



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA

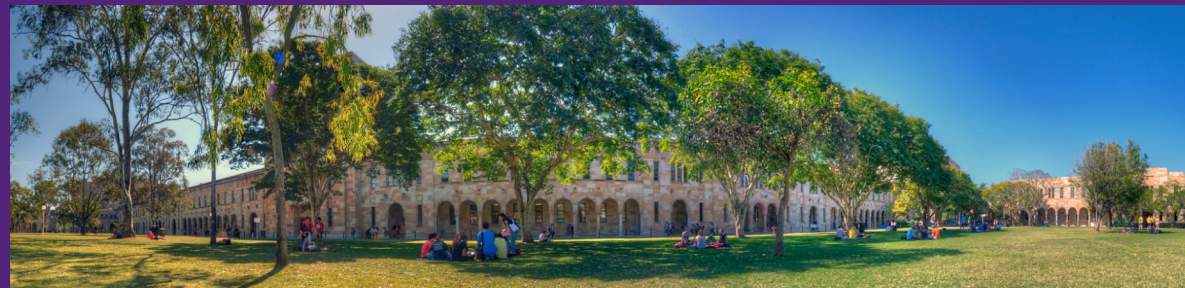
CREATE CHANGE

# A Corpus-Based Analysis of General Extenders in Irish English

Martin Schweinberger

<https://github.com/MartinSchweinberger/GenExIrE>

NPIE 7, Cork, June 8-10, 2022



# What are General Extenders?

## (1) Adjunctive general extenders

- a) She 's her bag packed and <,> washed out and <,> her pencil case filled **and so on** (ICE-Ire: S1A-001\$B)
- b) But because it 's a real Protestant thing they were never treated with the right respect **and all this sort of stuff** (ICE-Ire: S1A-005\$A)
- c) Sure people 're bringing in fruit for people **and stuff** (ICE-Ire: S1A-005\$A)

## (2) Disjunctive general extenders

- a) You 'd think it was hot **or something** <&> laughter </&> (ICE-Ire: S1A-003\$B)
- b) I don't know what he was writing whether it was the cars **or whatever** you know (ICE-Ire: S1A-012\$A)

## Common pattern

**Connector** (and | or) + [**quantifier** (all | some) +] **generic NP** (thing | that stuff) [+ **comparative PP** (like this | like that)]

# What are General Extenders?

Also referred to “coordination tags” (Biber et al. 1999: 115–116), “set marking tags” (Dines 1980), “vague category identifiers” (Channell 1994), “extension particles” (Dubois 1992), and “approximation markers” (Erman 1995:136)

Grammatically optional + little propositional meaning (→ pragmatic marker)

GEs have been found to

- have genre-specific use
- be social stratified along gender, age and socioeconomic dimensions in BrE (Cheshire 2007), AusE (Dines 1980), NZE (Stubbe & Holmes 1995), and Montreal French (Dubois 1992)

In Toronto English: *stuff* lexically replacing other NP elements (Tagliamonte & Denis 2010)

In BrE, GEs are undergoing grammaticalization (bleaching of morphosyntactic properties and loss of agreement with host noun) (Cheshire 2007)

# Research Questions

1. What forms of GEs and how frequently are GEs used in IrE?
2. Is the IrE GE system currently undergoing change (are GEs stratified)?
3. What factors impact GE use in IrE?



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA

CREATE CHANGE

# Data and Data Processing

# Data

Spoken data from the Irish component of the *International Corpus of English* (ICE-Ire)

- 100 conversations
- app. 200,000 words
- 167 speakers
- 507 GEs

(after processing, 971 GEs across spoken text types)

```
<S1A-005> <2014>
<I>
<$A> <#\>Yeah I think so
<$B> <#\>Mmm <#\>Where's the little boy live
<$A> <#\>Up the end of the cul-de-sac <,,>
<$B> <#\>He'd be a bit younger than <@>forename1</@> wouldn't he
<$A> <#\>Yeah he's only five <,,> <#\>He's the same size as him
though
<$B> <#\><{><[>Mmm</[>
<$Z> <X><#\><[>He's a</[></{> big kid<,,></X>
<$B> <#\><0>laugh</0> Are you feeling any older <@>forenameC</@>
<$Z> <X><#\>Very much very much so<,,> <#\><{><[>I'm a bit
stressed</[> </X>
<$B> <#\><[>I hope you've worked</[></{> out where we're going for
this <,,> our celebration tonight
```

# Data processing

- Loading the spoken section of the corpus
- Split data into utterances
- Retrieved all GEs using a regular expression\* (see below)
- Combine data with speaker and text information

\* "(sort|sorts) of (stuff|shit)((\\betc\\bb|etcetera etcetera|etcetera|et cetera)|(and|or) (this|that){0,4} {0,1}(whatever|so on|so forth|all|something|everything|thing|things|stuff|shit) {0,1}(else|like){0,4} {0,1}(this|that|these|those){0,5})"

# Variable coding

Dependent Variable(s)		
Variant	categorical	and stuff   or something   and things like that   etc.
Independent Variables		
Speaker	categorical	Speaker ids
Age	categorical	19-25   26-33   34-49   50+
Gender (Sex)	nominal	Female   male
Region	categorical	RoI   NI   Mixed
AudienceSize	nominal	Dyad   Multiple
ConversationType	nominal	Same   Mixed



# Analysis

**Boruta** (finding sig. predictors)

- What predictors have any kind of statistically significant impact on the use of general extenders?

**Conditional Inference Tree** (partitioning of the data)

- What factor combinations lead to statistically different general extender usage profiles.

**Configural Frequency Analysis** (types- and antitypes)

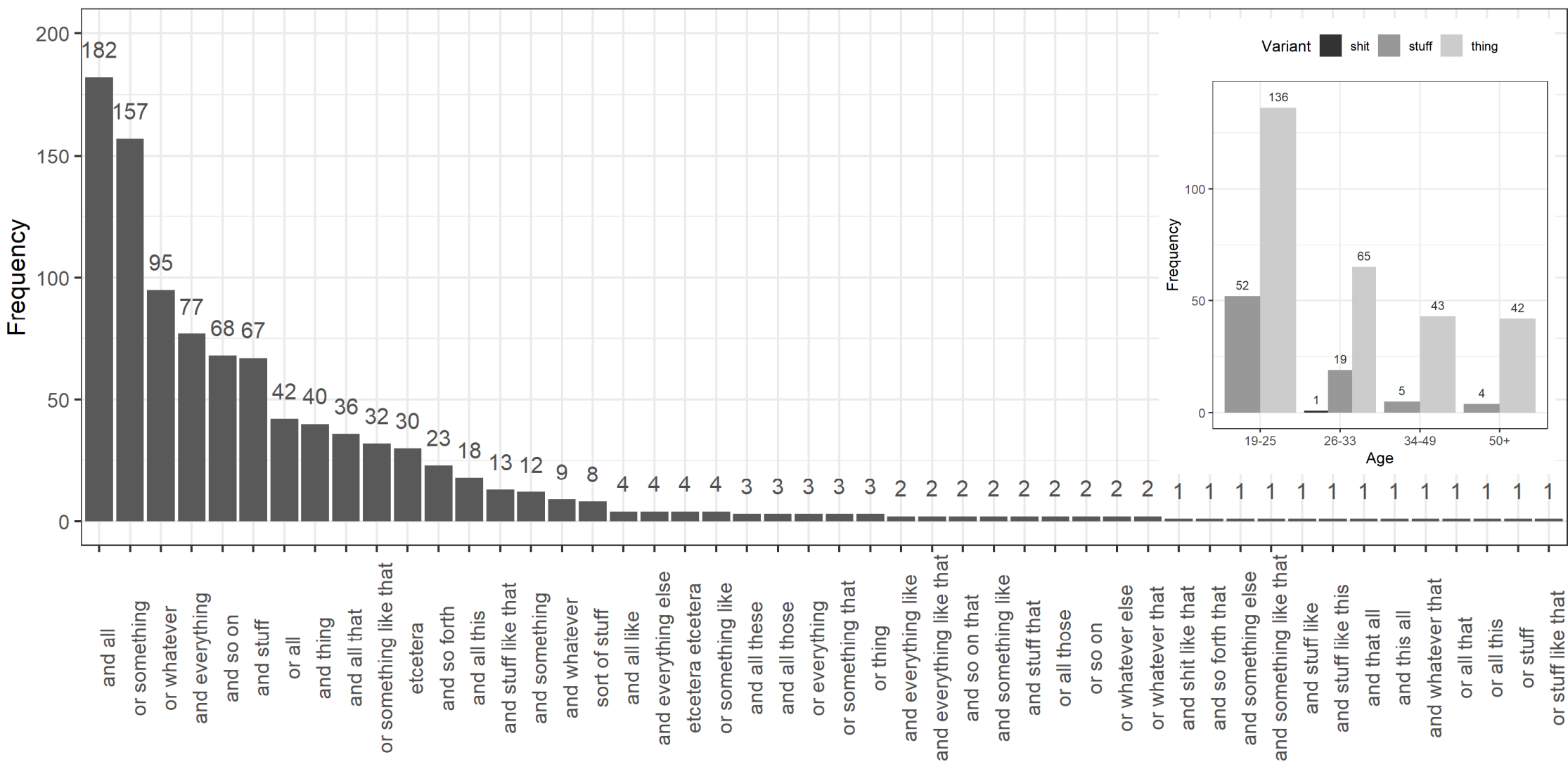
- What general extender types are over- and underused by social strata??



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA

CREATE CHANGE

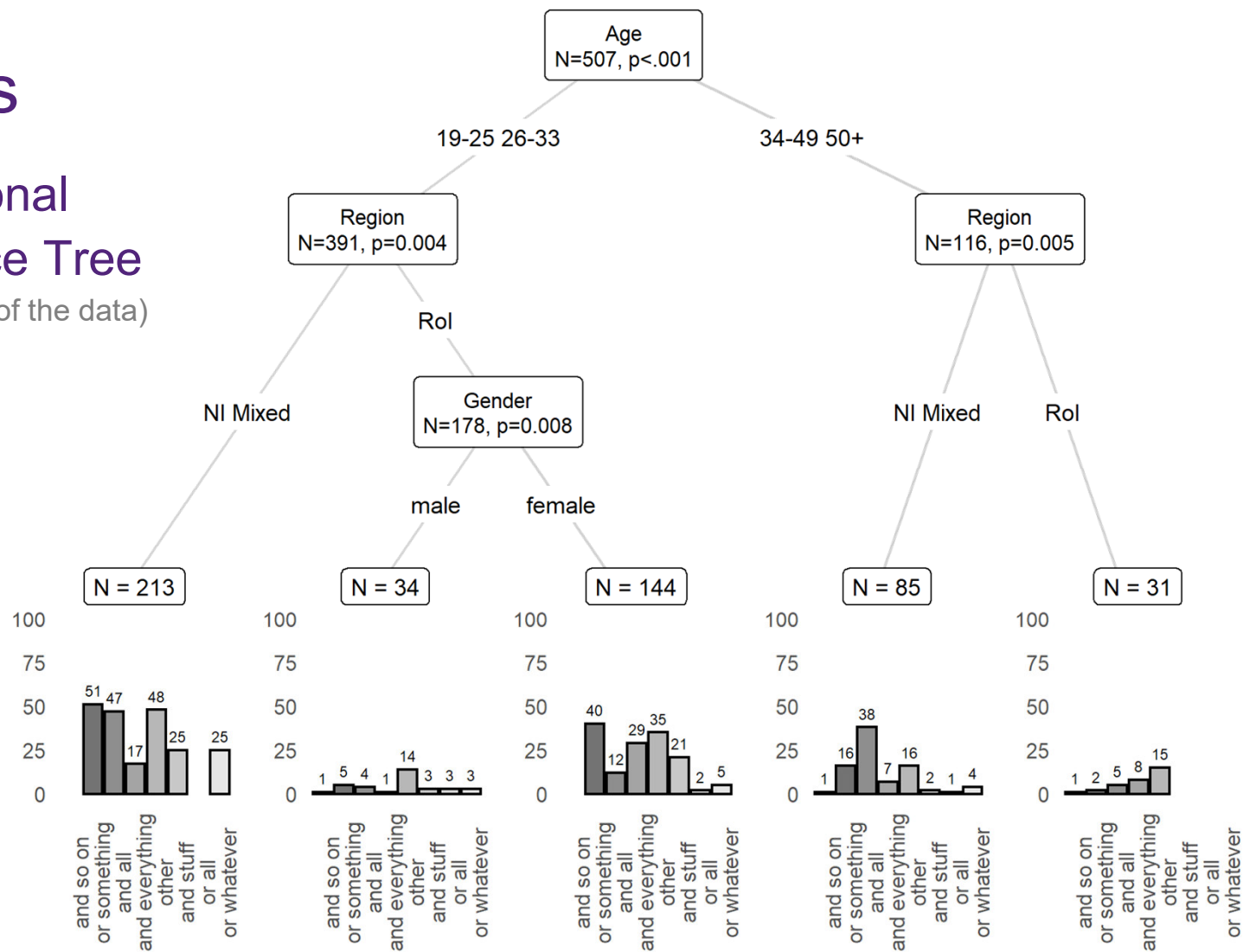
# Results | Findings



# Results

## Conditional Inference Tree

(partitioning of the data)



# Results

## Configural Frequency Analysis (CFA)

(over- and underuse)

Configuration	Frequency (obs.)	Frequency (exp.)	Q	$\chi^2$	z	p-value (z)
Rol male <i>or all</i>	3	0.6	0	10.74	3.28	0.0005***
Rol male <i>and so on</i>	2	0.3	0	10.66	3.27	0.0005***
Mixed female <i>or whatever</i>	8	2.8	0.01	9.46	3.08	0.0011**
Rol female <i>and all</i>	16	33.9	0.04	9.43	3.07	0.0011**



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA

CREATE CHANGE

# Discussion and Outlook

# Summary

## Findings

- *General extenders are stratified along social (not linguistic) dimensions: Age (young vs old), Region (RoI vs NI + Mixed), Gender (Men vs Women)*  
(shows that general extenders have social meaning)
- **Forms**
  - *and all* as well as *or something* are the dominant general extenders in IrE
  - Variants with *thing* dominant over variants with *stuff* | *shit*
- **Configurations**
  - Men from the republic underuse *or all* as well as *and so on*.
  - Women from the republic underuse *and all*.
  - Women that have ties to both the republic and Northern Ireland overuse *or whatever*.
- Study shows that variability in the use of general extenders is socially motivated.

# Discussion

Corpus-based analyses allow us to investigate a wide variety of linguistic phenomena both within and across varieties of English

- General extenders allow us to inspect mechanisms of ongoing change due to their interchangeability and variability as well as lack of overt social stigmatization (exceptions are, of course, general extenders which contain taboo words such as *shit*).
- Multivariate statistics allow us to determine the effects (or the impact) of various predictors simultaneously (at the same time) which is a major advantage because this safeguards from confounding (and thus wrong/inaccurate) results.



# Outlook

## Limitations

- Although ICE Ireland has the advantage that it is accompanied by detailed information about speakers, it is comparatively moderate in size which renders analyses of innovative variants and ongoing change among non-pervasive features difficult.
- Semantic features of the host element and discourse variables were not considered in this study.

## Future directions

- Apply analysis to GloWbE to analyse patterns of use across a range of varieties of English but social characteristics of users are difficult to ascertain.
- Future research will extend this study by adding information on semantic features of the host phrase.

# References

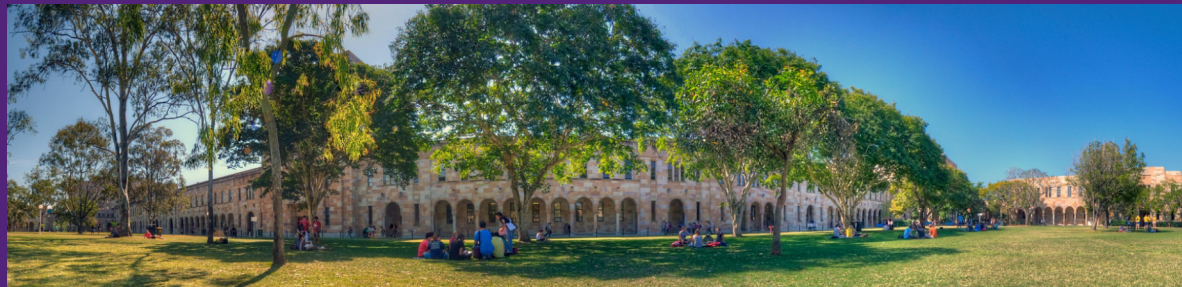
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Eric Finegan, & Randolph Quirk (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Channell, Joanna (1994). *Vague Language*. Oxford: Oxford University Press.
- Cheshire, Jenny (2007). Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11(2): 155-193.
- Dines, Elizabeth R. (1980). Variation in discourse – and stuff like that. *Language in Society* 9(1): 13-31.
- Dubois, Sylvie. (1992). Extension particles etc. *Language Variation and Change* 4: 179–204.
- Kallen, Jeffrey C., and John Kirk (2007). ICE-Ireland: Local variations on global standards. In Joan Beal, Karen Corrigan, and Hermann Moisl (eds.), *Creating and Digitizing Language Corpora*, 121-162. London: Palgrave Macmillan.
- Stubbe, Maria and Janet Holmes (1995). You know and other exasperating expressions: An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language and Communication* 15: 63–88.
- Tagliamonte, Sali A. and Derek Denis (2010). The stuff of change: General extenders in Toronto, Canada. *Journal of English Linguistics* 38(4): 335-368.

# Thanks a lot and stuff!

## A Corpus-Based Analysis of General Extenders in Irish English

Martin Schweinberger ([m.schweinberger@uq.edu.au](mailto:m.schweinberger@uq.edu.au))

NPIE 7, Cork, June 8-10, 2022



# Contact

Martin Schweinberger

T: +61 7 3365 6374

E: [m.schweinberger@uq.edu.au](mailto:m.schweinberger@uq.edu.au)

H: [martinschweinberger.de](http://martinschweinberger.de)

School of Languages and Cultures  
Gordon Greenwood Building  
The University of Queensland  
St Lucia Qld 4072 Australia



Australian Research Data Commons

