SP*i*LL

Short Papers in Language and Linguistics

Cardiff University

Editorial

Dear Readers,

Welcome to the very first issue of SPiLL! We are delighted to release this first edition of the Short Papers in Language and Linguistics Cardiff University postgraduate e-journal. It is packed full of exciting short research papers and blog-style pieces; showcasing original linguistics research at various stages of completion, and providing some excellent advice and tips for those embarking upon their PhD journeys. We really hope you enjoy, question and wonder at all the great work going on here at Cardiff.

This e-journal began its journey during a socially distanced coffee after the initial Covid-19 lockdown began to ease. Like everyone else, we were missing the hustle and bustle of the John Percival corridors and postgraduate offices. The opportunities to swivel round on your chair and ask that question, brain storm that idea or ask for that piece of advice had suddenly gone. As we all know, postgraduate research is quite the rollercoaster of an experience—we need these opportunities to ask, chat, (whinge), and discuss.

This year, perhaps more than ever, has highlighted the importance of reaching out, connecting with one another and discussing what we're up to. This postgraduate e-journal hopes to provide a space for those missing opportunities and conversations. It also hopes to connect postgraduates beyond the school of ENCAP. This issue contains papers from postgraduate colleagues in the Schools of Welsh and MLANG and our hope is that as this e-journal continues to grow, the list of schools from which students have submitted grows too. Please help us to continue spreading the word and encourage colleagues from across the university to get in touch and submit short papers. In our opinion, there is no such thing as 'too much' language and linguistics—we want to showcase as much research as possible!

As the e-journal continues to develop, we also hope to hold online discussions, share 'in-the-spotlight' research interviews, video snippets and 'in-question-with' author highlights. Be sure to keep your eye on our Twitter page (@SPiLL_Cardiff) to stay up-to-date with all our plans. Please contact us at mailto:spill@cardiff.ac.uk if you have any questions, ideas, feedback etc. Any and all thoughts very welcome! Our next Call for Papers will be released in January, ready for the release of issue 2 in the spring. We all know how time flies and it is never too early to start thinking about these things…so get thinking, get writing and get sending us your work!

Finally, we would like to say a huge thank you to Dr Gerard O'Grady who has supported this idea from the very beginning. Your words of encouragement gave us the confidence to embark on this journey. Thank you also to Dr Becky Munford who offered huge support with advertising and spreading the word about the e-journal as it first began. And, of course, a huge thank you to all those whose papers feature in this first issue, all those who have emailed, tweeted, messaged and spoken words of encouragement about this e-journal. We really hope it fills in a few of the missing gaps and provides a space to ask and answer those, 'so if you were me, how would you….?' questions.

We really hope you enjoy. Happy reading!

Ellen Bristow and Katharine Young
Co-founders and editors of SPiLL

What can I submit to SP*i*LL?

up to 1000 words on any of the following:

MA/PhD research topics & plans
Research methods
Ethics
Literature reviews
Data analysis
PGR tips and tricks
Book reviews

Inside this issue

The operationalization headache

**Jack Pulman-Slater (School of Welsh.
Cardiff University/Bristol University, AHRC)**

After you have decided which areas of speech or language you want to investigate, you then have to decide who is going to supply you with examples of it. Operationalization is the process of setting criteria against which participants can be included or excluded in a project. The word 'operationalization' has connotations of a stringent scientific process in which fuzzy concepts such as 'new speakers' can be tightened up in order to subject certain individuals to empirical linguistic enquiry. Empiricism is what Linguistics is all about and the end goal is often to use the data we have gleaned from a particular sample in order to conduct inferential statistics about the wider population. Even if statistics are not involved in our research, we still want to use our data to say something about the wider population. But how can we be confident in making statements about the listening comprehension skills of 'Dutch learners of Spanish as a third language' or the attitudes of 'North Walians to colloquial Southern Welsh expressions' if we're not confident about how we have operationalised these speaker types?

Let us begin with an example of an erroneous operationalization from Sociology. The Three Wales Model (Balsom 1985) divides Wales into three geographically distinct parts on the basis of national identity, party political allegiance and language. In 'British Wales', we have people who identify as British, are monolingually English-speaking and vote for unionist parties. In 'Welsh Wales', we have Welsh-identifying, monolingually English-speaking people who generally vote for socialist parties. In Y Fro Gymraeg (or Welsh Heartland), Welsh is the default community language, people identify as Welsh only and typically support Plaid Cymru. Balsom and his colleagues interviewed people via questionnaires and the respondents' potential replies seem simple given the root of the questions: British-identifying? Speak Welsh? Plaid Cymru or not? Except these are not the questions asked. The operationalization of 'Welsh speaker' is particularly problematic. Scully & Jones (2012:659) highlight that Balsom's preparatory research for the model (Balsom et al. 1983) in fact measured Welsh-speaking ability indirectly via means of a cultural attachment scale. This means that 'Welsh speaker' was operationalized using an additive scale which included the following variables:

THE 'WELSH SPEAKER' VARIABLES

1. SPEAKING WELSH
2. HAVING WELSH-SPEAKING PARENTS
3. HAVING SPOKEN WELSH AT HOME
4. BEING A RELIGIOUS NON-CONFORMIST
5. LIVING IN A 'WELSH HEARTLAND' AREA IN NORTH WEST OR UPPER-SOUTH WEST WALES
6. HAVING BEEN TAUGHT WELSH AT SCHOOL
7. READING A WELSH-ORIENTED DAILY PAPER
8. WATCHING WELSH TELEVISION NEWS

The problems with the scale are numerous and manifest; the scale probably says more about the personal prejudices of the researchers than is does about language. Whilst such an operationalization may give us information about language use or perhaps be an interesting cultural operationalization of Welsh speakers, it is not a purposeful linguistic operationalization of the speaker type 'Welsh speaker'. I cannot determine whether you speak a language by asking you if you are a religious non-conformist or if you live in a certain place.

However, this does not necessarily mean that we can always comfortably reject a priori certain criteria from our operationalizations. In the field of second language acquisition, we know that the factors influencing accent, for example, are numerous and potentially infinite:

age of onset, length of residency in a foreign country, personal/professional motivation, degree of concern for accent improvement, typological fit between first and second language to name but a few (see Piske et al. 2001 for an overview). Balsom et al. (1983) selected criteria that were simply irrelevant, but our problem in linguistics is that because language is a biological, cognitive and social phenomenon the list of criteria is extensive.

However, it must be borne in mind that our goal is to generate an acceptable operationalization for our research and not for all research. Consider the notion of a 'bilingual speaker'. This term has numerous potential definitions based on various parameters including age, chronology of acquisition, proficiency, language status and many more besides. These parameters create different types of bilinguals: compound, diagonal, functional, horizontal, symmetrical, vertical- the list goes on (Thomas & Webb-Davies 2017:16). When researching bilingualism, the operationalization required is that which is most pertinent and applicable to your individual research context. For example, Mooney (2017) examines phonetic transfer in French-Occitan bilinguals in Béarn (France) who he defines as speakers whose first language is Occitan and who learned French through emersion education from 6 years of age onwards. This research does not tell us about phonetic transfer in all bilinguals, instead it informs our understanding of this phenomenon through a very specific operationalization of this term.

Is the word 'operationalization' at best a bit of a red herring and at worst a mendacious piece of academic jargon? It depends what our expectations are. We need to avoid the attempt to strive for an operationalization that is infallible and allows us to speak with absolute certainty about the population at large. No operationalization in language research can do this. However, we have to generate some criteria by which to include or exclude participants otherwise language research would be a disorganised and meaningless free-for-all. But we ought to append our operationalizations with large warning labels and understand that they are only one of several serviceable possible options.

If you are worried that the operationalization you have set is too stringent or too wide in scope or includes possibly ambiguous criteria, then good! It probably is too specific, or too unspecific – but so is everyone else's! We can wrestle with these operationalizations until the bovine homecoming, because we will never create an entirely purposeful and scientifically infallible set of criteria within language research. Acknowledge the manifest strengths and potential weakness of your operationalizations, whether they are theoretical or were forced by the demands of conducting research during a global pandemic and government restrictions. Then write yourself a note and then save it for the viva.

References

Balsom, D. 1985. 'The three Wales model', in Osmond, J. (ed.) *The national question again*. Gomer Press: Llandysul

Balsom, D., Madgwick, P.J., van Mechelen, D., 1983. The red and the green: patterns of partisan choice in Wales. British Journal of Political Science 13, 299–325.

Mooney, D. 2017. Phonetic transfer in language contact: Evidence for equivalence classification in the mid-vowels of Occitan–French bilinguals. *Journal of the International Phonetic Association*, tt. 1-33.

Piske, T. et al. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics* 29(2), tt. 191-215.

Scully, R. & Jones, R. W. 2012. Still three Wales? Social location and electoral behaviour in contemporary Wales. *Electoral Studies* 31(4), tt. 656-667.

Thomas, E. & Webb-Davies, P. 2017. *Agweddau ar Ddwyieithrwydd*. Caerfyrddin: Y Coleg Cymraeg Cenedlaethol

A note on the study of linguistic interfaces: the case of clefts and the syntax-information structure-prosody interface[1]

Charlotte Bourgoin (KU Leuven/Cardiff University)

Keywords: clefts; prosody; information structure; linguistic interface; English language

With many sub-areas of study existing within the linguistics field, a number of cross-area interfaces have lent themselves to extensive study. One such interface lies at the intersection between syntax, information structure and prosody, and is associated with what are described as 'information-packaging constructions' (Huddleston & Pullum 2002). In my research, I focus on the case of *it*-clefts by investigating the interplay between their syntactically- and prosodically-coded meanings. Broadly speaking, this study is an attempt to elucidate how speakers manage information flow in spoken discourse.

The bi-clausal syntax of clefts, i.e. (1a), has traditionally been described as allowing the foregrounding of the post-verbal element in the matrix clause while backgrounding the information in the cleft-relative clause (Jespersen 1937; Huddleston & Pullum 2002: 1416-7).

(1a) It's their credibility that's in question. (LLC)

(1b) Their credibility is in question.

As information-packaging constructions, their information structure differs from that of their non-cleft counterpart, whose information is not presented in a complex sentence but in a simple one as in (1b). From a prosodic viewpoint, most studies argue that the clefted constituent generally carries either the unique information focus coded by a nuclear accent (Clark & Haviland 1977), or the main information focus marked by the 'strongest' of several nuclei (Prince 1978; Declerck 1984). Some authors even describe clefts as primarily construing focalisation (Givón 2001; Lambrecht 2001). However, a number of studies (see Delin 1990; Collins 2006; Kimps 2016) have challenged this systematic mapping by showing that the clefted constituent is not always new and/or in focus. What is at issue is thus the relation between the system of focus as part of information structure and the distribution of information.

Different approaches to the notion of focus exist in the literature. In the Hallidayan (1967) functional tradition, focus-marking in English is linked to intonation (Tench 1996), and information focus is associated with the element selected as the most salient new information of the information unit. What is presented as recoverable and non-recoverable is a rhetorically motivated speaker choice and does not necessarily correlate with what has or has not been mentioned before (Halliday & Matthiessen 2014: 118). With Halliday's definition of information focus, clefts are predicted to show a variety of focus assignments. In a different approach found in accounts such as Lambrecht's (2001), focus is not treated as a property of a given constituent but as a relation between that constituent and the proposition it is a part of. For Lambrecht (2001), the syntax of clefts construes the clefted element as the unambiguous argument focus necessarily requiring an accent and the cleft relative clause as the presupposed topic typically containing unaccented discourse-old information. These different approaches to focus make different predictions about the occurrence and interpretation of nuclear accents in clefts.

The goal of my research is twofold. Relying on a combination of auditory and instrumental analysis of 238 tokens of specificational *it*-clefts extracted from the London-Lund Corpus (LLC) (Svartvik 1990), I aim at (i) revisiting the notion of focus by providing a corpus-based account of its prosodic coding in relation with the distribution of information and (ii) clarifying the type of interplay there exists between the different components of the three-way interface by providing empirical evidence against any straightforward systematic mapping.

The analysis of prosodic foci shows that five patterns can be distinguished based on the presence and location of information foci. Full clefts typically exhibit a focal-focal pattern in which both the clefted

constituent and the cleft-relative clause carry one or more tonic syllables, but information focus may in some cases be absent in one or the other. When both carry foci, the analysis of the pitch excursion of each prosodic focus reveals that a hierarchy can be set up. In 87% of the tokens analysed, tonic syllables have a higher pitch on the clefted element than on an element of the cleft-relative clause. This is exemplified in figure 1 where the high (H*) accent on the clefted constituent *James* has a larger pitch excursion size than the low (L*) accent on gets in the cleft-relative clause.

This shows that although both parts of clefts may carry an information focus, the clefted constituent still tends to have a relatively stronger information focus. The two strategies generally attributed to clefts in the highlighting of the clefted constituent, i.e. syntactic and prosodic, are thus not systematically co-dependent. The analysis of discourse-familiarity reveals that the presence of a nuclear accent in the clefted constituent does not always correlate with discourse newness of that element. No straightforward relation can therefore be set up between discourse-newness and presence of prosodic focus.

As far as secondary accents are concerned, the presence of an onset in the tone unit of the clefted constituent is found in some cases to serve pragmatic purposes linked to the typical discursive functions of specificational clefts (O'Grady, 2014). While the high onset (marked by a caret) on *not* in (2) adds to the contrastive focus on the clefted constituent, the one on only in (3) makes overt the exhaustiveness implicature already triggered by grammatical form.

2) it's ^**not** phil\ologists# you want to c\onvince# (LLC)

3) it's ^**only** when they turn f\acing us# that you get the the the \underside# full \on as it were# (LLC)

Overall, this study provides further argumentation against analysing cleft syntax as solely construing focalisation. Instead, I show that a clear distinction should be made between the syntactically-coded highlighting of the clefted constituent, and focus-marking in the whole cleft construction as signalled by the prosodically-coded information structure. Cleft constructions illustrate how speakers balance their grammatical and prosodic choices in real-time in pursuit of their communicative and interactional goals.

References

Clark, Herbert H. & Susan E. Haviland. 1977. Comprehension and the Given-New Contract. In R. Freedle (Ed.), *Discourse production and comprehension*. Norwood, 1–40.

Collins, Peter. 2006. It-clefts and wh-clefts: Prosody and pragmatics. *Journal of Pragmatics* 38(10), 1706–1720.

Declerck , Renaat. 1984. The Pragmatics of *It*-Clefts and *Wh*-Clefts. *Lingua* 64, 251–289.

Delin, Judy. 1990. A multi-level account of cleft constructions in discourse. *Proceedings of the 13th Conference on Computational Linguistics 2,* 83–88.

Givón, Talmy. 2001. *Syntax. Volume* 11 Amsterdam: John Benjamins Publishing Co.

Halliday, M. A. K. 1967a. Notes on transitivity and theme in English, Part II. *Journal of Linguistics* 3, 199–244.

Halliday M.A. K. & Christian M. I. M. Matthiessen. 2014. *An introduction to Functional Grammar*, 4th edition. London: Arnold.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: CUP.

Jespersen, Otto. 1937. *Analytic syntax*. London: Allen & Unwin.

Kimps, Ditte. 2016. Non-prototypical English specificational cleft constructions and their prosodic patterns. International Workshop on non-prototypical clefts. KU Leuven.

Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39: 463–516.

O'Grady, Gerard. 2014. An investigation of how intonation helps signal information structure. In Wendy L. Bowcher & Bradley A. Smith (eds.) *Systemic Phonology: Recent Studies in English*. 27–52. London: Equinox.

Prince, Ellen. 1978. A comparison *of wh*-clefts and *it*-clefts in discourse. *Language* 54, 883–906.

Svartvik, Jan. 1990. *The London Corpus of Spoken English: Description and research*. Lund Studies in English 82. Lund University Press.

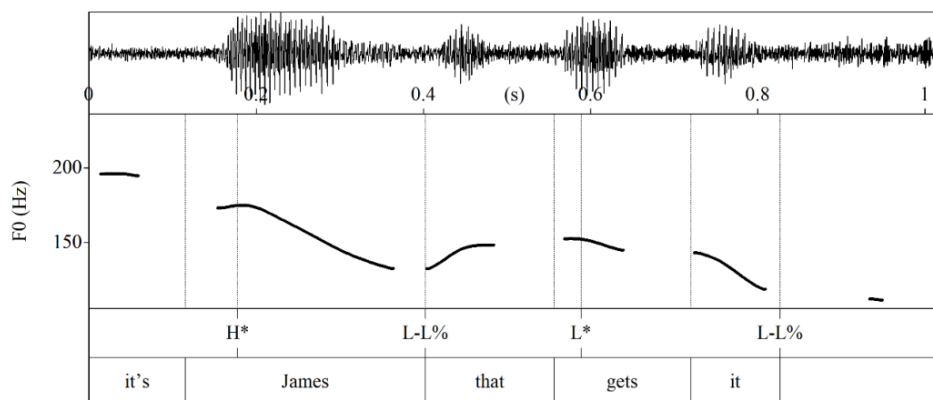Tench, Paul. 1996. *The intonation systems of English*. London: Cassell

Figure 1. Praat picture with ToBI annotation of *it's James that gets it* (LLC)

A comparative study of trilingual families in Helsinki and Cardiff areas: intergenerational language transmission experiences, influential factors for development of minority languages and children's perspectives

Kaisa Pankakoski (School of Welsh, Cardiff University)

Keywords: multilingualism, family language policy, bilingualism, early language acquisition, trilingualism

Introduction, theoretical framework, gap in research

Superdiversity and globalisation have led to the European capitals to grow exponentially. By 2035 a quarter of Helsinki citizens will speak a first language other than the official languages of the country. Cardiff is equally diverse: it has the highest population share (13%) of non-UK born residents in Wales. According to the latest census of 2011 just over 8% of Cardiff citizens announced speaking a main language other than the official languages of the country: English or Welsh.

My thesis looks at fourteen case study families in the linguistically diverse, officially bilingual capital city areas of Helsinki and Cardiff. The two cities are compared because of similarity in terms of having visible foreign language speakers, minority language populations and substantial government minority language support. However, they have different approaches to minority language promotion, foreign language learning, and heritage language support. In Finland, several initiatives help develop multilingual children's language repertoire such as free home language instruction, Swedish or Finnish as a second language tuition, and early start of modern foreign languages at age seven. On the other hand, the Welsh Government strategies (e.g. Iaith Pawb 2003, A Living Language: A Language for Living 2012, Welsh Medium Education Strategy 2010) have led to a sharp rise in Welsh speakers. The most recent—Cymraeg 2050—aims to increase the number of Welsh speakers back to one million by 2050. Therefore, comparison in the two areas is warranted to investigate the possible influence of language policy on the families' language transmission.

The multidisciplinary theoretical framework of my thesis combines insights from family language policy (King et al. 2008; Schwartz and Moin 2012; Curdt-Christiansen 2013; Smith-Christmas 2015) and language socialisation (used by several researchers, e.g. Barnes 2011; De Houwer 2015; Kheirkhah and Cekaite 2015; Guardado 2018). Each framework stresses the importance of the child which is why a central part of the data is obtained from interviewing children.

There are few studies that have looked at multilingual children's perspective or multilingual families' experiences overall. As far as I am aware, there is no trilingual transmission research in contexts with a community majority and minority language. Despite a small number of previous studies on transmission among trilingual families there is very little work which focuses on the factors which influence language transmission, parents' language strategies, or children's perceptions of their multilinguality. Therefore, this thesis will add to our knowledge in an emerging field of multilingualism and fill a gap in family language policy, language socialisation, and childhood bilingualism literature.

Research questions

The study aims to answer the following questions:

1. What are parents' self-reported experiences of raising trilingual children and how are these experiences shaped by sociocultural factors in Helsinki and Cardiff?

2. What are the children's perceptions of their identity, multilinguality and acquisition of their languages?

Research methods and analysis

A cross-cultural, comparative multiple-case study design was used. First, parents of the case study families and extended family members filled in an online questionnaire. Between 2017 and 2019 I conducted audio-recorded semi-structured interviews in seven Welsh and seven Finnish family homes with parents, and children aged 4-12. Each person could choose to be interviewed in English, Finnish, Swedish, Welsh, Spanish or French. Finally, I carried out audio or video recorded observations to examine how the languages work within the family environment.

The case study families in the capital city areas of Helsinki and Cardiff were limited to those:

a) where at least one child aged 4-12 regularly used and was exposed to three languages
b) out of the three languages two were official languages of the country (Swedish, Finnish, or Welsh, English); the children had also acquired an additional language or languages
c) who had lived in Wales or Finland for at least one year

After transcribing the 25.5 hours of multilingual interviews, I coded and analysed the data using NVivo software. The reflexive thematic analysis (Braun 2006), was chosen to keep a close link to the research questions and relevant literature reviewed in two literature reviews. The data from parents' interviews includes six main themes: Family circumstances; Parental language ideologies and motives to pursue multilingualism; Family language strategies; External factors affecting transmission; Description of linguistic behaviour and development of the children; and Experiences of raising trilingual children.

The study follows the ethics guidelines set by Cardiff University School of Welsh Research Ethics Procedures. Ethical approval was granted in March 2017.

References

Barnes, J. 2011. The influence of child-directed speech in early trilingualism. *International Journal of Multilingualism* 8(1), pp. 42-62.

Braun, A. 2006. *The effect of sociocultural and linguistic factors on the language use of parents in trilingual families in England and Germany.* University of Bedfordshire.

Curdt-Christiansen, X. L. 2013. Family language policy: Sociopolitical reality versus linguistic continuity. Springer.

De Houwer, A. 2015. Harmonious bilingual development: Young families' well-being in language contact situations. *International Journal of Bilingualism* 19(2), pp. 169-184.

Guardado, M. 2018. *Discourse, ideology and heritage language socialization: Micro and macro perspectives*. Walter de Gruyter GmbH & Co KG.

Kheirkhah, M. and Cekaite, A. 2015. Language maintenance in a multilingual family: Informal heritage language lessons in parent–child interactions. *Multilingua* 34(3), pp. 319-346.

King, K. A. et al. 2008. Family language policy. *Language and linguistics compass* 2(5), pp. 907-922.

Schwartz, M. and Moin, V. 2012. Parents' assessment of their preschool children's bilingual development in the context of family language policy. *Journal of Multilingual and Multicultural Development* 33(1), pp. 35-55.

Smith-Christmas, C. 2015. *Family language policy: Maintaining an endangered language in the home*. Springer.
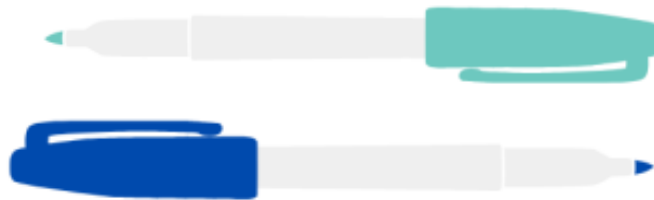
When Should You Give Up? Or, Failed Forays into Bayesian Statistics

Rowan Campbell, (ENCAP, Cardiff University, ESRC)

To set the scene: I'm working on my final analysis chapter, feeling like I'm on the home strait with the PhD, when I realise that the mixed effects models I've used before won't work with this linguistic feature. I'm looking at realisations of /t/ at the end of a small set of high-frequency words like that, but, at, which can be pronounced in Cardiff English as [t], [ʔ], a voiced 'tap' [ɾ] or [ɹ], or even elided completely. That's 4+ levels for a categorical variable, while GLMMs (generalised linear mixed models) only work with binary variables. So now what?

After much googling and reading of various forums to verify that there are no ways to do GLMMs with multinomial dependent variables, I started to look into Bayesian statistics. I've long wanted an excuse to delve into what I jokingly call the quantum of stats, so this seemed perfect timing … apart from the fact that I'm rapidly approaching my deadline and this would require an awful lot of extra learning.

Thankfully, the stats community who use R have a lot of resources online, so I thought I could at least have a look. I even came across a recent linguistics paper by Laurel MacKenzie using the MCMCglmm

package in R for GLMMs with Markov chain Monte Carlo methods, so I thought I'd give it a go. But it was like whack-a-mole – solving one error message would just reveal another one.

Ok. What about another package? Brms was highly recommended, but has a much more complex group of dependencies – in particular 'stan', which I just couldn't get to 'compile'. After a couple of days of googling and troubleshooting, I decided to just start from scratch – uninstall and reinstall R, and try it all again. I'm not sure if it was this, or the 'Getting Started' github page I finally came

across, but the practice model I ran based on this tutorial worked! Despite taking hours to run on my little old laptop.

In the meantime, Laurel Mackenzie had replied to my request for help after very kindly looking over my data and code and correcting some issues. I retried MCMCglmm – it worked! I then tried running brms on my own data – it worked! And I got this cool graph which neatly shows how the glottal [ʔ] and [t] have basically replaced each other over time (and Cardiff English 'CE' variants peaking then reducing). Please ignore the very large credible intervals though!
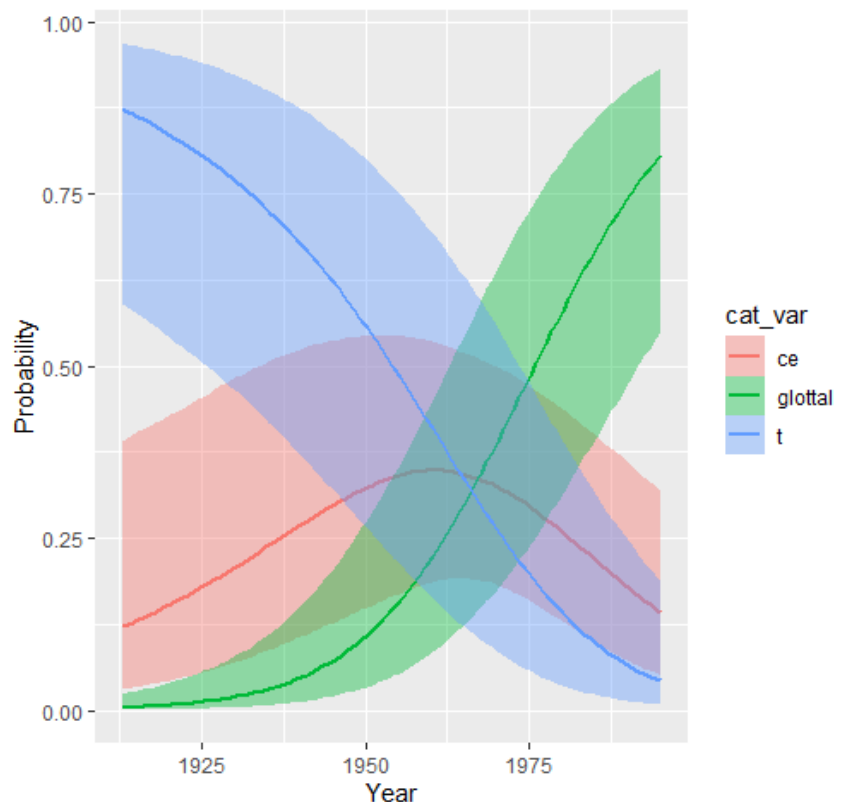
But then I ran into the next problem: how do I make sure my models are valid, and how do I present the data from them? Unlike frequentist models which estimate a single coefficient for each factor, Bayesian or MCMC models run thousands of iterations and provide the average of these. One way of showing model accuracy, used in MacKenzie's paper, is the confusion matrix. This is honestly as confusing as the title suggests. I stared and stared at it, and whenever I thought understanding was just on the edge of my consciousness, it slipped away. I've since read a useful explainer and it's actually fairly clear once you know what's going on, but maybe a name other than 'confusion matrix' would be more user-friendly to begin with!

So, I embarked on more tutorials trying to work out how to put my MCMC predictions into a confusion matrix… and my brain gave up. I was lacking too much of the nuts and bolts R knowledge to make sense of how to do this – it's probably as simple as turning the model output into an array, but I'm just not confident doing basic data wrangling in R (despite my supervisor's recommendation of this great book, R for Data Science!). Also, most tutorials seemed to be for continuous variables, and there was just too much new knowledge and different contexts for me to get my head around, so after two weeks I decided it was time to call it quits.

Conclusion: there are still more people I could reach out to and more things I could try, but in the interests of getting this thing finished, it's time to cut my losses and say goodbye to the Bayesian models ☹.

The 'sunken cost' fallacy is something I think is particularly pertinent to doing research – we run into problems and obstacles at every stage, and it's so difficult to gauge which avenues will be fruitful and which won't. If we continue on one particular route, we have to keep solving each problem we come across – but there's no guarantee that we can overcome the problem after that (or at any rate, that we can solve it in a timely manner). I think that leaving Bayesian behind is the first time I have managed to do this in the PhD – a previous example of this type of 'problem after problem' situation was when I was trying to install FAVE to automate the extraction of vowel frequencies from my interview data. I kept going with this and eventually got there, but it cost me a lot of time. If I were earlier in the PhD, would I keep going with Bayesian? Probably: the problems are definitely not insurmountable, but I think it's a sign of growth to let it go now and remember that the best PhD is a finished PhD.

Secondary conclusion: uninstalling and reinstalling R is always a good shout!

Defining Spanglish

Monique McClymont, (ENCAP, Cardiff University)

A short introduction to Spanglish

Language contact can cause linguistic conflict, linguistic interference, and over time, lead to language substitution and the appearance of new languages. In the United States, the contact between English and Spanish has resulted in the emergence of the linguistic phenomenon, Spanglish. As the name indicates, Spanglish is an intermix of Spanish and English; the term 'espanglish' (the Spanish translation of 'Spanglish') was coined in the late 1940s by Puerto Rican poet, Salvador Tío (Betti & Enghels 2018, p.351). Despite the fact that these languages first converged when Britain colonised the Americas in the early seventeenth century, it is difficult to determine with precision when Spanglish emerged (Stavans, 2003, p.38). In addition to being a communicative practice, Spanglish is a true demonstration of multiculturalism and, for many Latinos, a form of identity (Betti & Enghels 2018, p.352). Presently, there is a dearth of rules associated with Spanglish, and thus it manifests in different forms. Consequently, 'Spanglish' is currently a catch-all term used to describe the acts of combining English and Spanish in a given speech or text.

Although a bicultural and bilingual society is not uncommon, Spanglish has sparked greater controversy than most other languages in contact (Montes-Alcalá 2009, p. 97). Some argue that it attests to the richness of Spanish, whereas language purists (prescriptivists) argue that it is a sign of linguistic ignorance and pollutes the dominant language in the United States, English (Sayer 2008 p.96). Further, scholars debate its origin, its place in society and popular culture, and remarkably, its linguistic features and how to accurately define/label the phenomenon (see, for instance, Stavans 2003; Otheguy and Stern 2010; Fuller 2013). However, linguistic research is still considerably scant, given that Spanglish is not a new phenomenon and the fact that the United States has the second-largest Spanish-speaking population in the world; it is home to 41 million native Spanish speakers, which is 13 percent of the population of the United States, and 12 million additional bilingual Spanish speakers (López & Gonzalez-Barrera 2013).

Working towards a definition and a standard variety of Spanglish

Hancock (cited in Stavans 2003, p.5) states that Spanglish is very hard to define; Stavans (2003, p.5) on the other hand, argues that it is not impossible to define, but that people are simply not willing to define it. Spanglish has been referred to as a pidgin, that is, a simplified form of a language used for communication between individuals who do not share a common language (Muysken & Smith, 1994); Lipski (2008, p.69), however, argues that Spanglish does not fit this description as it is no less complex than a language.

Ardila (2005), on the other hand, considers Spanglish to be an Anglicised Spanish dialect. However, some versions of Spanglish consist of a higher level of English than Spanish, which contradicts this definition. Generally, a dialect is a relatively accurate definition as Spanglish emerged in a geographical location in which settlers developed a form of language mixing in order to effectively communicate in a highly bicultural society. Nevertheless, it is unlike a dialect with regard to the fact that, although subject to change, a dialect is a relatively stable variety of a language which one can learn, whereas Spanglish currently lacks a consistent structure and thus cannot be learned.

Interestingly, Lipski (2008, pp.223-224) does not attempt to pinpoint one definite version of Spanglish, but instead categorises three different levels which depend on one's level of bilingualism: the first requires minimal fluency in the second language and instead consists of borrowings to suit the phonological system of the second language; the second requires a high level of bilingual proficiency, loan translations (or calques), and a syntactic structure which results in a

high convergence between the languages; the third involves fluent bilinguals who can switch between the languages with ease in a single conversation or text.

A typical example of a loanword that can be used at the first and second levels is 'jangear', derived from 'to hang out' in English (Stavans 2003, p.152). This word is phonetically English whilst consisting of a regular Spanish verb suffix, -ear. Further, the letter 'h' in Spanish is silent and so the use of the letter 'j' creates a sound in Spanish which is similar to the sound of the letter 'h' in English. Additional examples of lexical borrowings are located in the first and only Spanglish dictionary, *Spanglish: The Making of a New American Language*, by Professor Ilan Stavans (2003). Included in this dictionary are words, termed 'hybrid words' by Stavans, which combine both languages and which are principally phonetically and orthography English words with Spanish grammar. Examples of hybrid words include 'flirteo', which means 'to flirt' and 'flodiar', also spelled 'flodear' or 'fludear', which means 'to flood' (Stavans 2003, p.124). This dictionary (which includes approximately 6,000 words) follows a rule structure and is the first evidence of a consistent Spanglish variety.

In relation to Lipski's third version of Spanglish, this level involves code-switching, which occurs when a word or phrase in one language is substituted for a word or phrase in another. There are two types of code-switching: intersentential code-switching involves switching between sentences and intrasentential code-switching involves switching within a single sentence, which is more elaborate and complex as proficiency in both languages is necessary in order to avoid violating the grammatical structure or either language (Montes-Alcalá, 2000, p.219). However, intrasentential code-switching may not always represent the complexity of Spanglish; in the command "Dame una hamburguesa sin lettuce por favor" (Heredia & Attarriba, 2001, p.164), for example, an English word is incorporated into a Spanish sentence, which exemplifies a Spanish dialect with loanwords and does not demonstrate a complex form of Spanglish, although it meets the criteria for an example of Spanglish and intrasentential code-switching nonetheless. Overall, Spanglish cannot accurately be described as a pidgin or a dialect but rather a code used by bilingual speakers which can manifest in many forms.

The Future of Spanglish

Notwithstanding Spanglish's notoriety for corrupting Spanish and English (Sayer 2008, p.96), the vitality of it persists in the streets and across various types of media (Betti & Enghels 2018, p.351). Variations of Spanglish appear in the television series Ugly Betty and Dora the Explorer, and popular Hollywood films including Madagascar and Happy Feet, among other forms of broadcasting. Spanglish is gaining increased recognition as a result of the platform offered by the media, a platform that is paving the way for standard Spanglish (Stavans 2014). According to Betti and Enghels (2018, p.352), however, "the streets will always be one step ahead",

which renders it difficult for scholars to standardise and conceptualise Spanglish. Although the future of the phenomenon cannot be accurately determined, over time, it may stabilise and develop into a new hybrid language, as was the case with Yiddish (Sayer 2008, p.99). Indeed, its continuous evolvement and further linguistic research may grant Spanglish a linguistically recognised definition and consistent structure.

References

Ardila, A. 2005. 'Spanglish: An Anglicized Spanish Dialect'. *Journal of Behavioural Sciences*, 27(1), pp. 60-81.

Betti, S and Enghels, R. 2018. 'Spanglish: Current Issues, Future Perspectives, and Linguistic Insights'. In: Stavans, I. (ed.) *The Oxford Handbook of Latino Studies*. Oxford University Press.

Fuller, J.M. 2013. *Spanish Speakers in the USA*. Bristol: Multilingual Matters.

Heredia, R.R. and Altarriba, J. 2001. 'Bilingual Language Mixing: Why do Bilinguals Code-Switch?' *Current Directions in Psychological Science*, 10(5), pp. 164-168.

Lipski, J.M. 2008. *Varieties of Spanish in the United States*. Washington, D.C.: Georgetown University Press.

López, M.H. and Gonzalez-Barrera, A. 2013. 'What is the Future of Spanish in the United States?' *Pew Research Centre*. Last accessed: 11.11.20. Available: <https://www.pewresearch.org/fact-tank/2013/09/05/what-is-the-future-of-spanish-in-the-united-states/>.

Montes-Alcalá, C. 2000. *Attitudes Towards Written and Oral Codeswitching in Spanish-English Bilingual Youths*. University of California: Santa Barbara.

Montes-Alcalá, C. 2009. 'Hispanics in the United States: More that Spanglish'. *Camino Real*, 1, pp. 97-115.

Muysken, P. and Smith, N. 1994. 'The Study of Pidgin and Creole Languages'. In: Arends, Muysken, and Smith (eds) *Pidgin and Creoles: An Introduction*. Philadelphia: John Benjamin.

Otheguy, R. and Stern, N. 2010. 'On So-Called Spanglish'. *International Journal of Bilingualism*, 15(1), pp. 85-100.

Sayer, P. 2008. 'Demystifying Language Mixing: Spanglish in School'. *Journal of Latinos and Education*, 7(2), pp. 94-112.

Stavans, I 2003. *Spanglish: The Making of a New American Language*. New York: HarperCollins.

Stavans, I. 2009. 'The Mestizo Tongue'. *Hispanicla.com*. Last accessed: 12.11.20. Available: <http://www.hispanicla.com/the-mestizo-tongue-8830>.
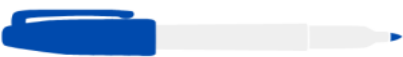
SESIYNAU

GWAITH AR WAITH WORK IN PROGRESS

SESSIONS

Mae myfyrwyr ymchwil Ysgol y Gymraeg yn cynnal sesiynau anffurfiol i gyflwyno rhywbeth y maent wrthi'n gweithio arno. Cynhelir y sesiynau yn Gymraeg neu yn Saesneg ar-lein unwaith y mis. Mae croeso i bawb fynychu.

Research students at the School of Welsh hold informal work in progress sessions to present an aspect of their work they're currently working. These monthly sessions are held online through the medium of Welsh or English. Everyone is welcome.

MULTILINGUAL PARENTS' LANGUAGE TRANSMISSION METHODS: DISCOURSE STRATEGIES

An interactive workshop with Kaisa Pankakoksi
16.12.20 at 13:00

AMRYWIAETH FFONOLEGOL CYMRAEG CAERDYDD

(Phonological variation in Cardiff Welsh)
Ianto Gruffydd presents data and findings from his PhD
20.01.20 at 13:00

LANGUAGE POLICY IN HIGHER EDUCATION

Assala Mihoubi presents findings from her fieldwork
17.02.21 at 13:00

COFRESTRWCH YMA I FYNYCHU / REGISTER HERE TO ATTEND:
TINYURL.COM/ARWAITH

The Cardiff Welsh rounded close back vowel /u(:)/

Ianto Gruffydd (School of Welsh, Cardiff University, Coleg Cymraeg Cenedlaethol)

Cardiff has a growing number of Welsh-medium Schools across the city, and the number of Welsh speakers in the city is increasing from census to census. The variety of Welsh spoken in Cardiff has not yet been documented or studied linguistically, and my research, which was undertaken in a Welsh-medium secondary school in the city (YUGC), aims to record Cardiff Welsh as a variety of Welsh for the first time. This paper aims present results on the Cardiff Welsh rounded close back vowel /u(:)/.

Cardiff English has been researched more extensively, however, and different research has studied the Cardiff English rounded close back vowel. Coupland (1988, p.26) noted that they were 'noteworthy for their lack of diphthongization and centralization when realized phonetically', while Collins and Mees (1990, p.94) specify that 'goose is also closer, somewhat advanced from back'. Therefore, both studies mention some degree of fronting from back in the Cardiff English rounded close back vowel.

In other varieties of English, GOOSE-fronting has become a supralocal feature. GOOSE-fronting is where the /u:/ vowel is produced with the tongue at a more advanced position towards the front of the mouth than was traditional.

In this paper, I will answer the following research questions: firstly, does fronting occur in the corresponding Welsh rounded close back vowel? And secondly, what linguistic and social factors affect the rounded close back vowels in Cardiff Welsh?

There were four parts to my methodology, firstly I undertook a 6-week period of ethnographic observation, to identify the social structures of the classroom. It also prepared the students for the sociolinguistic interview, as the students became acclimatised to my presence and to engaging with me as a researcher. Questionnaires were completed by the whole class for demographic information, and to identify my interview sample. Finally, the main methodological resource for collecting data was the sociolinguistic interview. I conducted 24 semi-structured interviews following a module of conversation structure (Labov 1973). The sociolinguistic interviews contained an interview designed to elicit every day speech, and a words list task to elicit more formal registers.

In regards to the acoustic methodology, all formants were measured with a script (Stanley 2017) in Praat (Boersma & Weenink 2020). Vowels were normalised to identify any outliers. The formants were also normalised, to eliminate any biological differences between boys and girls. Formants were normalised via the NORM suite website with the Watt & Fabricius (2003) method. The formants were exported to excel, where I coded for different linguistic and social factors. After coding, the tokens were capped at 10 tokens per word per speaker.

Social and linguistic factors were coded for in the factor-by-factor analyses. ANOVAs and T Tests tested for statistical significance. Next, I'll present some results from the factor-by-factor analyses.

The first factor presented is community of practice. There are 9 communities of practice in the research, *Bechgyn Meme Chat*, Meme Chat Boys; Cassie; Geeks; *Merched Chwaraeon*, Sporty Girls, *Merched Cymraeg*, Welsh Girls; *Merched Normal*, Normal Girls; *Merched Poblogaidd,* Popular Girls; *Merched Rhun,* Rhun's Girls and *Sbarteiniaid*, the Sbarteiniaid. The names were mostly what the students called the groups, but some were created by me.

Figure 1: Community of Practice
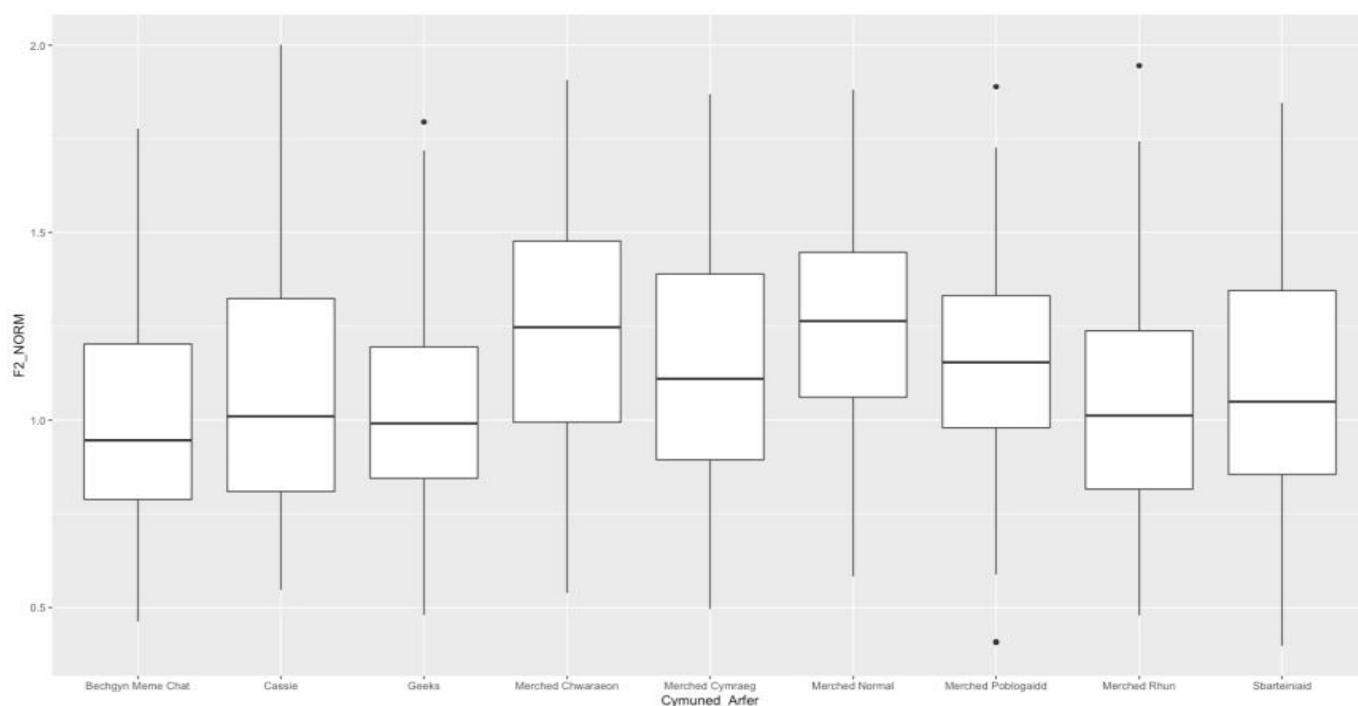
We can observe from Figure 1 that there's variation between the different communities of practice. The *Merched Chwaraeon* and the *Merched Normal* lead with the most fronted /u(:)/ vowels and highest mean F2, while *Merched Rhun* and *Bechgyn Meme Chat*, produced a considerably lower mean F2, An ANOVA test found Community of practice to be a statistically significant factor with p < 0.001.

The next social factor presented is Gender. Females often lead linguistic change or linguistic variation, and even though there are studies where males have done so (Baranowski 2017), it is typical that females lead in close back rounded vowel fronting (Kerswill 2005; Flynn 2012). Figure 2 shows that the same can be said for the Cardiff Welsh rounded close back vowel, as the girls recorded a more fronted /u(:)/ vowel, with a mean F2 of 1.26, whilst the boys produced a mean F2

Figure 2: Gender

of 1.16; .10 less than the girls. An ANOVA test indicated that this factor was statistically significant (p < 0.001).

Figure 3 shows preceding phonetic environment effect on the Welsh rounded close back vowel. *GOOSE* vowel research typically displays a fronting order of palatals > coronals > non-coronals > sonorants. Here we see a different order, with palatals conditioning the most fronting, producing a mean F2 of 1.34, followed by sonorants producing a mean F2 of 1.21, followed by coronals producing a mean F2 of 1.18, and non-coronals condition the least fronting with a mean F2 of 0.89. Therefore, the inhibiting effect of a preceding sonorant does not occur in Welsh as it does in English in regards to the rounded close back vowel. An ANOVA test indicated that preceding phonetic environment is statistically significant p < 0.001.

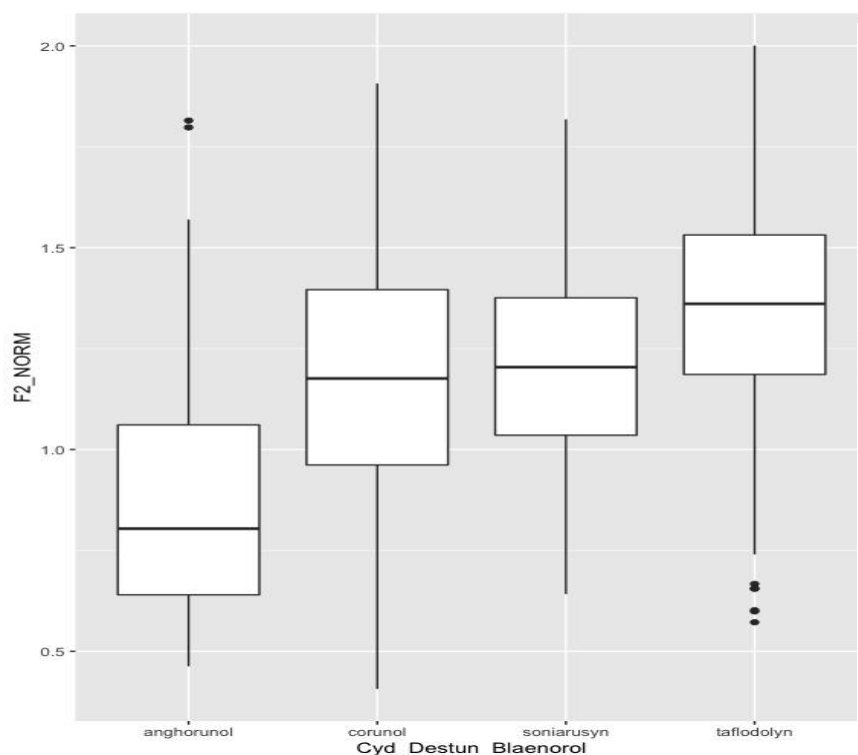The final factor analysis considers task effect. The sociolinguistic interview was divided into two tasks,

Figure 3: Preceding phonetic environment

Figure 4: Task

the interview task itself in which I attempted to elicit the most everyday type speech possible by the students, and the words list, where I attempted to elicit formal registers of the students' speech. Figure 4 displays the difference between both tasks. The Interview task produced a higher mean F2 of 1.27, compared to the Words List which produced a mean F2 1.03. This indicates that a fronting round close back vowel is part of everyday speech for the students. An ANOVA test indicated that task was statistically significant.

To conclude, and to answer my research questions, a fronted rounded close back vowel is common in YUGC Cardiff Welsh and part of everyday speech, as shown in the Task analysis above. Community of

Practice and Gender are two social factors that affect the Cardiff Welsh rounded close back vowel. As is often typical in language variation and change, female speakers are more likely to produce more fronted rounded close back vowel. Preceding phonetic environment is a linguistic factor that affects the Cardiff Welsh rounded close back vowel, and does so with a different fronting order to English: palatals > sonorants > coronals > non-coronals.
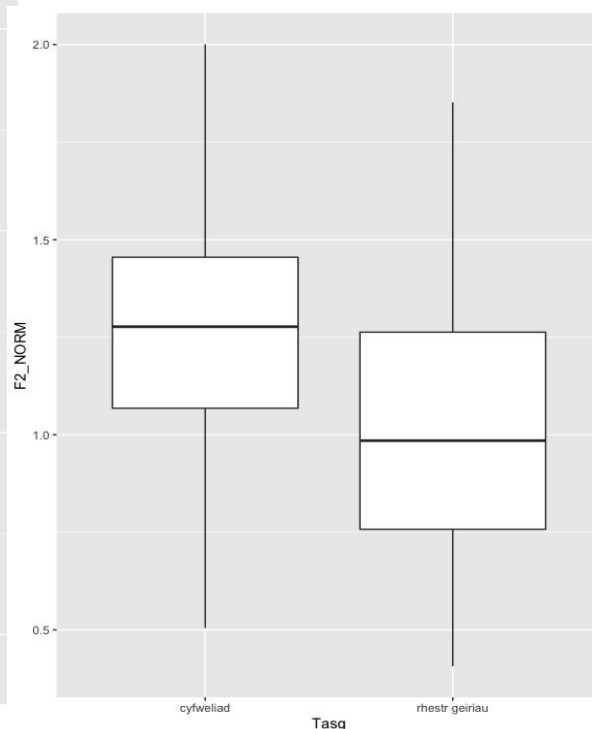
References

Baranowski, M., 2017. Class matters: the sociolinguistics of GOOSE and GOAT in Manchester English. *Language Variation and Change*, *29*(3), tt.301-339.

Boersma, Paul & Weenink, David (2020). Praat: doing phonetics by computer [Rhaglen gyfrifiadurol]. Version 6.1.16 ar gael: http://www.praat.org/

Collins, B. and Mees, I.M., 1990. The phonetics of Cardiff English. *English in Wales: Diversity, conflict and change*, pp.87-103.

Coupland, N., 1988. Dialect in Use: Sociolinguistic Variation in Cardiff. Cardiff: University of Wales Press.

Flynn, Nicholas. 2012. A sociophonetic study of Nottingham speakers. PhD thesis, University of York.

Labov, W., 1984. Sociolinguistic Patterns. 1972. *Trad. Alain Kihm*. University of Pennsylavia.

Matthews, T. 2005. Discourses of intergroup distinctiveness among adolescent girls. Unpublished PhD Thesis.

Mesthrie, R., 2010. Socio-phonetics and social change: Deracialisation of the GOOSE vowel in South African English 1. *Journal of sociolinguistics*, *14*(1), pp.3-33.

Paradis, J. 2001. Do bilingual two-year-olds have separate phonological systems? *International journal of bilingualism* 5(1), tt. 19-38.

Simonet, M. 2010. Dark and clear laterals in Catalan and Spanish: Interaction of phonetic categories in early bilinguals. *Journal of Phonetics* 38(4), tt. 663-678.

Stanley J. 2017. Formant Extraction in PRAAT. http://www.joeystanley.com/

Watt, Dominic and Anne Fabricius. 2003. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1 ~ F2 plane. *Leeds Working Papers in Linguistics and Phonetics 9*: 159–173.
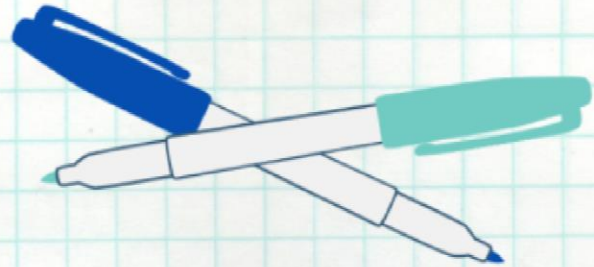
CALL FOR PAPERS

Call is now open for SPiLL issue (ii)

submit a short paper – no more than 1000 words ex. refs to spill@cardiff.ac.uk

submissions close 19th Feb 2021

WE LOOK FORWARD TO HEARING FROM YOU!

Collecting Linguistic Data in a Pandemic – A Corpus Method

Katharine Kavanagh (ENCAP, Cardiff University, ESRC)

Introduction

The Covid-19 pandemic and its ensuing social restrictions have impacted many researchers' work over the course of 2020. My PhD data collection has been no different, although I'm lucky enough that changes have been relatively minor due to a significant portion of my project being reliant on collecting corpus data from existing sources, rather than live fieldwork.[i] In this short paper I will detail my main corpus collection method as an indicative guide for anyone who may be considering corpus methods for the first time following their own disrupted research pathways.

My research requires two related and specialised corpora of promotional texts, pertaining to circus performance in the UK between 1 February 2019 and 31 January 2020. The first comprises show reviews (n=705), and is the main focus of this paper. The second will comprise publicity blurbs for the productions represented in the review corpus (and is therefore dependent on the contents of the first corpus).[ii] What follows is a step-by-step explanation of my process so far, and an invitation to join our regular online

Corpus Linguistics Reading and Training group for anyone interested in finding out whether corpus methods could aid their own research.

Circus Review Corpus (CircRC20)

The corpus was collected using a snowball sampling method, which is a way of accessing 'hidden' examples from a known population (Johnson, 2014). Unknown publications and productions were identified from a known population of specialist circus review publications, using a multiphase online search technique, until examples were exhausted. An Excel spreadsheet (Figure 1) was created to capture the publication and production titles, and to record the number of corresponding review texts. This collection process took the form of three phases, cycled in a repeating pattern ABCB-ABCB:

Phase A - Search by Production

1. Production titles were collected initially from all the combined UK reviews published during the collection period by two specialist seed publications, The Circus Diaries and King Pole.[iii] In subsequent cycles, new production titles revealed by Phase C were identified from the spreadsheet and used.

2. Each production title in turn was entered into the Google search engine alongside the word 'review'. If these search terms produced several unrelated items on the first page of results, they were supplemented with the performing company name and/or venue to narrow down the results to those relevant.

3. Each search was considered complete when three consecutive pages of results yielded no further hits for collection.

Phase B – Collection of files

1. Each 'hit' was opened, and the body text of the review was copied and saved as a plain text file (headlines and standfirst text, which are conventionally written by an editor rather than the review author, were omitted).

2. XML tags were added to the plain text file to assign an ID reference and to record metadata that will later be used for analysis (Kavanagh, 2019), identifying the production name, author name, publication name, publication date, performance venue and performance city.[iv] If the review included a star rating or additional 'verdict', tags were also added to record these (Figure 2).

Phase C - Search by Publication

1. Publication titles were collected from the new spreadsheet entries made during Phase A.

2. The website of each publication was accessed.[v] In most cases, it was possible to browse a chronological list of reviews for the time period and identify the circus productions for collection.[vi] On websites where

no such chronological list was available, a search was made on the site for each of the following terms: circus, cirque, acrobatic.

The ultimate application of this corpus collection will be an analysis of the linguistic representation of circus in public promotional texts, and its impact on the stakeholder groups of audience members and circus professionals, as part of a

Corpus-Assisted Discourse Studies methodology (Partington, 2008) that also draws on Critical Discourse Analysis (Fairclough, 1995).

	A	AR	AS	AT	AU	AV	AW	AX	AY
1	Publication	Everything I See I Swallow	Fatcrobat	Filament	Flight Paths	Gandey's Circus	Gibbon	Giffords Xanadu	Great Yarmouth Hippodrome
56	DVM Theatre								
57	East Midlands Theatre					1			
58	Eastern Daily Press								
59	Eastleigh News								
60	EdFringeReview.com	2							
61	Edinburgh Festivals for Kids								
62	Edinburgh Festivals Magazine	1		1					
63	Edinburgh Spotlight								
64	Essex Live								
65	Evening Standard								
66	Everything Theatre	1				1			
67	Exeunt					1			
68	Express & Star								
69	Fairy Powered Productions							1	
70	Family Life Bury								
71	Fest			1					

Reviews | Review ID | Blurbs | +

Figure 1: Excerpt from data collection spreadsheet


```
backbone the stage RC54                                    +

<meta ID="RC54" production="Backbone" author="Thom Dibdin" verdict="The laid-back
Australian circus troupe makes the defiance of gravity seem easy" stars="4" date="AUG 28
2019" venue="UNDERBELLY MCEWAN HALL" city="EDINBURGH" publication="The Stage"
url="https://www.thestage.co.uk/reviews/backbone-review-at-underbelly-edinburgh--cleverly-
choreographed-circus">

Backbone, the latest touring circus show from Gravity and Other Myths, is a wonderful
example of how to make something insanely difficult seem easy.

For the most part this is a cleverly choreographed and deceptively casual display of
balance, tumbling and people throwing each other around the stage. There is rarely ever
just one thing happening at any one time, with the 10 performers acting like cogs in a
particularly complex mechanism. This is all set to a thunderous, genre-hopping soundtrack
by Elliot Zoerner and Shenton Gregory.

At the start, all the equipment — and all of the performers — are laid out on the stage,
like the parts of a kit. Then a single acrobat is pushed up into the air by poles, until
they are left balancing high above the stage on the top of a single pole.

In Darcy Grant's production, the smallest performers get thrown the furthest and raised
the highest, while a final sequence leaves only the female acrobats standing under a
descending bag of stones. Whether or not it was intended as such, this acts as a potent
visual metaphor.

A final acrobatic mega-mix finale threatens to outstay its welcome but actually it just
adds a little bit of chaos to the mix.
</meta>
```

Figure 2: Sample of a tagged review file

publications was considerably higher than in the next most prolific publication. (King Pole n=37, The Circus Diaries n=53. Next most prolific publication n=26. Mean reviews per publication in corpus = 3.09)

[iv] The original url of the review was also included, as this may be a useful source of data for other investigators in the future.

[v] In a handful of cases, this required a free trial subscription that was subsequently cancelled.

[vi] For the purposes of this research, a circus production was defined as any production that self-identifies as circus (or has been identified as circus in other reviews). Theatre productions that use the concept of circus solely as a setting or design aesthetic were excluded.

References

Fairclough, N. 1995. *Critical discourse analysis: The critical study of language,* London, Longman.

Heritage, F. 2020. Applying corpus linguistics to videogame data: Exploring the representation of gender in videogames at a lexical level. *Game Studies,* 20(3).

Johnson, T.P. 2014. Snowball sampling: Introduction. *In:* Balakrishnan, N.C., T; Everitt, B; Piegorsch, W; Ruggeri, F; Teugels, J.L; (ed.) *Wiley StatsRef: Statistics reference online.* Chichester: John Wiley & Sons Ltd.

Kavanagh, K. 2019. XML mark-up: An annotation tool for discourse analysis. *Methods Blog* [Online]. Available from: http://walesdtp.ac.uk/methodsblog/2019/05/21/xml-mark-up-an-annotation-tool-for-discourse-analysis/ [Accessed 2 Nov 2020].

Partington, A. 2008. The armchair and the machine: Corpus-assisted discourse studies. *In:* Torsello, C.T., Ackerley, K. & Castello, E. (eds.) *Corpora for university language teachers.* Bern: Peter Lang. 189–213.

Shaw, P. 2009. The lexis and grammar of explicit evaluation in academic book reviews, 1913 and 1993. *In:* Hyland, K. & Diani, G. (eds.) *Academic evaluation: Review genres in university settings.* Basingstoke: Palgrave Macmillan. 217-235.

Taylor, C. 2013. Searching for similarity using corpus-assisted discourse studies. *Corpora,* 8**,** 81-113.
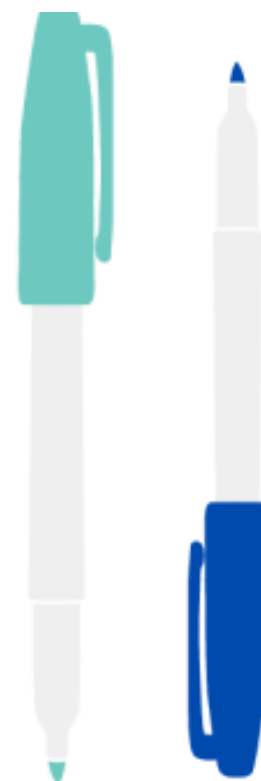
Reading Group

Students and staff who are interested in learning more about Corpus Linguistics and its applications, or in developing and sharing existing knowledge, are invited to join a regular online reading, discussion and training group hosted within Cardiff University's School of English, Communication and Philosophy.

This is held via Microsoft Teams, and can be joined by emailing the author on kavanaghk@cardiff.ac.uk. Papers we have addressed so far this term include Heritage (2020) and Taylor (2013).

PhD blog

Elisa Ramírez Pérez (ENCAP, Cardiff University/Bristol University)

Hello everyone!

My name is Elisa Ramírez Pérez and I am PhD student at ENCAP about to start my 3rd year. As some of you may know, my project falls under the field of English historical linguistics, since I am researching two copies of the Holy Gospels originally written in Latin but which were given Old English glosses (interlinear translations) in the 10th century. Some of you may be familiar with these manuscripts, as they are quite famous: the Lindisfarne Gospels and the Rushworth Gospels, housed at the British Library and Bodleian Library, respectively. These Old English glosses were mainly written in the late Northumbrian dialect, although some sections were glossed in the Old Mercian dialect. My project is, however, interested in the late Northumbrian sections, because previous scholarly research had revealed that the late Northumbrian dialect was much more advanced from a morphological viewpoint than any other 10th century Old English dialect. I started researching this dialect for my MPhil thesis, for which I focused on Matthew's Gospel in both manuscripts. The focus of my MPhil thesis was the morphological status of weak verbs (equivalent to regular verbs in Modern English) in these two texts. Since my MPhil thesis revealed that these verbs were indeed more morphologically simplified in the Northumbrian texts than in texts written in other Old English dialects, I decided to continue this avenue of research for my PhD. And this decision led me to Cardiff University, since there was no better supervisor for me than Dr Sara Pons-Sanz.

I know many of you know Dr Pons-Sanz, and therefore many of you also know how wonderful she is on all possible levels. The reasons I wanted to work with Sara are both academic and personal. Not only is she an expert in the Northumbrian dialect (among other topics), it was also thanks to her that I fell in love with historical linguistics 7 years ago, while I was an undergraduate student at the University of Westminster, where she used to teach before moving to Cardiff. Sara has been such an important person in my (academic) life (and by this I mean since I was a rather lost undergraduate student), I have come this far thanks to her fantastic and inspirational teaching, honest and sound advice, and constant encouragement and kindness. Again, I know many of you have had similar experiences with her.

My PhD journey at Cardiff (and Bristol) has been fantastic thus far. The first thing I welcomed when starting at ENCAP was how friendly and helpful people were. I had had quite a different experience during my MPhil, so it felt comfortable and unthreatening to be around other research students at Cardiff. One of my main regrets thus far has been not being able to develop more relationships and friendships with my fellow PhD students at ENCAP. For most of my PhD I have been living either in London or Bristol, so I did not spend much time in Cardiff where I could get to know other students and socialise. The times I have been in Cardiff, I have been barely able to socialise because of academic pressures, teaching times, and so on. This has meant that my PhD journey has been quite solitary thus far, in the sense that I have not had much time or opportunity to share time and experiences with other PhD students. Solitary but never lonely! Anyone who has seen me around knows that I am quite sociable and open, and like to talk (maybe too much!) with people. But academic priorities rule (and time lost in commuting does not help), therefore I have not had much opportunity to socialise with you, which I regret. I am trying to become more effective when working, and make time for socialising, now that it is possible to be in touch online!

Before finishing off, I would like to share one more aspect of my development as a researcher which is perhaps helpful to new PhD students. I remember while I was doing my MPhil (which lasted 9 months) I looked forward to my PhD hoping that I would have lots of time and opportunities to learn more about the topic I was so passionate about. While it is true that I have now had

much more time to read books and articles than I had in my limited time during the MPhil, it is also true that time flies. I sometimes cannot believe I am about to start my final year of PhD, and while I have made progress and have acquired new skills and competences in these last two years, I still have the feeling I have not learnt enough, and have not become as good a researcher as I thought I would. Expectations are good and bad. They drive you to become a better 'you', but they can also be psychologically draining if you do not meet them in the end. I am still trying to strike a happy balance in terms of what I expect of myself. Talking to friends, supervisors and even my internal reviewer about my expectations, successes, failures and fears has helped me greatly to manage my expectations and goals. If you feel comfortable sharing these with other people (especially

people who have been in your shoes before), I would encourage you to do so. We are really fortunate to have great people at ENCAP, not only from an academic point of view, but also from a human one, so I suggest you make the most of it!

This is all from me for the time being! I hope we can 'meet' in the next few months, whether online or in person.

All the very best for your PhDs,

Elisa

ps/ And now the best of this blog entry: pictures from the two beautiful manuscripts I am working with: the Lindisfarne Gospels and the Rushworth Gospels. Enjoy!
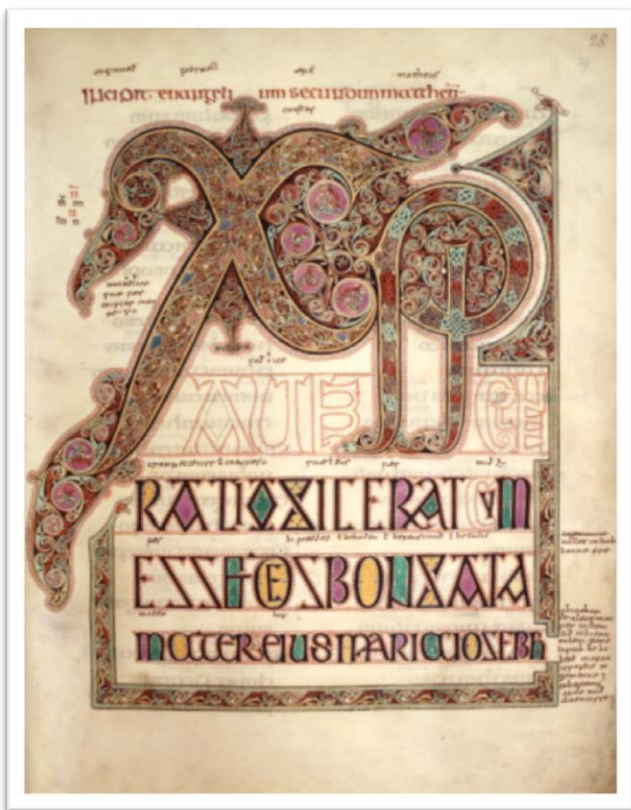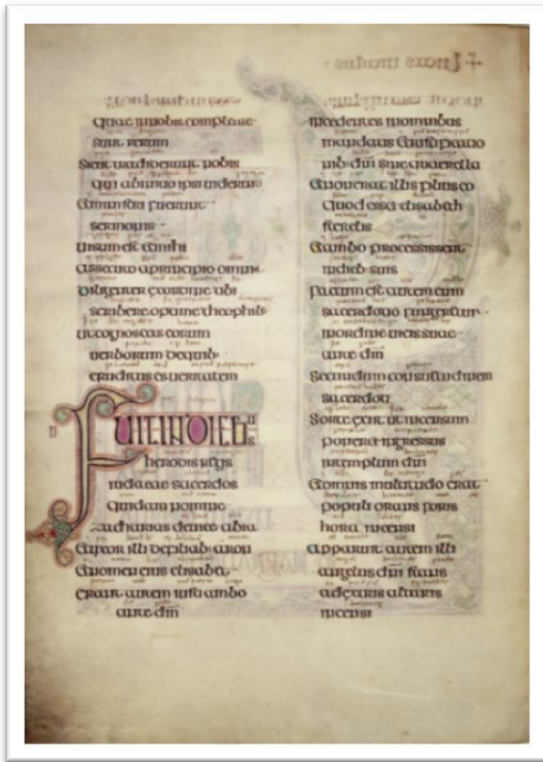

< Chi-rho page, Lindisfarne Gospels, f.29.r:
http://www.bl.uk/manuscripts/Viewer.aspx?ref=cotton_ms_nero_d_iv_fs001r

Marcus miniature, Lindisfarne Gospels, f.93.v: >
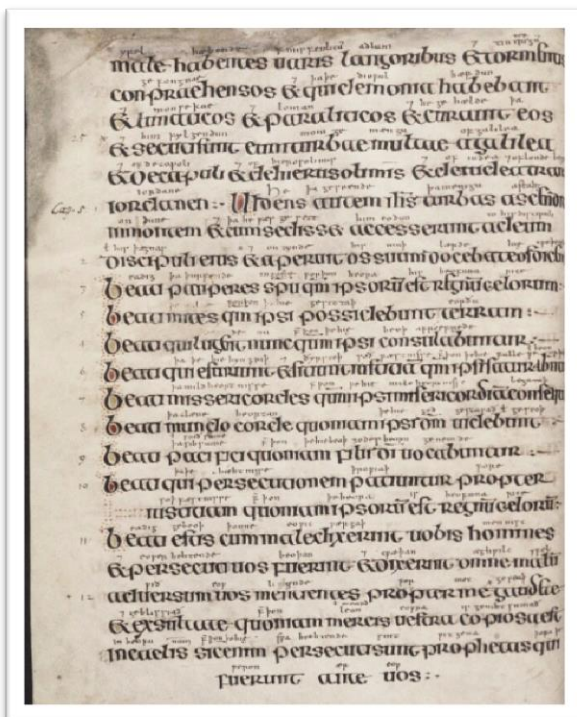http://www.bl.uk/manuscripts/Viewer.aspx?ref=cotton_ms_nero_d_iv_fs001r

< Latin text with interlinear, Old English glosses, Lindisfarne Gospels, f.139.v:

http://www.bl.uk/manuscripts/Viewer.aspx?ref=cotton_ms_nero_d_iv_fs001r

John miniature, Rushworth Gospels, f.126.v:
https://digital2.bodleian.ox.ac.uk/objects/a89f4ed5-d4e7-4d80-9966-1f4e4a3c8b82/

< Latin text, interlinear Old English glosses, Rushworth Gospels, f.006.v:

https://digital2.bodleian.ox.ac.uk/objects/a89f4ed5-d4e7-4d80-9966-1f4e4a3c8b82/