Spring 2021 | Issue 4

U-Lingua

The Undergraduate Linguistics Association of Britain's Quarterly Magazine

In The Field

Linguicism:
Linguistics' Problematic
Past and Present

A Hands-On Approach

Linguistics x Mathematics:
An Interview with Alex Bellos

The Real World

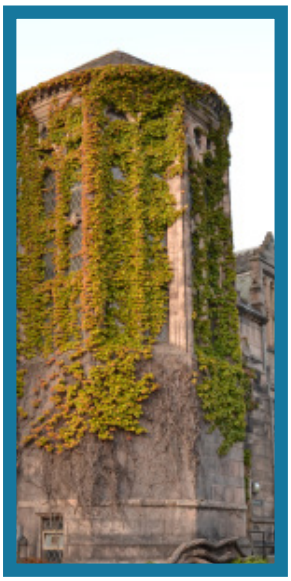
In Conversation with:
Speech Language
Therapist Eleanor

Behind The Bookshelves

Building Your Own Twitter Corpus:
A Step-by-Step Guide

ULAB

Contents



Editor's Letter

In this, our final issue of U-Lingua as an Editorial Team (before the election of the next Editor-in-Chief!), we have a fantastic mix of articles and interviews for you to enjoy! From an interview with Alex Bellos, author of *The Language Lover's Puzzle Book* to a series of articles discussing the intersection of racism and language, and from insights into being a Speech Language Therapist to instructions on how to build your own Twitter-based corpora, we've really got it all. While this academic term, like the last, has been very difficult for us all, the whole Editorial Team has worked diligently and, as always, I remain very proud and grateful for their efforts.

Almost a year ago, we first had the idea for a student-run, UK-based, language-themed magazine. Since, it's not unreasonable to say our little idea has flourished and grown into a highly impressive enterprise. We've started to establish a readership, fostered a reputation for exciting, relevant, and thoughtful content, and begun a project that has enormous potential for the future. As this is the last time I, as Editor-in-Chief, will write to you, our reader, I want to express my gratitude for your support. Thank you for clicking, for reading, and for coming along with us on this journey. We work so hard to ensure the quality of what we output, from editing to designing, and we're so happy you come back every three months.

If you've particularly enjoyed following U-Lingua since its inception, have a passion for language, and think a career in journalism might be something you'd enjoy, I thoroughly recommend that you get involved at the magazine. The new Editor-in-Chief will be elected this April at the Annual General Meeting of ULAB at the 2021 conference, so if you think you'd excel at leading us for the next year, you must put your name in the hat too! Anyone interested in joining U-Lingua is welcome to and encouraged contact me personally (details can be found on the ULAB website - see page 37).

T. R. Williamson
Editor-in-Chief, *U-Lingua*
Archivist, *Undergraduate Linguistics Association of Britain*
University of Cambridge

I Camembert-lieve it's the fourth issue of U-Lingua already! These last few months have mainly been spent preparing for the first ever online ULAB conference, planning some more epic online events and working hard to get all of the JoULAB submissions reviewed. As this is my last U-Lingua contribution as National Chair I'd like to say a huge thank you to all of the 2020-2021 ULAB National Executive Committee, the Local Committee, the Institutional Representatives Board, the U-Lingua Editorial Team, the JoULAB Editorial Committee, and all of the reviewers both for JoULAB and the conference. Your hard work and enthusiasm has made my job a whole lot easier and I brie-lly appreciate it! We've achieved so much this year despite everything and I couldn't be prouder of this team.

At the conference we will be holding our Annual General Meeting in which we elect a whole new ULAB National Committee, Board of Institutional Representatives and a host institution for the ULAB 2022 conference. If you're an undergraduate linguistics student and you think you'd enjoy being part of the ULAB committee then you should definitely consider running - just make sure you've registered and paid to attend the conference and you turn up to the AGM on the Sunday! I a-brie-ciate you reading this far and hope you enjoy this amazing magazine.

Cliodhna Hughes
National Chair, *Undergraduate Linguistics Association of Britain*
University of Edinburgh
Dairy Godmother

ULAB 2021 is almost upon us! I am so proud of all that ULAB has been able to achieve this year, and I'm so excited for our online conference. The Local Committee has been hard at work sorting out the MS Team and Discord platforms that we'll be using. Thank you to everyone who submitted to present at the conference, we can't wait to hear about your research! Aside from the inevitably brilliant student presentations to look forward to, we've got plenary talks, two panels, and a LaTeX workshop in store. We've also got some fantastic socials lined up on the Friday and Saturday, and on the Sunday we have the ULAB AGM. If you want to run for a committee position, or pitch to have your institution host ULAB 2022, that's your chance! Make sure you've registered and paid your conference fee. We can't wait to see you in April!

Beatrix Livesey-Stephens
Local Chair, *Undergraduate Linguistics Association of Britain*
University of Aberdeen

Meet Our Local Co

University o

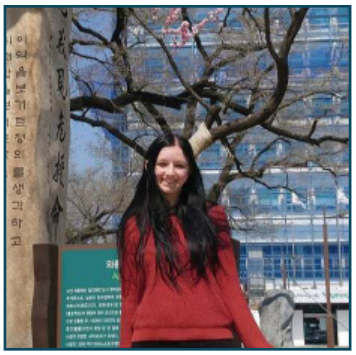
(The people who have b
ULAB2021 the bes

Beatrix Livesey - Stephens *Local Chair*

Beatrix (Bea) Livesey-Stephens is a third-year undergraduate Language & Linguistics student at the University of Aberdeen, where she is the President of their Linguistics Society. As the Convenor of AUSA Disabled Students Forum, she is particularly interested in the areas of Linguistics surrounding disability, such as the linguistics of sign language and Braille. She would also like to work on the ethics of NLP and continue to host UKLO markathons. When not doing Linguistics, she's making new consent workshops for CASE, or knitting.

Emily Lea Heggernes *Local Secretary*

Emily is a third-year student of linguistics at the University of Aberdeen. Her interests within the field include historical linguistics and sociolinguistics. In addition to this, she has a love for music and art. Currently, she is secretary of the University of Aberdeen linguistics society, as well as the local ULAB secretary.



Tina Wolff *Local Treasurer*

Tina is a second-year student in Modern Languages (French, Spanish, Arabic), Translation and Interpretation at the University of Aberdeen. She is particularly interested in neurolinguistics, language acquisition, and multilingualism in relationship to translation studies. She is part of the committee of the Linguistics Society at the University of Aberdeen, and the treasurer of the local committee of ULAB. Her hobbies include gymnastics, rock climbing, dancing, and studying too many languages (Dutch, Chinese, and Japanese), including Skyrim's Dovahzul in case it ever serves as useful.

Emma Johnson *Tech Coordinator*

Emma is a second-year studying International Relations and Language & Linguistics at the University of Aberdeen. She is originally from Germany, but learned Spanish while living in Latin America for a while. Emma is particularly interested in historical and sociolinguistics, however she is enthusiastic about discovering more areas of interest. When not studying she can usually be found at the beach, baking or reading.

T Our
ommittee
of Aberdeen

been working on making
st conference ever)

Sharon Methven
Social Media Coordinator

Sharon is a second-year Gaelic studies and linguistics and language student at the University of Aberdeen where she is on the committee of the Linguistics Society. She is especially interested in multilingualism and how it pertains to language maintenance and shift. Additionally, she is the president of the Mature Students Society.

Stephanie McWilliams *Liaison Officer*

Steph is the Liaison Officer for the ULAB Local Committee. She is a third-year Linguistics student at the University of Aberdeen and is also studying online Japanese Linguistics at Dalarna University. Her particular areas of interest are language acquisition and learning as well as phonetics.

Hafren Vaughan *Easy-Chair Manager*

Hafren, [hafɪən] or [havrɛn], is a first-year undergraduate Language & Linguistics student at the University of Aberdeen. She is too excited about all aspects of linguistics to decide on a particular interest at the moment. If not studying, she may be found learning languages, singing in Japanese – or really just obsessing over Japan. She also enjoys memorising random lists.

Emme Strickland *Ordinary Member*

Emme Strickland is a third-year undergraduate student at the University of Aberdeen studying international relations and linguistics. She is also the treasurer for the Aberdeen University Linguistics Society. Inspired by television personality and author Susie Dent, her favourite field within linguistics is lexicography. In her free time, Emme can be found collecting cinema tickets, mindlessly refreshing Twitter, and engaging in various forms of artistry, such as sketching.

In The Field


BLACK
LIVES
MATTER
WHITE SILENCE = WHITE VIOLENCE
BLACK LIVES MATTER

Linguicism:
Linguistics' Problematic Past and Present

Caitlin Wilson, third-Year MA Linguistics, University of Edinburgh

Nowadays, it may seem that linguistics is a subject open to all that strives to understand human language on a global scale. It brings together people from all over in its quest to fathom how languages are connected. Linguistics aims to make universal statements about the myriad of languages that separate us, so as to unite us. However, it would be foolish to believe that this has always been the case. Indeed, in the past, this subject was used not to unite but to separate even further. Historical comparative linguists, such as 18th century German philologists Friedrich Schlegel and Franz Bopp, had, for a long time, aimed to create a hierarchy of languages, establishing which were the 'best' and 'most complex' and which were 'simple' and 'less sophisticated'[1].

Western linguistics arose against a backdrop of Eurocentrism and colonialism[2]. Further down the line, the subject helped fuel Nazism and white supremacy, creating a clear distinction between us and them; our language(s) and theirs. Indeed, for many years, biological racism was the norm in scientific communities and was used to defend and justify colonialism and slavery. So called 'enlightened thinkers', notably Charles White[3], argued for polygeny, a belief that white and black people are two different species. These beliefs fuelled western linguistics into decades of Eurocentric research aiming to bolster the languages of white Europeans and degrade those of the countries they brutally colonised.

Modern day linguistic discrimination, or 'linguicism' is defined as when "ideologies and structures [...] are used to legitimate, effectuate and reproduce an unequal division of power and resources between groups [...] defined on the basis of language"[4]. Robert B. Moore's classic article, Racism in the English Language[5], breaks down the different types of linguistic racism. These range from obvious bigotry (racial slurs), colour symbolism (white = good/pure, black = bad/unclean), ethnocentrism (otherwise known as white racial framing), qualifying adjectives (e.g., describing a person of colour as "articulate"), and more.

Linguicism is not only about how we perceive languages compared to each other; it can also be found within a language. The promotion of prestige varieties in English, for example, is widely prevalent today across the globe. Varieties deemed more 'native-like' are taught and encouraged in second language English classes all over the world whereas as 'local' varieties are stigmatised[6]. African American English (AAE), one of the most widespread varieties of the language is frequently described as 'lacking grammar' or being 'less complex'[7]. However, as Rickford and King

as well as many others have shown, research has time and again shown that AAE is highly systematic and follows strict grammar rules[8].

Racism consistently proliferates our everyday lives and seems a mainstay in our societal dialogues. This is made apparent in studies of language variation. In certain studies, subjects are made to rate the intelligibility of the speaker as well as to make assumptions about the speaker's ethnicity, level of education, and job (e.g., in studies using the matched guise test)[9]. Stuch research reveals the preconceived ideas we all have about ethnicity, gender, and social class. And indeed, most of these assumptions tend to be based on lived experience. For example, it is indeed true that women tend to have a higher pitch which is why it is often associated with being effeminate and TH fronting (the pronunciation of [ð] [theta] as [f] or [v]) is found more frequently in working class circles rather than middle class settings[10]. The problem arises when we associate these social indications to stereotypes and use them to discriminate against people. A study by Levon found that listeners rate guises with higher pitch as "less competent", thus showing that people have inherent beliefs about women and intelligence[10]. Coupland and Bishop looked into social attractiveness based on linguistic variety in the U.K[11]. They found that Standard English and Standard Scottish English scored highly across the country, but variants described as 'Asian' were systematically placed near the bottom of the table. These attitudes towards language reveal social attitudes that people are less likely to admit having. Such prejudices are fed to us through our education system, our judicial system, and society. This is the very definition of structural racism and linguistic studies highlight that linguistic racism is very much part of it.

It should also be noted that the current global climate has not helped the situation. Recall Donald Trump and Mike Pompeo loudly describing COVID-19 as the "Wuhan/Chinese virus"[12]. Thus ensued a slew of racist attacks on Chinese people (and frankly, other people who seem Asian) across Europe and America[13]. Furthermore, the current pandemic has highlighted a strong undercurrent of vaccine hesitancy in ethnic minority communities in the UK[14]. It is naïve to think that this hesitancy is unwarranted given that people of colour have historically been used as subjects in medical experiments, one of the most famous being the Tuskegee Syphilis Study[15]. We must remember this historic mistreatment of ethnic minorities for many reasons, but especially to highlight that racism and linguicism are deeply intertwined and how difficult it is to decouple the two[16].

This problem is not limited to isolated instances of racism but still appears in linguistic research today. Indeed, David Frawley argues that modern research into Indo-European linguistics continues to attribute a non-Indian origin to India's languages, preferring to attribute it to "their own supposed Proto-Europeans"[17]. We find thus that these white Eurocentric ideas have not fully left the field of research. It us up to us to notice and call out any instances of such racism, be it intentional or not. We should constantly remind ourselves of our aims to bridge gaps between languages and cultures and not further these divides. Linguistics is a field for all.

References:

[1] Rabault-Feuerhahn, P. (2008). L'archive des Origins: Sanskrit, philologie, anthropologie dans l'Allemagne du XIXe siècle. Editions du Cerf: Paris.
[2] Frawley, D. (2007). The hidden racism of linguistics. World Affairs: The journal of international issues (11, 3), 142-152.
[3] Oxford Dictionary of National Biography https://www.oxforddnb.com/view/10.1093/ref:odnb/9780198614128.001.0001/odnb-9780198614128-e-1009488?rskey=igl204&result=1
[4] Skutnabb-Kangas, T. (2015). The Encyclopaedia of Applied Linguistics. Wiley Online.
[5] Moore, R. B. (1976). Racism in the English language. In The Production of Reality: essays and readings on social interaction. Ed: Jodi O'Brien, 119-126. Pine Forge Press: California.
[6] Heller, M. & Martin-Jones, M. (2001). Voices of authority : education and linguistic difference. Ablex: Conneticut.
[7] Slagel, J. (2020). Racism, linguistics and the law. The Oxford Student. https://www.oxfordstudent.com/2020/07/09/racism-linguistics-and-the-law/
[8] Rickford, J. R. & King, S. (2016). Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. Language (92, 4). 948-988.
[9] Kircher, R. (2016). 13 The Matched-Guise Technique. Research methods in intercultural communication, 196. Wiley Online, 597-598.
[10] Levon, E. (2014). Categories, stereotypes, and the linguistic perception of sexuality. Language in Society 43, 539–566
[11] Coupland, N., & Bishop, H. (2007). Ideologised values for British accents. Journal of Sociolinguistics, 11(1), 74-93.
[12] Rogers, K. (2020). Politicians' Use of 'Wuhan Virus' Starts a Debate Health Experts Wanted to Avoid. The New York Times. https://www.nytimes.com/2020/03/10/us/politics/wuhan-virus.html
[13] Townsend, M & Iqbal, N. (2021). Far right using coronavirus as excuse to attack Asians, say police. The Guardian.
https://www.theguardian.com/society/2020/aug/29/far-right-using-coronavirus-as-excuse-to-attack-chinese-and-south-east-asians
[14] Taylor, L. (2021). Fear, mistrust fuel COVID-19 vaccine hesitancy among UK ethnic minorities. Reuters. https://www.reuters.com/article/us-health-coronavirus-britain-vaccines-f-idUSKBN2A925Q
[15] Jones, J. H. (2008). The Tuskegee syphilis experiment. The Oxford textbook of clinical research ethics, 86-96.
[16] De Costa, Peter. (2021). Ask the expert: understanding linguistic racism. MSU Today. https://msutoday.msu.edu/news/2021/linguistic-racism
[17] Frawley, D. (2007). The hidden racism of linguistics. World Affairs: The journal of international issues (11, 3), 142-152.

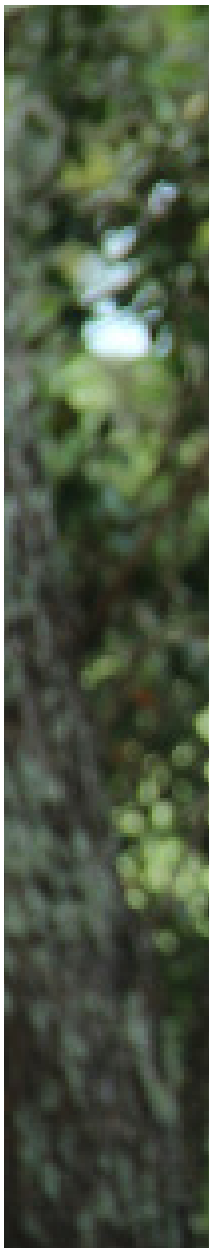
Bottom-Up Approaches in CDA:
Resistance to Māori Racism through Language

Elif Yildiz, second-year MA Linguistics, University of Edinburgh

In early February 2021, the parliament of New Zealand announced that ties were no longer a mandatory piece of business attire after a dispute with a Māori member of parliament who refused to wear one. He was instead sporting a *hei-tiki*, a pendant made of *pounamu* (greenstone) which is considered a *taonga*, the word for treasure in Māori language. His refusal to wear what he called a "colonial noose" is an active form of expression, a non-verbal political statement[1]. Such social (both linguistic and non-linguistic) practices have been studied under the lens of Critical Discourse Analysis (CDA), an interdisciplinary approach which focuses on the concept of discourse and the many intertwined relations of power with which it is dominated.

CDA is a paradigm adopted by idealist, often politically engaged scholars who seek to solve societal problems by deconstructing and hence exposing the hegemonic relations which cause them. But most of them, by only looking into the top-down manifestations of those power dynamics actually contribute to perpetuating the very things they 'critically' research. Indeed, there is a wide gap in the CDA literature on bottom-up relations of dominance. Most studies focus on its institutionalisation and mediatisation through text analyses of news or education policies while omitting 'bottom-up' relations of resistance, compliance, and acceptance[2][3].

This gap is not just found in CDA theory but also in its methodology. There are relatively few studies which ask participants from minority ethnicities and from stigmatized groups or indigenous language speakers their opinion and experience of power relations. Yet this discrepancy perpetuates the inequalities that CDA exposes and criticizes. In an attempt to give pride of place to bottom-up approaches, this article looks at two studies focusing on bottom-up responses to racism directed towards Māori people of New Zealand, and their language: te reo Māori.



The first study showcases a piece of research which asks Māori participants to account for the racism directed at them, and particularly to express their thoughts on why such racism occurs[4]. The socio-cognitive approach developed by Teun Van Djik helps to shed light on the agency of knowledge, the role that awareness of existing racism in society plays in shaping one's discourse, and mostly the close ties between ignorance and the perpetuation of race and racist talks[5].

The second study displays how L2 acquisition of te reo Māori by Pākehā, non-Māori White New-Zealanders of European descent, leads them to inadvertent critical awareness of racism[6]. The acquisition of an indigenous language can never be separated from wider socio-historical concerns of colonialism. Especially when the acquisition is sought by colonial settlers themselves, it is essential to frame those interactions within that greater context all the while staying away from simplistic and generalising tendencies. Hence the dialectical-relational approach initiated by Norman Fairclough provides us with a framework to observe the dialectical relations between structure and action in the acquisition of an indigenous language[7]. Structure here refers to the social structure of the 'real world' in which Pākehā people and Māori people are different. But through the notion of agency, we understand that by lear-

ning te reo Māori, Pākehā learners fight against racism, albeit inadvertently.

Both of these studies pertain to Kaupapa Māori Theory, which is a framework, a modus operandi, and a philosophy of doing Māori research, related to "being Māori" thereby respecting its principles and values. Furthermore, Kaupapa Māori takes for granted the legitimacy of te reo Māori and is concerned with "the struggle for autonomy over our own [Māori] cultural well-being"[8].

Accounting for racism against Māori in Aotearoa (the Māori designation for New Zealand) by Pack, Truffin and Lyons[9] exhibits how racist talks are discursively constructed both by Pākehā people, often referred to as the majority group, and by Māori minorities in the sense that their account of racism does not merely reflect an underlying mental schema but constructs racism[10]. The main reasons participants give for the racism they experience is the fact that racism is mediatised and institutionalised in New Zealand. However, another consensus among participants is the bottom-up explanation which states that people who are racist are "ignorant". Indeed, most participants described racist Pākehā as ignorant people who "don't know/understand/realize", thereby qualifying their racism as non-intentional. This notion of ignorance that participants referred to shows that such people diminish and construct racism as merely a judgement levelled at Māori people based on the assumption that there

exists a level playing field[11].

One participant said: "they seem to think that everyone is born on a level playing field, but they don't understand the racism that Māori children have to grow up in". Racist discourses are thus constructed as though, because the Māori socioeconomic status is lower than the Pākehā one, it is their own fault and their lack of individual achievement, which they consider to be morally deplorable thereby justifying their racism with a neoliberal rhetoric. Interestingly the participant who accounted for this is herself a Pākehā woman married to a Māori man. Yet she used the referent "they" instead of "we" to mark a separation due to her experiential awareness of racism (perhaps through her children).

Van Dijk[12] explains this type of account with the psycholinguistic model of social cognition which he defines as the mental construction, after witnessing an event, of a unique model of that situation. Such models are organized by a schema featuring categories such as Setting, Circumstances, Participants, Event – all of which are possibly accompanied by an evaluative modifier. These categories then reappear in the semantics of the sentences they use in discourse because that social cognition precisely plans their discourse. It is thus important when analysing text to remember that participants map their own biased recalls onto the general social structures they are describing. Another reason given for this ignorance is the lack of social contact between Pākehā and Māori, reinforced by media misrepresentation. Here we see a potential advantage of the self-building nature of discourse. The Māori participants of this study talk of Pākehā's ignorance and lack of education, suggesting that if they were better educated, they would be less racist. These discursive accounts thus display the potential of bottom-up responses to give impetus for change.

Similarly, the study featuring Pākehā learners of Māori shows how L2 acquisition of the indigenous language of Aotearoa by colonial settlers is a bottom-up initiative to overcome racism, despite this not being the goal for learners at first[13]. The growing efforts of indigenous language revitalisation around the world since the beginning of the 21st century go hand in hand with the growing attention to struggles for democracy and self-determination[14]. In fact, in the past twenty years 30 'Truth and Reconciliation Commissions' were established around the world"[15]. Perhaps individual efforts from CSL2 (Colonial Settler Second Language) learners can thus be considered as attempts of remembrance and even apology, paving way for better bicultural relations.

A participant reported that the biggest learning challenge for him was that "because [he is not] Māori, [he] struggles to [...] feel like [he] was allowed to learn". Interestingly, similar feelings are reported by non-autochthonous learners of Scottish Gaelic who feel unsure about their linguistic authenticity and appropriateness especially in an environment where they are among Gaels[16].

Yet, Te Huia[17] posits that it is specifically through the discomfort of being a Pākehā in the Māori dominant environment that is a Māori language class, that those Pākehā can truly learn the sources of that discomfort. Huygens[18] argues that after being exposed to some contact with the colonised, settlers may experience "the same emotions indigenous groups experience when undergoing critical conscientization: anger and blame at how much has been hidden".

While this critical awareness may have been initiated inadvertently (as reasons to start learning an albeit minority language are varied), this inadvertent critical awareness can be transformed into an active strategy which would lead

Pākehā to encourage fellow Pākehā to read and learn about Māori racism. Gradually working its way up to the civil society, this awareness can then help put pressure on political actors to reform the institutions and the media for them to acknowledge, and truly apologize for all the wrongs done and the continuing racism on the Māori people of Aotearoa.

References:

[1] BBC (2021). Maori MP ejected from NZ parliament for refusing to wear a tie. BBC News. https://www.bbc.co.uk/news/world-asia-55996688
[2] Van Dijk, T.A. (1993a). Elite Discourse and Racism. Newbury Park, CA: Sage.
[3] Van Dijk, T.A. (1993b) Discourse, Power and Access, in C.R. Caldas (ed.) Studies in Critical Discourse Analysis. London: Routledge.
[4] Pack, S., Tuffin, K., and Lyons, A. (2016). Accounting for Racism Against Maori in Aotearoa/New Zealand: A Discourse Analytic Study of the Views of Maori Adults. J. Community. Appl. Soc. Psychol. (26), 95-109.
[5] Van Dijk, T. A. (2008). Discourse and context: A Sociocognitive approach. Cambridge: CUP.
[6] Te Huia, A. (2020). Inadvertent critical awareness and Māori language learning in Aotearoa New Zealand: Pākehā learner perspectives. Settler Colonial Studies (10,2), 257-277.
[7] Fairclough, N. (2001). Critical discourse analysis as a method in social scientific research. In R. Wodak & M.Meyer (eds.), Methods of critical discourse analysis, 121–138. London: Sage.
[8] Jackson, A.-M. (2015). Kaupapa Māori theory and critical Discourse Analysis: Transformation and social change. AlterNative: An International Journal of Indigenous Peoples (11,3), 256-268.
[9] Pack, S., Tuffin, K., and Lyons, A. (2016) Accounting for Racism Against Maori in Aotearoa/New Zealand: A Discourse Analytic Study of the Views of Maori Adults. J. Community Appl. Soc. Psychol. (26), 95-109.
[10] Potter, J. (1996) Representing Reality: Discourse, Rhetoric, and Social Construction. London: Sage.
[11] Potter, J. and Wetherell, M. (1987) Discourse and Social Psychology: Beyond Attitudes and Behaviour. London: Sage.
[12] Robinson (eds) Handbook of Social Psychology and Language. 163-83.Chichester: Wiley.
[13] Te Huia, A. (2020). Inadvertent critical awareness and Māori language learning in Aotearoa New Zealand: Pākehā learner perspectives. Settler Colonial Studies (10,2), 257-277.
[14] McCarty, T. (2003). Revitalising Indigenous Languages in Homogenising Times. Comparative Education (39,2), 147-163.
[15] Whitlock, G. (2006). 'Active Remembrance: Testimony, Memoir and the Work of Reconciliation', in Rethinking Settler Colonialism, ed. A.E. Coombes 24–44. Manchester University Press: Manchester.
[16] McLeod, W., & O'Rourke, B. (2015). „New speakers" of Gaelic: Perceptions of linguistic authenticity and appropriateness. Applied Linguistics Review, (6,2), 151-172.
[17] Te Huia, A. (2020). Inadvertent critical awareness and Māori language learning in Aotearoa New Zealand: Pākehā learner perspectives. Settler Colonial Studies (10,2), 257-277.
[18] Huygens, I. (2011) Developing a Decolonisation Practice for Settler Colonisers: A Case Study from Aotearoa New Zealand. Settler Colonial Studies (1,2), 53-81.

Behind The Bookshelves


var b,d=this,e=this
undelegateEvents
toggleClass
previewDeviceButtons
keyEvent:function
maybeRequestFile
Backbone.View.extend
listenTo(c.collection
announceSearch
function(){c.overlay
render:function

Building Your Own Twitter Corpus: A Step-by-Step Guide

Imogen Davies is currently studying for a Linguistics MA (with a specialisation in Formal Linguistics) at Queen Mary University of London. Last year she graduated with a Linguistics BA from the University of Cambridge, where her dissertation research on the Adjective + 'as' construction in colloquial English involved creating a corpus of tweets in order to find and analyse examples of the construction. Here, she provides a step-by-step guide to building a corpus using Twitter for linguistics research.

When starting my undergraduate dissertation research on a particular construction in colloquial UK English, I realised that even though I'd heard it being used frequently, it was very difficult to find evidence of it in traditional corpora like the British National Corpus. I ended up turning to social media data – where people's writing tends to closely replicate the way they speak[1] – to create my own corpus. Creating a corpus using Twitter is easier than you might think, and you'll end up with a fantastic resource for research, whether you're interested in dialectal variation, discourse analysis, or computational linguistics. Read on to find out how to do it!

Step 1: Check Twitter will be suitable for your research

The first thing to do before you build your own corpus of tweets is to run some preliminary searches on Twitter for keywords related to the topic you want to study, to check that Twitter will be a good source of data for your corpus. There are about 500 million new tweets every day, so Twitter isn't going to show you every single tweet that matches your search terms, but searching is a good starting point to check that Twitter will have a decent range of examples of the phenomenon you want to study.

Top Tip:

If you're considering using your corpus to study particular linguistic constructions, try and get creative with Twitter's Advanced Search to verify that there are individual tweets with examples of the structures! I needed my corpus to contain examples of the word "as" being used after adjectives as a stand-alone intensifier, so my first step was searching for tweets that did contain the exact phrase "funny as" but didn't contain any of the many, many swear words that might come after "as".

Step 2: Apply for access to Twitter's live stream of public tweets

Once you're happy that Twitter is going to have data that's of interest to you, the next step is to apply for an account on Twitter's 'Developer Portal'. This can be linked to a personal account you already have, or you can set up a brand-new one. The application just involves answering some questions to check that you won't use your developer account for anything illegal or unethical. It's definitely beneficial if you already have a good idea of what research you'll do with your corpus so you can answer the questions thoroughly. After a few days, Twitter will approve your account. You'll then be given access keys for the 'Streaming API'. These are what allow you to connect to Twitter's live stream of all public tweets being sent.

Remember:

If you're creating a corpus for a dissertation project or something similar, make sure you check whether you also need to get ethics approval from your university for collecting and storing the Twitter data.

Step 3: Choose the types of data to collect

This is where it gets exciting, as you now get to decide exactly what kinds of data you'll be including in your corpus. Using the basic code that Twitter provides for connecting to the stream, you'll automatically just get the text of the tweet and its unique ID number. However, there's tons of other information associated with each tweet you collect, ranging from the time and location at which it was sent, to the number of likes and retweets it had when you collected it, to the username and bio of the user who sent it. Some of this information might be useful for your corpus – for example, if you want to make comparisons about language use between different geographical areas, or different dates within the timeframe that you're collecting tweets. If so, you can adapt

the basic code to make sure that when you connect to the live stream, you collect whatever information you need along with each tweet. It's totally possible to do this even if you're not familiar with coding, using the tutorials on the Twitter developer platform.

Step 4: Create filters

The other thing you need to plan at this stage is how to filter the stream. You almost certainly don't want to download absolutely any public tweet that's being sent, so you'll also want to adapt the code to include filters. When I built my corpus I was specifically interested in the use of the adjective + "as" construction in UK English, so I used location filters, language filters, and a keyword filter so that I only got UK tweets in English that contained the word "as". You could filter in lots of other ways too, like only getting tweets that are replies to other tweets if you wanted to look at conversational interactions, or only getting tweets from verified users if you wanted to see how celebrities and large companies use language on Twitter. The possibilities are pretty much endless!

Step 5: Connect to the stream

Once you've adapted the code, you're ready to use your access keys to connect to the stream and start collecting tweets! The text of all the tweets you're collecting (along with whatever additional data you've requested) will be saved as a .csv file on the device that's streaming.

Top Tip:

It's worth checking whether there are any short-term laptop loan schemes at your university that could get you temporary access to an extra laptop to use for streaming. This isn't essential, but does make it easier to do the streaming while juggling whatever other things you need to do on your own device.

Step 6: Start working with your corpus of tweets

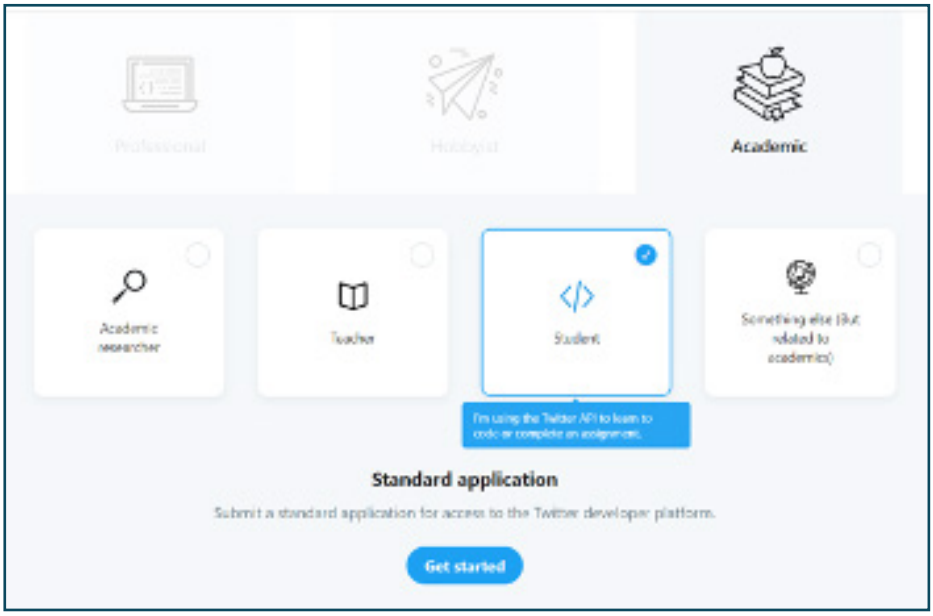
You can start looking at the tweets you've already collected while you're still connected to the stream. Definitely do this straight away, so you can check that your filters are working as you'd hoped, and see how quickly your corpus is growing. (As a guide, when I was filtering the stream to UK and Ireland tweets containing "as", I collected a little over 10,000 tweets per day.) Once you have enough tweets for your research purposes, you can simply disconnect and get to work exploring and analysing your corpus data! Personally I sorted through my data mostly using some tricks in Microsoft Excel, but there are lots of more sophisticated methods out there. If you want to try these, I'd recommend looking at the 'Social Media Research Toolkit' online[2], which has lots of free options for understanding and analysing Twitter data.

While this method is powerful and simple, it does have a shortcoming in that you can only collect tweets that were sent at or around the time that you are streaming, so you won't be able to get historical data this way. If you want a corpus of tweets from a particular time period, your best bet is to check the 'DocNow Catalog'[3], which has over 100 different free corpora of historical tweets. These are mostly focused around specific political events and movements in different countries - for example, the US 2020 election[4], Ireland's 2018 referendum[5], and the #BlackLivesMatter movement in the 2010s[6]. If you want to study how some aspect of language has changed over time, you could even use one of these historical corpora in addition to the streaming method to give you two time periods to compare.

Twitter (and social media more generally) is rapidly being recognised as a valid and highly useful source of data for many different subfields of linguistics, with recent Twitter linguistics research covering topics as diverse as sociophonetics[7], automatic sarcasm detection[8], discursive styles of politicians[9], and Welsh dialect syntax[10]. Creating your own corpus of tweets is a really exciting way to start investigating any of these topics - not to mention, a great way to turn scrolling through social media into a productive part of your degree. To anyone who gives it a go, good luck and happy corpus-building!

References:

[1] Tagliamonte, S. & Denis, D. (2008). Linguistic ruin? Lol! Instant messaging and teen language. American Speech, 83(1), 3-34.
[2] Ryerson University Social Media Lab. (2021). Social Media Research Toolkit. Retrieved from https://socialmediadata.org/social-media-research-toolkit/.
[3] Documenting the Now. (2021). DocNow Tweet Catalog. Retrieved from https://catalog.docnow.io/.
[4] Chen, E. & Ferrara, E. (2021). 2020 US Presidential Election. https://github.com/echen102/us-pres-elections-2020
[5] Littman, J. (2018). Ireland 8th amendment referendum vote. doi:10.7910/DVN/PYCLPE
[6] Giorgi, S., Guntuku, S. C., Rahman, M., Himelein-Wachowiak, M., Kwarteng, A. & Curtis, B. (2020). Twitter Corpus of the #BlackLivesMatter Movement and Counter Protests: 2013 to 2020. https://zenodo.org/record/4056563#.YEFtdmj7RPY
[7] Tatman, R. (2015). #go awn: Sociophonetic variation in variant spellings on Twitter. Working Papers of the Linguistics Circle of the University of Victoria, 25(2), 97-108.
[8] Cai, Y., Cai, H. & Wan, X. (2019). Multi-modal sarcasm detection in Twitter with hierarchical fusion model. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2506-2515.
[9] Breeze, R. (2020). Exploring populist styles of political discourse in Twitter. World Englishes, 39(4), 550-567.
[10] Willis, D., Leemann, A., Blaxter, T. & Gopal, D. (2021). Tweetolectology: Investigating the diffusion of morphosyntactic innovations using social media. Retrieved from http://tweetolectology.com/index.html.



The page you'll see when starting your application for a Twitter developer account, which gives you access to the live stream of public tweets.

Improving Probabilistic Models for Fake News Identification:

A Review of *Linguistic Indicators in the Identification of Fake News* by Dorota Marquardt in *Mediatization Studies* (Vol. 3, 2019)

Katie Shi, an undergraduate studying Quantitative Social Science and Linguistics at Dartmouth College, discusses a 2019 paper by Dorota Marquardt which aims to develop and improve tools for automated fake news identification, using qualitative and quantitative corpus analysis.

In many ways, the year 2016 marked a new global era of post-truth politics, where appeals to emotion and personal beliefs now seem to hold more power over public opinion than objective facts. This political culture has led to an exponential proliferation of 'fake news,' a phenomenon that continues to be driven by the growing influence of social media platforms as news sources[1][2]. Linguistic Indicators in the Identification of Fake News[3] aims to identify fake news in text by studying its linguistic features. Here, 'fake news' is defined as untrue information which can be verified to be false, and has the intent to mislead. Marquardt's paper is split into two sections. The first section is a quantitative analysis of two text corpora: one with real news articles, and one with articles deemed to be fake news. The second part is a sentiment analysis of both corpora, using the results of the quantitative analysis. Finally, the author discusses the efficacy of a probabilistic model for recognising truthful information, based on their findings.

Marquardt[3] presents two different text corpora for analysis. The first is a corpus of 30 real news articles which had been originally published by the Polish Press Agency (Polska Agencja Prasowa, PAP). The PAP is Poland's national news agency, and it's also the largest news source in the country. When creating this corpus, the author assumed that the probability of fake news articles from the PAP would be lower than from other publishers, given its reputation and status. The second corpus contains 30 fake news articles that students had collected during an internet journalism workshop at the University of Economics in Katowice. During the workshop, the students in this workshop found real and fake internet news pieces and compared the structure of each kind. The examples in the second corpus, then, are considered most representative of fake news. The criteria for an article to qualify as 'fake news,' however, are unclear.

Each text from both corpora was analysed for frequency according to the following variables: keywords, verbs, nouns-to-verbs ratio, and number of words per sentence. The keywords of an article were defined as the nouns, adjectives, pronouns, and other parts of speech most relevant to the article's subject. In table 1 are three articles from the real news corpus:

Article	Nouns, adjecti… pronouns, et…
"Kaczyński: strong West Pomerania is in the Polish national interest"	Kaczyński (5), str… (4), very (4), Ger… (4)
"Canadian Prime Minister on the Polish festival in Toronto: diversity is the source of strength"	Polish (20), Cana… (14), festival (11), Poland (10), Toro… (10), Prime minis…
"Cardinal Bagnasco: Europe should follow the path towards unity"	Europe (13), unit… church (6), Bagna… (5)

Keywords in t…

The author found that, on average, real news articles had a higher nouns-to-verbs ratio than fake news articles — an average of 4.27 for the real news corpus compared to 2.73 in the fake news corpus. Real news articles also had more words per sentence than fake news articles.

Marquardt highlights that a major component of the recent shift in news reporting is a

relegation of a statement's truthfulness for its emotional appeal with an audience. In natural language processing, sentiment analysis is a common text categorisation task where the sentiment - the positive or negative orientation of a text — is extracted[4]. For the sentiment analyses of the Marquardt (2019) corpora, the factors considered were the dominant positive and negative emotions of each article, the total number of words per text, and difficulty class (ranked from most to least simple on a point scale of 1-7). One article from the real news corpus, for example, had 28 positively-associated words and 16 negatively-associated words, while the dominant emotions were joy and usefulness (positive) verus sadness and unhappiness (negative). The positive and negative emotions of an article were determined using a sentiment analysis software; the software calculates the number of positively-associated words and negatively-associated words in a text, and the dominant positive and negative emotions are identified as well. The difficulty class was calculated based on the formula of the Gunning fog index, a readability test for English writing. The fog index indicates the (approximate) years of formal education a person needs to understand the text during a first read. The difficulty class for this study is for Polish texts — with a score of 1 indicating a simple text and a score of 7 indicating esoteric, specialist texts — but it's unclear how these levels were determined.

The author found that titles with stronger emotions — either more positive or more negative overall — tended to belong to a fake text rather than a real one. Out of the 30 texts in the real news corpus, only two had more dominant negative notions than positive ones. In the fake news corpus, seven texts were more dominantly negative than positive. As we might expect, real news articles also had details that could be verified (such as an individual's full name, or a cited source) which fake articles lacked. Additionally, while real news articles did have greater text difficulty than fake ones, what's most notable is how text difficulty varied by genre. For real news, political texts were usually simpler (ranked 3 or 4), while scientific and legal texts were usually the most difficult (ranked 6 or 7). Fake news articles didn't have this genre distinction, and no text from the fake corpus ranked above 5 on the difficulty scale. However, no statistical tests of significance were carried out on these findings for sentiment and text difficulty, so it's difficult to tell if these results mean anything more concretely.

ves, c.	Verbs	Noun-to-verb ratio	Mean sentence length (in words)
ong many	be (8), highlight (3), state (2), know (2), speak (2)	1.76	19
da nto ter (9)	be (26), say (8), celebrate (5), highlight (3)	3.16	22
y (8), isco	be (19), follow (3), say (3), highlight (3)	2.96	22

he true news corpus[3]

Based on these analyses, the author argues that the qualitative factors should be incorporated into probabilistic models predicting the level of truth in a text. To increase the probability of information being either true or false, the text complexity level should be analysed — if the text score is 6 or 7, the probability of that text's truthfulness may increase. Higher occurrences of keywords in an article's title, consistent sentiments between an article's subject and text, and the presence of verifiable sources and full names of individuals all correlate with a greater probability of truthfulness for that article. While these are interesting proposals, more articles would need to be included in the corpora, and more rigorous statistical analyses would need to be performed, before such definitive conclusions should be drawn.

In computer science, fake news identification has been an automated process that's concerned more with the diffusion of false texts, as well as the veracity of textual information, than with the linguistic features of the texts themselves[5]. Fake news classification, then, often comes from external comparison — an article's level of truthfulness is based on comparison with other news articles on the same topic. By examining fake news articles on a syntactic and discursive level, Marquardt[3] is able to pinpoint elements to consider when building more precise models for identifying false information.

Overall, the author's analyses of fake news articles are compelling. The factors that have been identified can not only be used in improving probabilistic models, but can also serve as strategies for critical engagement when consuming news media. The author mentions that the findings of this paper apply equally to Polish and English texts. It would be interesting to see how the qualitative and quantitative factors studied in this paper would change based on different languages — especially non-Western ones — and larger corpora that might include more types of texts besides the journalistic form.

References:

[1] World of the Year 2016 (2016). Oxford Languages. https://languages.oup.com/word-of-the-year/2016/
[2] Yerlikaya, T. (2020). „Social Media and Fake News in the Post-Truth Era: The Manipulation of Politics in the Election Process." Insight Turkey, vol. 22, no. 2: 177-96. Accessed March 8, 2021. doi:10.2307/26918129.
[3] Marquardt, D. (2019). Linguistic Indicators in the Identification of Fake News. Mediatization Studies, 3, 95-114. doi: 10.17951/ms.2019.3.95-114
[4] Jurafsky, D. and Martin, J. (2020). Speech and Language Processing, chapter 4. https://web.stanford.edu/~jurafsky/slp3/
[5] Vosoughi, S., Roy, D. & Aral, S. (2018). The spread of true and false news online. Science, Vol. 359, Issue 6380, pp. 1146-1151 DOI: 10.1126/science.aap9559

More from Our Last Issue on "Language and the Mind"

Investigating Phonetic Reduction:

Discoveries in Language and the Mind

Scott James Perry, Ph.D. Candidate, Department of Linguistics, University of Alberta

How do humans use sound to communicate? This is a core question at the centre of linguistic inquiry, and one that can be explored in a number of ways. One line of research in the speech sciences concerns the variability of spoken language, as the same word or sentence is never produced exactly the same way twice. Much of the variability in speech that we encounter in our everyday lives is due to *reduction*. Reduction involves parts of speech being altered or deleted in casual speech compared to how it would be produced by the same speaker enunciating more carefully. One example of reduction in American English is the word *yesterday* being pronounced as *yeh-shay* [**jɛʃeɪ**], which is quite different from the pronunciation of this word as given in a dictionary. Reductions of this sort are pervasive in everyday speech. This is one of those cases where "hearing is believing", so the reader is invited to hear examples of reduction at the Alberta Phonetics Laboratory's website.

The ubiquity of reduction in our everyday speech has led researchers to believe that the nature of reduction is due to how language is represented in the mind[1]. This has consequences for both the production and perception of speech, and we must therefore investigate spontaneous and casual speech to understand how we produce and perceive speech more generally.

As might be evident from the pronunciation of the word *yesterday* given above, reduced speech can be challenging to understand when taken out of context. In real-life conversations, however, these reductions rarely pose an obstacle to effective communication. Many people do not even notice these reductions, possibly because their minds can 'repair' the reduced speech, allowing them to perceive sounds that are not really there. Some reductions, however, are noticed more readily than others. Short, common phrases may be reduced often enough that these reductions become *lexicalized*, i.e., stored in long-term memory as the reduced form. A common phrase that is likely lexicalized for many English speakers is the phrase 'I don't know.' This particular phrase is often written down as 'I dunno,' indicating that people are likely aware of how they reduce this phrase.

While there are many exciting lines of research concerning reduction, one of the most fundamental questions that we as researchers can ask about reduction is '*why do we do it?*'. There have been several accounts of why reduction occurs, and some of these explanations are not mutually exclusive. One well-known account claims that speech follows the principle of economy of effort[2]. In general terms, people only speak with the minimum effort required for other people to understand the words they are saying. This balance leads to the speaker reducing parts of speech when they can do it and still being understood. A corollary of this account is that more predictable parts of speech will be reduced. This follows logically because listeners can exploit this predictability to narrow down which words are likely to have been spoken, even if the speaker was not as clear as they could have been. The *Probabilistic Reduction Hypothesis* (PRH) explicitly posits this prediction at the level of words[1], and says that words with a higher probability of occurring are more likely to undergo reduction, but similar predictions can be made of syllables[3] or multi-word utterances[4].

The PRH can only be true if probabilistic information regarding the relationships between words is stored in speakers' minds. While information regarding word frequency has long been included in models of language, information regarding how likely words are to occur together has only been included more recently. Ample evidence concerning how we produce, process, and perceive speech has shown that probabilistic variables in addition to frequency explain reduction in both experimental and naturalistic settings[1,3,5]. It is important to note here that various types of probabilistic information are represented in the minds of speakers, and that the variables used in any given study only approximate certain aspects of this information. The PRH is concerned with the lexical level, the level of words and how they relate to one another in terms of how often they appear before and after one another. We also have evidence that other types of information regarding relations between linguistic elements are represented in the mind and impact reduction, such as semantic, syntactic, and discourse level relations[6].

While research into reduction has significantly grown over the past 20 years[7], there is still a need for further research that investigates a bro-

ader range of languages. To take Spanish as an example, although it is a relatively well-studied language, research on reduction has generally been limited to the /s/ sound. We are currently undertaking research at the Alberta Phonetics Laboratory that investigates variation in stressed vowels in Spanish. This research involves using the Nijmegen Corpus of Casual Spanish[8], which comprises numerous recordings of spontaneous conversations between friends. The use of large data-sets of spontaneous speech is necessary for three reasons. First, we cannot control the language being produced, and therefore need to collect larger amounts of data to ensure that we have a suitable number of tokens of the sound(s) of interest that we can analyze. Second, as we cannot experimentally control factors known to influence the production of the sounds of interest, we must statistically control for these factors, which further raises the number of tokens needed for analysis. Third, using spontaneous speech as opposed to other forms of speech is important, as the majority of daily communication involves spontaneous speech, and there are notable differences in this casually-spoken speech and read speech.

For vowels, two common acoustic measurements of reduction are vowel duration and distance from the vowel space center. Our preliminary results[9] indicate that more frequent and more predictable words have stressed vowels that are shorter and more centralized in the vowel space, after controlling for the length of words and how fast participants were talking when they produced the word. This finding provides strong support for the PRH, as stressed vowels in Spanish are thought to be relatively stable and be reduced less than in other languages.

Findings of reduction consistent with the PRH inform us about the representations of language in the mind. Our research on Peninsular Spanish indicates that vowels are less acoustically salient when produced in words that are more predictable. These sorts of gradient effects on speech production indicate that the representation of sounds in the mind is more complex than is often assumed. While discrete abstract representations of sounds may or may not have a cognitive reality, we also have increasing evidence that there is a more complex story.

References:

[1] Jurafsky, D., Bell, A., Gregory, M., and Raymond, D., W. (2001). Probabilistic relations
between words: Evidence from reduction in lexical production. In Frequency and the Emergence of Linguistic Structure, Volume 45 of Typological Studies in Language, (Bybee, J. and Hopper, P., Eds.), 229–254.
[2] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Speech production and speech modelling, 403-439. Springer, Dordrecht.
[3] Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. Lang Speech, 47(1), 31-56.
[4] Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. The Mental Lexicon, 6(2), 302-324.
[5] Baker, R. E., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. Lang Speech, 52(4), 391-413.
[6] Clopper, C. G. & Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In Rethinking reduction (Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B., & Zellers, M., Eds.), 25-72.
[7] Warner, N. (2019). Reduced speech: All is variability. Wiley Interdiscip. Rev. Cogn. Sci, 10(4), e1496.
[8] Torreira, F., & Ernestus, M. (2010). The Nijmegen corpus of casual Spanish. In Seventh conference on International Language Resources and Evaluation (LREC'10) (pp. 2981-2985). European Language Resources Association (ELRA).
[9] Perry, S. J., Kelley, M. C., & Tucker, B. V. (2020). Word frequency, predictability, and lexical class influence different aspects of Spanish tonic vowel production. JASA, 148(4), 2474-2474.

The Real World

26 In Conversation with: Speech Language Therapist Eleanor


Décalque au crayon les différentes pièces
porte-nom sur du papier calque. Pour
couleur correspondants, utilise le systè
calque, repasse au crayon sur les con
imprimés sur le carton de couleur. Il ne
(A)
Pour assembler le menu, lis
les explications ci-dessous et
suis les illustrations :
98
le menu

In Conversation with:
Speech Language Therapist Eleanor*

In this edition, Blue Bates Cambridge speaks to a Speech Language Therapist about her experiences in the field over the last 20 years

Blue: It's really lovely to have you speaking to us, Eleanor. Thank you so much for agreeing to do this interview. My first question to you is: could you tell us a little bit about yourself and what you do?

Eleanor: I'm a principal speech and language therapist in North London. The area is very diverse in terms of ethnicity, sociology, economic status and the languages spoken, which makes my job and that of the team both interesting and challenging. The Speech and Language Therapy team is based with colleagues from occupational therapy and physiotherapy which is great as it encourages therapists to talk, work and learn together. I'm responsible for a team of about 40 therapists who provide intervention to children and families, children centres, community clinics and schools.

Blue: I'm curious to know: what inspired your interest in speech, language and communication?

Eleanor: I was one of those children who always wanted a job helping people. I looked at a whole range of caring professions including nursing, OT and physio. I wasn't even really aware of SLT as a profession and came across it by accident. I managed to make contact with an SLT and arranged to sit in on some sessions and that was when I decided that it was the direction I wanted to take. Helping people to talk and communicate just seemed like a great thing to be able to do.

Blue: I saw that you studied Speech Sciences at University College London. What kinds of things did you learn about at university and how did that prepare you for the field of Speech Language Therapy?

Eleanor: My time at UCL seems a very long time ago and one of the things I really enjoyed about the degree was that it touched on many different subjects – from linguistics to anatomy and physiology to statistics to specific modules on speech and language disorders in adults and children. And, of course, there were the student placements which in my memory were more about observation than working directly with patients.

Blue: You've worked within the NHS for a long time. How has your job changed since you first joined?

Eleanor: Things have changed hugely since I started work back in the early 90s. For a start, we give all our staff an induction programme, whereas on my first day at work I was told that I would be the one carrying out bilingual assessments as the therapist who was supposed to be doing them was ill! It was a real baptism by fire. Therapists are now provided with support and supervision in a much more structured way.

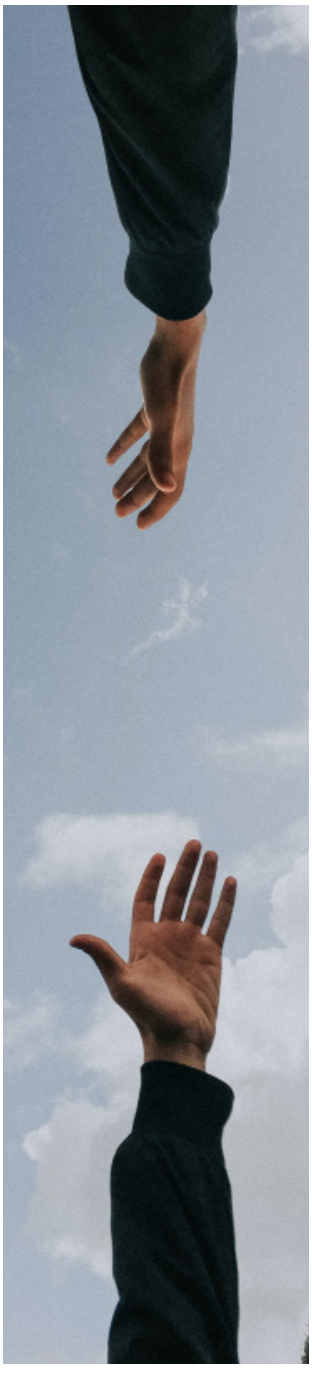
Now we think more about what works in terms of intervention, using evidence-based practice, and there is much more emphasis on ensuring parents, teachers and others are able to support children's speech language and communication needs. In the past the therapist was often seen as the only one who was able to bring about change – now it's much more of a team effort with the parents as integral part of that team

Blue: I have to admit, I'm a little out of touch with the more technical ins and outs of speech language therapy. However, in the time you've been working as an SLT, have there been any major breakthroughs in the field? Any new theories, methods or ideas which have shaken up the way you operate?

Eleanor: There are few that spring to mind: Hanen or Parent Child Interaction Approaches and Intensive interaction. These approaches now form the basis for intervention to pre-school children and those with social communication difficulties. In school-aged children with language disorders, approaches such as shape coding and Colourful Semantics are useful interventions, these show how words are put together into sentences so that children can communicate more effectively.

Blue: The impossible question: if you can give even a rough approximation, what does a day in your life look like?

Eleanor: As the manager of the service, my days are probably very different to that of one of the therapists. I can give you a flavour of my week. So, over this week I have delivered



training to the new community paediatrician registrars with my OT and physio colleagues, I've attended the SEND assessment panel. I've also had a budget meeting and supported a therapist in the writing of a report for a tribunal. We interviewed for and appointed two new therapists who we hope will start with the team in a couple of months' time. And lastly, I have visited a child in school to review his speech, language and communication needs. There have, of course, been numerous emails and telephone calls to respond to!

Blue: The global pandemic has affected almost every facet of our lives, fundamentally changing the way we communicate with one another. How has COVID-19 and the consequential shift to 'work-from-home' affected your role? How have you and your colleagues adapted to the shifting demands of the pandemic over the past year?

Eleanor: The pandemic has been an interesting time. Back in March last year we stopped seeing children and families face to face and moved to telehealth. There were a few exceptions, and those were children with feeding and swallowing difficulties.

Telehealth has definite pros and cons. It can be difficult for some of the families in the area we working to access telehealth for a whole number of reasons. There can be language barriers, and they may not have access to Wi-Fi and/or have a suitable device.

On the plus side, working with some of our children with ASD or social communication difficulties via virtual appointments has been helpful. It has meant that they don't have to come into clinic (which can be a strange and noisy place), they are in their own home environment in a comfortable environment, and we can see how the child communicates in that setting. We are able to see the sort of toys and activities that the child likes to do at home, and we can coach the parents using Parent Child Interaction Strategies.

I think our therapists understand the benefits and limitations of telehealth and have enjoyed the opportunity to work at home, but they also enjoy (and miss) coming into the base where they can meet other therapists and discuss things like the children and their day! This is when a lot of informal learning goes on. So, in terms of moving forward, we need to think carefully about the balance between home-working and telehealth for both the children and families we work with and for the staff.

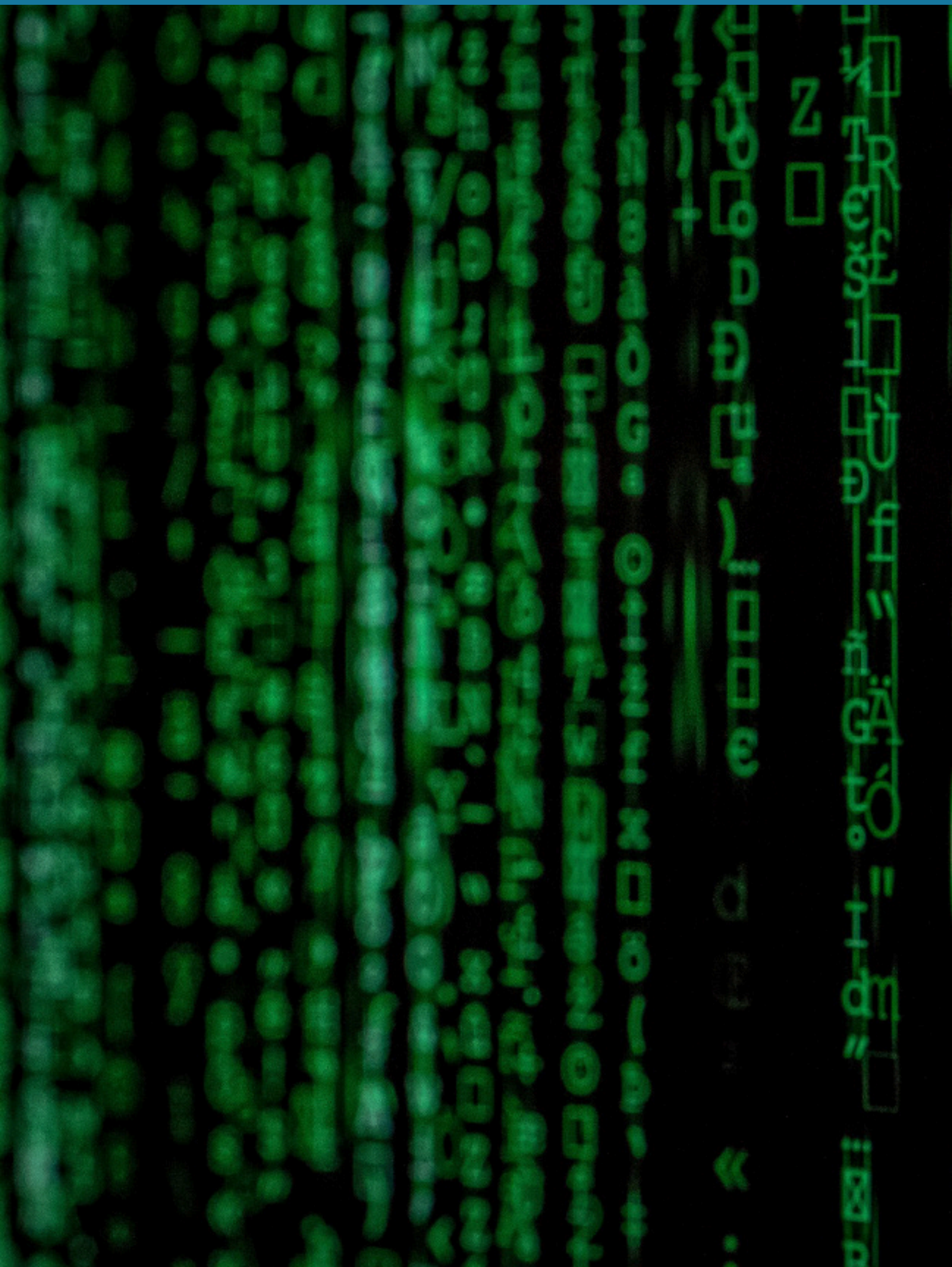
Blue: I can imagine that, within your field, there are many different specialisms under the umbrella of SLT. Which areas of SLT do you predominantly work in, and has this changed over your career? Why/why not?

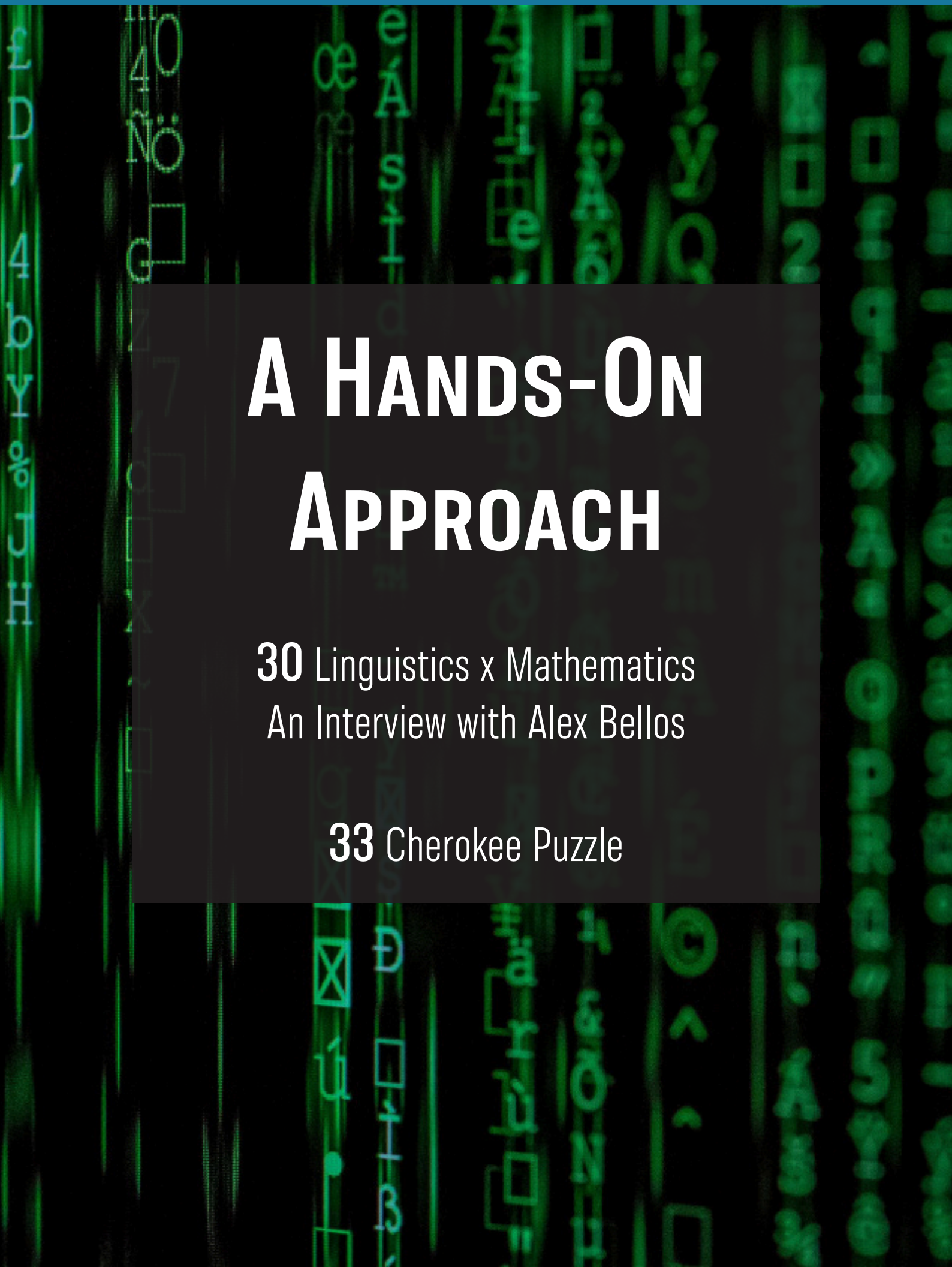
Eleanor: I have always worked with children with special educational needs in schools, so I guess that's my main area of specialism. I have worked also worked with the preschool age group and children in specialist schools, but I've always come back to the working in mainstream schools. In a mainstream primary school, you'll be working with children between five and eleven; they can present with all sorts of communication difficulties so really no two days are the same. I really enjoy working in schools with the teaching assistance and the class teachers to make sure they understand how they can support children with speech, language and communication needs.

Blue: If you had to convince someone to become a speech language therapist in 10 words or under, what would you say and why?

Eleanor: It's a rewarding, fun job and no two days are the same (and in when working with children you get to blow bubbles and play with fun toys!).

*Names have been changed for the purpose of anonymity





A Hands-On Approach

Linguistics x Mathematics

An Interview with Alex Bellos

Stephanie Jat is a third-year BA Linguistics student at the University of Cambridge. Here she talks to Alex Bellos, a writer and broadcaster. He currently runs a column in The Guardian for maths and language puzzles, and has recently published The Language Lover's Puzzle Book, a collection of puzzles on language and linguistics. In this interview he discusses his experience with the interaction between mathematics and language.

Were you always interested in linguistics, or was it a chance discovery related to your work in maths?

I was not consciously aware of an interest in academic linguistics, but I'm completely immersed in the world of languages and always have been. My mother speaks French and Hungarian at home and my father speaks German, Russian and French, so I grew up bilingual English-French, and then I got German A Level, and then for fun tried to teach myself Russian. I moved to Brazil as an adult, and now I speak fluent Portuguese. Maths is my background, but I find words really fun – I'm always playing around with them, making puns and other kinds of word games. My dad [David Bellos] is actually a linguist (Professor of Comparative Literature at Princeton), and one of his areas is *Oulipo*, which is an experimental literary group that tries to make literature from mathematical constraints. So, I've always been aware of the playful connections between maths and language. In fact, the first column I wrote for The Guardian was a column about language – it was a column called *Lingua Franca*, and was essentially about neologisms. Each week I would choose some area and talk about all the jargon and neologisms within it. I've always been interested in language, if not linguistics.

Alex Bellos, writer and broadcaster

Do you think taking a mathematical perspective to language helps you spot their similarities better?

I did my thesis on the philosophy of maths, about investigating its basic structure – what you'd call foundations – so, you need to have a language to try and understand the mathematics. As a student I was looking at the mathematical structure of maths, but mathematics is a language, so I was looking at metalanguages of metalanguages and meta-meta-languages and things like that. I think that I look at languages in a certain way and see patterns and connections that I can link to my background in maths.

Would you say there's a 'method' to tackling linguistics problems in the same way there are steps to completing maths problems?

Yes, there's definitely a method. When I started to do the Linguistics Olympiad problems, I noticed almost all of the problems start with some basic pattern recognition, whether it's looking at different elements of words that might be repeated or slightly changed. That kind of classification and noticing patterns I can do pretty well – that's the mathematics part. But then what you get with these Olympiad problems is that you've got to use some kind of linguistics brain, which is different from the maths brain – you have to make a leap of faith. You have to think: how would humans do it? The-

A.Davey from Portland, Oregon – ¸Namgis mural in Alert Bay, BC, Canada with Kwakwala text reading "Gilakas'la"

re's a kind of randomness to how languages work, which is not predictable, and sometimes you're looking for something which is so easy that you would never even have thought it was something you had to look for. I'm not sure that maths is a great help for doing that, because maths is incredibly logical. You're only allowed to assume what you can prove 100%. With these problems you need to make this extra leap and I find that fascinating and challenging to do. The other thing that I really liked about these problems, which is different to the maths problems that I'd done before, is that you were learning something very real. You feel that you're learning something real about the world, so it gives you this extra excitement.

What have you found to be the most useful analytical skill transferred from maths to linguistics?

When trying to solve Olympiad-style linguistics puzzles, I feel that what the maths brain brings is that maths is essentially just solving puzzles all the time. You get a sense for where the clues are, how the person who set the puzzle is trying to puzzle you, and where they're trying to put up the walls that you've got to break down. Also, the analytical skills would be of pure logical deduction, but there are many ways to logically deduce something – by contradiction, by induction, etc. – so you've actually got quite a good toolkit to break these things down. The skills you don't have is anything that's linked to meaning or the outside world, so sometimes you just need to use your common sense. Now, if you were talking about whether a maths brain equips me to learn languages, I'd say that the maths brain really does. Maths is a language that you need to learn how to speak, or how to write, how to use all its different symbols, and how to understand its very clear definitions and its rules. I mean, open any book of maths, and if you don't understand the language, you won't understand the book. It uses interesting symbols; you've got to know what the symbols are; you've got to know the rules about how they're put together. There's a community of people who talk to each other using it; there are people who speak it better than others; there are lots of similarities. When I was learning Portuguese, I was making puns and trying to write alliterative sentences. I was trying to push the boundaries of what you were allowed to do, and to me that's something which is a mathematical approach. Another thing that comes from maths is that I used to love being completely immersed in an environment where I absolutely did not understand anything, where you're actually forced to have to use your wits to try to communicate. I find maths a very creative discipline where you've got to try and find a proof, you just throw everything at it and often under pressure.

Are there any languages you would call particularly mathematical?

I'm not sure you could call one language more mathematical than the other – I don't really know what the criteria would be, but a brilliant question actually in The Language Lover's Puzzle Book about Kwakwala (from Vancouver Island) in the form of a word search. The language is formed from a few simple roots and loads of suffixes and it's all about mix-and-matching the word to the suffix, and that struck me as a really mathematically efficient way of organising a language.

Do you think a person's view of numbers can be affected by their aspects of their language?

I would say it does in different ways. In Japan they have what they call the *kuku*, which is the times tables learnt in a musical way without a tune, like a rhyme. What is interesting is

that when they get to a bit that doesn't scan very well, they just change the words slightly, just to make it as easy as possible to sing along to, and it seems to really make a difference. So, definitely the way that they have used words to learn numbers seems to work. I can remember I went to the BBC to do a radio documentary on Japan and numbers, and we just stopped adults in the street and gave them simple mental arithmetic questions. They all got them right, and they all said that they were singing this *kuku*. It had been so deeply implanted in their memory that they would still be able to use it thirty or forty years later.

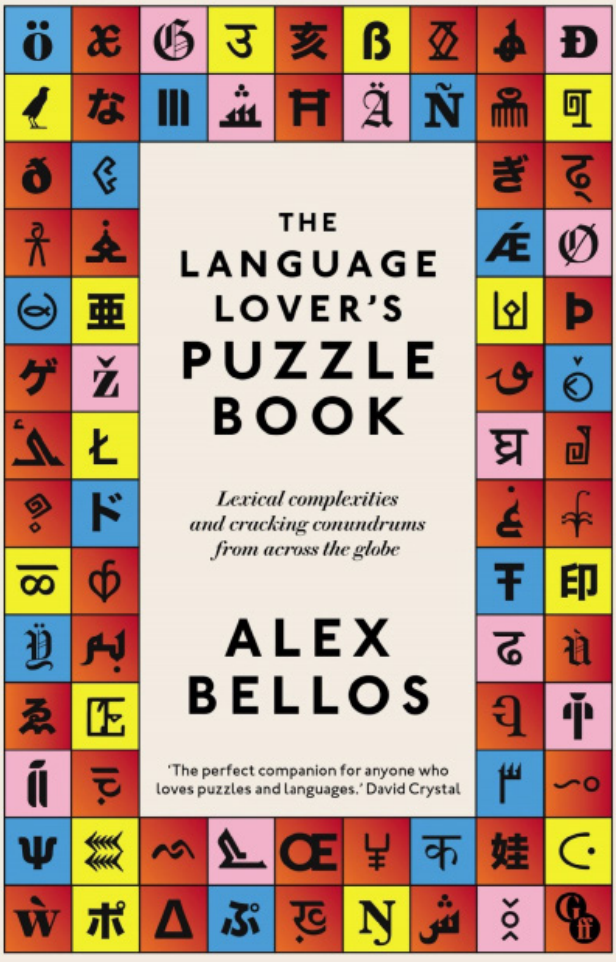
Have you heard of artificial mathematical languages, such as Lincos and Transcendental Algebra[1]? Might they be good attempts at creating a 'universal' language?

No – I don't think you'll find anyone who will say that there is a universal language that has succeeded on its initial terms at all. Arika Okrent in *Land of Invented Languages* tells the wonderful but essentially tragic stories of all the people who've tried to set up these mathematical universal languages, and they just don't work. They become far too complicated far too quickly, and far too rigid to ever be a way that you can speak to anyone else. What I think about something like Transcendental Algebra – I think that it was done with a tongue in the author's cheek, one reason being that he announced it in a universal language [known as Occidental] that nobody else spoke! If you really believed that this was the future of humanity, you'd at least announce it in a language that people spoke. It's interesting as an experiment to see how you can do it, but as something serious it's just absurd.

Finally – linguists often debate whether language is innate, but they also draw parallels between maths and language – what's your take on whether knowledge of maths is innate?

If I had to give a one-word answer, I would say yes – there's something about the mathematical structure of the world which seems to exist even though we can only discover it because we're human. Of course, there are people who can't manipulate numbers in arithmetic, but then, does that mean they can't do maths? For most people maths is numbers, and maths begins with numbers, because numbers are basic tools for maths – even though maths is so much more than that; and there have been cases of professional mathematicians who have elements of dyscalculia. Maths comes from the animal ability to distinguish size – almost all living things that have been tested show some kind of mathematical ability. Animals are given these tests, with two pieces of food on one side and one bit of food on the other side, and they all go to the side with two bits of food. Essentially, they're doing maths because they're working out "oh yeah, one's twice as big as the other". I think that most mathematicians would say that maths exists independently of humans.

Alex Bellos' new book, The Language Lover's Puzzle Book, is a collection of linguistic puzzles sourced from Linguistics Olympiad competitions around the world. Each puzzle is accompanied by a worked solution, giving an introduction to the problem-solving strategies used in deciphering linguistic data, and an explanation of the varied and fascinating linguistic phenomena on which the puzzles focus. It is available online and from bookstores, as are his previous books on mathematics puzzles and football.



References:
[1] Gilyarova, K. (2003). 1st International Linguistics Olympiad, individual paper. https://ioling.org/problems/2003

Cherokee Puzzle

by Liam McKnight

A computational linguist wanted to write a program to transliterate between the Cherokee syllabary and the Latin script (as used for English). They came up with a program X which transliterates perfectly well from Cherokee to Latin script. However, when trying to transliterate from the Latin script back into Cherokee, they came up with three simple (but different) methods by which this could be done. They wrote three different programs, A, B and C - one for each method.

Input	X	A	B	C
ᏌᏍᏏ	sasha	ᏍDᏍᏏ	ᏌᏍᏏ	ᏌᏍᏏ
ᏏᏥ	hama	ᏏᏥ	ᏏᏥ	ᏏᏥ
	hanah	ᏏᎾh	ᏏG	ᏏG
	nahnaa	ᎾᏏD	GᎾD	ᎾᏏD
ᏏᏌ		ᏏᏍD	ᏏᏌ	ᏏᏌ
	ahgaa	DhᏚD	DhᏚD	DhᏚD
	gashama	ᏚᏍᏏᏥ	ᏚᏍᏏᏥ	ᏚᏍᏏᏥ
	mhahnah	mᏏᏏh	mᏏᏏh	mᏏhG

They noticed that when passing one Cherokee word, GᎾ, through program X and back through programs A, B and C, they received not only GᎾ as an output in one or more cases but also ᎾᏏ. Both GᎾ and ᎾᏏ are valid Cherokee words.

They decided they would input some made-up Cherokee words to see what their programs did differently.

N.B. characters or groups of characters which cannot be transliterated into the Cherokee script are output unchanged. "D" and "G" are Cherokee characters; "h", "m", "n" and "g" are not.

(a) Fill in the table to the right (dark squares do not need to be filled in)

(b) How do programs A, B and C work?

Input	X	A	B	C
ᏍᎾᏏᏍ	?	?	?	?
	magsa	?	?	?
	?	ᏍDᏏᎾ	?	ᏌᏚᎾ

The linguist's friend also tried to write their own transliteration program (program D), which was quite similar to programs A, B and C, but it didn't work very well. After running a Cherokee string through program X and program D, the following results were produced (table to the right):

Input	Output
GᎾ	GnD
ᏌᏌ	ᏍDᏍD
ᏚᏌᏥ	gDᏍDmD
ᏏG	hDG

(c) How does program D work?

(d) Are there any strings in the Latin script which produce four different results when input to programs A, B, C and D? If so, give an example of one.

(e) Determine if the following statement is true for any of the programs A, B, C or D: "For any input string entirely in the Cherokee script which is passed through X and then through the program, the output will also always be entirely in the Cherokee script."

Cherokee is an Iroquoian language spoken by around 12,000 people in North America. Its script was invented by ᏍᎩᏯ (Sequoyah) in the early 19th century.

Meet Our Ed

T. R. Williamson *Editor-in-Chief*

Hello! I'm an MPhil (by Thesis) student at Cambridge, focusing on the interrelation of the mind and concepts with meaning in language for my research. As a co-founder of U-Lingua, and as its Editor-in-Chief, I elicit great satisfaction and excitement from the success we've already enjoyed as an Editorial Team! My role for the magazine involves overseeing all of the magazine's operation, including article idea approval, editing with Section Editors, formatting and designing with the Editorial Designer, publicity, and eventual publication every quarter.

Caitlin Wilson *Section Editor - In The Field*

I'm currently at the University of Edinburgh in the third year of an MA in Linguistics. My main interest within the field is phonology, but I'm also very interested in language acquisition and historical linguistics. My role for U-Lingua is as Section Editor for the news section of the magazine, In the Field. I commission, organise, and edit articles that discuss anything new happening in the world of linguistics.

Marion Willingham

Section Editor - Behind The Bookshelves

I'm Marion, a second-year undergraduate linguist at Gonville & Caius College Cambridge. I'm yet to specialise but have a longstanding interest in computational linguistics and corpus analysis - especially as their applications can cover topics well beyond traditional linguistics. I'm also really enjoying studying syntax and language typology. As section editor for Behind the Bookshelves I will be working to bring new and topical academic research to our readers in an easily digestible format.

Fancy joining the editorial team? Why not run for a position at the ULAB Conference's Annual General Meeting in a fe

ITORIAL TEAM

BLUE BATES-CAMBRIDGE

SECTION EDITOR - THE REAL WORLD

I'm Blue, a finalist at the University of Cambridge, reading (predictably) Linguistics. I've loved studying a broad range of disciplines within the subject, from the history of linguistics as a science to the processes and rules behind language attrition in heritage speakers. This is what drew me to U-Lingua; the magazine operates as a platform for students to informally explore topics tangentially related to their areas of study yet allows for a degree of creativity not often facilitated in formal study. My section — The Real World — looks into the applications of Linguistics outside the academic setting and into fields like NLP, translation and interpreting and forensic linguistics. I'm excited to continue seeking out opportunities for students to integrate Linguistics into their future career and beyond.

LIAM MCKNIGHT

SECTION EDITOR - A HANDS-ON APPROACH

I'm a third-year BA Linguistics student at the University of Cambridge. Within linguistics, I'm interested in theoretical phonology and morphology, as well as Celtic languages. I also enjoy solving and writing language puzzles, of the sort used in the UK Linguistics Olympiad. The Section I edit, *A Hands-On Approach*, focuses on language puzzles as well as analysis of real-world linguistic data more generally.

ANNE ESCHENBRUECHER *EDITORIAL DESIGNER*

Hello! I am Anne, an MA Computational Linguistics student at the University of Wolverhampton, currently finishing up my degree. After that, I will start a PhD, also in Computational Linguistics. I am particularly interested in automatic text simplification systems but also in all other areas of natural language processing. In my free time I enjoy reading fantasy, writing short stories and spending time outside in nature.

w weeks' time? More details can be found on the ULAB website. If you have any questions, do feel free to get in touch!

Editorial Team

Editor-in-Chief T. R. Williamson
Section Editor for News Caitlin Wilson
Section Editor for Academia Marion Willingham
Section Editor for Industry Blue Bates Cambridge
Section Editor for Puzzles Liam McKnight
Editorial Designer Anne Eschenbruecher

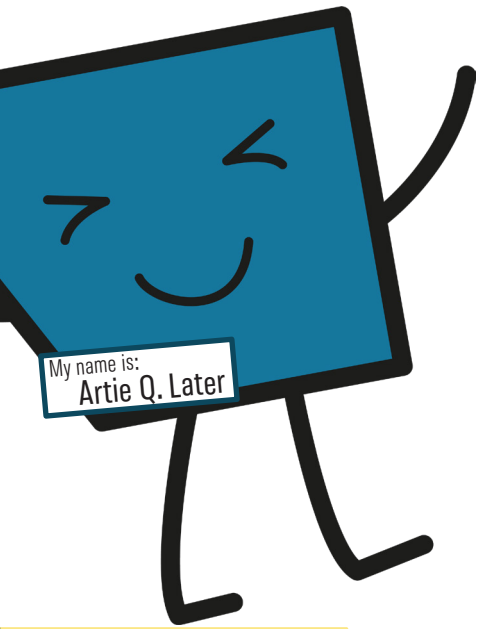
Authors

Linguicism Caitlin Wilson
Māori Racism Elif Yildiz
Build Your Own Twitter Corpus Imogen Davies
Fake News Identification Katie Shi
Investigating Phonetic Reduction Scott James Perry
Interview with Eleanor Blue Bates Cambridge
Linguistics x Mathematics Stephanie Jat
Cherokee Puzzle Liam McKnight

Photo Credits

Photos of Aberdeen Campus Kristina Callan Marcinkowska
Section Cover "In The Field" Jakayla Tonez on Unsplash
Story Photo 1 "Linguicism" Ehimetalor Akhere on Unsplash
Story Photo 2 "Linguicism" James Eades on Unsplash
Story Photo "Māori Racism" Callum Parker on Unsplash
Section Cover "Behind The Bookshelves" Markus Spiske on Unsplash
Story Photo "Twitter Corpus" Imogen Davies
Story Photo "Phonetic Reduction" Gerd Altmann on Pixabay
Section Cover "The Real World" Sigmund on Unsplash
Story Photo "Speech Language Therapist" Youssef Naddam
Section Cover "A Hands-On Approach" Markus Spiske on Unsplash
Story Photos "Linguistics x Mathematics" Alex Bellos

Join us on our social media!



Our Website

www.ulab.org.uk

Facebook

Undergraduate Linguistics Association of Britain

Twitter

@ULAB_Ling

Instagram

@ulab_ling

LinkedIn

Undergraduate Linguistics Association of Britain

Discord

www.ulab.org.uk/discord

Strava

www.strava.com/clubs/ulab