That's a good idea.
        LESLEY: It is.
        LUCY: They can kick that bit and that bit. And they can get me [ʌʉ]t!

(54)   Marie (3;10) and Eileen
        MARIE: Is this goi[n] to be the yummiest chocolate ever, is it mam?
        EILEEN: Mmm. I canna wait to taste it, Marie.
        MARIE: Me too. I'm mixi[n] it up. It's starti[n] to melt mam … We like
                 chocolate melti[ŋ], do we?

(55)   Charlie (2;11) and Amy
        CHARLIE: And *there's lions* and elephants and zebras and sheep and dogs
                  and—

AMY:	*Are there any hippopotamus*?
CHARLIE:	No, mammy.
AMY:	No hippos? I like hippos.
CHARLIE:	*There's monkeys.*
AMY:	*Oh, are there monkeys in there?*

To date, most analyses have concentrated on one linguistic variable in the acquisition of variation, but much more can be uncovered in comparing and contrasting a number of different variables, as noted in Chapter 1. As Tagliamonte and Molfenter point out, the principles of what is acquired when 'arise not only from the behavior of individual variables, but crucially on pointed comparison across them' (2007: 656). By comparing and contrasting a range of variables that include different levels and different complexities (e.g. Kerswill 1996, Chambers 1992), we will be able to assess to what extent variables behave differently in the acquisition of variation.

For each of the variables in question, we will conduct a series of analyses across a number of factors that may govern the observed variation:

(1) community, caregivers and children in overall distributions
(2) individual speakers
(3) ages of children
(4) gender
(5) caregiver/child pairs
(6) different situational contexts
(7) different linguistic constraints.

Comparison Across Community, Caregivers and Children

As indicated in Chapter 1, CDS has been shown to be qualitatively different from adult-to-adult speech, with modifications at the lexical, syntactic and phonetic levels including higher pitch, slower speech rate and simple sentence structure (e.g. Snow 1995). We also showed that CDS might differ with respect to variation: caregivers use fewer local forms in CDS when compared to adult norms (e.g. Foulkes et al. 1999, 2005; Roberts 2002). Put very simply, what we might expect is that when adults talk to each other, they would make a lot of use of *aye, r*[u:]*nd* and *be*[ʔ]*er*, but when caregivers talk to children, they would make more use of *yes* and *r*[ʌʉ]*nd* and *be*[t]*er*. This is said to arise from concern on the part of caregivers, and in particular female caregivers, with 'teaching' their children the 'proper' language. At the same time, not all communities studied showed this effect (e.g. Van Hofwegen & Stob 2011), nor did all variables (Smith et al. 2007, 2013). In order to test caregiver input vis-à-vis community norms, across each variable we compare the rates of use of particular features across community and caregivers (schematised in Figure 2.4).

2.4 Comparison between community and caregiver

2.5 Comparison between caregiver and child
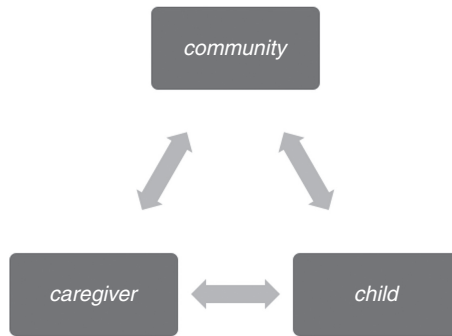
2.6 Comparison of community, caregiver and child

The variationist literature suggests that there is a statistically significant correlation between caregiver and child rates of use across a range of variables (e.g. Kerswill & Williams 2005; Patterson 1992; Nardy et al. 2013). In other words, if a caregiver uses lots of glottal stops, then their child will too. We test this in these data by comparing overall rates of caregiver input to child output (Figure 2.5).

These analyses provide a triangularisation between community, caregiver and child (Figure 2.6) that will allow us to compare and contrast what is the same, what is different and why across these three different groups.

We also test for the correlation between caregiver input and child output, both in overall distributions and across caregiver/child individual pairs (Figure 2.7).

Individual vs Group Norms

In sociolinguistic research, variation is generally treated as a community-based norm, with each individual conforming to a pattern of use. However, there may

2.7 Individual caregiver/child pairs
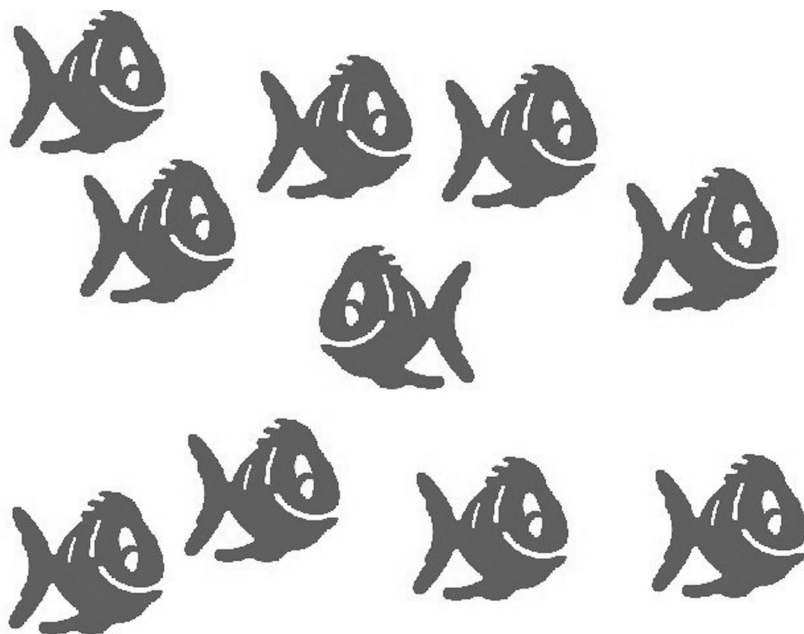
2.8 The individual swimming against the tide of group norms

be 'outliers' who do not conform in this way (e.g. Guy 1980), oddballs 'whose speech seems completely anomalous' (Chambers 2003: 93) and who are swimming against the tide (Figure 2.8). As Britain (2003) points out, it is important not to ignore such speakers if a full picture of the variation is to be gained. Thus, we will also make note of individual differences in use across the different variables.
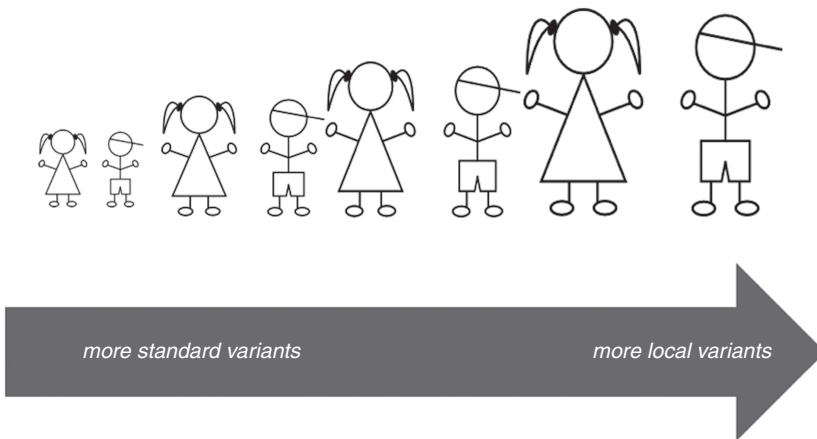
2.9 The effect of children's age on use of variants

Age of Child

Research has shown that caregivers decrease the use of specific CDS forms, such as vowel lengthening and exaggerated pitch, as a child matures (e.g. Huttenlocher et al. 2007; Kaye 1980; Phillips 1973). Similar conclusions are reached in the study of variation, where caregiver talk 'gradually becomes more similar in character to that of the interadult mode as the children get older' (Foulkes et al. 2005: 201), i.e. there are higher frequencies of the standard variant with caregivers in interaction with very young children, but these are replaced by increasing frequencies of the non-standard variant as the child gets older (Figure 2.9). Thus, for example, there may be lots of *yes* with younger children but more *aye* with older children.

In order to test this in our data, we investigate rates of use of the local form from (caregivers of) the youngest child to (caregivers of) the oldest in order to gain insights into the driving force behind CDS across the different stages of acquisition and its possible effects on the child's language. If use is conditioned by age, we might expect to see lower rates of the local forms with the (caregivers of) younger children, as in the stylised graph in Figure 2.10. If use is not governed by age, we would expect to see a flat line from youngest to oldest, as in the stylised graph in Figure 2.11. Alternatively, as per previous research (Smith et al. 2007, 2013), the caregiver/child pairs may fall into two main groups (Figure 2.12).

We can also use an age comparison to examine a key question raised in Chapter 1: do children acquire one variant first, and if so, which one – the standard or the local form? Or do they acquire both variants at once?
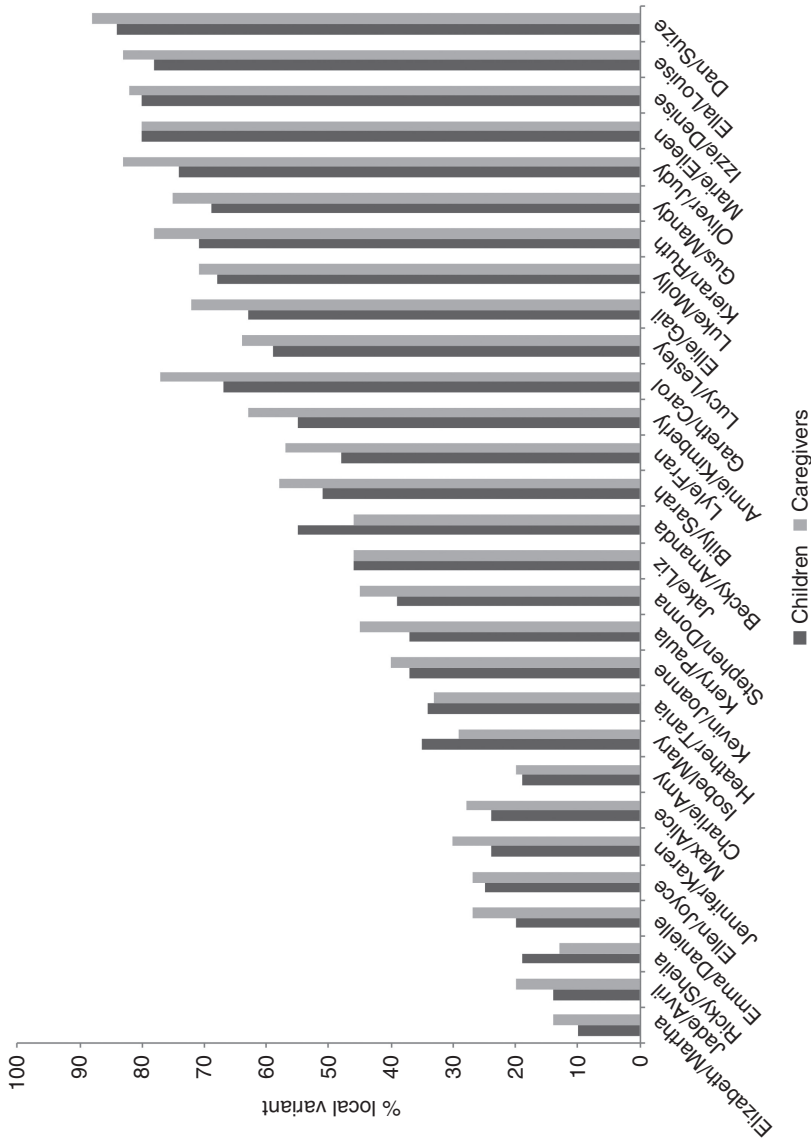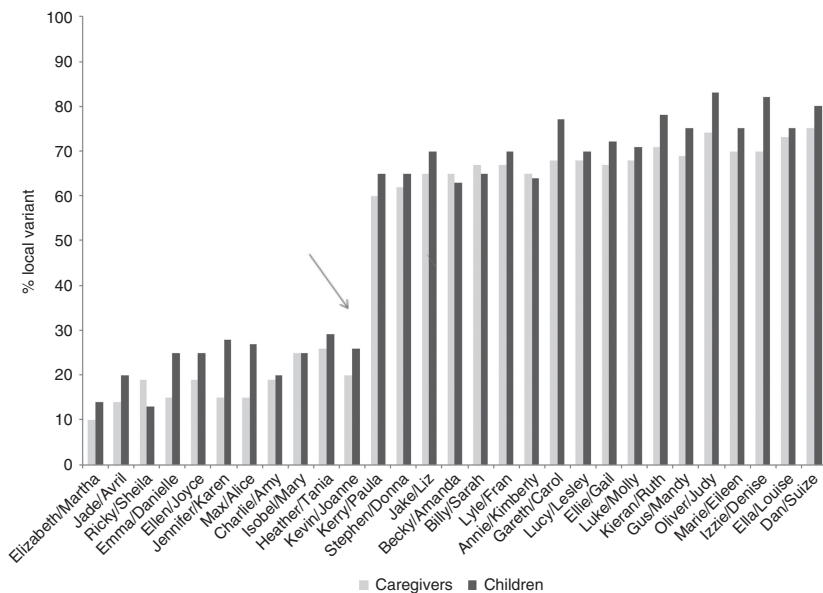
2.10 Stylised graph showing age differences
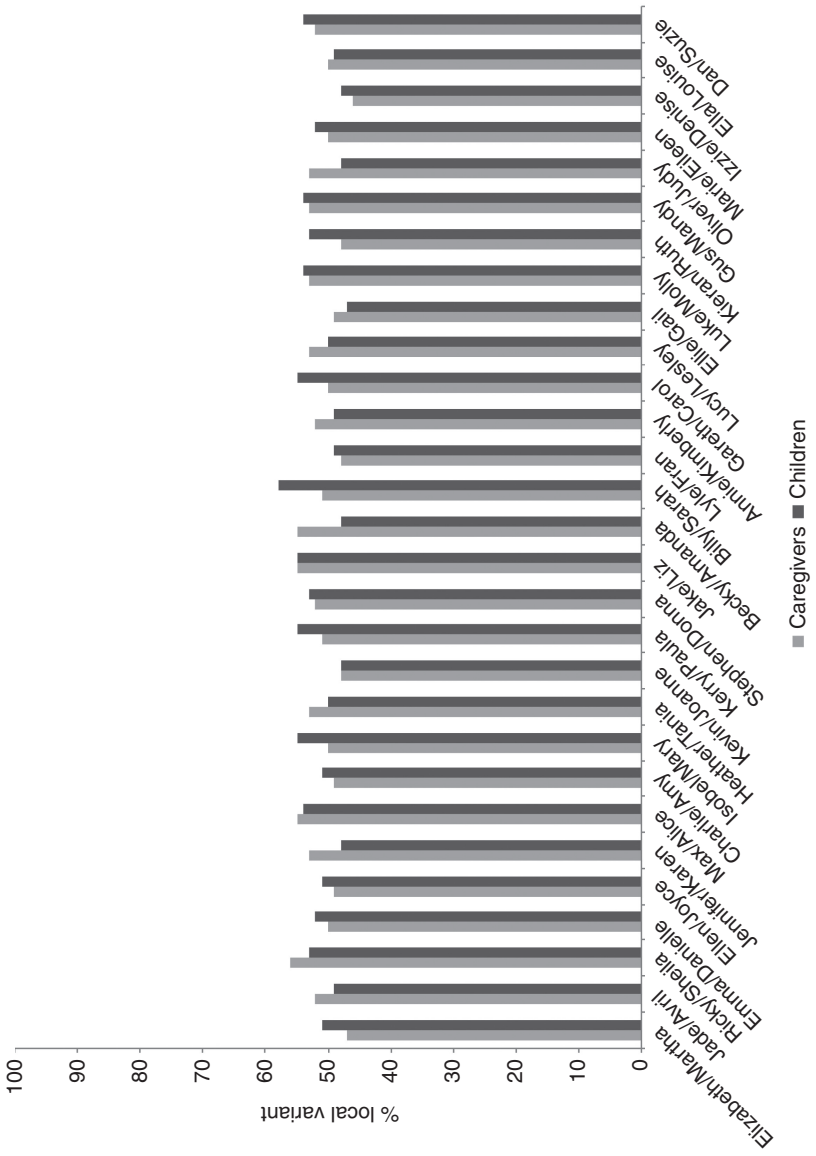
2.11 Stylised graph showing no age differences

2.12 Stylised graph showing split between (caregivers of) younger and older children

An optimal way to address that question is through longitudinal data, where a child's linguistic development is recorded over a set period of time (e.g. twelve months or two years). Our data in this study is not longitudinal but cross-sectional: the children are compared at a single point in time. However, because we have children of different ages in our sample, we can use the cross-sectional data as a proxy for maturational change. Thus, if the variable patterns as in Figure 2.12, we can hypothesise that one variant is acquired first and another added to that as the child grows older (as in Figure 2.13). If the variable patterns as in Figure 2.11, we can hypothesise that both variants are acquired at the same time (Figure 2.14).

Style-Shifting

As discussed in Chapter 1, acquiring a language demands much more than knowledge of its words, sounds and sentences. It also requires knowledge of how to talk to different addressees on different occasions, in different settings and on different topics. Although this is often characterised as a mapping of discrete voices for different contexts, the sociolinguistic literature demonstrates that the situation is much more fluid, with subtle but dynamic

2.13 Variants acquired sequentially

2.14 Variants acquired simultaneously

adjustments to a speaker's style across a range of situations. Despite this complexity, adolescents and adults systemically style-shift quite seamlessly, and the emergence of adult-like patterns of style-shifting may appear in the preschool years. In these data, we wanted to specifically test Labov's observation that 'linguistic variation is transmitted to children as stylistic differentiation on the formal/informal dimension' (2001: 437). To do this, we divided the recordings into different stylistic contexts that arise naturally from the caregiver's multifaceted roles in daily interaction with her child: *teaching* and *discipline* versus *routine, play* and *intimacy.*

   Rather than coding piecemeal for individual variables, where the wider context may be obscured, we coded every single utterance in the corpus for a particular context. This meant that the different contexts were not restricted to categorisation based on individual words, sentences or speaker turns, but instead covered the full context of use, which could include one or two interactions in a specific situation or multiple interactions across varying contextual fields. This is exemplified in the following interaction (56), where

Max (2;11) and his caregiver, Alice, move from a *teaching* context, to *discipline*, to *play*, to *routine* and back to *discipline* in a short space of time.

(56)   Max (2;11) and Alice

ALICE: We'll have to go up with Ryan's birthday present.
          What will you sing to him? What will you sing to
          Ryan today? Eh? How old's Ryan today, eh? Is he two?
          Is he four? What is he then? Eh? Is he seven?
MAX: No.
ALICE: Six?   *Teaching*
MAX: Mmm.
ALICE: I don't know what he is then. You tell mam.
MAX: Six.
ALICE: Six? He's not. He's going to be . . . What? Eight.
MAX: Eight.
ALICE: Eight.
          Get that out your mouth! Stop eating your glove!   *Discipline*
          Are you hungry, are you eating your glove?
MAX: [making noises]
ALICE: Are they nice and tasty? Hmm? You gonna eat   *Play*
          your gloves for your tea? Mmm. No?
OTHER: Are you going to walk Marcus, eh?
ALICE: Out of your pram and walk.   *Routine*
MAX: [Humming]
ALICE: Yeah.
ALICE: Don't take it out! Don't take it out! You might break it.
          Don't take it!   *Discipline*
OTHER: Put it in your pocket, okay?

Not only was this task extremely time-consuming, but how an utterance should be categorised was not at all times black and white. While some were very clear, others were not, as we will discuss.

Routine was fairly straightforward and included the rituals of everyday life such as eating breakfast (57), shopping (58) and getting dressed (59).

(57)   Ellie (3;5) and Gail

ELLIE: Mam, can I get some juice please?
GAIL: Are you wanting a carton or are you wanting blackcurrant?
ELLIE: Blackcurrant.

(58)   Max (2;11) and Alice

ALICE: Didn't you go and do the shopping today with grandma?
MAX: Yes.
ALICE: And what did you buy?
MAX: Nothing.
ALICE: What did you buy at the shops today?
MAX: Nothing.
ALICE: Did you go to the bakers?
MAX: Yes.

ALICE: What did you buy at the bakers?
MAX: Softies.
ALICE: Softies.
MAX: And a piece [biscuit] for me.
ALICE: A piece for you. What kind of piece did you choose?

(59)   Ellen (2;10) and Joyce
JOYCE: Right, on your trousers.
ELLEN: That's enough. That's enough, mam.
JOYCE: Enough what?
ELLEN: Enough clothes.
JOYCE: Enough clothes? You've got to get on your socks. Now, up your
       trousers.

It was much trickier to code for *teaching* and *play*, as these are often one and the same in caregiver/child interaction, teaching through play being a key component of the early years (e.g. O'Reilly and Bornstein 1993). For this reason, we created a strict protocol for how to code the data. Contexts were coded as *teaching* only in cases where the caregivers were *overtly* teaching the children, for example, asking them what colour something was (60) or about left and right (61):

(60)   Ellen (2;10) Joyce
JOYCE: What colour's that ball?
ELLEN: It's pink.
JOYCE: That's right. That one, fit colour's that one?
ELLEN: That's white.
JOYCE: *No*, what colour's that? It's p—.
ELLEN: Yellow.
JOYCE: *No*, p—.
ELLEN: Pit.
JOYCE: Pink.

(61)   Gus (3;9) and Mandy
MANDY: You sit down on the stool and I'll put them on 'cause you'll fall if
       I put your slippers on there. Okay, what foot's this? Can you
       remember?
GUS: Right.
MANDY: Right, that's correct. Well done. And what foot's this one?
GUS: Left.
MANDY: Left, well done.

Teaching segments often demonstrate some degree of help from the caregivers, such as 'sounding out' the start of the word, as in (62):

(62)   Gus (3;9) and Mandy
MANDY: Do you remember the name of the bird that was in our garden
       the day?

GUS: Yeah.
MANDY: What was the name of the little bird? What kind of bird was it?
GUS: Eagle.
MANDY: No, it was a cha—.
GUS: Chaffin bird.
MANDY: Cha—
GUS: A chaffin.
MANDY: A chaffinch.
GUS: A chaffinch.
MANDY: A chaffinch, that's right.

Segments where a book was being described either by the caregiver or by the child were also coded as *teaching*. Although in the former cases, it was mainly the child who was asking the questions, the manner of description made it clear that the caregiver was to some extent teaching the child new words (63):

(63)   Ellen (2;10) and Joyce
ELLEN: Once upon a time, what's this?
JOYCE: A witch.
ELLEN: A witch. What's this?
JOYCE: Where?
ELLEN: There.
JOYCE: It's a pumpkin.
ELLEN: What's that?
JOYCE: A broom.
ELLEN: What's that?
JOYCE: A cat.
ELLEN: What's that?
JOYCE: A dog.
ELLEN: What's that?
JOYCE: A astronaut.

Questions from the child that involved requests for information on issues of, for example, daily routine (64) were not counted as *teaching*, however.

(64)   Isobel (3;0) and Mary
ISOBEL: A train. A trainie. A Thomas train. Oh my Thomas! Where's my
          Thomas mammy?
MARY: He's round the side.

Contexts were categorised as *play* when they involved not only actual games (65) but also more general play, as in (66). They also included imaginary play (67):

(65)   Becky (3;2) and Amanda
AMANDA: No bones.
BECKY: No bones you've got. Ha ha ha ha ha.
AMANDA: That's nae nice. Right it's your go. Spin the wheel. I think you're
          cheating. Are you cheating?

BECKY: I didna ch— I didna play with you cheating!
AMANDA: No I ken you werena cheating, I was only joking.

(66)   Ricky (2;10) and Sheila
RICKY: You watch Spiderman to climb?
SHEILA: Okay, I'll watch Spiderman climb.
RICKY: Climb climb climb climb climb climb.
SHEILA: He can go high.
RICKY: Hang on. Higher, higher.
SHEILA: Is he gan to catch the baddies?
RICKY: The baddies hide up there.

(67)   Luke (3;5) and Molly
LUKE: Hey, mam. Here's one paw for you. Here's one paw for you. Here's
         one paw for you. Here you go. Here you go.
MOLLY: Thank you.
LUKE: I'm gan to put it on. I'm gan to put the paws on. You put the next
         paw on.
MOLLY: Come on then. Give us your hand.
LUKE: Okay. You put this paw on, right?
MOLLY: In there. Put your thumb in there. See? Okay.
LUKE: Do you need a paw?
MOLLY: Eh?
LUKE: Do you need a paw?
MOLLY: A paw?
LUKE: Aye.
MOLLY: Fit for?
LUKE: For my dog. Do you need a paw? Do you need a paw?
MOLLY: A paw.
LUKE: Aye.
MOLLY: Paw. Good doggy. Well done, clever dog.

Contexts were coded as *discipline* when the caregiver was scolding their child or admonishing them to behave better, as in (68) and (69). This involved a lot of the negative imperative forms 'don't' and 'dona'. The child's response was also considered to belong to the context of discipline.

(68)   Luke (3;5) and Molly
MOLLY: Just leave that. Just hold on. No dona, you'll break it. Don't do that.
LUKE: How?

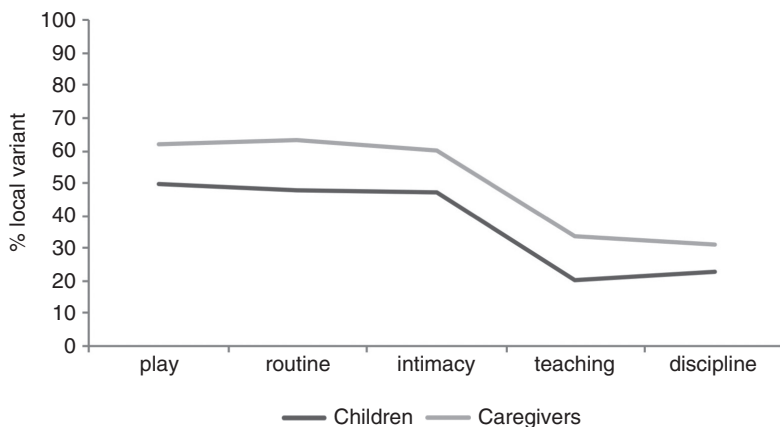(69)   Stephen (3;2) and Donna
DONNA: Dona stand on there, Stephen. Dona stand on that gate. Stephen,
         dona do that!
STEPHEN: Ow!
DONNA: Stephen, sit on your bum now. Stop lying on that thing. I know
         what you're trying to do. Sit up!

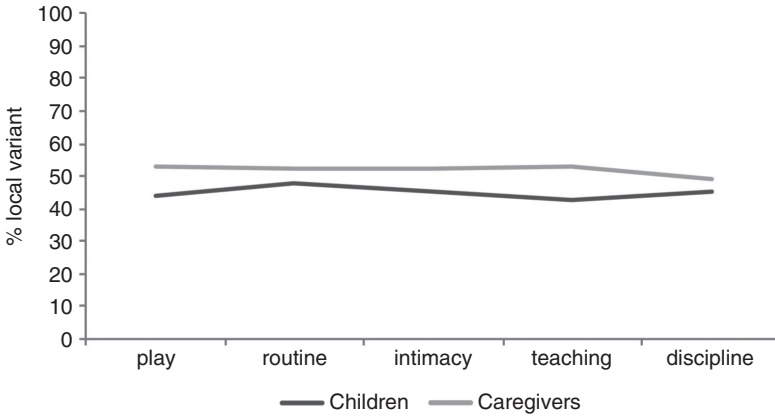*Intimacy* (70) made up the final context of use.

2.15 Stylised graph showing use of local variant in different situational contexts (style-shifting)

(70)   Jake (3;2) and Liz
          LIZ: I love you.
          LIZ: Ow you banged my head.
                  [kissing sounds]
          JAKE: I'm gonna kiss your head.
          LIZ: You gan to kiss it better, oh thank you.
          LIZ: Love you.
          JAKE: Another kiss.
          LIZ: Another ain?
          LIZ: Yuck.

In analysing these different situational contexts, if there is style-shifting we expect to see a pattern such as Figure 2.15, where there are higher rates of the local form in the informal contexts of *play*, *routine* and *intimacy* and lower rates in *teaching* and *discipline*. A pattern as in Figure 2.16, on the other hand, would suggest that there is little to differentiate the different contexts of use (i.e. no style-shifting according to situational context). In cases where we do not have enough tokens to divide the data in this way, we will collapse *play*/*routine*/*intimacy*[5] and *teaching*/*discipline* into two broad categories: formal and informal, as in Figure 2.17.

---

[5] As will be shown in Chapter 3, the overall low rates of tokens in intimacy contexts meant that they were often not included in the analysis.

2.16 Stylised graph showing use of local variant in different situational contexts (no style-shifting)
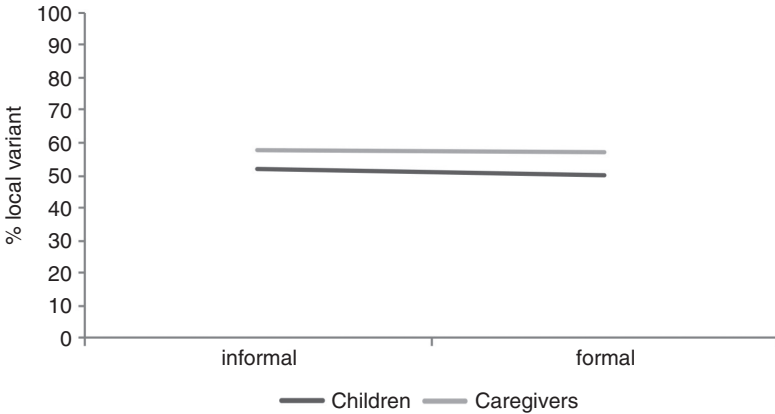
2.17 Stylised graph showing use of local variant in different situational contexts (formal vs informal)

Gender

Chapter 1 details findings on gender differences in caregiver speech with respect to variation, with, for example, higher rates of the non-standard form used with boys when compared to girls (Foulkes et al. 2005). Gender differences in child speech are not so clear cut, as also discussed in Chapter 1, one hypothesis being that gender differences do not arise until children move into the wider community (e.g. Roberts 1997). In these data, we test for potential
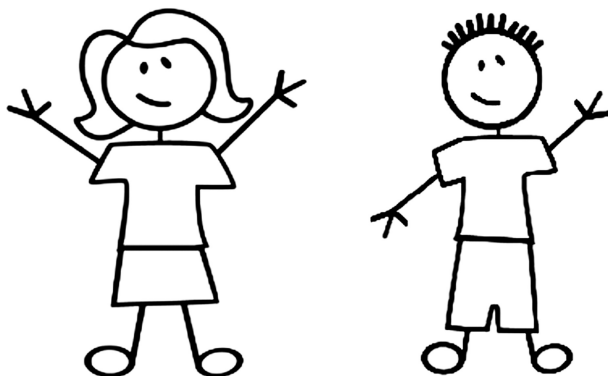
2.18 Comparison of girls and boys

gender effects in both caregiver and child speech across the range of variables that we examine (Figure 2.18).
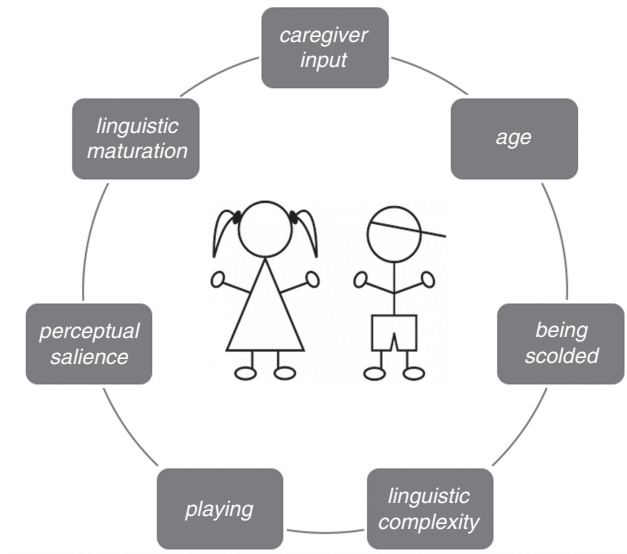
Linguistic Constraints

In addition to social constraints, a number of linguistic constraints will also be tested within each linguistic variable. This will include, where relevant, phonetic context, subject type and lexical collocation. Such analyses also relate to the 'complexity of the conditioning' (Kerswill 1996: 199) of a variable that might influence when linguistic constraints are acquired. To date, studies have largely concentrated on one variable in isolation, and thus systematic comparison across a range of variables has not been possible. This means that it has not been possible to test, for example, Chambers' (1992) principles of second dialect acquisition, according to which lexical replacements are acquired faster than phonetic or phonological rules and further, with phonological variables, simple rules are acquired earlier and faster than complex rules. In this study, we contribute to the question of how different types of variables are acquired by comparing multiple variables in order to establish which types are acquired at what stage of development (Figure 2.19).

The Interaction of Linguistic and Social Constraints

Roberts (1994: 30) points out that most studies of acquisition tend to 'focus on one type of constraint or the other'. However, variation is the result of multiple competing effects in the ambient language, as in Figure 2.20. By testing both linguistic and social constraints, we may be able to contribute to the debate on order of acquisition (e.g. Labov 1989). Simply put, do children learn social and
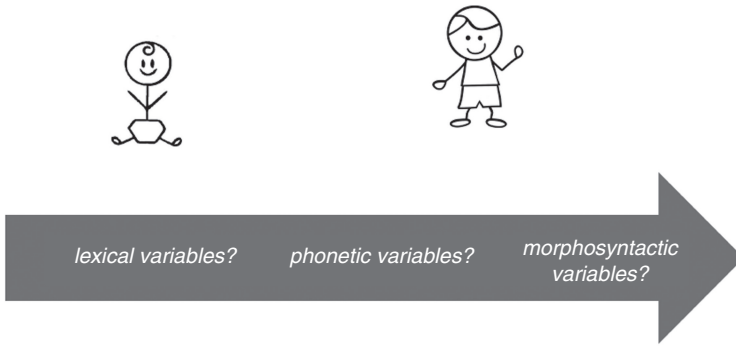
2.19 Order of acquisition of different types of variable

2.20 Potential influences on the acquisition of variation

stylistic constraints first, or grammatical ones or both at the same time (Figure 2.21)?

Developmental vs Sociolinguistic Variability

Before embarking on the analyses, we need to consider one more key issue in the investigation of the acquisition of variation: how to distinguish between

2.21 Order of acquisition of social and linguistic constraints

variation arising from developmental processes and variation that is part of community norms (e.g. Roberts 2005; Tagliamonte & Molfenter 2007; Miller 2015).

In (71), Luke is telling his mam, Molly, about a particularly exciting accident that he and his granny witnessed some time ago in:

(71)   Luke (3;5) and Molly
    LUKE: Em I seen— I seen helicopters on the road afore.
    MOLLY: Have you?
    LUKE: Aye. Me and granny goed and see helicopters flying afore and riding on the road and there was cars coming.
    MOLLY: Was there?
    LUKE: And the cars— and the racing cars bashed together.
    MOLLY: Oh no.
    LUKE: And the lorries took them away and so the man eh— the mannie took the— tooken— ticken eh racing cars in the garage so the mannies can fi— so the mannie can fix them.
    MOLLY: Uh-huh. Was the cars badly damaged?
    LUKE: They were all crashed.
    MOLLY: Oh dear.

In this extract, Luke uses an array of irregular verbs forms, including *seen* and *see* for standard *saw, goed* for *went* and three different forms for the preterite of take: *took, tooken, ticken*. Luke also uses yet another form in the mix: *taen* for preterit *took* (72):

(72)   Luke (3;5) and Molly
    LUKE: Mam, I havena nothing on my feet.
    MOLLY: Oh where's your trainers at?
    LUKE: Don't know, *taen* them off.
    MOLLY: Eh?
    LUKE: *Taen* them off.

MOLLY: Go get them then.
LUKE:　No, you.

An important part of the analysis is to distinguish which of these forms are bona fide sociolinguistic variables and which are the result of Luke's linguistic journey to adult-like attainment of forms. One way to distinguish these is to see what the adult speakers in the community are doing. In examining the data from the ambient language in the adult corpus, we find that *goed* for *went* is never heard in adult speech in this community. Instead, this is a very common overgeneralisation used by children in the early stages of language acquisition (e.g. Pinker 1994: 273). *Seen* for *saw*, on the other hand, is also a productive form in adult speech in the community, as demonstrated in the examples from the caregivers in our corpus in (73–75). In other words, it is a sociolinguistic variable.

(73)　Gus (3;9) and Mandy
　　　MANDY: *I seen* a couple of photos of you today.
　　　GUS:　　Did you?
　　　MANDY: At the playgroup.

(74)　Jake (3;2) and Liz
　　　LIZ:　How was Megan greeting this morning? *I seen* her greeting this morning when you went in. Fit was wrong with her?
　　　JAKE: I dona ken. No, she wasna greeting.

(75)　Billy (3;2) and Sarah
　　　SARAH: That's a robin.
　　　BILLY: Aye.
　　　SARAH: Mind *we seen* a robin in the garden?
　　　BILLY: Mmm.
　　　SARAH: And the robin's red, mind?

And what of the verb *take*? Luke's use of *tooken* and *ticken* are both non-target-like forms. In Buckie, you would not hear an adult use these. *Taen* on the other hand, is part of community grammar, and it can be used in both past participle (76–77) and preterite (78–79) contexts:

(76)　Annie (3;3) and Kimberly
　　　KIMBERLY: We're going to the big hall to see auntie K dressed up.
　　　ANNIE:　　And I wina cry.
　　　KIMBERLY: I hope no or you'll get *taen* home.

(77)　Lucy (3;4) and Lesley
　　　LESLEY: Lucy, you'd drive a saint mad. Is that what you're taking? That? Could you nae have just *taen* it the first time I asked you?

(78)　Annie (3;3) and Kimberly
　　　KIMBERLY: But I canna see fit it is. Fit is it?

OTHER: Tell Mum.
ANNIE: A king.
KIMBERLY: A king?
OTHER: She telt me it was a king.
KIMBERLY: Fit, afore she *taen* it out?

(79)   Dan (4;2) and Suzie
SUZIE: So, fa picked you up fae school?
DAN: Granda.
SUZIE: And then far did he take you?
DAN: Granny Pat.
SUZIE: She was— oh, she *taen* you?
DAN: Yeah. What did you said?
SUZIE: Granny *taen* you?
DAN: Yes.

In most cases, what is part of community grammar and what is developmental is quite straightforward: if we find it in the adult data, it is a sociolinguistic variable; if we do not find it, it is likely that it is part of developmental processes. However, some cases are less straightforward, particularly those cases where a form is both developmental and also part of community grammar. For example, a child is known to go through a stage where 3rd person singular *-s* on lexical verbs is sometimes present and sometimes not, as in (80–81) from Harris and Wexler (1996: 11).

(80)   a.   It only write in the pad.
        b.   My finger hurts.

(Eve, 2;0 file 14)

(81)   a.   Patsy need a screw.
        b.   This goes in here.

(Peter 3;3; file 8)

However, variation in 3rd person marking in adult speech, where *-s* can be absent, is amply attested, particularly in AAVE, as demonstrated in (82) from Cukor-Avila and Bailey (2015: 189).

(82)   She *bring* 'em in every mornin'.

A similar example comes from finite *be* deletion. Children are noted to go through a stage where both copula and auxiliary *be* are absent in utterances, as in (83) from Vainikka (1993: 272):

(83)   I ø popping balloons

(Nina, 2;0 file 5)

Such forms start to disappear from around twenty-eight months (Brown 1973). However, finite *be* deletion is also well-attested in adult data in some communities, as demonstrated in (84) from Labov (1969: 717).

(84)   a.   He ø fast in everything he do.
        b.   We ø on tape.

The possible entanglement between variation arising from stage of development and that of community norms is not restricted to morphosyntax. Phonetic forms, too, may present problems in analysis. For example, coronal stop deletion, or more commonly (-t,d) deletion, in word-final consonant clusters (85) is amply attested in the variationist literature for adult speech (e.g. Bayley 1994; Fasold 1972; Guy 1980; Tagliamonte & Temple 2005).

(85)   We was walking down Micklegate and we *grabbed* him and *grab*Ø this lad as
        well.

                                                            (Tagliamonte & Temple 2005: 281)

At the same time, consonant cluster deletion is described as 'the most common and longest lasting stage' (Shriberg & Kwiatkowski 1980: 138) in first language acquisition, where word-final consonant clusters with /t/ or /d/ may not be fully acquired until around 4;0 (e.g. McLeod et al. 2001b; Stoel-Gammon 1987). Thus, deciphering whether the child's deletion of /t/ or /d/ in coda clusters as in (84) is the result of developmental or community norms may be problematic (see also Smith et al. 2009).

A number of researchers have tackled this potential problem head on, and through careful analysis have been able to disentangle the competing influences – developmental vs community variation. For example, Green (2011: 39) examines finite *be* deletion in the speech of African-American children. The results led her to conclude that children start with no *be* in non-past contexts but slowly realise from the input that these can occur variably. In other words, this is an *insertion*, rather than a deletion process in the acquisition of this variable. She thus foregoes the description of *deletion*, as her data shows 'no evidence that the BE form was ever present' (Green 2011: 45). Miller (2013) demonstrates, through a longitudinal comparative study of Nina and Sarah from the Brown (1973) corpus, that non-agreeing *don't* in the speech of these children is the same in the initial stages due to developmental processes, but Sarah's continuing use in the later stages (86) is due to caregiver input, where her mother and father (87) also used non-agreeing *don't*.

(86)   a.   She *don't* get no one to play with.

                                                                      (Sarah 147; 4;10)

        b.   It *don't* have no wings!

                                                                      (Sarah 117; 4;07)

(87)   She *don't* have no pennies.

                                                                      (Sarah's father)

Beyond disentangling developmental vs community-based variation, such analyses may have implications for first language acquisition research more

broadly (e.g. Miller 2013). While such analyses are beyond the scope of the current study, it is important that we note potential overlap between developmental and systematic adult-like variation in examining the data at hand.

We have now provided a qualitative, descriptive account of the data for analysis. Before we turn to the variationist analyses of individual variables, we first provide an overall view of the corpus as a whole.