
Narrative Structures in Colloquial English

JAMIE BAILEY

MPhil Theoretical and Applied Linguistics Dissertation
University of Cambridge

Contents

Abstract	iii
Statement regarding access to resources	iv
Declaration of originality	iv
Statement of length	iv
Acknowledgements	v
List of abbreviations	vii

NARRATIVE STRUCTURES IN COLLOQUIAL ENGLISH

1	Introduction	1
2	Outline of the narrative structures	2
2.1	Structural properties	2
2.2	Discourse properties	4
3	Dialect survey	4
3.1	Methodology	4
3.1.1	Overview of responses collected.	4
3.1.2	Stimulus selection and presentation	5
3.1.3	Demographic information	6
3.1.4	Geographic information	9
3.1.5	Survey distribution	9
3.1.6	Methods of analysis	10
3.2	Data analysis	11
3.2.1	Interstructural correlations	11
3.2.2	Social distribution	16
3.2.3	Geospatial distribution	20
3.2.4	Effect of syntactic factors on acceptability	22
3.2.5	Effect of exposure on acceptability	23
3.2.6	Effect of temporal–aspectual complexity on acceptability.	25
3.2.7	Summary of survey findings	26
4	Analysis of SCRs and PaPRRs	26
4.1	Analysis of SCRs	27
4.1.1	A survey of previous analyses	27
4.1.2	The size of SCR clauses.	30
4.1.3	Non-presuppositional subjects and anti-locality	31
4.2	Analysis of PaPRRs	36
4.2.1	Previous analyses	36
4.2.2	Extraction from PaPRRs with non-presuppositional subjects	36
5	Analysis of EBICs	40
6	The Suppressed Grammaticality Hypothesis	44

7	Conclusion 46
	References 48
	Appendix 52

Abstract

This thesis discusses the distributional and syntactic properties of three presentational constructions present in colloquial registers of some dialects of English. I hypothesise, following Barbiers (2005), that the structures are available to a large group of speakers, but that the reported acceptance of the structures is dependent on the exposure speakers receive to them as part of their linguistic input, which I term the Suppressed Grammaticality Hypothesis (SGH). I make use of an acceptability judgement survey to establish the nature of interspeaker variation in the UK and Ireland, and I argue that the results of the survey support the SGH. In addition, I draw a parallel between the restricted syntactic distribution of the presentational structures and those of apparent Complex NP Constraint violations in Scandinavian, Hebrew and other languages, building on work by Sichel (2018) to attribute this common pattern of variation to the same syntactic principles, namely Diesing's (1992) proposal that only presuppositional DPs undergo movement to a derived syntactic position, whereas non-presuppositional DPs remain in a lower position. I further argue that one of the three structures fills the paradigmatic gap in English expletive–associate constructions, such that dialect speakers have the ability to form non-progressive monoclausal presentational structures, as well as the progressive structures to which speakers of the majority dialect are limited.

Statement regarding access to resources

The COVID-19 pandemic did not significantly impact on my ability to access resources or collect primary data during the completion of this project, and the content of this dissertation is not materially different as a result.

Declaration of originality

I declare that this dissertation is the result of my own work and includes nothing which is the output of collaboration, except where specifically indicated in the text.

Statement of length

I declare that this dissertation does not exceed the word limit (20,000 words, excluding translations and glosses) that has been set for the MPhil degree in Theoretical and Applied Linguistics by the Faculty of Modern and Medieval Languages and Linguistics.

Acknowledgements

I'm not quite sure how I imagined the circumstances in which I would be writing up and handing in my MPhil thesis when I started my fourth year at Cambridge eight months ago, but it had certainly not occurred to me that I might be doing it in the midst of the worst global pandemic in more than a century. For many reasons other than the pandemic or my degree, though, this year has been one of the hardest I've had—and one that I've learnt the most from—in lots of ways. Thankfully, I've had lots of people on hand to help me through it.

First and foremost, this project would not have been possible without the 1,030 kind people that took ten minutes out of their lives to listen a random stranger, who in most cases they've never heard of and will probably never hear about again, say fifty funny sentences and then tell me what they think about them. The same goes for the many people who shared the survey online, spreading it far and wide. Out of the respondents, a fair few sent me emails or messages to convey feedback, express enthusiasm or just ask questions out of interest, which helped me improve the survey in the middle of the data collection process. The fact that so many people put a little bit of time aside to help me was truly heartwarming, and probably the most rewarding part of the whole project.

My huge thanks go to David Willis, who kindly agreed to supervise this dissertation, for his patience, insight, reassurance, and guidance on how to make sure my work was conducted with maximum rigour. I am also greatly indebted to my undergraduate Director of Studies and supervisor, Theresa Biberauer, for her boundless enthusiasm, unwavering support and the belief she has in her students, through which she has helped me grow immeasurably throughout my four years at Cambridge. My thanks also go to the SyntaxLab crew, including Ian Roberts and Michelle Sheehan, for valuable input and feedback, as well as many an enjoyable pint of ale in the Pickerel Inn over the years. Tam Blaxter has been instrumental in helping me understand and use the geospatial analysis techniques used in this thesis, for which I am very grateful to her. Jamie Douglas was also very kind to provide me with an opportunity to talk through some of the ideas raised in his PhD and discuss their potential application to my work, so I am very thankful to him for his time and ideas. Thanks to Justin Malčić, for the unfathomable amount of linguistic (and other) discussion we've had over the years and for putting up with all my questions—and answering them too! And thanks to all the other staff and students, of which there are too many to mention, who have made my experience in (D)TAL such an enjoyable and stimulating experience.

I must also thank various kind linguists and non-linguists who provided me with the (near-)native speaker judgements for Dutch (Kim Groothuis), German (Emma Pruin), Hebrew (Hagit Borer, Eitan Grossman, Yonathan Nativ, Omer Preminger, Ur Shlonsky, Ivy Sichel, Ghil'ad Zuckermann), Swedish (Ulrika Andersson, Eva Henderson, Sophie Michalski), Welsh (Owen McArdle), and Yorkshire (Chloe Allenby and family).

The Elite Athletes of Cambridge University Lightweight Rowing Club and Cambridge University Women's Boat Club have featured heavily in my life this year, and a special mention must go to all the lads (and Eytan) of CULRC, under the sublime leadership of Teague Smith, for being an absolute pleasure to work with throughout the year until March, as well as fellow Maggie woman Jo Matthews for her support and friendship through many a breakdown, both personal and vehicular, throughout the whole process. The experience taught me so much more than how to be a better cox, so I thank coaches Katy Knowles, Rob Weber, Paddy Ryan and Matthew Temple for all their input into me, and teaching me how to *choose* to respond to adversity. In this vein, a special mention must also go to Lance Badman for his refreshing and honest humour which has kept me connected to the world outside the ten-foot brick walls surrounding me throughout my time in this city. More recently, the Granta Fan Club and friends have been instrumental in keeping me sane throughout the lockdown by

allowing me to continue bossing people to do exercise while myself sitting completely still and drinking Yorkshire Tea.

I am truly blessed to have such wonderful friends. My co-residents in the humble dwelling of 9RT, Will Crisp, Mercedes MacDonald and John Perry, have provided endless hours of enlightened discussion, moral support, laughter and food throughout the year, and my thanks go to them for their patience and understanding of my early nights and early mornings throughout the all too short five months we spent together. Thank you to Anna Odorici for being the best college sister I could ask for through four years, showing me the true meaning of 'cuisine' by supplying me with authentic Italian parmesan, salami and wine, and always paying me back; to the Rachael Speed for keeping me grounded and connected to what's important, her endless provision of carrots and always being on hand to boost morale with her infectious grin which never fails to lift my mood; to Rebecca Marrow for her Fun Facts™, her personal life management consultancy services and her enduring friendship and support; to Ben Farnworth for two years of the best company full of stories and laughter, not to mention his unrivalled pie-making skills; and to Sophie Michalski for being my rock through thick and (very) thin, and for minding the rudder. (But most importantly thanks to Selwyn College Buttery for their wonderful chips.)

Lastly, the biggest thanks must go to my sister and to my parents, for their unending support and love, especially given all the challenges I have presented them with through the pandemic. If nothing else, this year has taught me to take nothing for granted, and nothing more so than family.

List of abbreviations

ACC	Accusative case
BIC	Bare infinitival construction
CIP	Clause-internal phase
CNPC	Complex NP Constraint
EBIC	Existential bare infinitival clause/construction
InfPRR	Inf(initival)P reduced relative
NOM	Nominative case
NS	Narrative structure
NRS	Narrative relative structure (i.e. SCR or PaPRR (or InfPRR))
OCR	Object contact relative
PaPRR	Past participle reduced relative
PrPRR	Progressive participle reduced relative
RR	Reduced relative
SCR	Subject contact relative
SES	Socioeconomic status
SGH	Suppressed Grammaticality Hypothesis
SPRR	Stative participle reduced relative

NARRATIVE STRUCTURES IN COLLOQUIAL ENGLISH

1 Introduction

This dissertation examines three structures which occur in colloquial registers of some dialects of English. These are (i) Subject Contact Relatives (SCRs), (ii) Past Participle Reduced Relatives (PaPRRs; previously referred to as Narrative Relatives (Bailey 2019)), and (iii) Existential Bare Infinitival Clauses (EBICs; previously Narrative Infinitival Clauses (Bailey 2019)). The three are exemplified in (1).

- (1) a. There's a dog has gone past the window. Subject Contact Relative (SCR)
b. There's a dog gone past the window. Past Participle Reduced Relative (PaPRR)
c. There's been a dog go past the window. Existential Bare Infinitival Clause (EBIC)

Of these, SCRs have received the most attention in previous literature (Jespersen 1909–40; Erdmann 1980; Harris & Vincent 1980; Nagucka 1980; Prince 1981; van der Auwera 1984; Quirk et al. 1985; Lambrecht 1988; Huddleston & Pullum 2002 and others in the descriptive literature; Doherty 1993, 1994, 2000; Henry 1995; den Dikken 2005; Montgomery 2006; Haegeman et al. 2015; Haegeman 2015; Williamson 2016 in the generative literature), while PaPRRs only received their first detailed treatment in Bailey (2019), and EBICs have, to my knowledge, received no mention at all. Although these three structures exhibit significant structural differences on the surface, I will argue that there is reason to consider them as a group to which at least a partially unified syntactic analysis can be applied. As such, I will refer to them collectively as Narrative Structures (NSs).

There are two main properties exhibited by NSs. The first is their common presentational discourse role, which makes the examples in (1) analogous to the monoclausal presentational structures found in other Germanic languages, given in (2).

- (2) a. Es ist ein Hund am Fenster vorbei-gegangen. German
EXPL is a dog on.the window past-gone
b. Er is een hond langs het raam gelopen. Dutch
EXPL is a dog past the window walked
c. Det har gått en hund förbi utanför fönstr-et. Swedish
EXPL has gone a dog past outside window-the

A dog has gone past the window.

The second property shared by the NSs is their restricted syntactic distribution. In those dialects which feature NSs, they may only occur in a limited set of contexts, namely when the subject of the relative clause (RC) is an argument of an existential or intensional verb in the matrix clause, although Doherty (1993, 2000) reports that SCRs at least are licensed in a broader set of contexts in certain more liberal varieties.

Given these observations, a syntactic analysis should aim to account for the dialectal variation shown by the structures and the common syntactic properties they share, including both the presentational discourse function and the syntactically restricted distribution, the details of which are briefly fleshed out in §2. In order to achieve these aims, I make use of a grammaticality judgement

survey to establish the nature of interspeaker variation among the population of the Anglo-Celtic Isles, the results of which I present and discuss in §3. In §§4–5, I provide syntactic analyses for the three NSs: §4.1 deals with the analysis of SCRs, §4.2 seeks to extend this analysis to PaPRRs, and §5 contains the analysis of EBICs. In §6, I place the factors determining grammatical variation within the broader context of the nature of speakers’ grammars, building on work by Barbiers (2005) to formulate the Suppressed Grammaticality Hypothesis, which governs the extent to which speakers accept stimuli presented to them. §7 concludes.

2 Outline of the narrative structures

2.1 Structural properties

In this section I provide a descriptive outline of each of the three NSs. All three structures commonly begin with a semantically impoverished matrix clause, which for descriptive purposes I will refer to as the *introducer*. The most frequently attested introducer is expletive *there*’s; temporal–aspectual variations thereon are also licit. At least five forms of introducer are standardly recognised within the literature on SCRs (e.g. Doherty 1993, 1994, 2000; Henry 1995), exemplified below.

- | | | | |
|-----|----|--|-------------------------------------|
| (3) | a. | There was a dog went past the window. | copular existential |
| | b. | i. We had a dog went past the window. | <i>have</i> -possessive existential |
| | | ii. You get a dog goes past the window. ¹ | <i>get</i> -possessive existential |
| | c. | It was a dog went past the window. | <i>it</i> -cleft |
| | d. | They think that was a dog went past the window. | copular sentence |
| | c. | I knew a dog went past the window. | <i>know</i> -complement |

Although these five contexts seem to constitute the core of structures that license NSs, with examples formed in other syntactic contexts generally deemed ungrammatical, there is some degree of variation in this regard. For instance, Doherty (1993, 2000) further identifies that some speakers of liberal dialects accept SCRs with introducers containing a universal quantifier, ‘free-choice’ *any*, intensional contexts and predicate nominals (see Doherty 2000:87–9 for examples). Furthermore, it will be observed in §5 that EBICs with possessive existential introducers appear to be acceptable in the majority dialect (experiential *have* constructions), and that *know*-complements receive a radically different interpretation in EBICs to that that they receive in SCRs and PaPRRs.

While the introducer appears to be what is shared between the NSs, the subordinate clause is where the three diverge. Subject Contact Relatives resemble the Object Contact Relatives (OCRs) available in the majority dialect in that they appear to be a full restrictive relative clause that lacks an overt relative pronoun or complementiser. There are no temporal or aspectual restrictions on the verbal complex contained within the subordinate clause of an SCR. Compare the dialectally and syntactically restricted SCRs to the universally available OCRs in the following examples.

- | | | | |
|-----|----|---|------------------------------------|
| (4) | a. | The dog who/that Mary likes is called Fido. | Object Relative |
| | b. | The dog who/that likes Mary is called Fido. | Subject Relative |
| | c. | There’s a dog who/that likes Mary. | <i>there</i> ’s + Subject Relative |

¹ The grammaticalisation of *have* + *got* in English becomes problematic when rendering such an example as (3b(ii)) in the past tense.

- | | | | |
|-----|----|--|---|
| (5) | a. | The dog Ø Mary likes is called Fido. | Object Contact Relative |
| | b. | * The dog Ø likes Mary is called Fido. | Subject Contact Relative |
| | c. | There's a dog likes Mary. | <i>there's</i> + Subject Contact Relative |

Past Participle Reduced Relatives are somewhat resemblant of SCRs in that they contain a restrictive relative clause lacking an overt relative pronoun or complementiser with the same syntactic restriction on introducer elements. In addition to this, however, they also lack overt auxiliary verbs which appear higher than the past participle, i.e. finite auxiliaries, modal auxiliaries and perfect *have*. In this respect, they resemble majority dialect progressive participle reduced relatives (PrPRRs) and even more closely the marginally acceptable stative participle reduced relatives (SPRRs; i.a. Douglas 2016:231–3), which are both exemplified in (6). However, these two structures are not subject to the same syntactic restrictions as PaPRRs. The somewhat obvious point should also be made that PaPRRs obligatorily contain an active past participle, and therefore perfective aspect.

- | | | | |
|-----|----|--|-------------------------------------|
| (6) | a. | The dog going past the window is called Fido. | Progressive participle RR |
| | b. | The man just arrived at the office is called Fred. | Stative participle RR |
| | c. | * The girl just eaten the apple is called Lucy. | Past participle RR |
| | d. | There's a girl just eaten an apple. | <i>there's</i> + Past participle RR |

The third of the three NSs differs from the other two in that it does not resemble a relative clause. Instead, it is more akin to other Bare Infinitive (BI) constructions found in the majority dialect, namely causative, perceptive and experiential structures, exemplified in (7). The fact that the 'subordinate' part of the structure may only contain a bare infinitive means that tense and aspect must be marked in the introducer, with a consequence of this being that EBICs cannot contain progressive aspect. Unlike its majority dialect lookalikes, the grammaticality of examples with simple past or simple present tense is uncertain, as shown in (8). This point will be picked up on in §5.

- | | | | |
|-----|----|--|-------------------------------|
| (7) | a. | We've made a dog go past the window. | Causative BIC |
| | b. | We've seen a dog go past the window. | Perceptive BIC |
| | c. | We've had a dog go past the window. | Experiential BIC ² |
| | d. | There's been a dog go past the window. | Existential BIC |
-
- | | | | |
|-----|----|---|--------------------------|
| (8) | a. | * There's being a dog go past the window. | Illicit progressive EBIC |
| | b. | ? There was a dog go past the window. | Illicit simple past EBIC |

In addition to these three core structures, a fourth structure appears to exist which resembles the PaPRR but has non-finite perfective *have* as its highest overt auxiliary. Following the approach to the clausal spine proposed by Harwood (2013, 2015, 2017) and others, according to which the *have* auxiliary would appear in Inf, I will term this the InfP reduced relative (InfPRR). This structure is exemplified in (9).

- (9) There'll have been a dog have gone past the window.

The circumstances in which this structure can appear are limited and rare, as both a modal verb and perfective aspect must be present, necessitating at least modality if not also perfectivity in the

² An alternative reading of this is as a causative BIC; see Donnelly (2016) for a comparison of the two.

introducer. Furthermore, this combination of modality and perfectivity may also be conveyed by a (shorter) PaPRR. These two facts combine to produce the result that such an InfPRR may almost never be uttered. Nevertheless, both I and an additional linguistically trained speaker of a similar dialect to mine judge this example to be grammatical. Although I have included an InfPRR in the dialect survey conducted, I will leave discussion of this structure aside in this dissertation given its relative rarity. Whether it bears more similarity to SCRs or PaPRRs, or indeed forms a separate type on its own, remains to be determined (though see Bailey 2019:23–4 for a little more speculative discussion).

2.2 Discourse properties

As mentioned above, NSs function as presentational structures, perhaps analogous to their counterparts found in various other Germanic languages. While speakers who are unfamiliar with such constructions may more readily compare them to their full relative structural counterparts ((10b)) in the majority dialect, their interpretation is arguably more similar to a standard declarative sentence ((10c)). The situation may be compared to the interpretation of expletive–associate constructions, as exemplified in (11).

- (10) a. There was a dog went past the window.
 b. There was a dog that went past the window.
 c. A dog went past the window.
- (11) a. There was a dog going past the window.
 b. There was a dog that was going past the window.
 c. A dog was going past the window.

Thus instead of setting up an object in the discourse and assigning an attribute to it, as happens in the (b) examples above, the presentational NSs refer simultaneously to the event as a whole.

In order to establish the relationship between each of these structures given their common discourse function, I turn in the next section to the results of a quantitative dialect survey, which will allow me to shed light on the correspondence between each type of NS, and hence glean insights into their structure.

3 Dialect survey

In this section I present the results of an online dialect survey conducted over a four-month period between January and May of 2020. In §3.1 I outline some details of the methodology and related issues, before moving onto the analysis in §3.2.

3.1 Methodology

3.1.1 Overview of responses collected

The online survey conducted took the form of an acceptability judgement questionnaire. After having answered some questions designed to determine their social and geographical characteristics, participants were asked to categorise 50 aural stimuli given the following options, thus forming a four-point Likert scale: (i) “This sounds completely normal”; (ii) “This sounds OK but I don’t think I would say it”; (iii) “This sounds slightly strange and I definitely wouldn’t say it”; (iv) “This sounds

completely wrong”. Complete responses were submitted by 1030 participants in total, of which 1003 declared themselves to be native speakers. Due to error, the 27 non-native speakers (comprising 2.6% of responses) were not removed from the data before statistical analysis. For the purposes of the geospatial analysis, 821 respondents identified themselves as having grown up at a specific location in the UK or the Crown Dependencies of Guernsey and the Isle of Man,³ which share a common postcode system, and 84 respondents provided such locations within the Republic of Ireland. Therefore the geospatial analysis was performed on 905 responses.

3.1.2 *Stimulus selection and presentation*

A total of 50 aural stimuli were presented to each participant. Each stimulus was a temporal and/or aspectual variant of one of four active and two passive events. In the analysis, each event is distinguished by an alphabetic prefix in the stimulus label. Events were chosen such that there was an even distribution of transitive, unaccusative and passive verbs and human and non-human subjects. Care was taken to ensure that each event only contained non-dialectal words that are frequently used in everyday spoken English so as to minimise the effect of word frequency on acceptability judgements. The events are listed below as expletive–associate constructions.

- (12)
- a. A: There’s a woman going up the hill.
 - b. B: There’s a girl eating an apple.
 - c. C: There’s a cat finding a mouse.
 - d. D: There’s a bird falling out of the tree.
 - e. E: There’s a man being arrested.
 - f. F: There’s a dog being trained.

A list of 90 possible structures to be tested was drawn up, from which 50 were selected for inclusion in the survey, in order to reduce the time taken for each respondent to complete the survey and hence increase the number of complete responses. The selection allowed for the eventual testing of the effect of structural or temporal–aspectual complexity on acceptability. Each active event appeared in 10 stimuli (2 SCRs, 2 PaPRRs, 1 InfPRR, 2 EBICs, 1 majority dialect PrPRR, 1 universally ungrammatical example and 1 marginal SPRR), and each passive event in 5 (3 PaPRRs, 1 PrPRR and 1 SPRR), each of which had a different structure and/or temporal–aspectual specification. The manifestations of events A (active) and E (passive) are given below.

- (13)
- a. A01: There’s a woman gone up the hill.
 - b. A02: There’s been a woman go up the hill.
 - c. A04: There was a woman went up the hill.
 - d. A06: There’ll be a woman go up the hill.
 - e. A07: There’s a woman been going up the hill.
 - f. A10: There’s a woman would have gone up the hill.
 - g. A11: There’ll have been a woman have gone up the hill.
 - h. A12: There’s a woman going up the hill.
 - i. A13: There’s being a woman go up the hill.
 - j. A14: The woman gone up the hill is called Wendy.

³ No responses were received from Jersey.

- (14) a. E01: There's a man been arrested.
 b. E07: There's a man been being arrested.
 c. E09: There'll have been a man been being arrested.
 d. E12: There's a man being arrested.
 e. E14: The man been arrested is called Simon.

A full list of all 50 stimuli, together with their acceptance scores, is given in the Appendix.

Given the generally spoken nature of NSs, it was decided that aural stimuli would more accurately represent an authentic context in which such structures would be produced than written stimuli. All stimuli were recorded using my voice. I received a number of emails from participants after they had completed the survey, as well as comments on social media posts, some of which commented on the potential influence of the accent with which the stimuli were spoken on the acceptability judgements they gave. My accent has features characteristic of the Hambleton district of North Yorkshire; during the recording I did not aim to replicate any form of Standard English accent, but I also tried to reduce the strength of the characteristics of my local accent.

A particular unforeseen consequence of the aural nature of the stimuli was that a significant difference was observed between RRs headed by past participles (i.e. PaPRRs and SPRRs) with respect to whether the participle bore *-en* or *-ed* morphology, and hence whether or not the past participle was acoustically similar to the present participle with *-ing*. A similar point was frequently brought up in the feedback of respondents who were at times unable to distinguish between the participles *been* and *being*. Specifically, stimuli B01, B14, D01 and D14 with *-en* participles consistently received higher acceptability scores than their counterparts in events A and C with *-ed* participles, which is presumably attributable to respondents interpreting them as *-ing* participles and therefore erroneously identifying the stimulus as a PrPRR, which is fully grammatical to speakers of the majority dialect. Fortunately, the *-en/-ing* error affected transitive and unaccusative verbs and human and non-human subjects equally, so statistical analysis of these categories was not severely hampered. However, to reduce the effect of this skew on the results, the stimuli mentioned above were excluded from the analysis at several points, as well as their counterparts. Where this has been necessary, an asterisk indicates the absence of stimuli from events B and D in the analysis (e.g. PaPRRs*).

Stimuli were presented to respondents in one of two set orders, which I will henceforth refer to as Set A and Set B. This was done so as to test for the effect of prior exposure in the form of other NSs on the acceptability given to a stimulus. The order of stimuli was randomised within each set. It is possible that a random order in one set and its reverse order in the other would have provided opportunity for more structured analysis in this respect. Each set was presented to an equal number of participants who started the survey, although among those who submitted a complete response, 509 respondents were given Set A while 521 respondents were given Set B.

3.1.3 Demographic information

Respondents were asked to provide demographic information which would allow sociolinguistic analysis to be conducted. These were year of birth, gender, the level of education of both the respondent and their mother (the latter being intended to act as a proxy for socioeconomic status), occupation and self-identified socioeconomic status, as well as whether or not the respondent was

linguistically trained.⁴ Socioeconomic status could be described as one of four options: lower/working class, lower middle class, upper middle class and upper class.

A significant skew towards younger respondents was present in the results, perhaps due to the distribution methods employed in order to maximise the survey’s reach (see §3.1.5). Figure 1 shows a histogram of the age of respondents.

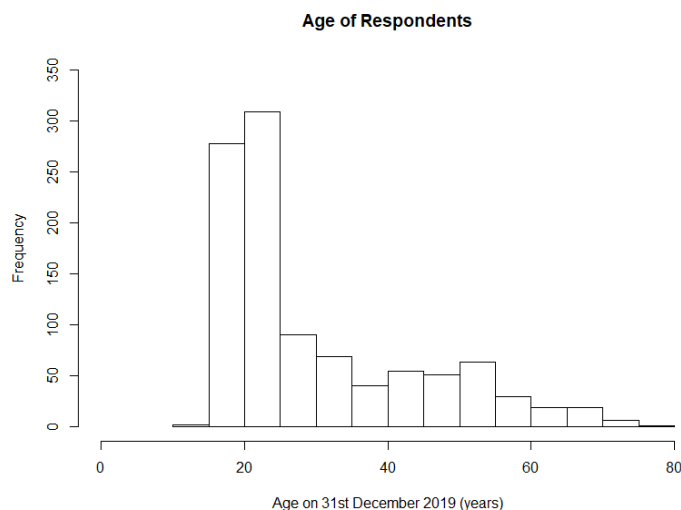


Figure 1. Age of respondents.

More females submitted complete responses than males. Of the 1030 responses received, 596 (57.9%) were from self-identifying females, 405 (39.3%) from self-identifying males, and 29 (2.8%) from people who did not identify as either male or female. Figure 2 displays this information.

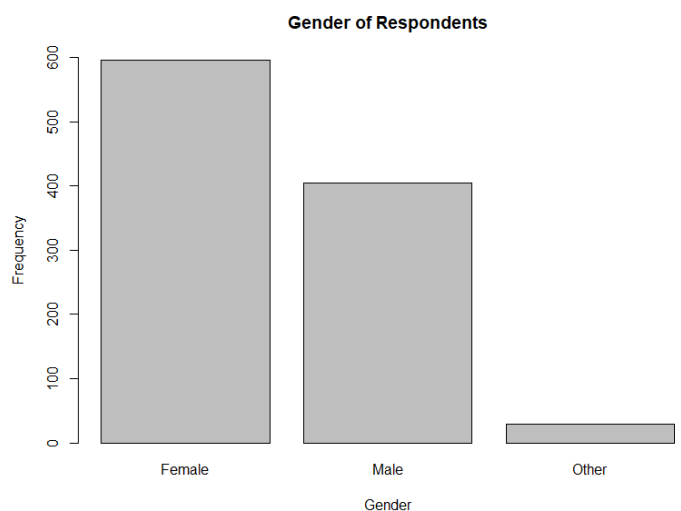


Figure 2. Gender of respondents.

The self-identified socioeconomic status of respondents is shown in Figure 3. Some feedback from respondents highlighted the simplicity of the set of categories chosen with which socioeconomic status could be classified, and indeed the distinction between lower, middle and upper classes is

⁴ A question asking for respondents’ level of linguistic training was introduced late in the data collection process in response to feedback received from a respondent, which resulted in only 170 responses including this data.

widely regarded in society as outdated (see Block 2015 for an overview within the context of applied linguistics). However, given that the analysis of the effect of socioeconomic status on the acceptance of the stimuli was not a priority of the project, and that many people still readily identify with at least one of these categories regardless of their opinion on the categories' validity, this system was deemed adequate for the present purposes.

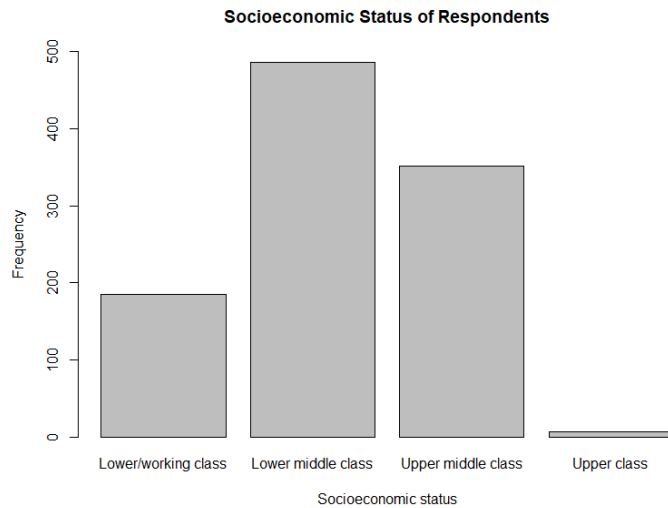


Figure 3. Socioeconomic status as self-identified by respondents.

Respondents' level of education is displayed in Figure 4. Respondents were presented with the options 'Below GCSEs', 'GCSEs or equivalent', 'A-levels or equivalent', 'Undergraduate degree', 'Postgraduate degree' and 'Higher than postgraduate degree'.

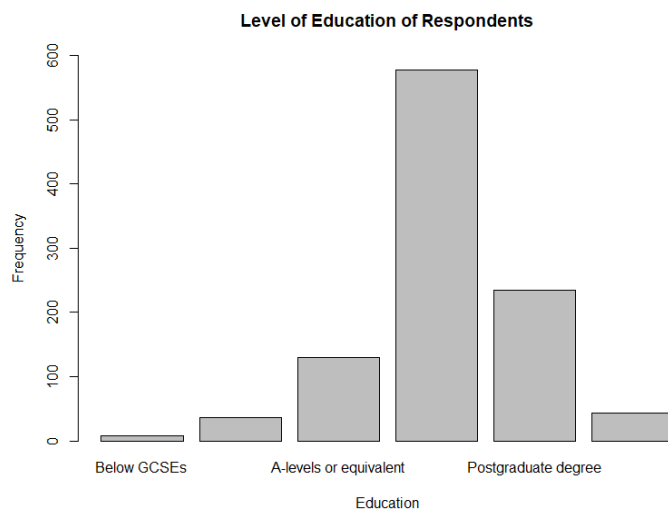


Figure 4. Level of education of respondents.

Although maternal education and occupation were also probed, I leave descriptive analysis of these two factors aside here, for reasons to do with the peripheral nature of interest in maternal education and the complexity of data processing for occupation.

3.1.4 *Geographic information*

In order to facilitate geospatial analysis respondents from the UK, Ireland, the Isle of Man, Jersey and Guernsey were asked to provide the location of their upbringing during the first seven years of their life. The relationship between the age of a child and the stage of syntactic acquisition is still largely unclear and fiercely debated, so the age of seven was selected under the tentative assumption that most syntactic acquisition occurs before this age. Respondents to the survey reported in Bailey (2019) were asked to specify the nearest town, the county and the country in which they spent this period of their life. However, they varied as to the level of specificity with which they gave their answers, resulting in varying degrees of accuracy of speaker localisations. To combat this inaccuracy, the present survey required respondents to state the first part of their childhood postcode, such that the postcode ‘CB2 1TP’ would be given as ‘CB2’.

This had various benefits and drawbacks. The major advantage of this methodology is that the increased and uniform specificity was particularly useful in urban areas. For instance, as opposed to giving a response of ‘London’, respondents were able to specify which of hundreds of postcode districts they lived in, allowing a much more accurate and spread out geographical representation of where responses should be localised to for geospatial analysis.

However, while useful in urban areas, the use of postcodes generally hindered accuracy of localisation in rural areas. The Scottish Highlands, for instance, have relatively few postcode districts for a wide area, so clusters tend to form in these sparsely populated areas instead of the more heavily populated urban areas. An option might have been to request respondents’ full childhood postcodes, but the present methodology maintained a more comfortable degree of privacy for the respondents, particularly so in Ireland where every address has a unique Eircode. Given the number of responses from each type of area, though, the present methodology is arguably still advantageous overall with respect to the accuracy of respondent localisations.

Another issue is that postcodes are a relatively recent introduction. In the UK they were phased in between 1959 and 1974, and I received feedback from a number of respondents to the effect that the respondent did not have a postcode when they grew up because they had not been introduced yet. The situation is more extreme in Ireland, where Eircodes were only introduced in 2014. After a number of comments about the issue, the survey was amended to include an instruction to provide the smallest level of geographical specification possible, e.g. street name or village, to enable maximum accuracy of localisation. Nevertheless, a manual check of postcode data prior to geospatial analysis resulted in several postcodes being entered by hypothesis, e.g. a response of ‘Dublin’ would be represented as ‘D01’, despite there being a high chance that this was not the true representation of the respondent’s location of upbringing.

In addition to the location of each respondent’s upbringing, the locations for the upbringing of each respondent’s ‘Primary caregiver’ and ‘Secondary caregiver’ were also probed. However, due to the complexities involved in analysing the effect of such data on acceptability, this factor will not be discussed in the results.

3.1.5 *Survey distribution*

I judged social media the most effective way to reach a large number of potential respondents, notwithstanding the inevitable demographic skew that would characterise the data as a result of this. My personal social media accounts were used to advertise the survey, and respondents were encouraged to share the survey after completing it, especially with members of their family from a different generation. On Facebook, I made efforts to advertise the survey in Facebook groups with large numbers of members for people who were interested in language and dialect, in the case of a

British-centric group I used, or in their local or national identity, in the case of some Hiberno-centric groups I found. Towards the end of the data collection period I also attempted to contact Young Farmers groups in Wales and a similar institution in Ireland, but these attempts were unsuccessful. On Twitter I made use of contacts who were able to distribute my survey to an audience with a general interest in linguistics, and encouraging people to share the survey after completing it resulted in steady stream of responses which lasted for about a week. Both the use of Twitter in this way and targeted Facebook groups enabled me to reach an audience outside of my immediate social network, and introduced the majority of respondents above the age of thirty into my sample.

After an initial period of data collection, I compared the density of responses received to date to the population density of the UK and Ireland. This comparison was used to produce a map which highlighted underrepresented areas in my data, enabling me to target sharing of the survey by utilising various geographically advantageous contacts. The use of such a map in posts advertising the survey appeared to be an effective means of encouraging people from the underrepresented areas to complete the survey, by exploiting a sense of loyalty to their local area.

3.1.6 *Methods of analysis*

All statistical and geospatial analysis was conducted using RStudio, and all maps were produced using QGIS.

The primary means by which amassed acceptability judgements will be presented takes the form of a percentage score. A stimulus judged as fully acceptable by all respondents would receive a score of 100%, while a stimulus judged as fully ungrammatical by all respondents would receive a score of 0%. The intervening judgements on the four-point Likert scale detailed in §3.1.1 were assigned scores of 33.3% and 66.7% accordingly. The acceptability for a given stimulus is determined by the average of these percentage scores over all respondents in a given sample. In other words a score of 100% indicates that 100% of respondents found the stimulus 100% acceptable, but lower scores could either be due to some respondents finding the stimulus less acceptable to a large degree, or all respondents finding the stimulus less acceptable to a smaller degree. This methodology assumes that the four-point Likert scale is linear, but given the qualitative nature of the responses as they were presented to participants, this is by no means a valid assumption. An additional issue may be that only the best answer may be taken to represent a positive judgement of the stimulus, with the other three conveying at least some degree of unacceptability. The decision to present the options as such was motivated by the observation that levels of unacceptability may be distinguished, ranging from mild unacceptability or marginality to severe ungrammaticality, whereas there is no such distinction between e.g. ‘quite grammatical’ and ‘extremely grammatical’; rather an example is either fully acceptable or it is not.

One way to improve this methodology would be to present participants of a separate pilot survey with each of the four response options and ask them to place them on a linear scale, using the results from this to more accurately represent the results from the main survey. However time constraints prevented such a study from being undertaken, and so I proceed under the assumption that the four-point Likert scale may be taken to be linear.

The maps presented were created using a Kernel Density Estimate algorithm. Each response was given a new value which corresponded to the mean of the 5% of the data points which were the closest to that response. This data was exported to QGIS, where Voronoi polygons around each data point were computed.

3.2 Data analysis

This section presents the results of the quantitative dialect survey conducted. Although the collection of sociolinguistic information forms a large part of this section, the sociolinguistic properties of the NSs are only of peripheral interest for the project and are not considered in the syntactic analysis presented in §§4–5. However, given that an aim of this dissertation is to introduce PaPRRs and EBICs in particular to the generative literature, the data is nevertheless presented, with many of the issues raised being left for further investigation.

3.2.1 Interstructural correlations

A Spearman’s rank test was performed on the data to establish the correlation between the acceptability of two given groups of stimuli. However, given that acceptability judgements are given in the form of a four-point Likert scale, the *jitter* function was used to create an arbitrary ranking of the respondents which minimised the extent to which stimuli were scored identically by multiple respondents, providing a semi-continuous ranking of respondents based on minor deviances from their responses. In order to reduce the high degree of uncertainty created by ordering the respondents on such an arbitrary basis, the Spearman’s rank test was repeated 1000 times for each test performed, and a mean of all the correlation values was taken. This had the effect that exact *p*-values could not be computed for the correlation tests, and so they are not displayed here.

Aggregated Spearman’s rank correlations between the acceptability scores given to various structural groups, calculated using the method just detailed, are given in Table 1. Table 1a gives correlations over the entire structure groups, while Table 1b gives correlations between the temporally–aspectually simplest stimuli, i.e. stimuli 1, 2 and 4 for events A and C only. A score of 0.3 is deemed a weak correlation, 0.5 an average correlation and 0.8 a strong correlation.

	PaPRR*	EBICs*	SPRRs*
SCRs*	0.730	0.517	0.416
PaPRRs*	–	0.509	0.465
EBICs*	–	–	0.337

Table 1a. Spearman’s rank correlations between structure groups (all stimuli).

	PaPRRs*	EBICs*	SPRRs*
SCRs*	0.668	0.390	0.307
PaPRRs*	–	0.509	0.377
EBICs*	–	–	0.265

Table 1b. Spearman’s rank correlations between temporally–aspectually simplest stimuli within structure groups.

It is clear that the strongest correlations are between PaPRRs and SCRs. The next strongest correlations are between EBICs and the PaPRRs/SCRs, while the stative participle RRs show only a very weak correlation with any of the NSs. The result that PaPRRs and SCRs seem to be covariant with each other to a greater extent than with EBICs may be taken as evidence that they are more closely structurally related to each other than they are with EBICs. Such a strong correlation between the two suggests that the particular (formal) characteristic of a given speaker’s grammar—perhaps a formal feature—which permits SCRs also permits PaPRRs; if this feature is present then both SCRs and PaPRRs are licensed in the grammar, and if not then the speaker finds both ungrammatical. This hypothesis is reinforced by the surface structural similarity displayed by the two constructions. Meanwhile, the weak correlation between stative participle RRs and the NSs suggests that these two

groups of structures are not formally related. In Bailey (2019), I rejected the possibility that PaPRRs could be structurally related to phasally derived SPRRs (see §4.2), and this evidence seems to support such an analysis.

Given that I have hypothesised a closer structural relationship between PaPRRs and SCRs, the next question concerns the nature of this relationship, as compared with the nature of the relationships between each of these two structures and EBICs, and between each of the NSs and SPRRs. Table 2 shows the mean acceptability scores of all stimuli in a given structure group, as well as the scores for the temporally–aspectually simplest stimuli within each group. A full table of acceptability scores for all stimuli can be found in the Appendix.

	All stimuli	Simplest stimuli*
SCRs	29.0%	41.3%
PaPRRs*	40.4%	58.2%
EBICs	33.3%	34.7%
SPRRs*	52.0%	46.7%

Table 2. Mean acceptability scores by structure. Values for simplest stimuli and for all PaPRRs and SPRRs exclude data from the stimuli mentioned in §3.1.2.

Overall, PaPRRs seem to be judged most favourably among the NSs, with SCRs and EBICs showing generally similar levels of acceptability. The fact that all EBICs have a higher acceptance than all SCRs yet the reverse is true for the simplest stimuli may however result from the fact that the group of SCRs contained on average more temporally–aspectually complex stimuli than the group of EBICs (see e.g. Gibson (1998) and Divjak (2017) for discussion of the effect of structural complexity on acceptability).

The following plots show the relationship between the acceptance of two (groups of) stimuli within individual speakers. The *jitter* function has been used in order to convey more clearly the areas of the plot which have a high density of points.⁵ The red line shows where the two (groups of) stimuli would be judged as equally acceptable; a point below this line would indicate that the speaker represented by that point judged the structure represented by the *x*-axis to be more acceptable than the structure represented by the *y*-axis, and a point above the line would indicate the reverse.

⁵ The *jitter* function appears to generate square-shaped clusters in the right-hand plot of Figure 5, for instance, because the scores of only two stimuli have been amalgamated on each axis, whereas many more stimuli are taken into account by the left-hand plot of Figure 5. The clusters on the left-hand plot are put into much smaller square shapes due to the fact that each axis represents many more sentences. The number of clusters on each axis is determined by the formula $3n + 1$, where n is the number of stimuli over which an average acceptance score has been found.

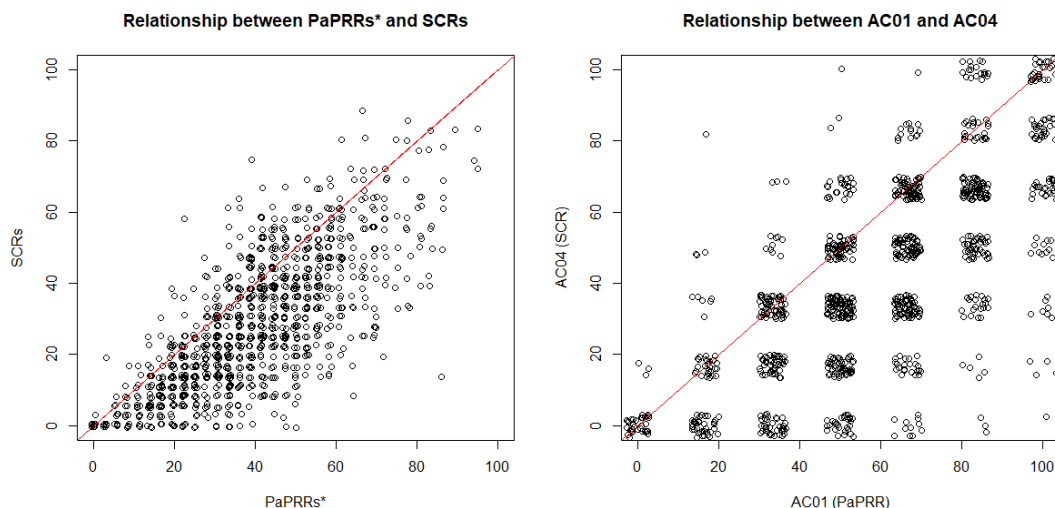


Figure 5. Relationship between all PaPRRs* and SCRs (left) and between the simplest such structures (right).

Figure 5 shows that most speakers more readily accept PaPRRs than they do SCRs, suggesting that speakers whose grammars allow SCRs may necessarily have to allow PaPRRs as well. Thus an implicational relationship may be hypothesised between these two structures, along the lines of the implicational hierarchy in (15):

- (15) Subject contact relatives (SCRs) > Past participle reduced relatives (PaPRRs)

Note that, although not every respondent's data point conforms to this generalisation, the judgements of an individual informant are known to be subject to a high degree of error, as observed by authors such as Schütze (1996/2016) and Cowart (1997). As such, although the relationship between the two structures may not be reliably represented in the data for any individual speaker, the fact that the number of responses received was so large means that the overall trend shown by Figure 5 is statistically reliable.

We can compare the relationship between SCRs and PaPRRs to those between each of these structures and EBICs, shown in Figures 6 and 7 respectively.

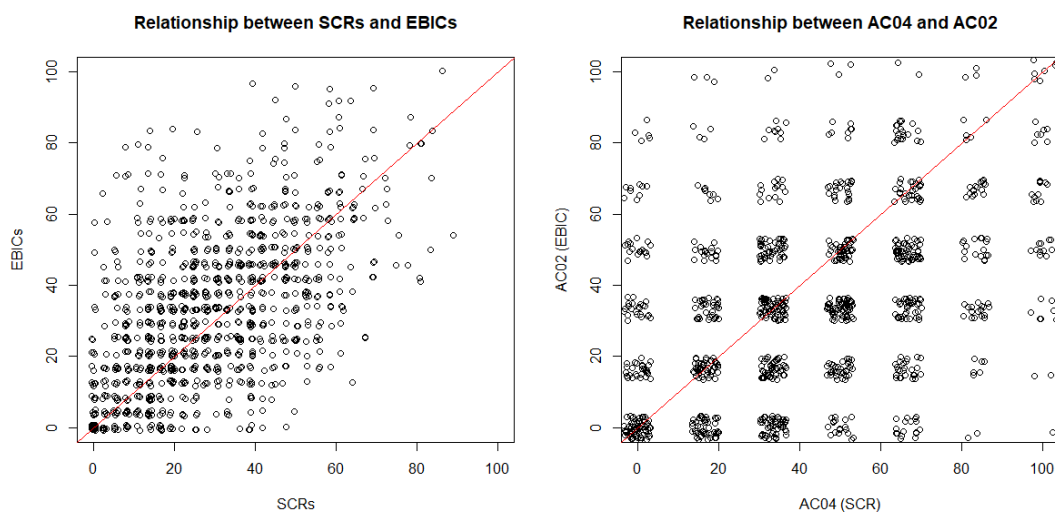


Figure 6. Relationship between all SCRs and EBICs (left) and between the simplest such structures (right).

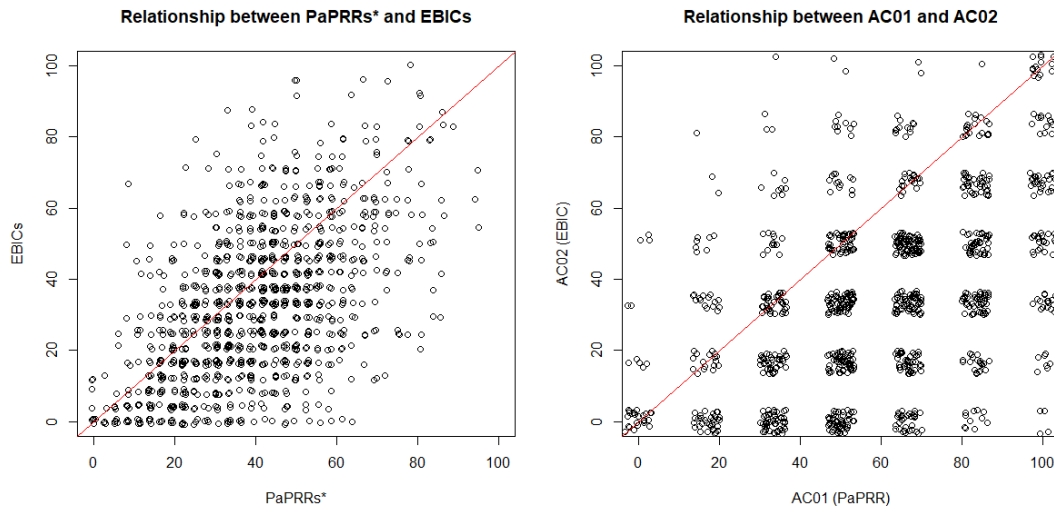


Figure 7. Relationship between all PaPRRs* and EBICs (left) and between the simplest such structures (right).

Very little sign of an implicational relationship between SCRs and EBICs is shown in Figure 6, although many speakers seem to show a preference for PaPRRs over EBICs in Figure 7. Nevertheless, the strength of the correlation between these latter two structures in Table 1 remains less than that shown between SCRs and PaPRRs. It must also be borne in mind that such plots as Figures 5–7 cannot show any kind of structural relationship between two (groups of) stimuli *per se*, but instead provide an indication of the relative reported levels of acceptability of two structures within one grammar. It is only when data from such plots is combined with information about covariance given by correlation scores that it can be deduced whether and how two structures are syntactically related. Thus the high correlation and reasonably strong implication between SCRs and PaPRRs may be taken to indicate that these two structures are in some way syntactically related, while the less clear relationships and weaker correlations between these two and EBICs may be taken as evidence that this same relationship does not exist for these pairs of structures.

Let us now compare the acceptance of each of the NSs to SPRRs.



Figure 8. Relationship between all SCRs and SPRRs* (left) and between the simplest such structures (right).

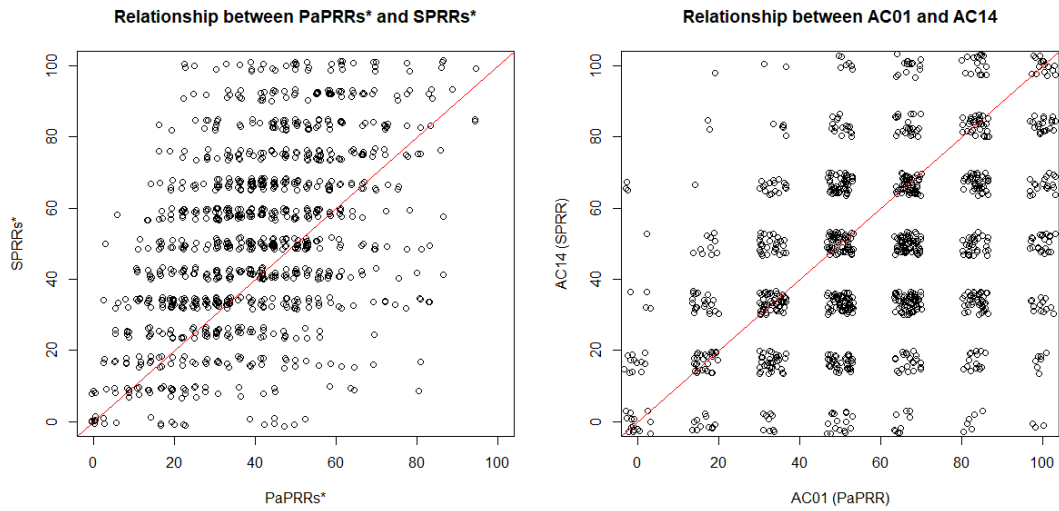


Figure 9. Relationship between all PaPRRs* and SPRRs* (left) and between the simplest such structures (right).

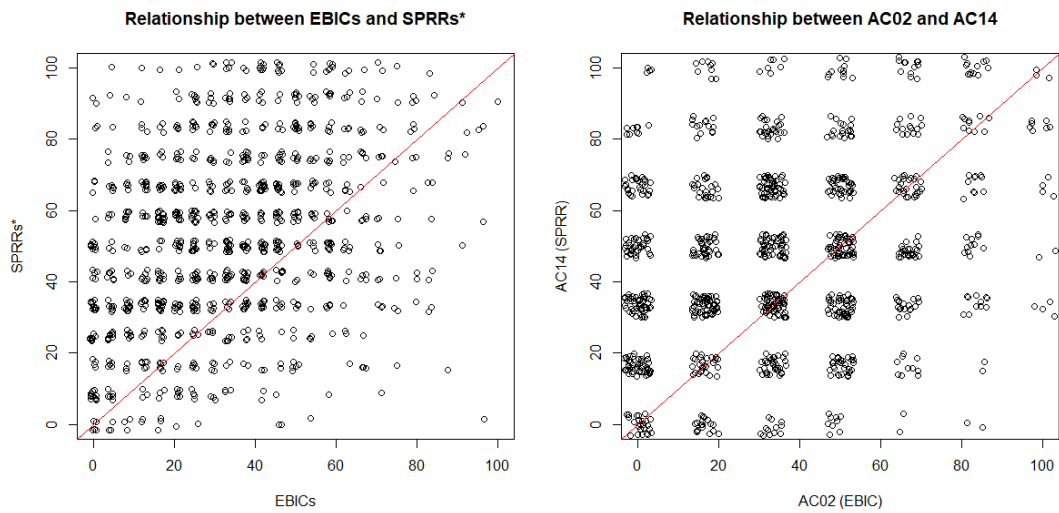


Figure 10. Relationship between all EBICs and SPRRs* (left) and between the simplest such structures (right).

There appear to be preferences for SPRRs in the left-hand plots for all three structures, although the reverse is true in the direct simplest structure comparison for PaPRRs, and no clear reference is discernible in the direct simplest structure comparisons for either SCRs or EBICs. Notably, the data for SPRRs in the left-hand plots include the passive stimuli E14 and F14, which both contain a past participle *been* where the present participle *being* is also licensed in the majority dialect. The passive stimuli also have higher acceptability scores overall than the active stimuli A14 and C14, so this could explain the mismatch between the two sets of plots.

Of importance is the fact that no clear correlation is evident between SPRRs and any of the NSs, with all correlation scores involving SPRRs being lower than all scores between two NSs, as shown in Table 1. Furthermore, the acceptance of SPRRs is neither consistently higher nor lower than that of any of the NSs in Figures 8–10. Therefore the hypothesis that NSs are structurally related to SPRRs does not find support from the survey results displayed here.

To summarise, the dialect survey conducted has suggested that SCRs and PaPRRs appear to be syntactically related to each other, such that their acceptance by a given speaker is covariant and that speakers will tend to judge PaPRRs as more acceptable than SCRs. On the basis of this close relationship, I will henceforth refer to these two structures collectively as the *Narrative Relative*

Structures (NRSs). In addition, EBICs also display a certain degree of covariance with SCRs and PaPRRs, albeit to a lesser extent. Meanwhile, none of the three NSs show a strong degree of covariance with phasally derived SPRRs. These findings are summarised in Figure 11.

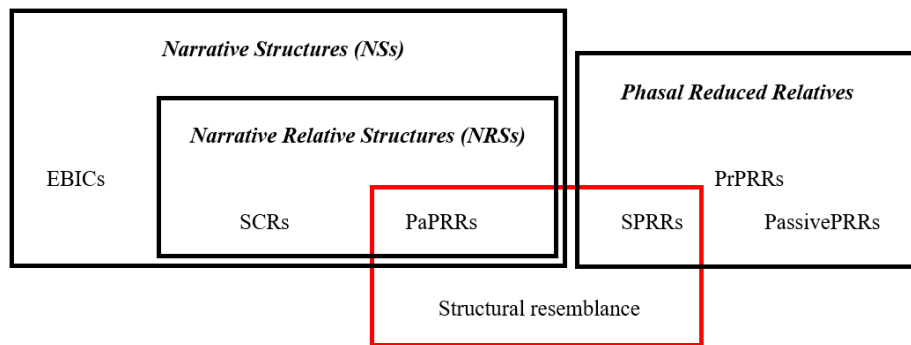


Figure 11. Schematic diagram of relationships between NSs and phasally derived reduced relatives.

3.2.2 Social distribution

The effects of age, education, gender and self-defined socioeconomic status (SES) on respondents' acceptance of stimuli were tested using regression models, together with which of the two orders of sentences the respondent was presented with (i.e. Set A or Set B). A stepwise regression was first conducted to establish which variables to include within each regression model, and the estimates for and significance (indicated by asterisks) of the effect of each factor on each structure are shown in Table 3, along with the results for the four active universally grammatical expletive–associate constructions provided for comparison. Blank cells indicate that the stepwise regression did not recommend the factor for inclusion within the model. Gender was never deemed significant by the stepwise regression model, and so it is omitted from the table.

	SCRs		PaPRRs*		EBICs		E–A cons.
	All	AD04 ⁶	All	AC01	All	AD02	
Age	0.17***	0.38***		0.20***		-0.12*	-0.14***
Education							
Level 2	-2.50	-0.64					
Level 3	0.70	3.58					
Level 4	2.36	6.45					
Level 5	6.01	10.80					
Level 6	4.54	11.11					
SES							
Level 2	-2.58	-2.88	-2.45	-4.91*	-0.86	-0.96	
Level 3	-7.13***	-10.25***	-7.28***	-12.74***	-5.28**	-6.55**	
Level 4	-19.28**	-22.49*	-21.39**	-23.94*	-18.10*	-17.75	
Set		-2.68	2.40*		-1.58	-3.95**	-1.25*

Table 3. Effect size estimates and significance for the effect of social factors on stimulus acceptance. Education levels correspond to the education categories outlined in §3.1.3 in the order 'below GCSEs' through to 'higher than postgraduate degree'. SES levels correspond to the socioeconomic categories outlined in §3.1.3 in the order 'lower/working class' through to 'upper class'.

⁶ AD04 is the amalgamation of the scores for A04, B04, C04 and D04. AC01 is the amalgamation of A01 and C01 only.

The regression models show that socioeconomic status, particularly if respondents self-defined as upper middle class, generally had a significant effect on the acceptance of all NSs, in each case such that higher levels of education corresponded with lower acceptance of NSs. Meanwhile, the acceptance of SCRs also appears to be affected by some extent by level of education, such that higher education correlates with higher acceptance, although this result is not statistically significant. Interestingly, age had a significant effect when the simplest (and therefore most frequent) examples of each structure were tested, but not on the body of examples of each structure as a whole. In this respect, there was a difference between the structures such that increased age generally correlated with higher acceptance of (the most frequent) SCRs and PaPRRs, but lower acceptance of the most frequent EBICs. The contrast between the figures for all structures and just the most common structures merits some discussion; the assumption would be that the figures for all structures are more statistically reliable given the larger dataset involved. However, the most frequent structures are those that speakers would most likely be able to relate to, and so are perhaps likely to receive the most accurate judgements. Notably, age also had a highly significant effect on the acceptance of expletive–associate constructions, a result that is somewhat surprising. This may call into question the validity of the rest of the results, given that expletive–associate constructions are assumed to be universally grammatical to all speakers of English. For brevity, the present discussion will only focus on those correlations which show statistical significance.

Figure 12 plots the acceptance of each of the NSs against respondents’ socioeconomic status. Note that the values 1 to 4 correspond to the socioeconomic categories outlined in §3.1.3 in the order ‘lower/working class’ through to ‘upper class’.

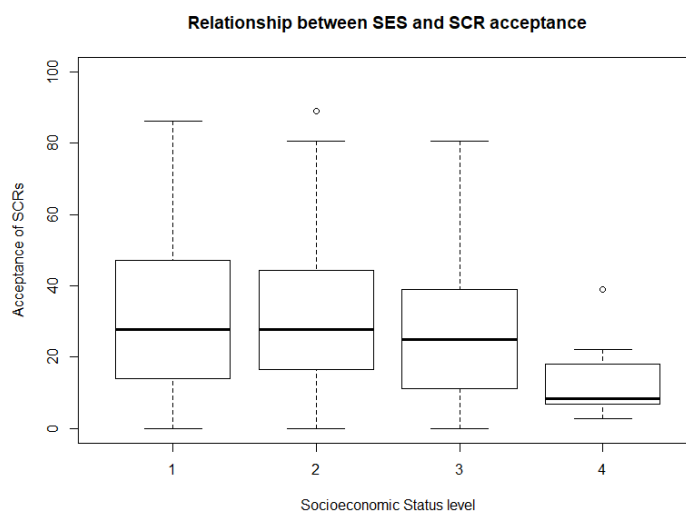


Figure 12a. Relationship between acceptance of SCRs and socioeconomic status.

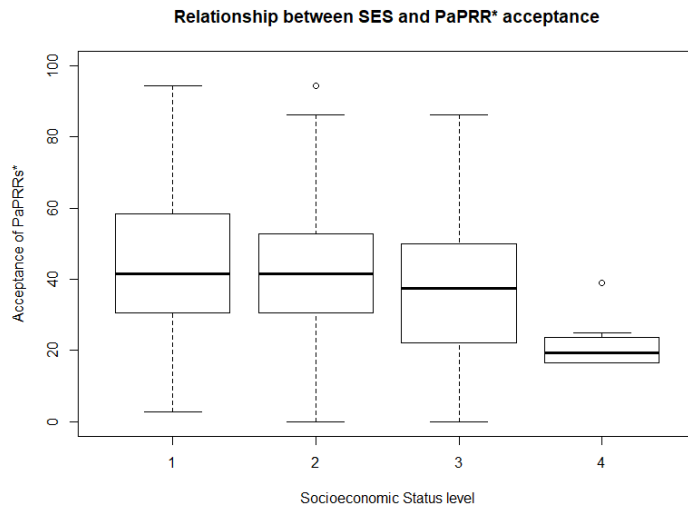


Figure 12b. Relationship between acceptance of PaPRR*s* and socioeconomic status.

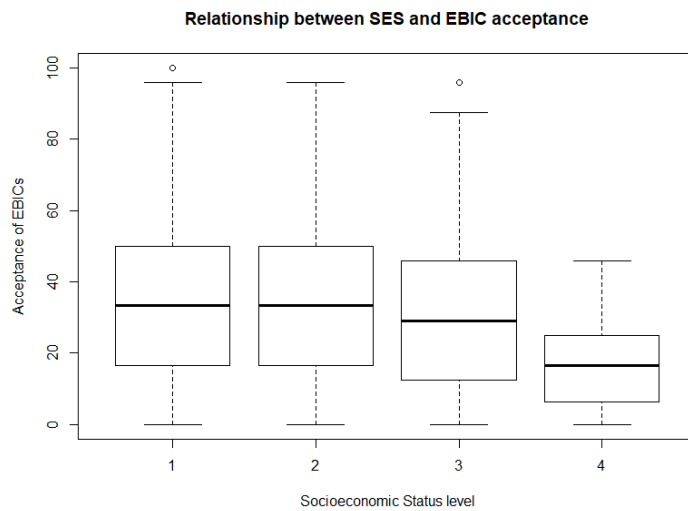


Figure 12c. Relationship between acceptance of EBICs and socioeconomic status.

I conclude from Figure 12 that all three NSs show a significant relationship between their acceptance and the socioeconomic status of respondents, such that respondents of lower socioeconomic status more readily accept the structures. It should be noted that only 7 respondents out of the 1030 (0.7%) identified themselves as ‘upper class’, so the results of this category cannot be taken as anything more than suggestive.

The effect of age on the temporally–aspectually simplest NSs is shown in Figure 13. For an explanation of why the results form horizontal clusters, see footnote 5.

NARRATIVE STRUCTURES IN COLLOQUIAL ENGLISH

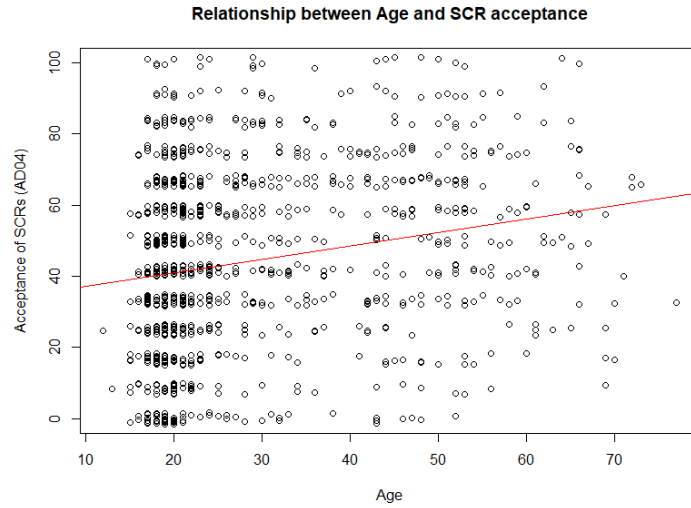


Figure 13a. Relationship between acceptance of temporally–aspectually simple SCRs and age.

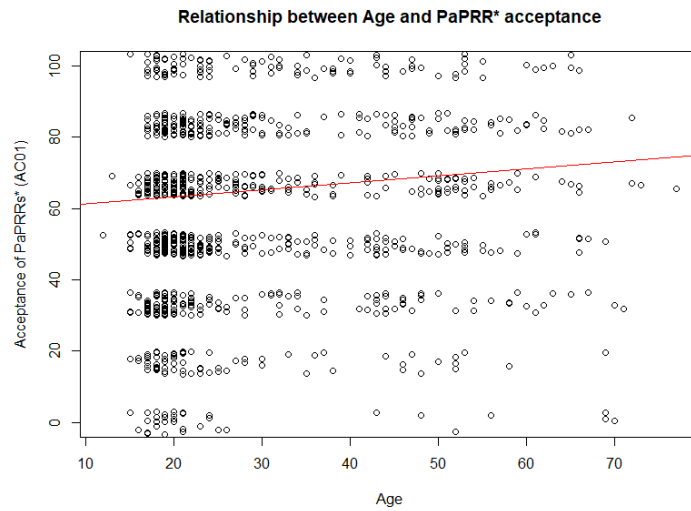


Figure 13b. Relationship between acceptance of temporally–aspectually simple PaPRRs* and age.

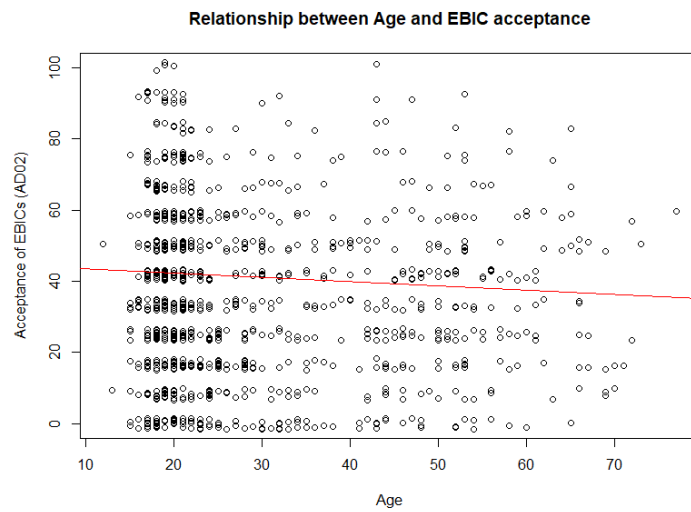


Figure 13c. Relationship between acceptance of temporally–aspectually simple EBICs and age.

These plots suggest that increased age corresponds with higher levels of acceptance of SCRs and PaPRRs but lower levels of acceptance of EBICs, although, as noted above, this relationship is not significant when more complex stimuli are also taken into account. It should also be acknowledged that there is a heavy skew in the sample towards younger speakers, and as such these results may not be representative of the wider population. As such, no firm conclusions can be drawn from this data, particularly with regard to any potential diachronic change in the acceptance of NSs. However, if we *were* to take these results as indicative of diachronic change, notwithstanding the unrepresentative sample, this would perhaps be a further suggestion that SCRs and EBICs are not syntactically related to each other, as one would appear to be increasing in acceptability through time while the other would be decreasing.

3.2.3 Geospatial distribution

The maps in this section show the geospatial distribution of the three NSs within the Anglo-Celtic Isles (figures used are average acceptances for all stimuli of a given structure). In each case, the left-hand map shows the absolute acceptance of the structures on a gradient colour scale between 0% and 100% acceptance, while the right-hand map shows the variation on a scale between the minimum and maximum acceptance values for the given structure. As such, the left-hand maps are suited to comparing the structures to each other, while the right-hand maps are suited for comparing different localities with each other with respect to a single structure. Note that as no responses were received from Shetland, it has been omitted from the map in the interest of space efficiency.

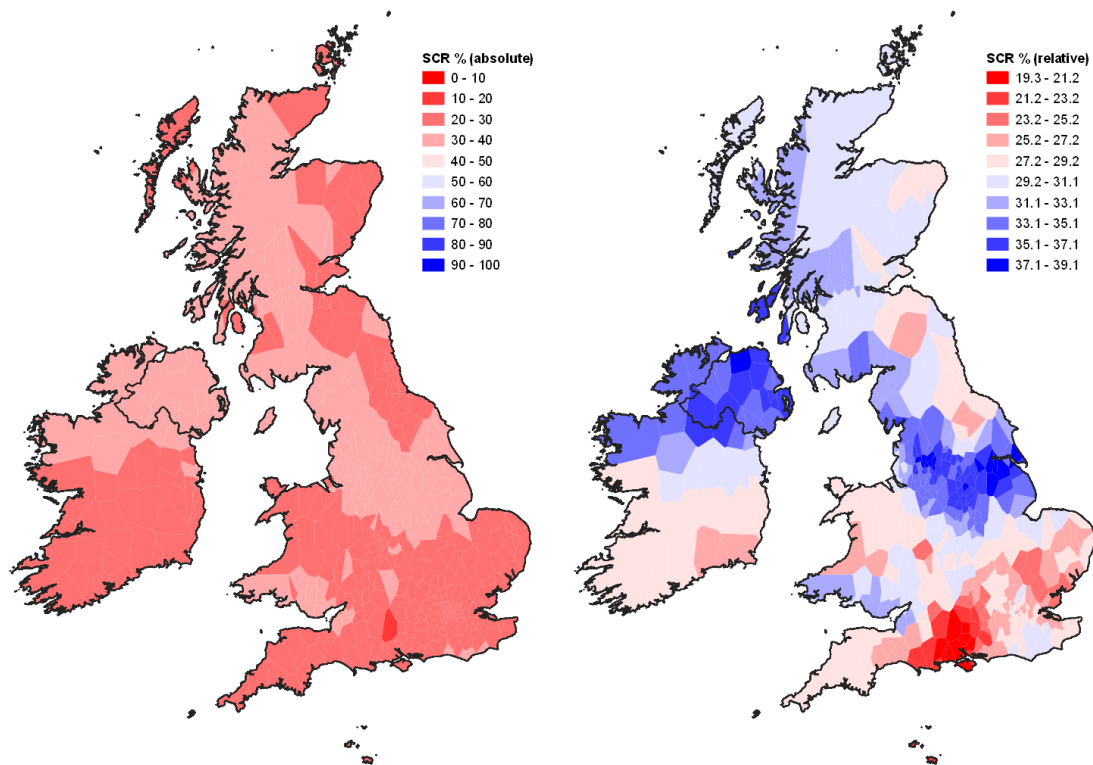


Figure 14. Geospatial distribution of SCRs.

NARRATIVE STRUCTURES IN COLLOQUIAL ENGLISH

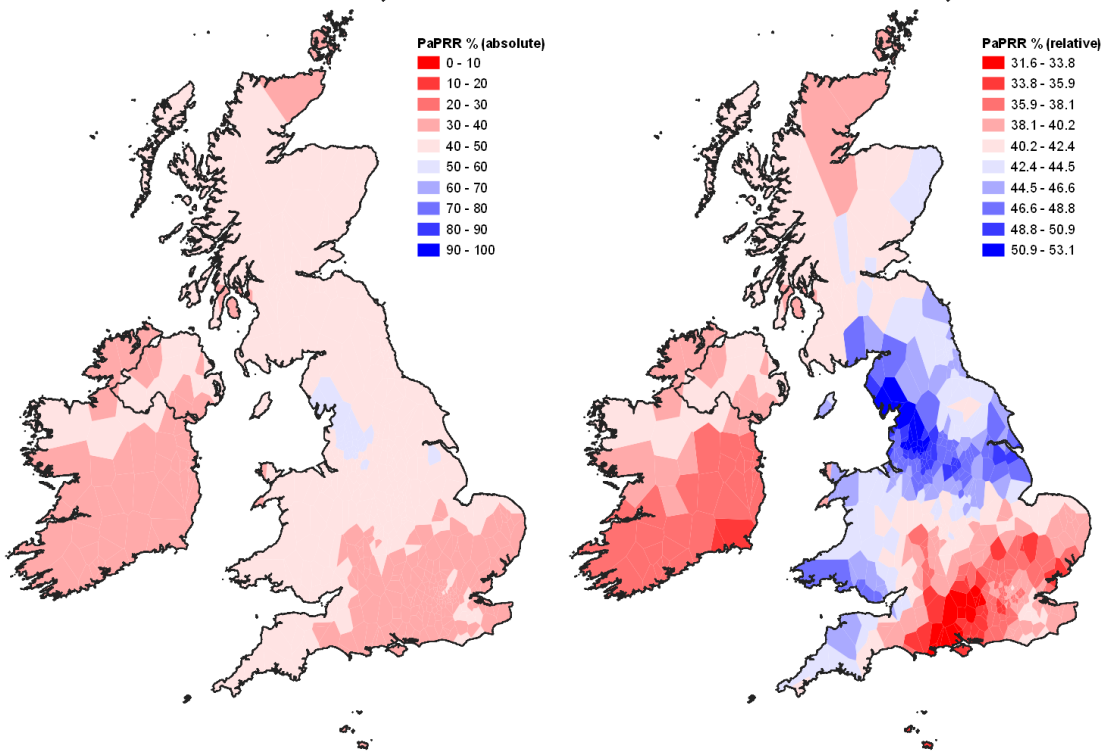


Figure 15. Geospatial distribution of PaRRs*.

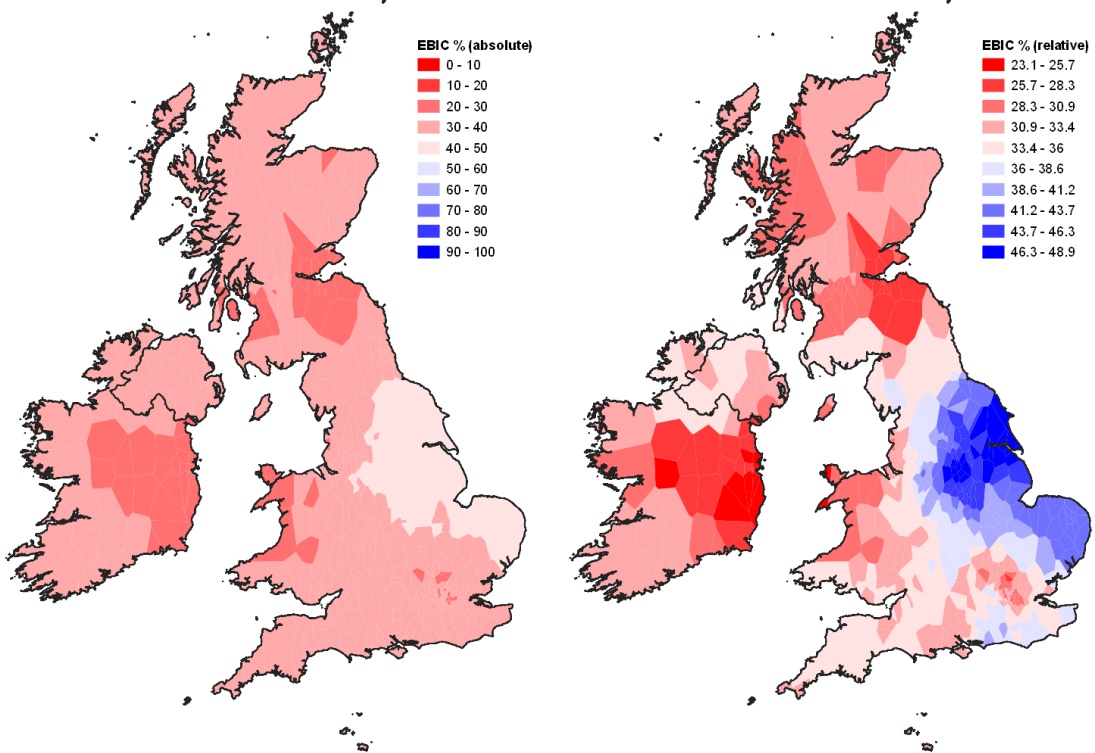


Figure 16. Geospatial distribution of EBICs.

Firstly, we can observe that the overall distributions of each structure, shown in the left-hand map of each pair, reflects the figures of overall acceptance in the left-hand column of Table 2, i.e. that PaRRs are more widely accepted than EBICs, and EBICs more so than SCRs (although note again that the SCR stimuli on average contained more temporal–aspectual complexity than the EBICs, and

that this pattern was reversed when only the simplest stimuli were compared). Even when we compare the differences between structures regarding which areas judge them more favourably, the vast majority of areas at least come very close to conforming to this hierarchy (with the exception of eastern parts of Suffolk showing a slightly higher acceptance of EBICs than of PaPRRs).

Turning our attention to a comparison of the right-hand maps of each pair, we can see that SCRs and PaPRRs show a striking similarity in terms of which areas of the Anglo-Celtic Isles most readily accept them. In particular, an area comprising the northern Midlands, southern and eastern parts of Yorkshire, Lancashire, (western) Cumbria and perhaps a small part of southwestern Scotland (although responses from this particular area were far from plentiful) is shown to form a core zone where the NRSs are judged most favourably, with the southern coast of Wales also accepting the structures more readily than average. The acceptance of PaPRRs is relatively higher in northern Lancashire and southern Cumbria when compared to the cities of southern Lancashire than the acceptance of SCRs. Particularly western areas of Scotland also show a somewhat lesser difference between the acceptance of SCRs and PaPRRs. Meanwhile, a core area in southern England comprising Dorset, Hampshire and Wiltshire and extending to the counties immediately to the north of London most strongly rejects both sets of structures, although seemingly more so relative to the rest of the Anglo-Celtic Isles with PaPRRs than with SCRs. A major difference between the acceptance of such structures relative to the rest of the language area, though, is seen in Ulster and northern parts of Connacht. Here SCRs appear to be accepted as readily as in the core north Midlands area identified in England, but speakers from Ulster and Connacht seem to accept PaPRRs on the other hand less readily than average, although these structures are still judged more favourably here than in Munster and Leinster. It is notable that an entire chapter is devoted to SCRs by Henry (1995) in her monograph on dialect variation with respect to Belfast English; the results of this survey seem to confirm that this city and its hinterland form one of the hotspots of SCRs within the Anglo-Celtic Isles. However, we again must bear in mind that even in these northern Irish provinces PaPRRs are generally more readily accepted than SCRs.

The other significant observation to be made from Figures 14–16 is that the geospatial distribution of EBICs differs significantly in a number of ways from that of the NRSs. Although the northeastern English Midlands, including southern and eastern Yorkshire, form part of the region which most readily accepts EBICs as well as the NRSs, in the case of EBICs the rest of this region consists of nearly the entirety of Yorkshire (the western fringes of the county were poorly represented in the data due to low population density), a significant part of the East Midlands and, albeit to a slightly lesser extent, northern and eastern areas of East Anglia. It may be observed that there is a not insignificant correspondence between the areas which most readily accept EBICs and the counties which were subject to the heaviest Norse settlement during the Viking Age. This is suggestive of influence from Old Norse, although this is a point I will not develop further here.

It seems clear that the geospatial distributions of NRSs on the one hand and EBICs on the other are substantially different, while those of SCRs and PaPRRs bear striking similarities. These findings lend further support to the developing hypothesis that SCRs and PaPRRs have a close syntactic relationship that is not shared with EBICs. Furthermore, all regions show a considerable preference for PaPRRs over SCRs, which is consistent with the conclusions from §3.2.1.

3.2.4 *Effect of syntactic factors on acceptability*

This section tests for the effect of argument structure and humanness of subject on the acceptance of NSs. A Shapiro-Wilk test performed on judgement data for the relevant groups of stimuli indicated that none of the data was normally distributed, and so a *t*-test to compare such parameters as argument structure and humanness of subject is of limited validity. Therefore, Tables 4 and 5 show the

percentage acceptance scores of the temporally–aspectually simplest examples of each type of NSs with respect to argument structure and humanness of subject respectively.

	Transitives	Unaccusatives	Passives
SCRs	38.5%	51.8%	–
PaPRRs	68.5%	75.7%	69.1%
EBICs	28.4%	44.2%	–

Table 4. Comparison of acceptance of temporally–aspectually simplest NSs with respect to argument structure.

	Human	Non-human
SCRs	44.2%	46.0%
PaPRRs	79.7%	64.6%
EBICs	35.8%	36.8%

Table 5. Comparison of acceptance of temporally–aspectually simplest NSs with regard to humanness of subject.

In Bailey (2019:6) I observe that PaPRRs show an overwhelming bias towards unaccusative verbs in the corpus of examples collected, but that spontaneous transitive examples are nevertheless still attested. The findings presented here suggest that this preference for unaccusativity holds for both SCRs and EBICs as well. A correspondence between unaccusativity and presentationality has commonly been suggested (see e.g. Lonzi 1986; Lambrecht 2000; Irwin 2018), so this result is perhaps expected. However, the observation that passives appear to pattern with transitives as opposed to unaccusatives is somewhat unexpected, given the derivational similarities posited to exist between passives and unaccusatives.

It seems that PaPRRs are the only structure to show an effect in terms of whether or not the subject is human; there is no immediately obvious reason as to why this might be. Note that this data includes the stimuli B01 and D01, about which I have assumed that many respondents mistakenly judged them as if they were majority dialect expletive–associate constructions. However, even if we choose to look at the set of PaPRRs with examples A07–D07, we still observe that the stimuli with human subjects have a significantly higher average acceptance (44.3%) than those with non-human subjects (32.6%).⁷ It may be that this effect is due to the relative frequency with human and non-human subjects appear in PaPRRs, although it is unclear why this reasoning does not also apply to SCRs and EBICs.

3.2.5 *Effect of exposure on acceptability*

As mentioned in §3.1.2, the stimuli were presented to respondents in one of two possible randomised orders, with a roughly equal number of respondents receiving each order. The respondents who received each order were grouped into Set A and Set B accordingly. The purpose of this was to test the difference between how a stimulus was judged by respondents in each set depending on how far through the survey the stimulus was presented, and therefore how much (immediately) prior exposure respondents had received to similar structures.

Firstly, a linear regression model was run comparing the difference in the average acceptance of each stimulus with the difference between the positions in which that stimulus appeared in each order.

⁷ For these stimuli, however, we also observe that transitive examples (40.6%) seem to be more readily accepted than the unaccusative examples (36.3%). Although worthy of investigation, I leave this issue aside here.

For instance, if stimulus X was the 43rd sentence presented to respondents in Set A, but the 13th sentence presented to those in Set B, the difference in position would be 30. This model was applied both to all stimuli presented to respondents and to only the NSs as a group. Both models found a statistically significant effect ($p < 0.001$) of the difference in stimulus position on the difference in acceptance between each set. For the NSs, the difference in percentage acceptance increased by an average of 0.31% for each increment of difference in position. A visual representation of this correspondence is shown in Figure 17.

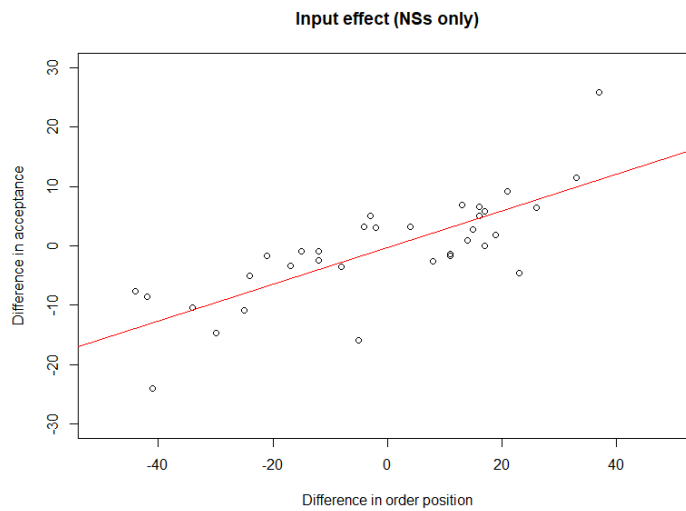


Figure 17. Effect of difference in stimulus position on difference in acceptance of stimulus.

This plot shows that there is a strong relationship between how many NSs the respondent has been exposed to during the course of the survey before the point at which they are asked to judge a given stimulus and how favourably they judge that stimulus; specifically, the greater the prior exposure to NSs within the survey, the greater the acceptance of subsequent NSs. A further representation of this correspondence is shown in Figure 18, which displays the relationship between each of the acceptance levels of the stimuli in the two sets.

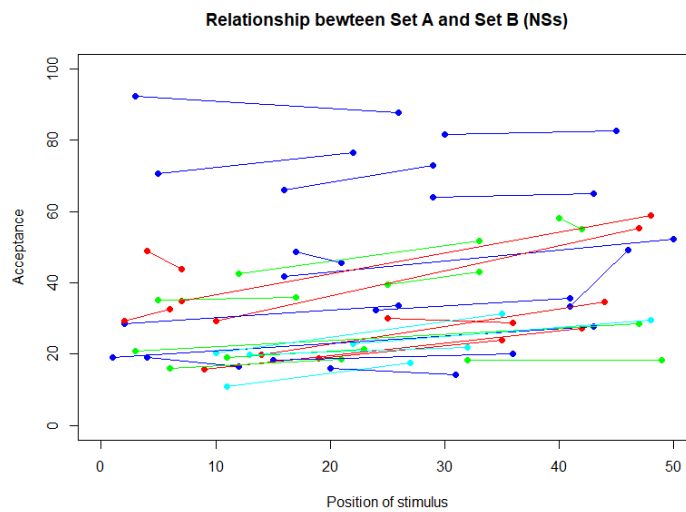


Figure 18. Relationship between acceptance of stimuli in Sets A and B depending on stimulus position. SCRs are shown in green, PaPRRs in blue, InfPRRs in cyan and EBICs in red.

A drawback of the survey design meant that there were insufficient numbers of non-NS stimuli to establish whether this effect is unique to NSs or whether it may apply to other marginally acceptable or dialectally variable structures. A correlation similar to that shown in Figure 17 was also found for SPRRs, which I hypothesise to be syntactically unrelated to NSs, but the fact that the sample of such structures was limited to six stimuli meant that the result was not statistically significant ($p = 0.260$). Nagata (1988) shows that Japanese stimuli presented to participants were given lower acceptability ratings when they were presented a second time, suggesting that the opposite effect to that observed in the present survey would be expected. This suggests that the observed increase in acceptability of the stimuli as the survey progressed is a property specific to NSs.

The correspondence observed in Figure 17 does not explicitly rule out the possibility that respondents judge sentences more favourably as they progress through the survey due to reasons other than the direct effect of exposure to particular grammatical structures. Therefore, a further linear regression model was run testing for a correspondence between the acceptance of each stimulus in each set and the position it appeared in the order of stimuli in that set. The results are presented in Figure 19.

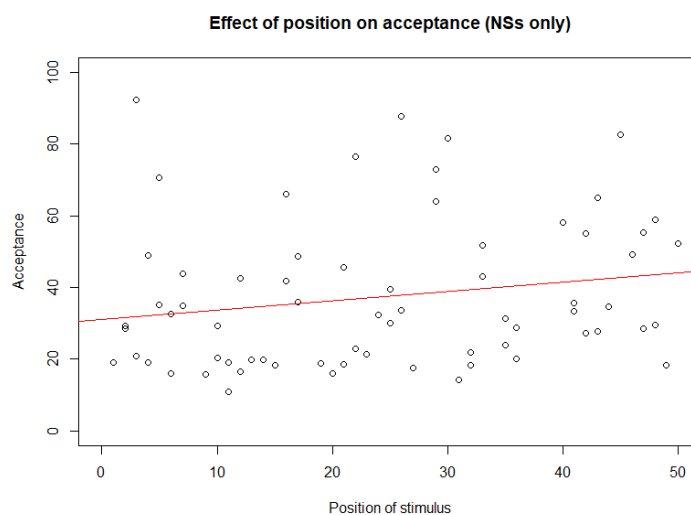


Figure 19. *Effect of stimulus position on acceptance of stimulus.*

Although the correlation that was found was positive, in that each subsequent position in the survey order on average increased the percentage acceptance of the stimulus by 0.26%, the result was not statistically significant ($p = 0.119$). The contrast between the correlations presented by Figure 17 and Figure 19 may tentatively be taken to suggest that, although the acceptance of stimuli increases with (immediate) prior exposure to similar structures in the survey, this is not a direct result of the respondent's progression through the survey, such that acceptance scores automatically increase as the respondent nears completion. As mentioned above, the lack of comparable marginal structures whose acceptance is however expected not to vary with prior exposure during a survey means that this effect cannot be confirmed as characteristic of NSs in particular, but we can nevertheless observe that respondents become more willing to accept the stimuli as they are exposed to more examples of a similar structure, which indicates that acceptance of the structures may be dependent on input.

3.2.6 *Effect of temporal–aspectual complexity on acceptability*

Stimuli of varying degrees of temporal–aspectual complexity were included in the survey to allow for testing of this effect. However, it has been observed (e.g. Schütze 1996/2016; Sprouse 2007) that

stimuli presented without context are susceptible to such an effect simply because of the amount of imagination required on the part of the respondent to posit a situation in which such a sentence would be appropriate (compare e.g. “A dog went past the window” to “A dog would have had to have been going past the window”). Therefore, although stimuli without overt perfective, progressive and passive auxiliaries are generally deemed more acceptable by respondents than those with such auxiliaries, I will not comment further on this pattern here, given that such a result may be expected. I refer the reader to the Appendix for detailed acceptance data for each stimulus.

3.2.7 *Summary of survey findings*

The dialect survey conducted has shed light on a number of issues which would otherwise have remained obscure and left the syntactic analysis of NSs in the following sections subject to speculation and the arguably dubious methods of introspection. Instead, we have reliable data, both qualitative and quantitative, on which to proceed in our effort to provide an analysis of the structures in question.

The main findings of the survey are summarised below:

- (i) The acceptance of SCRs and that of PaPRRs correlate more strongly with each other than that of either group with that of EBICs.
- (ii) The acceptance of SCRs by a speaker tends to imply the acceptance of PaPRRs by that same speaker.
- (iii) The acceptance of PaPRRs does not strongly correlate with that of SPRRs.
- (iv) Unaccusativity is preferred over transitivity for all structure groups tested.
- (v) SCRs and PaPRRs are most favourably judged in an area comprising a large part of northwestern England and the north Midlands, and SCRs are additionally judged favourably in Ulster and northern Connacht. Meanwhile, NBICs are most favourably judged in Yorkshire, the East Midlands and northern and eastern parts of East Anglia.
- (vi) The effect of immediately prior exposure to NSs has a positive effect on the acceptance of subsequent NSs within the survey (although it remains to be tested whether or not this effect applies to marginal structures more generally).

Given these findings, the primary factor that should be taken into account in the syntactic analysis is that SCRs and PaPRRs form a group to the exclusion of EBICs (and wholly unrelated to SPRRs), as is shown by the correlation in their acceptance, both geospatially and otherwise. This will form the basis for the following syntactic analyses. In addition, we should aim to explain why PaPRRs seem to be more widely accepted than SCRs, and unaccusatives more so than transitives.

4 **Analysis of SCRs and PaPRRs**

I established in §3 that the results of the dialect survey conducted suggest a closer syntactic relation between SCRs and PaPRRs than exists between either of these structures and EBICs. Therefore this section seeks to unify them in terms of their syntactic analysis. After using the distributional similarities between SCRs and PaPRRs to conclude that the SCRs under examination are indeed true relative structures, I will following Diesing (1992) and Sichel (2018) in assuming that the site from A'-extraction of non-presuppositional DPs takes place is lower than that of presuppositional DPs, and hence that presuppositional DPs are prevented from undergoing such extraction due to spec-to-spec anti-locality (Erlewine 2016, forthcoming). This results in the restrictions on the syntactic distribution of SCRs. I will then extend this analysis to PaPRRs by postulating the existence of a left periphery at

the edge of the higher phase of PaPRR clauses through which A'-extraction occurs, but which, as in the case of SCRs, is too local to the site of presuppositional DP extraction for such DPs to be extracted and form PaPRRs.

4.1 Analysis of SCRs

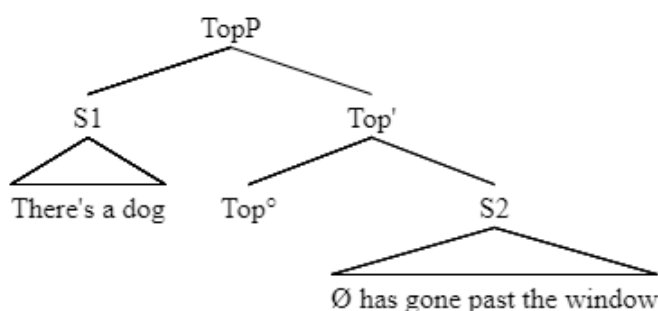
4.1.1 A survey of previous analyses

Out of the three NSs, SCRs are the structures that have been given the most attention in previous literature. This is perhaps somewhat surprising given that one of the conclusions of the dialect survey is that PaPRRs are more readily accepted, and therefore presumably more commonly attested, than SCRs. Previous analyses of SCRs may be grouped into three categories. The first, which is advocated primarily by Henry (1995) and den Dikken (2005), suggests that SCRs are not relative clauses at all, but are in fact monoclausal topic-comment structures, which make use of left-peripheral functional projections (i.e. TopP) to syntactically encode the observed discourse properties of the structures. The second posits that SCRs are indeed biclausal structures with restrictive contact relative clauses, which finds favour with authors such as Doherty (1993, 1994, 2000), Haegeman et al. (2015) and Haegeman (2015). The third, whose main proponent is Lambrecht (1988), is a variation on the second, in that a biclausal relative clause structure is assumed, but that the relativised nominal is simultaneously the object of the matrix clause and the subject of the embedded clause in a way unlike straightforward relativisation; Lambrecht terms his analysis the *presentational amalgam construction* (cf. Kluck 2011).

I will first show that neither the topic-comment analysis nor the amalgam analysis can account for SCRs as discussed in this dissertation within the broader context of NSs, which leads to me to support the true contact relative clause analysis. I will then suggest that all three analyses are nevertheless at least partially correct with respect to the set of structures they set out to analyse, and that the cohesion of the group of structures currently referred to as SCRs needs to be re-examined.

Henry (1995) and den Dikken (2005) propose an analysis of SCRs which holds that the example given at the start of this dissertation in (1a) would have the topic-comment structure given in (16):

(16)



Here the DP *a dog* has raised out of the phrase labelled S2 and has been joined by *there's* or whichever introducer is present in the sentence in S1. The introducer could be seen as some kind of topic or focus particle, considering that the contracted form *there's* is somewhat preferable to the uncontracted form *there is*. Henry (1995:131) remarks that topic-comment structures more generally have been identified as a particularly salient feature of Belfast English, which she argues lends support to this analysis, and also highlights the fact that resumptive pronouns can appear in the subject gap left by the topic fronting, resulting in monoclausal examples like (17).

(17) There was a woman she went to Spain last year.

However, Haegeman (2015; see also Haegeman et al. 2015) raises a number of problems with this approach, which I briefly list here. The first argument is that the topic–comment analysis predicts the non-constituency of what I have been taking to be a relative clause, but this is not borne out, as is shown by the simple fact that RCs in SCRs can be coordinated, as in (18).

(18) There was [a dog went past the window] and [a woman went up the hill].

The second argument is that the raised subject nominal can be shown to c-command the material in S2. For instance, it can license negative polarity items, as shown in (19), which is a contrast to what is possible in paratactic structures such as that proposed by Henry and den Dikken.

(19) There’s nobody knows how to do anything these days.

Note that both of these two arguments assume the constituency of S1 in (16), which is not necessarily a given if *there’s* is base-generated as a topic or focus particle in the left periphery but *a dog* raises from SpecTP. However, if S1 does indeed turn out to be a constituent then these are two very serious challenges to the topic–comment analysis.

The third argument made by Haegeman is that while the topic–comment analysis predicts SCRs to be root phenomena (Emonds 1976; Hooper & Thompson 1973; Haegeman 2012), with den Dikken stating that they are unembeddable, this prediction is not true. In fact, SCRs can occur as temporal clauses, conditional clauses and complement clauses to factive verbs, all of which are incompatible with root clause phenomena, in a direct contradiction to den Dikken’s assertion.

Lastly, under certain conditions discussed by Doherty (1993, 2000), SCRs can be stacked with other relative clauses, as exemplified in (20). As Haegeman (2015:142) observes, the second (*wh*-)relative clause presumably appears in S2, but its antecedent is in S1. It is unclear how this would work in the structure proposed by Henry and den Dikken.

(20) There’s a woman lives in Ghent [who knows all about this stuff].

To these four main problems flagged by Haegeman (2015) I would like to add an insight which falls directly out of the results of the dialect survey discussed in §3. I have concluded that the distributional similarities exhibited by SCRs and PaPRRs suggest a close syntactic relationship, such that the most desirable situation would be that the only difference between the two would be the size of the relative clause. In this spirit, if a topic–comment structure were to be applied to SCRs, then we should also expect it to be applicable to PaPRRs. However, assuming that the topic–comment structure is derived from that of a standard declarative sentence, this would lead us to expect PaPRRs to be derived from ‘standard declarative sentences’ like that in (21a). We may also expect that resumptive pronouns should be possible too, as in (21b).

(21) a. * A dog gone past the window.
b. * There’s a woman she gone up the hill.

I will not waste space by explaining how this analysis is unfeasible given the simple lack of tense in (21a) if it expected to be a declarative sentence. Suffice it to say that analysing the SCRs under discussion using the topic–comment structure in (16) brings with it serious problems.

My rejection of the amalgam analysis is along similar lines to my final argument for the rejection of the topic–comment analysis. Specifically, given the observed distributional similarities between SCRs and PaPRRs, we should expect any analysis of SCRs to also be applicable in a broad sense to PaPRRs. Indeed, the amalgam analysis runs into exactly the same problem as the topic–comment analysis in this respect. The amalgam analysis should lead us to expect the relativised nominal to simultaneously be the object of the matrix clause and the subject of the embedded clause, and therefore that each clause should also be able to stand freely on its own. However, although the matrix clause in (1b) can form a licit independent clause, the same is not true for the embedded clause, which, again would be expected to be (21a). The amalgam analysis is demonstrably inadequate to account for PaPRRs, and thus also inadequate to account for SCRs.

While I have concluded that neither of these analyses can account for the SCRs discussed in this dissertation, it is worth noting that there are some inconsistencies in the data which each analysis is an attempt to account for; indeed, it seems that the reporting of disagreement between speakers about the acceptability of structures is characteristic of the literature on SCRs. A particularly salient example is the reported existence of SCRs with resumptive subject pronouns in Belfast English by Henry (1995). While many of the SCRs discussed by Henry are licit in my dialect, those with resumptive subject pronouns are not, to the extent that as many SCRs as I have heard uttered by myself and others in my presence, I have never noticed an example with such a resumptive pronoun. Another instance of inconsistency may be the contrast in judgements between the two examples given in (22): Doherty (1993, 2000) states that (22a) is permitted in Hiberno-English, whereas Henry (1995) states that (22b) is ungrammatical in Belfast English.⁸

- (22) a. I'm looking for someone speaks Irish well.
 b. * They were looking for the girl had won the prize.

A third inconsistency appears to be the difference in whether or not SCRs are reportedly able to be paraphrased by full relative structures; Henry (1995) and Williamson (2016) suggest that this is the case, while Bailey (2019) notes a pragmatic difference between PaPRRs and their analogous majority dialect full relative structures, which I judge here to also apply to SCRs. These facts could suggest that the SCRs studied by Henry and those being reported on here are in fact two different structures; Henry's topic–comment analysis could indeed hold for SCRs in Belfast English, where topic–comment structures are otherwise common and resumptive pronouns are grammatical, but the SCRs present in other dialects could be true contact relative clauses. In this sense, both analyses may be correct with respect to the data on which they are based. This hypothesis finds potential support from the results of the dialect survey. Recall from §3.2.3 that while northern parts of Ireland showed a comparatively very high acceptance of SCRs, the same was not true of PaPRRs, which received slightly lower than average acceptance scores in this area. The crucial point here is that while the geospatial distributions of SCRs and PaPRRs were generally the same in Great Britain, they differed in Ireland. The suggestion that the topic–comment analysis could hold for SCRs in Ireland but not for those in Great Britain would explain this divergence in the correspondence between the geospatial distributions of the two structures.

With regard to the amalgam analysis, it is true that some of the examples cited by Lambrecht (1988) cannot conceivably be analysed as true relative clauses. The example in (23a) is such an example, while (23b) is a similar example that I have collected anecdotally.

⁸ This difference could be due to the difference in definiteness shown by the two nominals; it is commonly stated in the literature that relativised nominals are preferably indefinite.

- (23) a. I have one of my uncles (*who/*that) was an engineer and he told me...
 b. It's the first thing that goes through my mind (*which/*that) is...

It seems therefore that while the amalgam analysis cannot account for the SCRs discussed in this dissertation, there are many examples that can only be accounted for by such an analysis. A natural consequence of this conclusion is that many structures will exist which can successfully be analysed as either an amalgam or a non-amalgamative relative structure.

To summarise, while rejecting the topic–comment analysis and the amalgam analysis as potential candidates for the unified analysis of SCRs and PaPRRs, I have suggested that these analyses are appropriate for at least some of the data that they aim to account for, but that there is significant overlap between the strings that all three structures produce. Specifically, I have proposed to distinguish ‘Irish SCRs’ (topic–comment structures) from ‘British SCRs’ (true relative structures) based on their divergent distributional patterns, and suggested that the amalgam mechanism does apply to some structures in English which cannot be accounted for by the true relative analysis.

For the reasons above I conclude that ‘British SCRs’ are to be analysed as true contact relative clauses, and it is the details of this analysis which I will discuss in the next section.

4.1.2 *The size of SCR clauses*

I have concluded that SCRs, in British dialects at least, seem to be true biclausal relative structures, much like the object contact clauses (OCRs) available in the majority dialect. However, there has been debate surrounding the size of contact clauses in the literature. Doherty (1993, 2000) argues that both subject and object contact clauses are bare TP⁹ structures, whereas Douglas (2016) and Cinque (2020) argue that there is a C-domain present, albeit a truncated one. In this subsection I evaluate the evidence for and against these two possibilities.

Doherty (1993, 1994, 2000) argues in favour of a bare TP analysis of both OCR and SCR clauses. He observes that SCR clauses are subject to adjacency requirements such as bans on stacking¹⁰ and extraposition. He argues that this is accounted for by the TP analysis by virtue of an A'-chain derived without syntactic movement, with the external relative head directly A'-binding the internal subject gap. The adjacency restrictions follow from the Bijection Principle (Koopman & Sportiche 1982), which states that there must be a one-to-one correspondence between syntactic operators and variables. Given that the analysis supposes the existence of the operator outside the relative clause and the variable inside the relative clause, the extension of the Bijection Principle suggests that there must also be a one-to-one correspondence between matrix and contact relative clauses. Doherty cites additional evidence gleaned from adjunction diagnostics to suggest that contact clauses lack the structure required for adverbial adjunction, as suggested by his examples given in (24).

- (24) a. That's the man *(who) years ago Mary used to know well.
 b. That's the girl *(who) just yesterday was talking about you.

Douglas (2016) accounts for these adjunction restrictions by different means. He uses data such as the examples in (24) to suggest that contact clauses cannot have a C-domain the same size as those in full *wh*-relative clauses. However, he argues they do have a C-domain that is somewhat truncated. He observes that while infinitival zero-relatives seem to exhibit both A- and A'-properties, contact

⁹ IP in Doherty's terminology.

¹⁰ As is mentioned by Doherty (1993, 2000), two SCR clauses are able to stack, but the stacking of three or more clauses is banned. This possibility is attributed to McCawley's (1981, 1988) *pseudo-relative* construction.

clauses and infinitival *for*-relatives show only A'-properties. Following Longenbaugh (2016), he argues that the hybrid A-/A'-properties result from the existence of a composite probe seeking both A'- and A-related features, but that the existence of a C-domain in contact clauses (and infinitival *for*-relative clauses) would entail the existence of an A'-probe and hence explain why these structures only display A'-properties. Therefore he concludes that contact clauses have a single C-projection which hosts an A'-probe, but which still does not permit adverbial fronting, presumably due to the lack of TopP or FocP projections (Rizzi 1997).

Notably, the C-domain had not yet been split at the time of Doherty's (1993) thesis, so he was not equipped with the theoretical toolkit to come up with an analysis along such lines. However, Douglas does not explain the adjacency restrictions which Doherty observes to hold for both subject and object contact clauses. To account for this, I turn to Cinque's (2020) raising derivation of relative clauses. In this approach, the overt head raises from its base position to the left edge of the relative clause, where it is spelt out at PF. This contrasts to the matching derivation, whereby the overt head is external to the relative clause and the internally raised head is deleted under (partial) identity. Cinque states the eight key characteristics of the raising derivation to be: (i) (obligatory) reconstruction, (ii) sensitivity to strong islands, (iii) refractoriness to resumptive pronouns, (iv) possible occurrence in anti-pronominal contexts, (v) incompatibility with *wh*-pronouns, (vi) no extraposition, (vii) no stacking, and (viii) sensitivity to weak islands. Properties (iii), (v), (vi) and (vii) are immediately notable for being characteristic of the SCRs under discussion here, suggesting that the adjacency restrictions used by Doherty (1993, 2000) to argue in favour of a bare TP analysis of SCRs can be accounted for by the raising derivation. Given this conclusion, I proceed on the basis that SCRs contain relative clauses with a single projection in the C-domain whose specifier hosts the raised subject.

4.1.3 *Non-presuppositional subjects and anti-locality*

The most challenging and mysterious puzzle concerning the syntax of SCRs is the restricted set of syntactic contexts in which they are licensed, a fact frequently remarked upon but almost never tackled in the literature. In this section I propose to draw a parallel between the syntactic restrictions on SCRs and those on Complex NP Constraint violations observed in Scandinavian and other languages in order to shed light on a potential explanation for these restrictions.

As I have observed in previous sections, the contexts in which SCRs can appear grammatically include the copular existential, possessive existentials, *it*-clefts, copular sentences and *know*-complements, among certain others (see Doherty 2000:87–9). Contexts other than these result in ungrammaticality. Furthermore, Bailey (2019:5) notes that, as is true of arguments of existential verbs more generally, the relativised noun phrase in PaPRRs at least is usually indefinite, although some exceptions do exist. However, the vast majority of examples with definite nominals in my collected corpus of NSs are *it*-clefts or introduced by a demonstrative, suggesting that existential introducers may only select relative clauses headed by indefinite nominals.¹¹ Interestingly, this set of

¹¹ There are two counterexamples to this generalisation:

- (i) There's the M2 and W2 coxes are going.
- (ii) There's all the crap come out of the cupboard upstairs.

However, (i) may be analysed as an amalgam (see §4.1.1), and (ii) may have been uttered in the context of a list, so whether or not this sentence is a counterexample to the generalisation that the relativised nominal must be indefinite is unclear. Even if it is a true counterexample, this stands against the 187 indefinite examples that were attested, and as such may arguably be treated as anomalous.

contexts seems remarkably similar to those which license apparent violations of the CNPC which are found in languages such as Scandinavian (Allwood 1976, 1982; Engdahl 1997) among certain others. A contrast between English and Swedish is demonstrated in (25–28): while both the English examples (25b) and (26b) are ungrammatical, only the Swedish example in (28b) is ungrammatical, whereas (27b) (which directly corresponds with English (25b)) is grammatical.

- (25) a. There are lots of people that speak English.
 b. * Which language_i are there lots of people that speak *t_i*?
- (26) a. He talked to a woman that speaks English.
 b. * Which language_i did he talk to a woman that speaks *t_i*?
- (27) a. Det finns många som talar svenska.
 EXPL find.3SG.MID many REL talk.3PL Swedish
There are lots of people that speak Swedish.
 b. Vilket språk_i finns det många som talar *t_i*?
 which language find.3SG.MID EXPL many REL speak
Which language are there lots of people that speak (it)?
- (28) a. Han har pratat med en kvinna som talar svenska.
 he have.3SG talked with a woman REL speak.3SG Swedish
He talked to a woman that speaks Swedish.
 b. * Vilket språk_i har han pratat med en kvinna som talar *t_i*?
 which language have.3SG he talked with a woman REL speak.3SG
 Intended: *Which language did he talk to a woman that speaks (it)?*

We can see that both of the English examples and the second Swedish example all conform to the CNPC, in that (*wh*-)extraction from the complex NP is blocked. However, the first Swedish example in (27) violates the CNPC in allowing extraction of *vilket språk* (‘which language’) out of the relative clause headed by *kvinna* (‘woman’).

Cinque’s (2010b:82) characterisation of the syntactic contexts which license this kind of CNPC violation states that the head of the relative clause must be “indefinite and nonspecific”, the matrix clause verb must either be existential or of the type *know*, *see*, *meet*, *look for*, *have*, etc. (although quite what the definition of this class of verbs is Cinque leaves undiscussed), and the complex NP must be the subject of the relative clause. Particularly with regard to the first two of these criteria, there is a great deal of similarity between these contexts and those required for the licensing of SCRs. Therefore it may be revealing to pursue an analysis of the restrictions on SCRs in line with those on the CNPC violations observed in Swedish.

In accounting for CNPC violations of the same type in Hebrew, Sichel (2018) proposes that subextraction from complex NPs is only allowed if the nominal is a non-presuppositional indefinite, which means that only verbs which can take non-presuppositional arguments permit subextraction from that argument. Following Diesing’s (1992) analysis of the syntax of indefinite nominals, Sichel proposes that, while presuppositional object DPs undergo movement from their base-generated position, non-presuppositional object DPs remain *in situ*. This may happen either overtly or covertly. Crucially, the movement of a DP renders subextraction impossible—cf. Wexler & Culicover’s (1980) Freezing Principle, although Diesing attributes this restriction to Subjacency (Chomsky 1986).

Therefore in English, for instance, presuppositional object DPs appear in their base position overtly, but undergo covert movement to a higher position. It is from this higher copy that subextraction would take place, but the fact that the DP has undergone (covert) movement means subextraction is impossible. Meanwhile, non-presuppositional object DPs remain *in situ*, and so subextraction is possible (the CNPC violations are reported to be marginally acceptable by some English speakers; see Kush, Omaki & Hornstein 2013 for quantitative judgement data, albeit with an alternative analysis). In German, on the other hand, presuppositional object DPs undergo the same movement operation before spellout, so the effect of presuppositionality on the possibility of subextraction is readily observable (although the CNPC violations under discussion are not possible in German). With regard to subjects, Diesing (1992:22) suggests that presuppositional subjects are, in English at least, interpreted in SpecTP, while non-presuppositional subjects may be interpreted in SpecTP or SpecvP.¹² This finds a parallel in Boeckx's (2001) proposal that indefinites are forced to lower by the covert insertion of an expletive *there*_{LF} after spellout, giving rise to the ability of indefinites to undergo A-movement reconstruction in a way that (definite) arguments usually cannot.

While Sichel's (2018) analysis of CNPC violations makes use of the contrast between the higher position of presuppositional arguments and the lower position of non-presuppositional arguments, it does so indirectly. The crucial mechanism relied upon is the restriction that subextraction may not take place from a moved element, however that restriction is implemented theoretically. Clearly such a mechanism cannot be employed for an analysis of SCRs given that no subextraction takes place, although the similarities between the syntactic contexts of both constructions is still suggestive that the presuppositionality of the argument has something to do with licensing such constructions.

Diesing's (1992) proposals, particularly with regard to object movement, suggest that the copy from which an element is extracted for *wh*-movement or focus fronting is not necessarily the same copy as the one that is pronounced at PF. Instead, it seems that the copy from which extraction occurs is the one which is interpreted at LF, which is also the head of the movement chain which gives rise to the presence of the multiple copies. However, Boeckx (2001:519) notes that the position of interpretation cannot necessarily be equated with the highest copy in a chain, due to the observation that scope interpretations can occur with respect to copies other than those which have received Case. I would at this point like to assume that the copy from which the subextraction that occurs in the CNPC violations takes place is not the highest copy in a chain by default, but the one which is interpreted at LF, noting that in many cases these two properties may nevertheless be shared by the same copy. However, in cases where these two elements are not the same, this would have the effect that the constituent from which an element is subextracted for *wh*-movement or focus fronting is not the highest copy in the chain. There is also no *a priori* reason why subextraction for *wh*-movement and focus fronting should take place with respect to the copy which is interpreted at LF while movement for relativisation should take place with respect to the copy which is spelt out at PF, given that all of these operations are instances of A'-movement. Taken together, these considerations result in the hypothesis that relativisation of a non-presuppositional argument occurs from SpecvP, whereas that of a presuppositional argument occurs from SpecTP, and that both of these operations would take place from the copy which is interpreted at LF, as I argue to be the case for other types of A'-movement like *wh*-movement.

However, this hypothesis alone does not provide an account of why SCRs only seem to be licensed with non-presuppositional subjects. For the other half of this explanation I turn to Douglas' (2016) explanation of the restrictions on subject relativisation in contact clauses in terms of anti-

¹² I adapt Diesing's terminology from SpecIP and SpecVP to SpecTP and SpecvP for the sake of consistency within this dissertation.

locality, and specifically spec-to-spec anti-locality as conceived by Erlewine (2016, forthcoming). According to Douglas, finite zero-relatives in English (like *for*-relatives of the type *Here's an apple for you to eat*) have only one projection in the C-domain, as compared to full *wh*- and *that*-relatives which have a much more fully articulated left peripheral structure. For relativisation to occur the relativised DP must raise to the specifier of an appropriate head within the C-domain, which in *wh*- and *that*-relatives does not pose a problem for subject DPs raising from either SpecTP or SpecvP. However, due to the impoverished C-domain in zero-relative clauses, the raising of a subject DP from SpecTP to SpecCP would violate spec-to-spec anti-locality in Erlewine's (2016, forthcoming) sense, as XPs undergoing movement must cross at least one maximal projection other than the one in which they originated.

Erlewine (forthcoming) picks up on an observation by Baier (2017:368) that analyses which rely on spec-to-spec anti-locality are “fragile”, in that they are “very sensitive to minor differences in clause structure, both within a single language and cross-linguistically”. This property may be exploited by introducing small structural manipulations which would render anti-locality obsolete, in order to test the validity of such analyses. The English and Welsh examples in (29–30) neatly demonstrate this sensitivity, in that once the subject raises from a lower projection, anti-locality no longer comes into play and the resulting sentence is grammatical. In English expletive–associate constructions the subject DP overtly remains low in SpecvP, with SpecTP filled by expletive *there*. As is shown by the contrast in (29), subject extraction is permitted in expletive–associate constructions, in which the subject stays in SpecvP, but not in canonical declarative sentences, in which the subject DP is in SpecTP (examples adapted from Erlewine forthcoming).

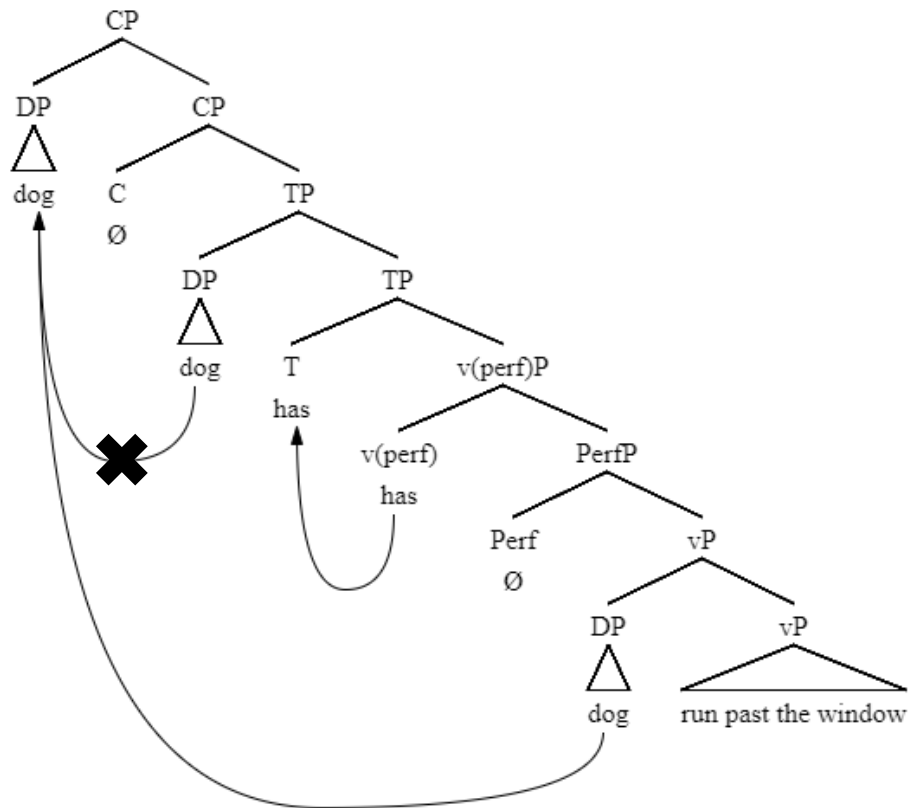
- (29) a. * What do you think [_{CP} __ that [_{TP} <what> is [_{vP} __ in the red box]]]?
 b. What do you think [_{CP} __ that [_{TP} there is [_{vP} <what> in the red box]]]?

In Welsh, whose subjects stay in a low structural position across the board, resulting in standard VSO word order, subject contact relatives are possible, as shown by the example in (30).

- (30) Oedd yna ddyn aeth i=r siop.
 was EXPL man went to=the shop
There was a man (who) went to the shop.

Given these observations, it seems reasonable to suggest that presuppositional subject DPs that need to undergo movement for relativisation from their covert position in SpecTP are prevented from doing so by spec-to-spec anti-locality, while non-presuppositional subject DPs that need to raise from the lower SpecvP position do not violate anti-locality in the process. This results in the limitations to the syntactic contexts in which SCRs may appear. This analysis of the internal structure of the relative clause is summarised in (31), with the cross indicating the movement of the higher presuppositional subject that is banned by spec-to-spec anti-locality.

- (31) There's a dog has run past the window.



A problem for this analysis may be that, while this suggestion provides an account of the restrictions when the relativised nominal is the object of the matrix verb, there is nothing to prevent the relativised subject also being the matrix subject, as in (32).

- (32) a. * [Some dogs went past the window] were barking.
 b. * There were [some dogs went past the window] barking.

This is a problem whose solution is far from obvious. However, Doherty (2000:87) identifies among the additional SCR-licensing contexts in liberal dialects (e.g. Appalachian English and Hiberno-English) ‘free-choice’ *any* and universal quantification, giving rise to the reported grammaticality of the examples in (33).

- (33) a. She ... gave me all the change was in the house.
 b. Everyone lives in the mountains has an accent all to theirself.
 c. Anyone sees you with that hole in your sweater will think you're broke.

While this suggests that SCR clauses may appear in the subject position of a matrix clause under the right conditions, many more of the examples that Doherty (2000:87–9) provides would be problematic under the current analysis. Having said this, recall the distinction I made in §4.1.1 between ‘Irish’ and ‘British’ SCRs, exemplifying the need for greater clarity with regard to which dialect examples come from and which analysis of SCRs is supposed to apply in that dialect. It may therefore be that the examples in (33) do not appear in the dialects under discussion in this

dissertation; to establish whether this is correct or not would require more detailed empirical investigation.

In this subsection I have advanced the study of SCRs by identifying the similarity of the contexts in which both SCRs and CNPC violations are permitted, and linking this to the non-presuppositionality of the element undergoing A'-movement in each case, this being reflected in the height of the syntactic position from which the element is moved. In the next subsection I turn to how an analysis along the lines sketched above may be applied to PaPRRs, so that our goal of pursuing a (partially) unified analysis of the two constructions may be fulfilled.

4.2 Analysis of PaPRRs

4.2.1 Previous analyses

Although PaPRRs have received very little attention in the literature, both in comparison to SCRs and in absolute terms, I first wish to reject an analysis put forward by Huddleston & Pullum (2002:1395f) before proceeding. They argue that the sentence in (34a) is derived from the ungrammatical monoclausal sentence in (34b) with obligatory cliticisation of *has* onto the expletive element *there*.

- (34) a. There's a man been shot.
 b. * There has a man been shot.

Bailey (2019:11–3) presents a number of arguments against this kind of analysis, which seeks to align the construction with similar presentational structures in other Germanic languages (see §1). Firstly, the supposedly obligatory cliticisation of *has* onto *there* would be a very specific rule, and even more so in the absence of more general rules of the same type involving either the element in T cliticising onto *there* in SpecTP or the more general co-occurrence of *there* + *has*. Secondly, the cliticisation of *is* and *has* only result in the same morpheme (i.e. -'s) in the simple present tense; when the construction appears in other tenses it is clear that the introducer is unambiguously a form of *be* and not *have*. Thirdly, Harwood (2013, 2015, 2017), Ramchand & Svenonius (2014) and Douglas (2016) all show that in expletive–associate constructions, the subject DP must follow the perfect auxiliary *been* and precede the progressive auxiliary *being*, which is supported by a range of empirical evidence suggesting that the subject occupies a specifier position at the edge of the clause-internal phase (CIP) (i.e. between the perfect and progressive aspectual layers); however, the structure in (34b) does not conform to this robust generalisation. Given these three fundamental problems with Huddleston & Pullum's (2002) analysis, and the fact that they only analyse one sentence in the space of a footnote, we can confidently reject this analysis as an account for PaPRRs.

Given the rejection of this analysis of PaPRRs, and also the aforementioned distributional similarities between PaPRRs and SCRs, in the next subsection I move on to examine how properties of the analysis of SCRs proposed in §4.1 may be applied to PaPRRs as well.

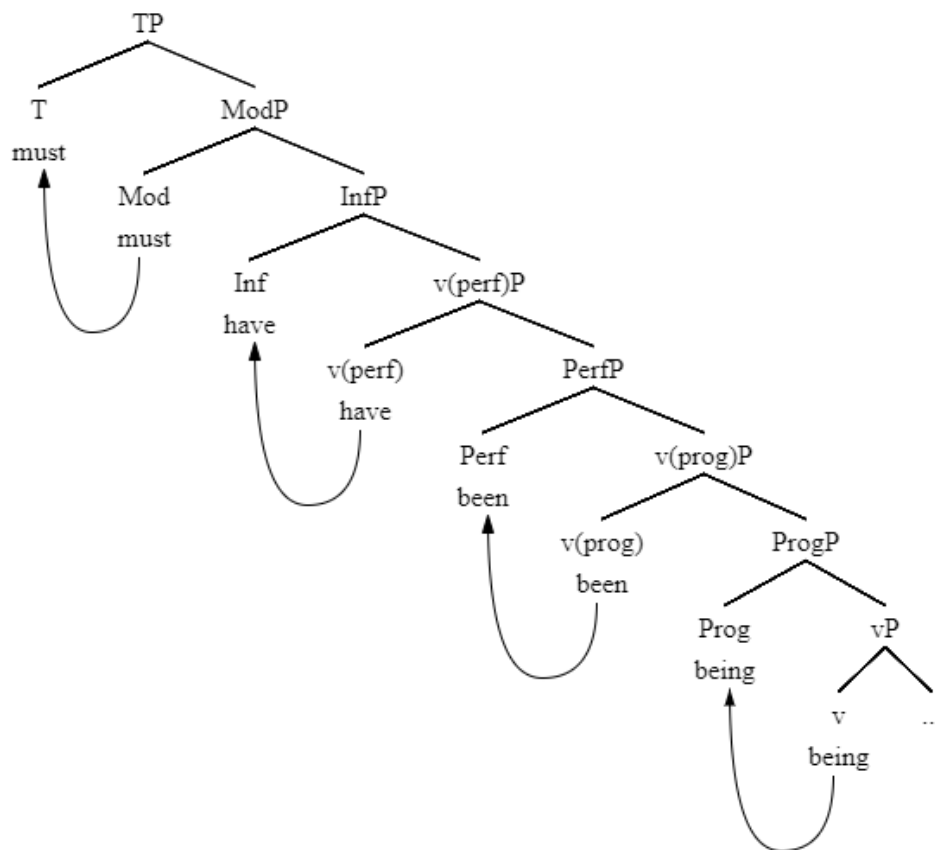
4.2.2 Extraction from PaPRRs with non-presuppositional subjects

It has been established that the results of the dialect survey—specifically (i) the strong correlation between the acceptance of SCRs and PaPRRs; and (ii) the geographical similarities in their distribution within Great Britain—make an analysis which seeks to exploit the similarities between SCRs and PaPRRs desirable, and so the analysis of PaPRRs should be highly resemblant of that of SCRs.

The crucial postulation relied upon in the account proposed for SCRs was that non-presuppositional DPs are extracted from SpecvP while presuppositional DPs are prevented from being extracted from SpecTP due to anti-locality, which, as I have noted, is sensitive to small changes in clause structure (Erlewine forthcoming). Considering the difference that presumably exists between the clause structure of SCRs and PaPRRs, however, there is no guarantee that an analysis of SCRs based on anti-locality can automatically be transferred to PaPRRs. The size and nature of PaPRR clauses is therefore fundamental to the applicability of the analysis of SCRs to PaPRRs. On these grounds, let us consider perspectives on the formation of reduced relative clauses.

I follow Harwood (2013, 2015, 2017) in assuming the structure of the English auxiliary system in (35). A paired layering approach is adopted, such that each aspectual projection has a corresponding vP shell, which serves as the base-generation site of the relevant auxiliary before it raises to agree with the higher aspectual projection. This ensures that auxiliaries do not raise to the base-generation site of another. It is also assumed that what you see is what you get (WYSIWYG), such that if aspectual projections are only present in the derivation if that aspect is contained within the sentence.

(35)



Douglas (2016) and Harwood (2017) argue that reduced relative clauses only contain structure from within the lower CIP, i.e. below and including the progressive aspectual layer (ProgP). He establishes this on the basis that auxiliaries higher than the present participle do not surface in reduced relative clauses, as shown in (36).

(36) The dog (*was) being chased was called Fido.

However, this analysis pays little attention to the mechanism by which relativisation is achieved, which is an even more serious problem when we consider Douglas' (2016:191–4) observation that there are more differences between full and reduced relative clauses than just their structural size, as suggested by the pair of examples in (37).

- (37) a. Penelope (seems/*is seeming) to be happy.
 b. The girl (*seems/seeming) to be happy is called Penelope.

The fact that a progressive reduced relative is licit in a context in which the progressive is generally deemed ungrammatical in matrix clauses suggests there are more fundamental differences between the two clause types than structural size alone.

On the other hand, Cinque (2010a, 2020) proposes that participial relative clauses are full TPs,¹³ although he suggests they lack a C-domain. He cites evidence from German that elements that usually agree in case with the subject bear nominative case in relative clauses regardless of the case assigned to the overt relative head, as shown in (38).

- (38) Wir sahen die [einer / *einen nach dem anderen angekommenen] Studenten.
 we saw the one.NOM one.ACC after the other arrived.ACC students.ACC
We saw the students (who had) arrived one after the other.

This observation, argues Cinque, following Fanselow (1986), is evidence for the presence of a PRO inside the reduced relative clause with which *einer* agrees, which suggests that such clauses may be as large as TP. However, the problem with this account is that it is unclear why overt auxiliaries (such as *waren* 'were' in the case of (38)) are absent from the relative clauses. Indeed, under this analysis reduced relative clauses do not seem to be very 'reduced' after all.

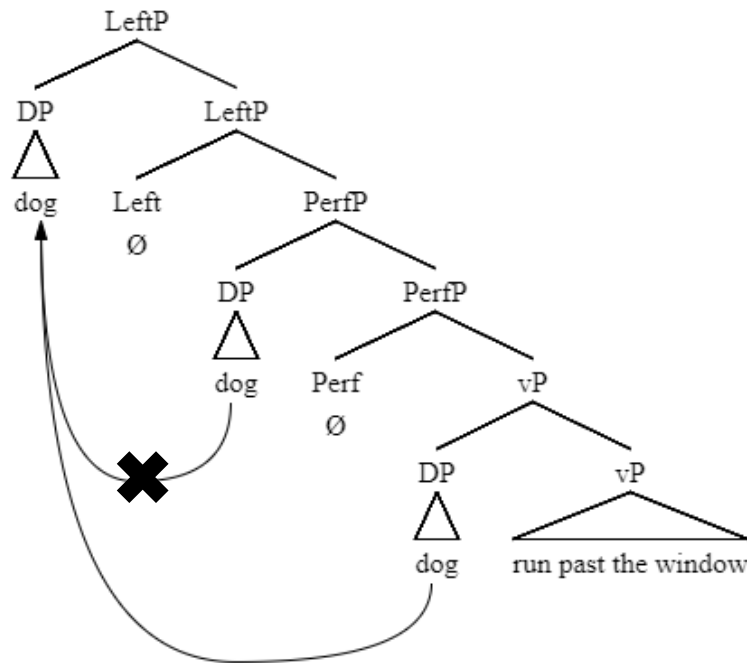
The difference between the reduced relative clauses analysed by Douglas and Cinque and PaPRRs is that the spellout domain of the former type is exclusively within the CIP, whereas this is not true of the latter type. As such, I refer to the former as 'phasal' reduced relatives. Furthermore, phasal reduced relatives are not subject to any distributional restrictions and are fully grammatical in the majority dialect, whereas this is not the case for PaPRR clauses. This means that even if there was a consensus on the size of phasal reduced relatives, it is not clear that this would necessarily be reflected in the size of PaPRR clauses.

Douglas (2016:227–9) considers the possible existence of a clause-medial left periphery at the edge of the CIP, in order to account for the acceptability of high adverbs in reduced relatives. This is an idea that has been pursued in the context of a range of domains (Watters 1979; Horvath 1986; Laka 1990; Travis 1991; Tuller 1992; Ndayiragije 1999; Belletti 2001, 2004; Jayaseelan 2001; Duffield 2007, 2013; Batllori & Hernanz 2013; Kandybowicz 2013; Wiltschko 2014; Biberauer 2017). If this is on track, then it would appear that all (verbal) phases might have some kind of left periphery, and the left periphery would become a property of the phase as opposed to of the clause. Although PaPRR clauses are not determined by the size of the CIP, they are still arguably biphasal structures, with a clause-internal phase extending as high as the progressive layer and the lower portion of what would otherwise be the clausal phase containing the rest of the clause; it seems unlikely that the higher portion of the relative clause would form a part of the phase headed by an element in the matrix clause. If the assumption that left peripheries are properties of phases is correct—and it is just that: an assumption—then we might postulate the existence of a left periphery at the edge of the phase

¹³ I use TP for the sake of consistency within this dissertation, but Cinque uses the label 'IP'.

which contains the participle of the PaPRR clause. If this left periphery is what the single C-projection postulated to be present in SCRs constitutes,¹⁴ then perhaps this left periphery need only consist of a single projection at the edge of PaPRR clauses too. I will refer to this single projection as LeftP for disambiguation. Given that, in Harwood’s (2013, 2015, 2017) paired layering approach, Perf is analogous to T in that it is the head to which lower auxiliaries raise for feature checking purposes as opposed to the head in which an auxiliary is base-generated before it raises, we might assume that SpecPerfP is an appropriate position to host (presuppositional) subject DPs. As such, anti-locality would prevent the extraction of presuppositional DPs to SpecLeftP from SpecPerfP, in a similar fashion to the way in which it limits the extraction of presuppositional DPs from SCRs. Meanwhile, non-presuppositional DPs are allowed to raise unimpeded from SpecProgP (or perhaps the specifier of the left-peripheral projection in the CIP) to SpecLeftP, resulting in the same asymmetry that we see in SCRs. This analysis of the relative clause is summarised in (39).

(39) There’s a [dog run past the window].



This analysis raises interesting issues, particularly the possibility of a left periphery at the edge of every phase, no matter its size, and whether or not the subject would (have any reason to) raise to SpecPerfP, given that the motivation for movement to SpecTP in terms of (nominative) case assignment probably isn’t applicable for SpecPerfP. If a presuppositional subject can raise to SpecPerfP if PerfP is the highest projection in the embedded clause (apart from the hypothesised LeftP) then it seems logical to suggest it could also raise to SpecInfP when this is the highest projection in the embedded clause, allowing for the grammaticality of InfPRRs mentioned in §2.1.

Also worth discussing is why phasal reduced relatives are *not* subject to the syntactic restrictions that apply to SCRs and PaPRRs. I suggest that the reason the subject occupies a specifier at the left edge of the clause-internal phase in the first place is because it must raise here in order to be accessible for subsequent movement operations once the lower phase has been spelt out, rendering it

¹⁴ It would be unclear why infinitival zero-relatives lack a C-domain completely if a (reduced) left periphery was present at the edge of every phase.

inaccessible to the computational system. Until this point, however, I suggest that no syntactic operations apply differently to presuppositional and non-presuppositional subjects, which means there are no distributional restrictions on phasal reduced relatives.

Having provided a unified analysis of SCRs and PaPRRs, as the results of the survey suggested was necessary, I turn in the next section to the analysis of EBICs, which the results suggested had a greater degree of independence in terms of the nature of their syntactic analysis.

5 Analysis of EBICs

The expletive bare infinitival construction (EBIC), repeated in (40), shows a number of properties that distinguish it from both SCRs and PaPRRs. This is supported by the findings of the survey that EBICs correlate less strongly in several ways with SCRs and PaPRRs than the latter two do with each other. From a descriptive perspective too, both SCRs and PaPRRs are evidently both forms of relative clause (see §4), whereas an analysis of EBICs seems to present more obstacles to overcome.

(40) There's been a dog run past the window.

A notable property of EBICs is that they do not readily seem to appear in simple tenses, as shown below.¹⁵

- (41) a. * There is a dog run past the window.
b. ? There was a dog run past the window.

Regarding introducers other than the existential copula, I also briefly note that EBICs with *have*-existential introducers are in fact available in the majority dialect, being commonly referred to as experiential *have* constructions. In addition, the interpretation that EBICs with *know*-introducers receive is radically different to that of NRSs with *know*-introducers. *Know*-introducers with NRSs introduce an individual into the discourse and assign a property to them, as noted by Henry (1995:125). However, those with EBICs convey that the speaker knows of an event that has taken place, and not the individual themselves. This difference suggests that the restricted set of syntactic contexts which seems to be common to EBICs and NRSs is in fact determined by different factors in each group of structures, and this is reinforced by the fact that verbs like *meet* and *invent* do not license EBICs, as shown below, despite Henry's (1995:125) inclusion of these verbs which "introduce an individual into the discourse" in the same class as *know*.

- (42) a. I've known a man go past the window.
b. * I've met a man go past the window.
c. * I've invented a robot go past the window.

These observations all support the viewpoint that fundamental differences exist between NRSs and EBICs, and that the analysis that unites SCRs and PaPRRs should not also be extended to EBICs.

A successful syntactic analysis of the EBIC should aim to take into account: (i) the suggestion by the survey that EBICs are less closely tied to SCRs and PaPRRs, as just mentioned; (ii) the degraded

¹⁵ This judgement is given by the author. Notably, in a survey I previously conducted (Bailey 2019), 20.5% of respondents accepted a sentence structurally identical to (41a), as compared to the 30.8% who accepted a present perfect counterpart. This is a point to which I return later.

acceptability of EBICs in the simple present and simple past tenses; (iii) the lack of any progressive auxiliaries being able to appear in the existential introducer; and perhaps (iv) the surface similarity shown by EBICs to other bare infinitival constructions in English. I will analyse EBICs as non-progressive structural counterparts to the expletive–associate constructions available in the majority dialect, suggesting that an agreement relation between the lexical verb and the closest verbal head that c-commands it is banned, which results in the insertion of a dummy *be* auxiliary to satisfy feature checking requirements.

I will begin the discussion of EBICs by observing that expletive–associate constructions in the majority dialect are only licit if they contain the progressive aspect, and also that the introducers of EBICs cannot contain progressive aspect, as demonstrated in (43–44).

- | | | | |
|------|----|--|--|
| (43) | a. | There is a dog running past the window. | licit progressive E–A construction |
| | b. | * There has a dog run past the window. | illicit non-progressive E–A construction |
| (44) | a. | There’s been a dog run past the window. | licit non-progressive EBIC |
| | b. | * There’s being a dog run past the window. | illicit progressive EBIC |

Therefore, in order to express progressive aspect in an EBIC, the progressive morphology must appear below the subject DP. This results in string equivalence between progressive EBICs and the expletive–associate constructions found in the majority dialect and commonly assumed to be derived as a monoclausal construction. Considering that the expletive–associate construction is only available with progressive aspect in the majority dialect, the potential for a connection between this construction and existential BICs warrants further investigation.

Harwood (2013, 2015, 2017), Ramchand & Svenonius (2014) and Douglas (2016) all use evidence from expletive–associate constructions to diagnose the size of the clause-internal phase (CIP), such that the CIP boundary corresponds with the progressive aspectual layer. The crucial observation is the position of the subject with respect to the aspectual auxiliaries. In a sentence which includes all auxiliaries in the English aspectual system, we can pinpoint the precise location of the subject with respect to each aspectual projection, which these authors take to be the position of the CIP boundary, under the assumption that the specifier of the phase head is used as an escape hatch for the subject to raise from its base-generated position in the lower phase to its surface position in the higher phase, at least in the absence of an expletive element in SpecTP. The diagnostics are shown in (45), with the subject underlined.

- | | | |
|------|----|---|
| (45) | a. | * There must have been being screened <u>a film</u> . |
| | b. | * There must have been being <u>a film</u> screened. |
| | c. | There must have been <u>a film</u> being screened. |
| | d. | * There must have <u>a film</u> been being screened. |
| | e. | * There must <u>a film</u> have been being screened. |
| | f. | * There <u>a film</u> must have been being screened. (with the intended interpretation) |

Therefore, in terms of the approach to the aspectual hierarchy proposed by Harwood (2013, 2015, 2017) and adopted by Douglas (2016), we may identify the subject in expletive–associate constructions to appear in SpecProgP.

It is apparent that this conclusion also holds of EBICs. Although the EBICs exemplified so far in this dissertation do not contain progressive aspect, I note that this may be because if progressive aspect were contained within the sentences, they would arguably be string equivalent—or perhaps structurally equivalent—to the expletive–associate constructions like (45c) available in the majority

dialect. The examples in (46) suggest are not incompatible with the hypothesis that in these constructions too the subject appears in SpecProgP.

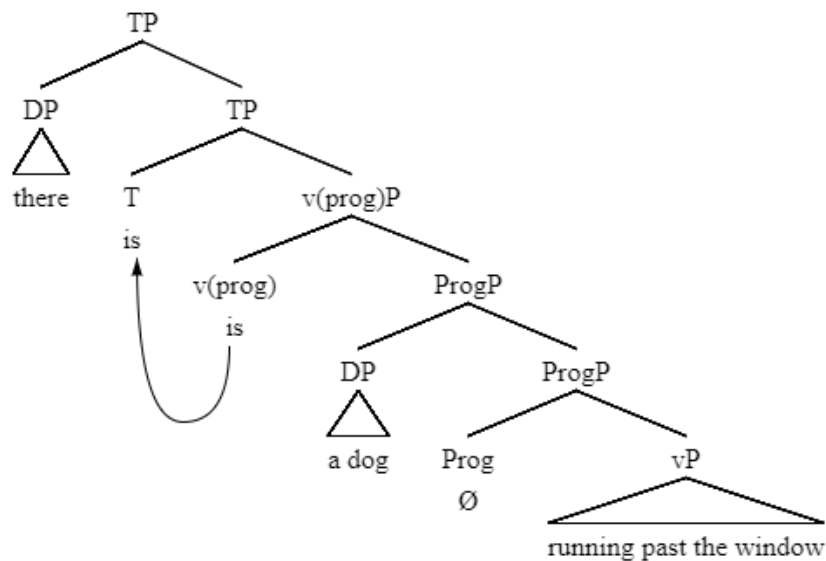
- (46) a. * There must have been screened a film.
 b. There must have been a film screened.
 c. * There must have a film been screened.
 d. * There must a film have been screened.
 e. * There a film must have been screened. (with the intended interpretation)

There are some conceptual advantages to analysing EBICs and expletive–associate constructions as the progressive and non-progressive sides of the same coin, primarily that if this analysis is accurate each structure fills the gap in the paradigm of the other. There seems to be no reason why EBICs should not be able to deal with progressive sentences, just as there seems to be no reason why expletive–associate constructions should only be able to deal with progressive sentences. That is, it is unclear why (47a) should be allowed while (47b) is ruled out, given the derivational similarity between the two.

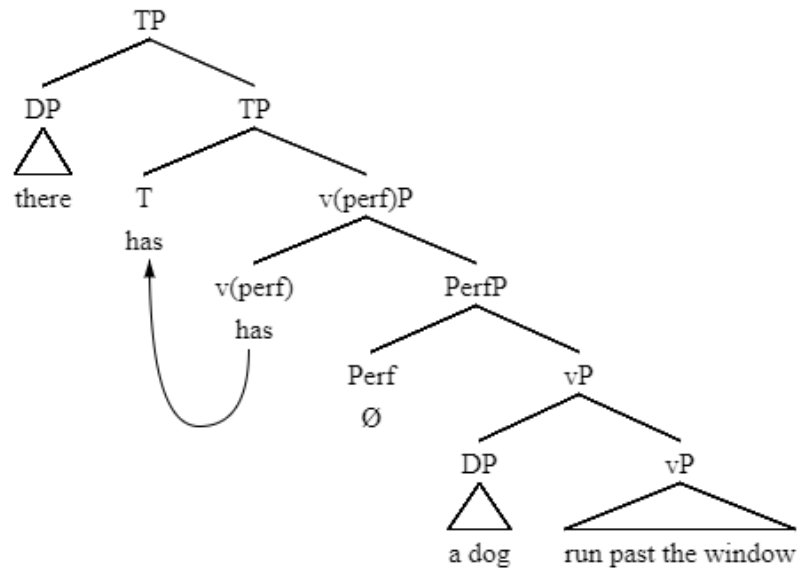
- (47) a. There is a dog running past the window.
 b. * There has a dog run past the window.

Simplified derivations of (47a–b) are given in (48a–b) respectively.

- (48) a.

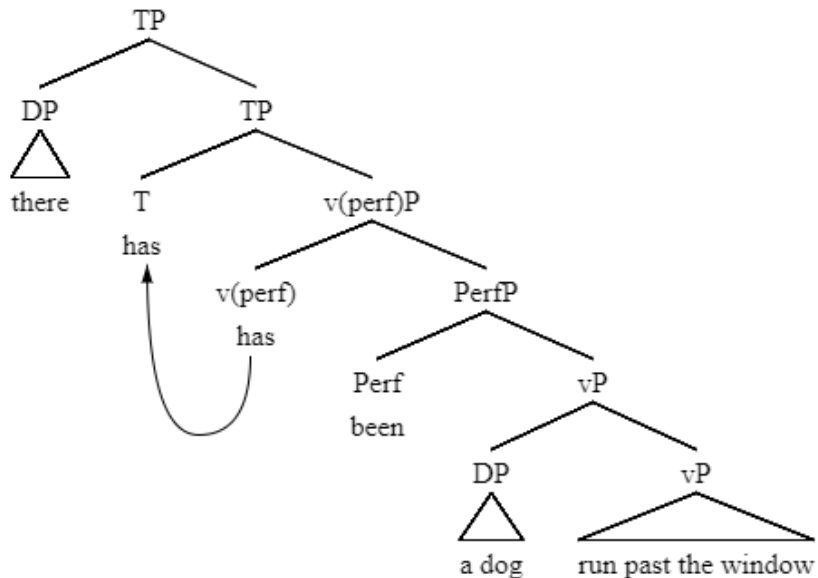


b.



Based on the observations we can make about the differences between the derivations in (48), we might hypothesise that something about the presence of the subject in an intervening position between Perf and *v* blocks the lowering of the perfective morphology to the lexical verb, about which it is well established that it cannot raise out of the extended verbal projection of English, following the loss of V-to-T movement in Early Modern English. However, it seems by inserting a dummy auxiliary *be* in Perf so that the perfective morphology may be spelt out, we arrive at a structure that is string equivalent to an EBIC, as shown in (49).

(49)



Dummy auxiliary insertion is by no means unknown to English, being manifest in *do*-support. In that case, Biberauer & Roberts (2009) argue that the insertion of a dummy auxiliary *do* in T—which happens in negative, interrogative or emphatic sentences, or those with locative or negative inversion, ellipsis or VP-fronting—occurs when T bears a superset of the features that are borne by V.

Specifically, the extra feature borne by T is [+AFF(ECT)], the presence of which results in the discourse effect associated with *do*-support contexts. Auxiliary *do* bears such an [AFF] feature, as well as the usual T-features such as [uV], [uΦ] and [iTENSE], but the null T-element which appears in unemphatic affirmative declarative sentences only bears the usual T-features and not [AFF]. In this case, the Subset Principle rules out the selection of the null T-element on the grounds that it an option is available with a greater level of specification, i.e. auxiliary *do*, and therefore *do* is inserted in such contexts.

Extending this to the insertion of auxiliary *be* in EBICs, we may want to assess the possibility of the existence of an extra feature on Perf which prevents the selection of an underspecified null Perf-element and hence the derivation of sentences such as those in (47b). I will refer to this hypothesised feature as [NAR(RATIVE)]. By hypothesis, this feature would exist in EBICs, and would force the insertion of auxiliary *be*, which would bear [NAR]. Those speakers in whose grammars [NAR] was not present would not accept EBICs as grammatical structures, accounting for the variation attested. In terms of EBICs without perfective aspect, it could be that [NAR] is borne by heads higher than Perf too. Recall that my personal judgement is such that I reject simple tense EBICs, but this is not the picture reflected in the data collected Bailey (2019), in which a simple present EBIC received an acceptance score of 20.5%, compared to the 30.8% given to a perfective EBIC and the 4.6% given to an ungrammatical control stimulus. The judgement that a simple tense EBIC is less acceptable than a perfective EBIC may suggest that fewer speakers' grammars can have [NAR] on T than on Perf.

An alternative proposal may be made in terms of Relativised Minimality (Rizzi 1990), to the effect that some head bearing properties shared by *v* and Perf intervenes between the two, blocking the agreement of the two verbal heads (presumably by the mechanism of affix lowering, which I assume to be commonly employed in majority dialect constructions). A potential problem for both of these types of analysis, however, concerns the use of the *be* auxiliary in simple tense contexts. In these cases, it is unclear why *do* would not be inserted, given that this is the auxiliary usually selected for insertion in T. One possible answer to this would be that EBICs have a more inherent connection to the copula–predicate relationship than the contexts in which *do*-support appears. Perhaps the lower phase/event domain/predicational layer has a greater separation from the higher phase, in the form of some kind of predicational property that is shared by EBICs and expletive–associate constructions but not declarative sentences in which *do*-support may be licensed.

In summary, I have hypothesised that EBICs and expletive–associate constructions are derived by the same mechanism, but that aspectual heads above the progressive layer are blocked from interacting with the lexical verb, which forces the insertion of a dummy *be* auxiliary in non-progressive constructions.

6 The Suppressed Grammaticality Hypothesis

The analyses presented in §4 and §5 detailed how the NSs may be licensed in dialect speakers' grammars. In this section, I wish to account for the gradient nature of the attested interspeaker variation by developing the Suppressed Grammaticality Hypothesis.

Barbiers (2005) analyses dialectal variation regarding word order in three-verb clusters in Dutch. Out of the six logically possible orders in which three elements can appear, two are never attested, and these he rules out as universally unavailable for syntactic reasons. However, the acceptance of the remaining four possible orders by Dutch speakers varies on a geographic basis, with most speakers reporting only two or sometimes three of the four orders to be acceptable. He further observes that speakers who accept a higher number of structures tend to reside at or near a linguistic boundary between two areas which generally accept different orders, and as such that the areas where a higher number of orders is accepted are transitional zones between dialect areas.

Based on this observation, Barbiers formulates the hypothesis that “[a]ll orders that the grammatical system allows are, in principle, available for each speaker in the Dutch language area, but which orders a speaker actually uses or reports to occur in [their] dialect depends on the input from the environment” (Barbiers 2005:255). He further states that “[t]his is a specific instantiation of the general generative hypothesis that universal grammar provides a large number of options from which a restricted set is selected during the process of language acquisition on the basis of actual input.”

I would briefly like to consider whether this is a hypothesis that can also apply to NSs, given the nature of interspeaker variation in terms of the structures’ acceptance. The key question to answer is what governs whether or not a speaker accepts the NSs, and to what extent. There are two possible—though not mutually exclusive—mechanisms which could regulate acceptability of structures. The first and more obvious governs whether or not a structure is available in principle to a speaker, and, assuming the Borer–Chomsky Conjecture that parametric variation is reducible to featural differences between items in the lexicon (Baker 2008), this is presumably dependent on whether or not a certain feature is present in the grammar. If this mechanism dictates whether speakers accept or reject NSs, we might expect that the structures are either ‘in’ or ‘out’ of the speaker’s grammar, regardless of the exposure they receive to the structures over a short period of time, and, particularly if this was the only mechanism responsible for interspeaker variation, that structure of all complexities would be judged equally, as long as processing demands did not become too high (although see §3.2.6 for why judgement data is not always reliable in this respect). This may also lead to difficulties in the interpretation of the structures for those who did not have them in their grammar, and a prediction might be that if the feature/structure was acquired in later life by a speaker in whose grammar it was not formerly present,¹⁶ the structure may become readily accessible to them with a discrete change.

The alternative hypothesis, which is what I refer to as the Suppressed Grammaticality Hypothesis (SGH) is along the lines of Barbiers’ (2005) proposal about Dutch dialects, i.e. that the syntactic nature of the grammar (in terms of the featural properties of lexical items) would in principle make the structures available for many more speakers than the number that report them to be acceptable, and the extent to which a speaker accepts such structures depends on their exposure to them in the input. What induces a speaker for whom the structures are available but not readily acceptable to have the structures in principle may be determined by other factors in the input, and direct evidence may not be necessary for the child acquirer to postulate their existence nonetheless (cf. the ‘poverty of the stimulus’ argument developed by Chomsky (1965, 1980)). This hypothesis might lead us to expect that acceptability judgements form a continuous scale instead of discrete ‘in’ or ‘out’ judgements (though there may be other reasons for judgements to vary in this way); that a change in the frequency of the structures in the input may lead to a change in their acceptance by a speaker; and that less complex constructions of the same type may be more readily accepted by speakers than their more complex counterparts (*modulo* the discussion in §3.2.6), based on the probability that more complex structures are likely to occur less frequently in the input.

Importantly, the featural variation hypothesis and the SGH are not mutually exclusive, and indeed I would like to propose that both factors may be at play in regulating the acceptance of NSs by speakers of English. Gradient variation may be due to both featural variation between speakers (such that the featural makeup of some speakers’ grammars licenses NSs whereas that of others’ grammars do not, however this featural variation may be determined) and suppressed grammaticality leading to

¹⁶ Although the idea of grammatical change in L1 does not conform to traditional generative ideas on language acquisition, the nature of such change is the subject of an emerging body of generative literature (see e.g. Hicks & Domínguez 2020). However, this discussion is clearly far beyond the scope of the present work.

marginal judgements within a single speaker. In other words, some speakers do not have NSs in their grammar, while others have the capability to have NSs but the degree to which they accept them varies according to the SGH. The combined effect of these two mechanisms of variation is the gradient effect we see across the Anglo-Celtic Isles, although the contribution of each mechanism cannot easily be isolated from the other.

7 Conclusion

Given the absence of any previous detailed literature on PaPRRs or EBICs, this dissertation represents an introduction of these two structures to the generative literature, hopefully with the effect that more future work will take them into consideration. The potential these structures have to shed light on the issues I have discussed—relativisation, truncation and presentationality as well as the contrast between the syntax of standard/written and colloquial/spoken varieties being among the most fundamental—is something I hope others will see fit to exploit.

I started by presenting the results of the large-scale dialect survey I conducted, and inferring from them that SCRs and PaPRRs should be given similar syntactic analyses given their distributional similarities. Meanwhile, the results suggested that EBICs were less closely connected to the former two structures, and that phasal structures like SPRs were less closely connected still. In addition, the difference between the distribution of SCRs and PaPRRs in northern Ireland relative to Great Britain lead me to hypothesise that the SCRs found on each island may be structurally different and therefore only weakly equivalent, with particularly the topic–comment structure available in northern Ireland able to generate sentences which are not grammatical in Great Britain. In these ways, the results of the dialect survey have had a direct influence on the nature of the syntactic analyses presented, both in terms of defining factors which it is necessary for the analysis to take into account or appropriate for it to ignore and forcing the adoption of a finer-grained typology of SCR-like structures available in the Anglo-Celtic Isles.

I proposed that the lower position from which non-presuppositional DPs are extracted for relativisation in SCRs is sufficiently far away from the landing site of the movement operation so as to avoid a violation of spec-to-spec anti-locality, whereas this is not the case for presuppositional DPs, which are obligatorily extracted from a higher position. The proposal that only non-presuppositional DPs may licitly be extracted accounts for the restricted syntactic distribution of SCRs, in that only matrix verbs which can select for a non-presuppositional object are able to serve as SCR introducers. This is a parallel to the CNPC violations observed in Scandinavian, Hebrew and other languages: only non-presuppositional DPs are available for subextraction, and this results in the limited syntactic contexts in which such violations are possible. I extended this analysis to PaPRRs, suggesting that the once the subject has raised from its base-generated position to the left edge of the clause-internal phase, it then obeys the same rules as in full clauses: presuppositional DPs raise to a higher position, whereas non-presuppositional DPs remain in the lower specifier at the edge of the clause-internal phase. Assuming the presence of a single left-peripheral projection at the edge of the higher phase in PaPRR clauses, this results in the constraints imposed by anti-locality as we also observe in SCRs.

Turning to the analysis of EBICs, I proposed that these structures have a structural connection to the expletive–associate constructions available in the majority dialect, given that each structure seems to fill the other’s paradigmatic gap, and that aspectual heads in the clausal phase are blocked from interacting with the lexical verb, and so a dummy auxiliary *be* is inserted to satisfy feature checking requirements. I argued that SCRs and PaPRRs should be given a similar analysis to account for their distributional similarities, while pursuing a different analysis for EBICs.

Despite the successes of the analyses presented, there are aspects of the structures that have not been captured in this work. Questions about why PaPRRs are more widely accepted than SCRs, why

the NSs seem to be banned from some speakers' grammars when they are licensed in others', and the nature of the relationship between EBICs and SCRs/PaPRRs that yields the relatively strong correlation observed between the two groups of structures still remain unsolved. However, this thesis may lay the foundations for future analyses, such that insights into the unanswered questions may draw on the data presented here.

References

- Aelbrecht, Lobke & William Harwood (2015). To be or not to be elided: VP ellipsis revisited. *Lingua* 153, 66–97.
- Allwood, Jens (1976). The Complex NP Constraint as a non-universal rule and some semantic factors influencing the acceptability of Swedish sentences which violate the CNPC. *University of Massachusetts Occasional Papers in Linguistics* 2, 1–20.
- Allwood, Jens (1982). The Complex NP Constraint in Swedish. In: Elisabet Engdahl & Eva Ejerhed (eds.), *Readings on Unbounded Dependencies in Scandinavian Languages*, 15–32. Stockholm: Almqvist & Wiksell.
- Baier, Nico (2017). Antilocality and antiagreement. *Linguistic Inquiry* 48, 367–377.
- Bailey, Jamie (2019). *There's a Verb Gone Missing: Past Participle Reduced Relatives in Colloquial English Narrative Structures*. B.A. dissertation, University of Cambridge.
- Baker, Mark (2008). The macroparameter in a microparametric world. In: Theresa Biberauer (ed.), *The Limits of Syntactic Variation*, 351–374. Amsterdam: John Benjamins.
- Barbiers, Sjef (2005). Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics. In: Leonie Cornips & Karen P. Corrigan (eds.), *Syntax and Variation: Reconciling the Biological and the Social*, 233–264. Amsterdam: John Benjamins.
- Batllo, Montserrat & Maria-Lluïsa Hernanz (2013). Emphatic polarity particles in Spanish and Catalan. *Lingua* 128, 9–30.
- Belletti, Adriana (2001). Inversion as focalization. In: A. Hulk & J. Y. Pollock (eds.), *Subject Inversion in Romance and the Theory of Universal Grammar*, 60–90. Oxford: Oxford University Press.
- Belletti, Adriana (2004). Aspects of the low IP area. In: Luigi Rizzi (ed.), *The Structure of CP and IP: The Cartography of Syntactic Structures* (vol. II), 16–51. Oxford: Oxford University Press.
- Biberauer, Theresa (2017). The Final-over-Final Condition and particles. In: Michelle Sheehan, Theresa Biberauer, Ian Roberts & Anders Holmberg (eds.), *The Final-over-Final Condition: A Syntactic Universal*, 187–296. Cambridge, MA: MIT Press.
- Biberauer, Theresa & Ian Roberts (2009). Subjects, tense and verb-movement. In: Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric Variation: Null Subjects in Minimalist Theory*, 263–302. Cambridge: Cambridge University Press.
- Block, David (2015). Social class in applied linguistics. *Annual Review of Applied Linguistics* 35, 1–19.
- Boeckx, Cedric (2001). Scope reconstruction and A-movement. *Natural Language & Linguistic Theory* 19(3), 503–548.
- Bošković, Željko (2014). Now I'm a phase, now I'm not a phase: On the variability of phases with extraction and ellipsis. *Linguistic Inquiry* 45(1), 27–89.
- Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam (1980). On cognitive structures and their development: A reply to Piaget. In: M. Piattelli-Palmarini (ed.), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.
- Chomsky, Noam (1986). *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, Noam (1999). Derivation by phase. *MIT Occasional Papers in Linguistics* 18.
- Cinque, Guglielmo (2010a). *The Syntax of Adjectives: A Comparative Study*. Cambridge, MA: MIT Press.

- Cinque, Guglielmo (2010b). On a selective “violation” of the Complex NP Constraint. In: J.-W. Zwart & M. de Vries (eds.), *Structure Preserved: Studies in Syntax for Jan Koster*, 81–90. Amsterdam: John Benjamins.
- Cinque, Guglielmo (2020). *The Syntax of Relative Clauses: A Unified Analysis*. Cambridge: Cambridge University Press.
- Cowart, Wayne (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage.
- den Dikken, Marcel (2005). A comment on the topic of topic-comment. *Lingua* 115, 691–710.
- Diesing, Molly (1992). *Indefinites*. Cambridge, MA: MIT Press.
- Divjak, Dagmar (2017). The role of lexical frequency in the acceptability of syntactic variants: Evidence from that-clauses in Polish. *Cognitive Science* 41, 354–382.
- Doherty, Cathal (1993). *Clauses Without That: The Case for Bare Sentential Complementation in English*. Ph.D. dissertation, University of California, Santa Cruz.
- Doherty, Cathal (1994). The syntax of subject contact relatives. *Proceedings of CLS 29*, 55–65. Chicago: Chicago Linguistic Society.
- Doherty, Cathal (2000). *Clauses Without ‘That’: The Case for Bare Sentential Complementation in English*. New York, NY: Garland/Routledge.
- Donnelly, Ciar (2016). *Analytic Causatives: The Special Case of ‘Have’*. B.A. dissertation, University of Cambridge.
- Douglas, Jamie (2016). *The Syntactic Structures of Relativisation*. Ph.D. dissertation, University of Cambridge.
- Duffield, Nigel (2007). Aspects of Vietnamese clausal structure: Separating tense from assertion. *Linguistics* 45(4), 765–814.
- Duffield, Nigel (2013). On polarity emphasis, assertion and mood in Vietnamese and English. *Lingua* 137, 248–270.
- Emonds, Joseph (1976). *A Transformational Approach to English Syntax*. New York: Academic Press.
- Engdahl, Elisabet (1997). Relative clause extractions in context. *Working Papers in Scandinavian Syntax* 60, 51–79.
- Erdmann, Peter (1980). On the history of subject-contact clauses in English. *Folia Linguistica Historica* I/1, 139–170.
- Erlewine, Michael (2016). Anti-locality and optimality in Kaqchikel agent focus. *Natural Language & Linguistic Theory* 34(2), 429–479.
- Erlewine, Michael (forthcoming). Anti-locality and subject extraction. To appear in *Glossa*.
- Fanselow, Gisbert (1986). On the sentential nature of prenominal adjectives in German. *Folia Linguistica* 20, 341–380.
- Gibson, Edward (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68, 1–76.
- Haegeman, Liliane (2012). *Adverbial Clauses, Main Clause Phenomena and the Composition of the Left Periphery*. Oxford: Oxford University Press.
- Haegeman, Liliane (2015). A note on English subject contact relatives. *MIT Working Papers in Linguistics* 77, 133–146.
- Haegeman, Liliane, Andrew Weir, Lieven Danckaert, Tijs D’Hulster & Liisa Buelens (2015). Against the root analysis of subject contact relatives in English. *Lingua* 163, 61–74.
- Harris, Martin & Nigel Vincent (1980). On zero relatives. *Linguistic Inquiry* 11(5), 805–807.
- Harwood, William (2013). *Being Progressive is Just a Phase: Dividing the Functional Hierarchy*. Ph.D. dissertation, University of Ghent.

- Harwood, William (2015). Being progressive is just a phase: celebrating the uniqueness of progressive aspect under a phase-based analysis. *Natural Language & Linguistic Theory* 33(3), 523–573.
- Harwood, William (2017). Reduced relatives and extended phases: A phase-based analysis of the inflectional restrictions on English reduced relative clauses. *Studia Linguistica* 72(3), 428–471.
- Henry, Alison (1995). *Belfast English and Standard English*. Oxford: Oxford University Press.
- Hicks, Glyn & Laura Domínguez (2020). A model for L1 grammatical attrition. *Second Language Research* 36(2), 143–165.
- Hooper, John & Sandra Thompson (1973). On the applicability of root transformations. *Linguistic Inquiry* 4, 465–497.
- Horvath, Julia (1986). *FOCUS in the Theory of Grammar and the Syntax of Hungarian*. Dordrecht: Foris.
- Huddleston, Rodney & Geoffrey K. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Irwin, Patricia (2018). Existential unaccusativity and new discourse referents. *Glossa* 3(1), 24. DOI: <http://doi.org/10.5334/gjgl.283>
- Jayaseelan, Karattuparambil (2001). IP-internal topic and focus phrases. *Studia Linguistica* 55(1), 39–75.
- Jespersen, Otto (1909–40). *A Modern English Grammar on Historical Principles* (vol. III). Heidelberg: Carl Winters Universitätsbuchhandlung.
- Kandybowicz, Jason (2013). Ways of emphatic scope-taking: From emphatic assertion in Nupe to the grammar of emphasis. *Lingua* 128, 51–71.
- Kluck, Marina (2011). *Sentence Amalgamation*. Ph.D. dissertation, Utrecht University.
- Koopman, Hilda & Dominique Sportiche (1982). Variables and the Bijection Principle. *The Linguistic Review* 2, 139–160.
- Kush, Dave, Akira Omaki & Norbert Hornstein (2013). Microvariation in islands? In: Jon Sprouse & Norbert Hornstein (eds.), *Experimental Syntax and Island Effects*, 239–264. Cambridge: Cambridge University Press.
- Laka, Itziar (1990). *Negation in Syntax: On the Nature of Functional Categories and Projections*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Lambrecht, Knud (1988). There was a farmer had a dog: syntactic amalgams revisited. *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 319–339.
- Lambrecht, Knud (2000). When subjects behave like objects: An analysis of the merging of S and O in sentence-focus constructions across languages. *Studies in Language* 24, 611–682.
- Longenbaugh, Nicholas (2016). *Difficult Movement*. Paper presented at GLOW 39, April 5–8, University of Göttingen.
- Lonzi, Lidia (1986). Pertinenza della struttura tema–rema per l’analisi sintattica. In: H. Stammerjohann (ed.), *Theme–Rheme in Italian*, 99–120. Tübingen: Narr.
- McCawley, James (1981). The syntax and semantics of English relative clauses. *Lingua* 53, 99–149.
- McCawley, James (1988). *The Syntactic Phenomena of English* (vol. II). Chicago: University of Chicago Press.
- Montgomery, Michael (2006). The morphology and syntax of Ulster Scots. *English World-Wide* 27, 295–329.
- Nagata, Hiroshi (1988). The relativity of linguistic intuition: The effect of repetition on grammaticality judgements. *Journal of Psycholinguistic Research* 17(1), 1–17.
- Nagucka, Ruta (1980). Grammatical peculiarities of the contact-clause in Early Modern English. *Folia Linguistica Historica* I/1, 171–184.
- Ndayiragije, Juvénal (1999). Checking economy. *Linguistic Inquiry* 30(3), 399–444.

- Prince, Ellen (1981). Toward a taxonomy of given-new information. In: Peter Cole (ed.), *Radical Pragmatics*, 223–254.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey N. Leech & Jan Svartvik (1985). *A Grammar of Contemporary English*. London: Longman.
- Ramchand, Gillian & Peter Svenonius (2014). Deriving the functional hierarchy. *Language Sciences* 46, 152–174.
- Rizzi, Luigi (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.
- Rizzi, Luigi (1997). The fine structure of the left periphery. In: Liliane Haegeman (ed.), *Elements of Grammar*, 281–337. Dordrecht: Kluwer Academic Publishers.
- Ross, John R. (1967). *Constraints on Variables in Syntax*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Schütze, Carson T. (1996/2016). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Berlin: Language Science Press.
- Sichel, Ivy (2018). Anatomy of a counterexample: Extraction from relative clauses. *Linguistic Inquiry* 49(2), 335–378.
- Sprouse, Jon (2007). *A Program for Experimental Syntax: Finding the Relationship between Acceptability and Grammatical Knowledge*. Ph.D. dissertation, University of Maryland.
- Travis, Lisa (1991). *Derived Objects, Inner Aspect, and the Structure of VP*. Paper presented at NELS 22, University of Delaware.
- Tuller, Laurice (1992). The syntax of postverbal focus constructions in Chadic. *Natural Language & Linguistic Theory* 10(2), 303–334.
- van der Auwera, Johan (1984). More on the history of subject contact clauses in English. *Folia Linguistica Historica* 5/1, 171–184.
- Watters, John (1979). Focus in Aghem. *Southern California Occasional Papers in Linguistics* 7, 137–197.
- Wexler, Kenneth & Peter Culicover (1980). *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.
- Williamson, Sara (2016). Subject contact relatives: A cross-dialectal approach. *Calgary Working Papers in Linguistics* 29, 41–60.
- Wiltschko, Martina (2014). *The Universal Structure of Categories: Towards a Formal Typology*. Cambridge: Cambridge University Press.
- Wurmbrand, Susi (2012a). *Agree(ment): Looking up or down?* Paper presented at MIT.
- Wurmbrand, Susi (2012b). The syntax of valuation in auxiliary–participle constructions. *Coyote Working Papers: Proceedings of the 29th West Coast Conference on Formal Linguistics*, 154–162.
- Wurmbrand, Susi (2013). QR and selection: Covert evidence for phasehood. *Proceedings of NELS* 42, 277–290. Amherst, MA: University of Massachusetts, GLSA.

Appendix

- Response A** This sounds completely normal
Response B This sounds OK but I don't think I would say it
Response C This sounds slightly strange and I definitely wouldn't say it
Response D This sounds completely wrong

Stimuli marked with * are those whose participles are likely to have been misheard as a present participle, resulting in an erroneous interpretation of the stimulus as a majority dialect structure (see §3.1.2).

#	Stimulus	Response frequency				% score
		A	B	C	D	
A01	There's a woman gone up the hill.	368	423	194	45	69.4%
A02	There's been a woman go up the hill.	155	264	309	302	42.1%
A04	There was a woman went up the hill.	143	353	320	214	47.1%
A06	There'll be a woman go up the hill.	221	256	280	273	47.1%
A07	There's a woman been going up the hill.	159	243	313	312	41.5%
A10	There's a woman would have gone up the hill.	39	162	323	506	24.7%
A11	There'll have been a woman have gone up the hill.	22	116	347	545	20.9%
A12	There's a woman going up the hill.	985	32	9	4	98.0%
A13	There's being a woman go up the hill.	5	26	181	818	8.0%
A14	The woman gone up the hill is called Wendy.	208	379	322	121	55.1%
B01	There's a girl eaten* an apple.	835	105	66	24	90.0%
B02	There's been a girl eat an apple.	77	156	366	431	29.4%
B04	There was a girl ate an apple.	97	333	321	279	41.4%
B06	There'll be a girl eat an apple.	25	113	360	532	21.4%
B07	There's a girl been eating an apple.	137	338	369	186	47.1%
B10	There's a girl would have eaten an apple.	20	109	348	553	20.3%
B11	There'll have been a girl have been eating an apple.	16	62	268	684	14.2%
B12	There's a girl eating an apple.	1017	10	3	0	99.5%
B13	There's being a girl eat an apple.	3	15	150	862	6.1%
B14	The girl eaten* the apple is called Emily.	363	139	282	246	53.4%
C01	There's a cat found a mouse.	170	315	314	231	47.1%
C02	There's been a cat find a mouse.	61	158	345	466	27.3%
C04	There was a cat found a mouse.	71	238	410	311	35.6%
C06	There'll be a cat find a mouse.	36	117	321	556	21.5%
C07	There's a cat been finding a mouse.	55	231	426	318	34.1%
C10	There's a cat would have found a mouse.	25	86	317	602	18.3%
C11	There'll have been a cat have found a mouse.	59	136	361	474	26.2%
C12	There's a cat finding a mouse.	636	258	104	32	81.8%
C13	There's being a cat find a mouse.	1	11	82	936	3.5%
C14	The cat found a mouse is called Felix.	125	222	364	319	38.3%
D01	There's a bird fallen* out of the tree.	620	293	91	26	82.1%
D02	There's been a bird fall out of the tree.	146	329	336	219	46.3%
D04	There was a bird fell out of the tree.	249	368	263	150	56.5%
D06	There'll be a bird fall out of the tree.	59	203	376	392	31.0%

NARRATIVE STRUCTURES IN COLLOQUIAL ENGLISH

D07	There's a bird been falling out of the tree.	53	204	397	376	31.2%
D10	There's a bird would have fallen out of the tree.	21	89	292	628	17.2%
D11	There'll have been a bird have fallen out of the tree.	54	163	316	497	26.0%
D12	There's a bird falling out of the tree.	958	56	13	3	97.1%
D13	There's being a bird fall out of the tree.	6	28	161	835	7.6%
D14	The bird fallen* out of the tree is black.	330	358	259	83	63.6%
E01	There's a man been arrested.	442	397	156	35	73.7%
E07	There's a man been being arrested.	22	105	318	585	19.2%
E09	There'll have been a man been being arrested.	30	76	225	699	15.1%
E12	There's a man being arrested.	1000	28	1	1	98.9%
E14	The man been arrested is called Simon.	264	332	306	128	57.0%
F01	There's a dog been trained.	330	388	228	84	64.5%
F07	There's a dog been being trained.	48	131	321	530	23.5%
F09	There'll have been a dog been being trained.	23	112	259	636	17.9%
F12	There's a dog being trained.	944	68	14	4	96.5%
F14	The dog been trained is called Fido.	322	265	286	157	57.7%