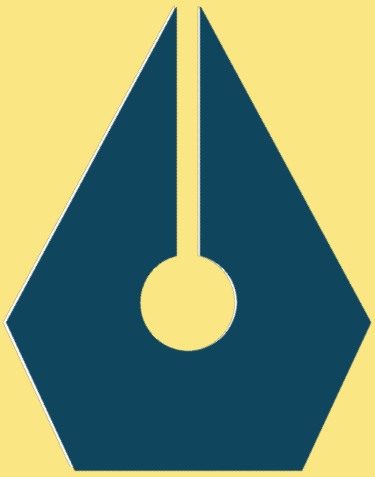
Winter 2022 | Issue 7

ISSN 2752-6453

U-Lingua

The Undergraduate Linguistics Association of Britain's Quarterly Magazine

THE CHANGE ISSUE



A Hands-on Approach

How To... Rebuild A Language

Your Turn: Translating Poetry

Words, Words, Words

Big Little Words:
The Weary Self-Awareness of Contemporary Poetry

Hot off the Press

"Secret" and Constructed Languages:
How We Change Our Language

Emergence of a Dialect:
The Origin of African American English

Behind the Bookshelves

Toxic Data:
Biases in Natural Language Processing

Wordslut:
A Feminist Reinvention of the English Language

Curiouser and Curiouser

The Next Step: Getting into Postgraduate Study

Anatomy of a Linguist

The Philosophical Value of Big Data:
A Plea for Intention Tracking

Through the Ages

Inventions and Reinventions

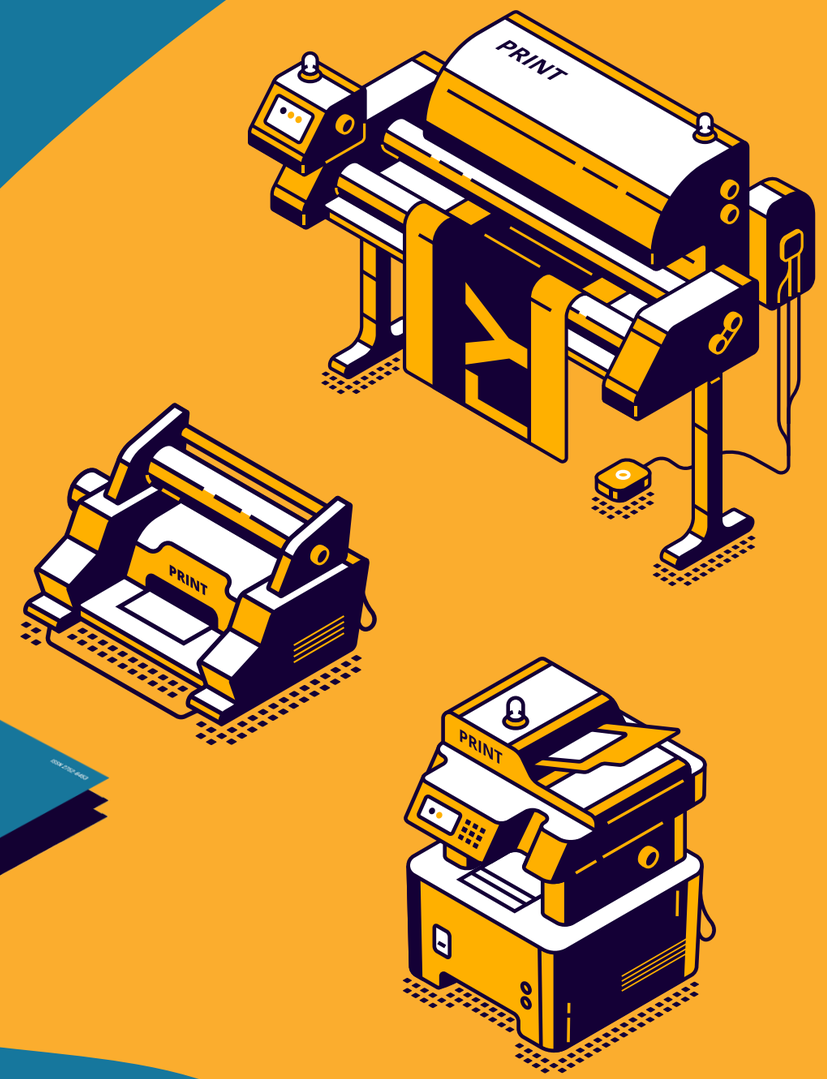
ULaB

U-Lingua

is going for print ! ! !

This (and some subsequent) issues of *U-Lingua* is getting the chance of being printed and distributed to you! We'd love to be able to send these to your Departments and Schools — let us know if you're interested in getting a copy by filling out the following form:

https://tinyurl.com/2p8nn2n6



Contents

6

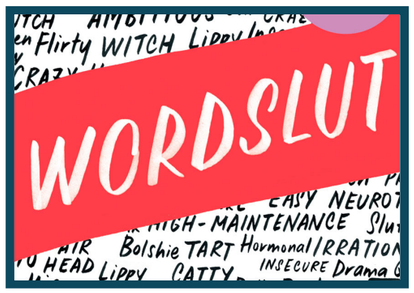
Editor's Letter
Updates from National Chair
Updates on the 2022 Conference
A Special Note from *JoULAB*

8

Meet Our
National Committee

12

Hot off the Press
(News and Culture)
14 "Secret" and Constructed Languages:
How We Change Our Language
18 Emergence of a Dialect:
The Origin of African American English



20

Behind the Bookshelves
(Academia)
22 Toxic Data: Biases in Natural Language Processing
24 *Wordslut*: A Feminist Reinvention of the English language

26

Through the Ages
(Historical)
36 Inventions and Reinvention

34

Words, Words, Words
(Literature)
Big Little Words:
The Weary Self-Awareness of Contemporary Poetry

36

Curiouser and Curiouser
(Outreach)
The Next Step: Getting into Postgraduate Study

30

A Hands-on Approach
(Puzzles)
30 How To...
Rebuild A Language
32 Your Turn:
Translating Poetry

38

Anatomy of a Linguist
(Food for Thought)
The Philosophical Value of Big Data:
A Plea for Intention Tracking

42

Opportunity, Contributors,
Social Media

Editor's Letter

Happy New Year! As 2022 kicks in, a lot of us will be thinking of resolutions and other changes we want to make to our lives. Fittingly, this issue looks at how change can manifest in language.

David Crystal says that "language has no independent existence apart from the people who use it". Prescriptivists often disseminate their fear of modern change in language: we've probably all heard the idea that slang, or "text speak" is ruining English, most of us deal with being teased for some type of non-standard language. But change itself can be improvement, not deterioration. In this issue, we look at how language has changed, and can change, and what changes we can make the most of to make our language more accessible and inclusive to its speakers. Some highlights include a whistle-stop tour of secret and constructed languages, ways in which speakers mould their language to their needs; a review of *Wordslut* and how women can reinvent English for themselves; an etymological look at English words and how the vocabulary has developed across time and cultures.

Linguistic change can come in these large, social and global scales, but also in more minute, personal or local ways: Our columnists speak of next stages of education, what we can do as individuals to re-juvenate a dying language.

It has been a really exciting quarter at *U-Lingua*, with the possibility of printed copies of the next issue being distributed (thank you to the LAGB Student Committee)! Issue 7 would not have been possible without my fellow Section Editors, Columnists and Writers; and our incredible Editorial Designer. Thank you as well, to our readers, for coming so far with us! Your interaction with us on the (relatively) new instagram account (@ulinguamagazine) has been really encouraging and welcome! If any of you have suggestions or things you want to see more of, let us know in the comments or dm! I hope you enjoy reading this issue, and that 2022 brings you all the best!

S. C. Jat
Editor-in-Chief, *U-Lingua*
University of Cambridge

Updates from National Chair

Since October, we have launched the very first ULAB buddy scheme, which had a whopping 49 sign-ups! We have also been busy undertaking "Project Poster", a scheme thought up by our Journal Editor, T. R. Williamson, to get a ULAB poster in every university that teaches linguistics in the UK, so look out for some ULAB posters in your area!

Lastly, we have been working hard behind the scenes to organise the first in-person ULAB events in over two years! We are planning on having two dinner parties in January, one in Edinburgh and one in London, and all linguistics students are welcome to come to whichever one is closest or more convenient for them. More details will be announced on our social media in the next few weeks! I look forward to hopefully meeting lots of you grate people in person very soon!

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

Updates on the 2022 Conference

Preparations for our 2022 conference are well underway, and we now know the conference will be taking place 9th-11th April in Edinburgh! Submissions are now open, so if you have done any research in your undergraduate degree please consider submitting an abstract to us before 23rd January! More information about submitting to the conference can be found on our website, under conferences.

We are excited to announce that we have been successful in securing funding from both the University of Edinburgh and the University of Aberdeen. We are currently deciding who to invite to give plenary talks and workshops, and working with the department at Edinburgh to make preparations for catering and room bookings for those of you who will be able to join us in-person — we are still hoping for a hybrid conference but will of course keep updating our plans in line with Scottish government guidelines, and we will make sure to keep you all updated. We can't wait to welcome you to Edinburgh in April, either virtually or in person!

Caitlin Wilson
Local Vice Chair, *Undergraduate Linguistics Association of Britain*
University of Edinburgh

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

A Special Note from *JoULAB*

A period of joy and good tidings has already begun in the undergraduate calendar. No, it's not the conclusion of Christmas; it's dissertation season! Final-year undergraduates across the UK and beyond will, I'm sure, be relishing the opportunity to begin the New Year by conducting independent linguistic research. The prospect can be daunting, intimidating, and even overwhelming, but fun can be found whilst completing your dissertation (if you look hard enough!).

Alongside organising an annual conference and publishing a quarterly magazine, ULAB also publishes a journal, *JoULAB*. It's the only undergraduate journal in the world taking submissions in all areas of linguistics, and it really wants you to submit! If you're writing a dissertation and would like to see it published, consider us once it's done — your work will be amazing and we can't wait to see it! For more information, check us out on Twitter and Instagram: @ULAB_Journal for both.

T. R. Williamson
Journal Editor, *Undergraduate Linguistics Association of Britain*
University of Cambridge

Meet Our National Committee

Cliodhna Hughes

National Chair

Cliodhna (she/her) is a fourth-year Linguistics student at the University of Edinburgh, particularly interested in language evolution, phonetics, speech processing and first language acquisition. She enjoys doing linguistics outreach, attending student conferences, and getting involved in just about anything linguistics-related! Outside of linguistics, her hobbies include glass fusing, playing underwater hockey and making unbrielievable cheese puns.

Hafren Vaughan

National Vice Chair

Hafren, [hafɹən] or [havrɛn], is a second-year undergraduate Language & Linguistics student at the University of Aberdeen. She loves everything to do with language, but is particularly interested in theoretical linguistics - especially morphology and syntax! When she's not studying, you will probably find her learning Japanese: memorising Kanji... or singing karaoke.

Beatrix Livesey-Stephens

National Treasurer and Accessibility Officer

Beatrix (Bea) is a final-year undergraduate Language & Linguistics student at the University of Aberdeen. As a disabled person and ULAB's Accessibility Officer, she is particularly interested in areas of linguistics concerning disability. Bea is also interested in translation and in critical discourse studies about consent and dissent. She would like to further explore the ethics of natural language processing, and she lives in fear of theta roles. When not doing Linguistics, she's making new consent workshops for CASE, or knitting.

Louis Van Steene

National Secretary and Webmaster

Louis is an MPhil (by Thesis) student at Magdalene College, Cambridge. After trying to learn a range of languages with highly variable success over the years, from Icelandic to Japanese, his interest in linguistics initially grew as a way to tie them all together. A Platonist at heart, his current interests primarily lie in formal syntax and phonology and their applications. Further intellectual pursuits span the philosophical and the psychological, but when not studying or learning languages, he enjoys literature, music and lists without Oxford commas.

Riley Crouch

Social Media Coordinator and Local Chair

Riley (they/she) is a third-year Linguistics student at the University of Edinburgh. They're primarily interested in pragmatics and discourse analysis, particularly looking at language about or used by trans people. Outside of Linguistics, you'll find her trying to make a new vegan recipe or drawing fun lines on their face.

Roma Dhasmana

Social Media Coordinator

Ro is a second-year Linguistics and Spanish undergraduate at the University of Aberdeen. Her main interests lie in second language acquisition and syntax. When not studying, she can be found singing in choirs, listening to horror podcasts, and thinking about cats.

T. R. Williamson

Journal Editor

Tom is an MPhil (by Thesis) student at the University of Cambridge. His research interests lie primarily at the intersects of mind and meaning, with his current thesis investigating whether the semantics of motor action encoded within certain English idioms affects psycholinguistic processing. As the Editor of *JoULAB*, it is his role to supervise all its activities from start to finish; particularly, he heads up submissions with Cliodhna, reviews with Liam, social media with James, copyediting with Rachel and the Associate Copyeditors, and publication.

Stephanie Jat

Magazine Editor-in-Chief

Steph is an MPhil (by Thesis) Linguistics student at Pembroke College, Cambridge, originally from Hong Kong. After spending too many hours doing UKLO puzzles, she decided to try her hand at the degree and decided she loved it. Her main interests lie in stylistics and its interaction with linguistic theory, specifically in the interface between poetic license and grammaticality. When not reading, she can be found trying not to burn down the kitchen, lost on a hike, or living her X-factor dreams from the safety of her living room.

Lydia Wiernik

Archivist

Lydia (they/them) is a second-year undergraduate studying Linguistics at the University of Edinburgh. As a quadrilingual and counting, they have always been fascinated by language, but Linguistics is where their passion lies. Their research interests include prosody and Deaf linguistics, with a focus on the cognitive processes behind them. When not sorting through the ULAB archives, they can be found archiving for Dr. Dan Gunn's soon to be published biography. In their spare time, they are an avid bug collector and Keats enthusiast, and will lecture about both unprompted if you're not careful.

Caitlin Wilson

Institutional Representative Coordinator

Caitlin (she/her) is a fourth-year undergraduate at the University of Edinburgh. She is focused on theoretical syntax and language preservation. When she isn't drawing syntax trees, she helps with fieldwork on the University's Nilotic Languages research team. Passions outside of Linguistics include terrible y/a fiction, being a bad runner, and Oxford commas.

Eloise Parr

Events Coordinator and Opportunities Coordinator

Eloise Parr (she/her) is a first-year PhD linguistics student at the University of Birmingham. She graduated from Coventry University with a BA (Hons) in English and then a MA in Applied Linguistics from the University of Birmingham. Her research is using corpus-assisted metaphor analysis to explore pregnancy metaphors. When not doing linguistics, Eloise can be found reading, tending to her plant babies, or unironically enjoying a Zoom quiz!

HOT OFF THE PRESS

14
"Secret" and Constructed Languages:
How We Change Our Language

18
Emergence of a Dialect:
The Origin of African American English

"Secret" and Constructed Languages:

How We Change our Language

Gabriel Bonté, Year 3,
English Language, English Literature and French Literature,
Faculté des Lettres de Sorbonne Université

Poland, 1887. L. Zamenhof, disappointed by the difficulties he encountered when learning Volapük as an international or universal language, decided to invent his own: Esperanto, drawing inspiration from Latin, French, German, English, Russian, Polish, Greek, and Hebrew. Esperanto. Esperanto, which means *hope*, was created with the goal of being a simple and accessible language, that would bring people from across the world together and facilitate understanding between cultures.

United States, 1960s. Gay men and other members of the queer community started using slang borrowed from sailors and showmen, who had themselves been inspired by old pidgins of tradesmen, such as the Mediterranean Lingua Franca. This way of speaking was called *Polari*, a term derived from the Italian word for "speaking", *parlare*. Even though it is mainly a lexicon using English syntax, word order and grammatical words, the dialect of Polari, when spoken by seasoned speakers, sounded like a different language to outsiders, probably because of its roots in Italian, Romani, Yiddish vocabulary, and cockney rhyming slang[1]. This allowed gay men to recognize fellow members of their community or, on the contrary, to hide from people who would attack and oppress them if they knew of their identity[2].

Both cases, while seemingly unrelated, illustrate a similar mechanism: the creation and use, as well as the deformation, modification, or complete rebirth of language. Human language continues to puzzle those who seek to explain it (i.e., us, linguists) and those who seek its origins (evolutionary linguists). So-called "secret languages" and constructed languages might prove to be useful insight in the field of language evolution since they allow us to witness the birth of a language. However, secret languages and attempts at universal languages differ in two major ways. Firstly, the former are designed to avoid understanding from outsiders, whereas the latter are designed to help understanding among very different groups of people. Secondly, the former usually use the structure of an existing language, while the latter try, more or less successfully, to come up with an original morphology and structure.

I/ Secret Languages

Secret languages, despite or, more likely, because of their secrecy, keep being born and used in various contexts. Some disappear, some evolve, but they seem to be a recurring phenomenon regardless. They almost always rely on an existing language for their structure, because, unlike other languages, instead of having hundreds of years to evolve, they need to arise, evolve and become functional very quickly, in order to be able to reach different people who need to speak a language not understood by outsiders, but do not have the time or resources to learn a foreign language, much less to invent a completely different one.

a/ Features and Origins of Secret Languages

Studying specific instances of secret languages can provide insight into their goals. Polari, like most secret languages, seems to be "only a vocabulary"[3]; it is often considered to be spoken English, only as a variation or sociolect of a language that already existed. It seems that it is enough for secret languages to only replace lexical words and not grammatical ones, since they are only referential: the grammar of a sentence, deprived of any lexical meaning, would not be enough for someone who does not speak the language to understand its meaning. Furthermore, using a grammar already known and shared by the group makes things easier to learn. Although Polari is a shared lexicon created and enriched by a specific group rather than an actual language it does have some morphological particularities. These include frequent affixation (the suffix *-ette* was employed extensively to mean that something is small), clipping (*Martini*, meaning "ring", was shortened to *marts*, and came to mean "hands"), compounds (*fake*, meaning "accessory", could be associated with *aunt nelly*, meaning "ears", so that *aunt nelly fakes* meant "earrings"), or the use of derivations of acronyms (*alamo*, meaning "attractive", was derived from the acronym *LMO*, or *Lick Me Out*). Because of its nature, users frequently invented and transformed words as well.

Another interesting instance of a secret language is Boontling, which was used exclusively in the small town of Boonville, California. It originated from language games played by children of the town, and is, once again, more of a lexicon relying on the general structure of English. The town was very secluded, and adults started picking up these children's slang, which allowed Boontling to become an actual dialect for the inhabitants — that is, until more outsiders came into the town and the tradition eventually was lost. While it lasted, from the 1890s to the 1960s approximately, it relied on anagrams, acronyms, compounds, foreign words, but also and perhaps more interestingly, on names or events associated with the town. For instance, *to high-heel* meant "to arrest", because a constable had one leg shorter than the other and wore a high heel to hide it. Taken from their original context, and through various levels of associations, these came to be words of their own[4].

b/ Goals of Secret Languages

But the most long-lasting and widely used secret language I came across in my life — not just while doing research for this article, but, like many young French people, in my school and in the streets of my city — is Verlan. This French back-slang's principles are seemingly very simple: it relies on inverting the syllables of a word, so that, for instance, *louche* (here, "shifty") becomes *chelou*. The meaning can also change or shift with verlanisation: *chelou* does not just mean "shifty" anymore, it evolved to mean a wide range of things, such as strange, bizarre, suspicious, weird or abnormal, and can be applied to many different contexts or situations, from an unusual weather to an aggressive person. Verlan originated in the 1970s and 1980s in French suburbs, where marginalized young people, especially poor young men, used it to avoid detection by outsiders, rich people, the police or even the adults around them. However, the language quickly became mainstream, with advertisers or singers using it to appeal to the youth. The countercultural anti-language became absorbed by the dominant culture. But Verlan did not simply disappear: It reinvented itself. In the 1990s, Verlan started to be used again, in similar contexts, even though it was more widely understood. It still filled the purpose of making speech less intelligible to figures of authority, but most of all, its aim was to establish a sense of belonging to a community, and to establish one's identity within that community[5]. Sociolinguistically, Verlan's appeal is perhaps less in its secrecy than in its way to form a community around itself, as well as in the endless creation it allows. Compared to other secret languages, Verlan allows for more derivation, because it does not rely on a fixed lexicon which has to be added upon for new words to become common, but rather, on a rule that can be applied to anything, in any context. Now that it is not so secret anymore, there is also a playfulness in trying to make it harder to understand, by re-verlanizing words, for example. There is almost a dialogue between Verlan and mainstream society: the former always tries to escape understanding from the latter, which makes it very dynamic.

Now that we have established the uses of these languages, let us dive into how they are structured. We should not be too quick to dismiss them just because they are "only lexicons": A lexicon does not exist as an abstract and fixed concept; it interacts with other parts of language.

c/ Structural and Morphological Particularities of Secret Languages

Diving deeper into the structural and morphological particularities of secret languages can therefore reveal their unique properties. While all the secret languages presented here rely on the structure of an existing language (English or French), they still exhibit specific properties. Polari, beyond its status as a lexicon, presented frequent affixation, clipping, compounds, or using derivations of acronyms. Boontling had less structure of its own, relying on anagrams, acronyms, compounds, foreign words, but also, and perhaps more interestingly, on names or events associated with the town. As for Verlan, while the key to its "secret" is so simple a child could understand it, when used with key words in the sentence, and with words that without even being "verlanised" are already not common (for example slang words or loans), the sentence becomes almost unintelligible to a non-initiated listener, or at least it does at first. Nowadays, although Verlan is still used in specific contexts or within a certain group of people, it is widespread enough that most people understand it. It also relies on truncation, elision, and phonemic variations which makes "decoding" harder than just reversing the word again. Verlan has a remarkable adaptability and resilience, and, unlike the other two secret languages, it is still used. This is visible in the dialect itself: Some words, which had become too mainstream or too understandable, were re-verlanised, or rather, doubly verlanised. For instance, *femme* ("woman") became *meuf*, which became *femeu* when re-verlanised. Furthermore, new slang words including foreign slang words were verlanized[6].

Therefore, beyond serving a function (secrecy, the establishment of community, or protection from another community), these three secret languages are playful[5]. The creation of these dialects is mediated by the instinctive joy of playing with words, encrypting them, and making associations between seemingly unrelated words and concepts. While the dialects studied cannot be understood as fully distinct languages, some of the mechanisms involved in creating them might be involved in the creation of fully functional languages.

II/ Constructed languages

a/ Goals of Universal Languages

What, then, of the actual creation of languages? Why are universal languages so unappealing once they are created? With its two million speakers, Esperanto is the most widely spread international auxiliary language, and is still far from representing the universality it aims at. International auxiliary, or, more simply put, "universal" languages, in order to serve their purpose, should be ethnically and politically neutral, easy to learn, and fully functional as human languages[7]. However, being politically and ethnically neutral is nearly impossible: Esperanto itself may be easy to learn for some people but does nonetheless disregard many cultures and languages. Choosing a script, such as the Latin alphabet for Esperanto, is already an excluding decision: while it makes things easier for people who can already read it, people who use Cyrillic script or Chinese characters will not have that advantage.

b/ Attempts at original structure and grammar

However, despite their arguably unrealistic premises, one could argue that constructed languages, and more specifically languages constructed with the goal of becoming universal languages, because they are "made from scratch", present a more interesting perspective in the study of the creation of a language. Just like studying a monkey in the zoo will not teach you about monkeys in the wild, studying a language created by linguists who have extensive prior knowledge of modern grammar will not give us insights to the origins of language. Furthermore, it seems that, in order to be a fully functional human language, the universal language must accept change and evolution: Esperanto itself comes from dissatisfaction with Volapük, that was itself a constructed language designed to be universal and has engendered many languages derived from it.[7] To quote the linguist Marina Yaguello, "Languages seem destined to split up into dialects; ironically enough, artificial languages aiming for universality are themselves also victims of this conflict"[7]. Some universal languages, like Kotava or Loglan, have strived for a more logical or more universal structure than Esperanto; but their success remains quite low. Constructed languages can achieve some success, but usually, even this success comes through the creation of a community. Anyone who has tried to learn a foreign language knows that it is easier and more enticing to learn it when practicing with other people, and constructed languages are no exception. However, the very existence of a community of people speaking a language together eventually brings about changes in that language. One of the most interesting and conclusive applications of constructed languages is in education, where we can observe the similarities between the learning and the creation of a language inside a community of pupils[8]. Most constructed languages designed to be universal languages, such as Esperanto, find some success through the internet: Esperanto has become more popular when it became easier to learn through online platforms. This allows small communities of people to get together, learn a language, and, most likely, make this language evolve.

We create new languages or dialects to form a sense of community; whether this is expressed through the creation of a secret language to avoid outsiders or through the creation of a utopian universal language depends on whether you have the time and resources to think about the technical aspects of languages. But the one similarity remains: Community, and our recurring need for new forms of communication, and the manipulation of our languages. It seems that we cannot refrain from making languages evolve.

References:

[1] Taylor, H. (2007). Polari: A sociohistorical study of the life and decline of a secret language. *University of Manchester Dissertation.*
[2] Green, J. (1997). Language: Polari. *Critical Quarterly*, 39(1), 127-131.
[3] Baker, P. (2003). *Polari-the lost language of gay men*. Routledge.
[4] Rawles, M. R. (1966). „ Boontling": Esoteric Speech of Boonville, California. *Western Folklore*, 25(2), 93-103.
[5] Lefkowitz, N. J. (1989). Verlan: talking backwards in French. *The French Review*, 63(2), 312-322.
[6] Boyer, H. (1997). Les mots des jeunes: observations et hypothèses. *Langue française*, (114).
[7] Olsen, N. (2003). Marketing an International Auxiliary Language: Challenges to a New Artificial Language. *Journal of Universal Language*, 4(1), 75-89.
[8] Adger, D., & van Urk, C. (2020). Three conlang projects at three. *Language invention in linguistics pedagogy*, 49.

Emergence of a Dialect:

The origin of African American English

Andrew Tobin,
fourth-year MA Linguistics,
University of Edinburgh

The last four centuries of British colonialism have seen the rise of a wide range of new varieties of English, as existing dialects mixed in the newly-established colonies, speakers of other languages learned English, and emerging cultural groups established their own linguistic identity.

African American English (AAE) is one of these new varieties, having emerged in the context of the Atlantic slave trade, and survived and evolved as a distinct dialect since then. AAE provides an interesting example of these new dialects, showing the effect of several major drivers of dialect emergence: Substrate influence, contact between different varieties and languages, and internal innovations.

A crucial factor in the emergence of AAE is its origins as a *non-native* dialect. It is the result of English being forced upon speakers of a diverse set of mother-tongues, and the West African origin of the first African Americans is reflected in the phonology and phonotactics of modern day AAE. In particular, most West African languages lack the dental fricatives [θ ð][1] and have phonotactics that restrict consonant clusters in the syllable coda[2]. As a consequence, AAE often replaces standard English's dental fricatives with other sounds, and frequently reduces coda consonant clusters[3].

this /dɪs/
test /tɛs/

Similar patterns are found in other African-influenced varieties. For example, in Ndyuka, an English creole in Suriname, think becomes *denki*, replacing the dental fricative with a stop, and adding a final vowel to prevent the coda cluster.[4]

Another African influence can be seen in prosody, with early recordings of ex-slaves revealing a prosodic rhythm that is strongly syllable-timed, compared to the stress-timed rhythm of their white contemporaries[5]. This almost certainly reflects influence from the syllable-timed languages of West Africa, an influence that can be seen today in African dialects of English such as Nigerian English[6].

External influences further helped to mould the emerging African American dialect in the 18th century, with the influx of Caribbean slaves into the American South. Unlike the comparatively small farms of the early United States, the Caribbean saw slaves working on massive plantations[7], and it was this plantation context, with minimal contact between blacks and whites, that saw the rise of major creoles such as Haitian Creole and Jamaican Patois[8]. These arriving Caribbean slaves brought tense-aspect features typical of creoles to AAE, for example the use of *been* to mark the past tense[3].

She been told him she needed the money[8].

As well as the Caribbean, many distinctive features of AAE can be tied directly to the non-standard dialects of American English spoken by the white settlers with whom early African Americans interacted. These include phonological features such as non-rhoticity and vocalisation of postvocalic /l/[3], as well as grammatical features such as double negatives[9]. Some of these features, once shared between white southerners and African Americans, survive today only in AAE, for example the use of unmarked past tense:

Mama work-∅ all her day. [referring to past][10]

In 1808, the United States banned the importation of slaves into the country, which informs us that Caribbean and African influences on AAE largely come from before this year. But over 200 years later, the dialect has not faded away, despite the 20th century rise of a pseudo-standard form of American English, based on the speech of white non-southern accents[11]. In fact, the dialect has continued to diverge and establish its own linguistic identity, with many of the features that define AAE today being innovations of the 20th century.

Following the Great Migrations of African Americans out of the American South, the creole-influenced tense system of early AAE transformed, gaining nuances not found in its creole ancestors[9]. Be, once used as a generic copula (where standard English would use *is*, *am*, *are*) developed a **habitual** meaning, describing events that are typically or habitually the case. More recently, it has evolved further, gaining an additional usage as a marker of durativity, typically with an intensive quality.

early AAE (generic copula)	*He be full*[12] 'he is full right now'
post-WW2 AAE (habitual)	*We be going to bingo*[12] 'we go to bingo regularly'
modern AAE (durative/intensive)	*Hey baby, this be Heywood!*[9] (shouted at someone)

Similar to *be*, stressed *been* has developed from its creole-like past-tense meaning to a more nuanced meaning of **remote past**, to describe things that have been true for a long time. For instance, *they* ***been*** *married* in AAE has changed from meaning "they got married" to its current meaning of "they have been married for a long time".

Instead of leading to the decline of AAE, contact with white dialects may have spurred this century of innovation in the dialect. When African Americans began to acquire the "standard" grammar of white dialects, this freed up the old particles to find more nuanced and expressive meanings, and as markers of African American identity[9].

"[The African American component of tense] is freed from the drudgery of every-day grammatical work"
William Labov[9]

A similar phenomenon has taken place with phonology — traditional AAE pronunciations, now in competition with standard American pronunciations, have taken on an expressive role. For example, the traditional monophthong PRICE vowel (e.g. *lie* /la:/) has been largely lost in many African American dialects outside of the South, but survives in particularly emotive or affective expressions: *high* is typically pronounced /haɪ/, but in the phrase *getting high* may be pronounced /ha:/[9].

The story of African American English demonstrates the importance of the social context of an emerging dialect's first speakers in driving its linguistic features. In the case of AAE, its origins as the second-language dialect of West Africans leaves traces in its phonology, and the early interactions of African Americans with creole-speaking Caribbeans and white Southerners drove its tense system and grammar. From this foundation, the dialect has survived and changed, developing features distinct from both the white dialects that surround it, and the African and Caribbean languages that influenced it.

References:

[1] Maddieson, I. (2013). *Presence of Uncommon Consonants*. The World Atlas of Language Structures Online.
[2] Wolfram, W., & Thomas, E. (2008). *The Development of African American English*. John Wiley & Sons.
[3] Winford, D. (1998). On The Origins of African American Vernacular English—A Creolist Perspective: Part II: Linguistic Features. *Diachronica*, 15(1), 99-154.
[4] Huttar, G. L., & Huttar, M. L. (2003). *Ndyuka*. Routledge.
[5] Thomas, E. R., & Carter, P. M. (2006). Prosodic rhythm and African American english. *English World-Wide*, 27(3), 331-355.
[6] Gut, U. (2002). Prosodic aspects of standard Nigerian English. *Gut and Gibbon, eds*, 167-78.
[7] Mufwene, S. S. (1996). The founder principle in creole genesis. *Diachronica*, 13(1), 83-134.
[8] Bickerton, D. (1981). *Roots of language*. Karoma.
[9] Labov, W. (1998). Co-existent systems in African-American vernacular English. *African-American English: Structure, History and Use*, 110-153.
[10] Poplack, S. (2006). How English Became African American English. *The Handbook of the History of English*, 452-476.
[11] Bonfiglio, T. P. (2010). *Race and the rise of standard American* (Vol. 7). Walter de Gruyter.
[12] Cukor-Avila, P., & Bailey, G. (1995). Grammaticalization in AAVE. *Annual Meeting of the Berkeley Linguistics Society*, 21(1), 401-413

BEHIND THE BOOKSHELVES

22
Toxic Data:
Biases in Natural Language Processing

24
Wordslut:
A Feminist Reinvention of the English Language

Toxic Data:

Biases in Natural Language Processing

Nicholas Daines is a second-year student of Linguistics at the University of Edinburgh. In this article he talks about biases in Natural Language Processing and how positive change is possible.

Natural language processing is more impressive and more relevant than ever. Anyone studying a foreign language is likely well-acquainted with the absurd usefulness (and crutch) of Google Translate, DeepL, or the like. Both people and governments increasingly demand more active speech policing by social media giants, a task accomplished overwhelmingly by machine learning (ML). Word processors and email services now ubiquitously offer spell-check or nuggets of stylistic wisdom. The assumption that ML should deliver pure and neutral results is now laughable to most. Machine learning, it turns out, is very good at reflecting and amplifying our own biases.

Bias in NLP can derive from many sources. What people write down is not an accurate reflection of the real-world frequencies with which such things occur. The samplings we make of available data aren't neutral, either. People tend to select data that conform to the result they're looking for. Those biases result in the construction of a non-neutral dataset. But we also interpret our data in biased ways: We overgeneralize, we inappropriately frame correlation as causation, and once we've done all that, we trust the outcome simply because a computer said it.

Language models can be biased in many ways, but some biases have garnered particular attention. Take gender, for example: In word-embedding models, *He is a doctor* has a higher likelihood than *She is a doctor*[1]. If we ask the model to make analogies, we get results like *Man is to computer programmer as woman is to homemaker*[1]. Similarly, NLP models happily engage in and reproduce stereotypes about racial or religious groups, e.g. associating Islam with terrorism[8]. Such outcomes are not surprising: They reflect patterns in source data.

Those examples are based on stereotypes that we can easily recognize, but NLP bias is not always so neat or discrete. Something we might like to do is to identify hateful or toxic speech online, so that we can remove or deemphasize it. One way to achieve this is through a blacklist of terms, but this is not very sophisticated — hate speech is more than keyword recognition. Such simple models also quickly result in unintended consequences; the parallel constructions *I am a deaf person* or *I am a transgender person* are much more likely to be rated as toxic than *I am a tall person*[2]. Using a model that interprets certain disabilities, sexualities, genders, or religions as inherently toxic can lead to systematic devaluing of related content[3].

It is not just simple "blacklist" models that result in such miscategorizations. Studies have shown common academic datasets to be biased against African American English, flagging AAE Tweets as "toxic" 2.2 times more than average[4]. To understand why this happens, consider how the training data for such categorization tasks tends to be generated: we select some big datasets (perhaps Tweets), then we crowdsource their annotation — i.e. volunteers categorize each Tweet or comment as "toxic" or "non-toxic". The annotations in a dataset are then dependent on the attitudes of the annotators[5]. For example, the Stanford Question Answering Dataset (SQuAD), intended for use in computer comprehension of written works, was generated from the answers given by volunteer annotators in response to reading-comprehension questions about Wikipedia articles. Volunteer corpus-annotators and question-answerers are, of course, not a representative sample of Twitter- or language-users. Simply look at maps of the geographical distribution of crowdsourcing volunteers (of e.g. Amazon's Mechanical Turk, used by the SQuAD project), and this is clear. If we use inadequately constructed or biased datasets in training speech-regulation algorithms for social media applications, we can end up with discriminatory results.

Another manifestation of bias is in the quality of results. Consider language-identification, a necessary step in most applications of NLP. Recognizing standard American or British English is no problem, because we have so many datasets for those. Other varieties have more trouble: Tweets written in transliterated Hindi, or Indian English more broadly, are not easily classified by these systems. Having woefully inadequate NLP models for certain linguistic regions has proven dangerous: In the Philippines and Myanmar, where Facebook dominates, lack of competent moderation has been implicated in the incitement of political violence[6].

How can bias be addressed? Ameliorative measures fall into two broad categories: Debiasing the datasets and debiasing the models themselves. MIT recently took down its "Tiny Images" dataset for visual processing after learning that many of its annotations were racist, misogynistic, or otherwise harmful[7]. However, most datasets needn't be deleted, but rather improved. Fewer annotated datasets are based on Indian or African varieties of English, as India and Africa have lower smartphone- and internet-penetration rates than North America, UK, Australia, etc. Expanding and improving datasets from underrepresented regions could improve model fairness. NLP service providers can also address model-embedded bias directly, by using algorithms to identify and reverse it. Algorithms like "Hard Debias" have been shown to remove gender-stereotyped relationships (e.g. *female* : *receptionist* or *female* : *homemaker*) without removing linguistically-relevant gender information (e.g. *actress* : *female* or *queen* : *female*)[8]. Applying the same principles more broadly could help remove unwanted associations from language models.

Discussion and treatment of these issues can be difficult. In February 2020, ethical AI researchers Timit Gebru and Margaret Mitchell were fired from their positions at Google after refusing to retract a paper that raised concerns about the risks of massive language models[9]. The authors warned that the massive scale of Google's NLP datasets makes it nearly impossible to screen them for harmful bias. Google claimed that the paper did not meet its standards. Later that year, Google employees signed an open-letter in support of Gebru and Mitchell, and staged a walkout in protest.

Things can change. The open-source community "Hugging Face", which Mitchell recently joined, is working to advance fairness in NLP by democratizing access to it[10]. Emily Bender, a co-author on the troubled Google paper, has recommended that datasets come with "data statements": Contextualizing characterizations of the data that help researchers understand its bias and generalizability[11]. Bender also notes what might be the most promising factor of change: the need for companies to "protect themselves from future embarrassment". While the incentives of corporate sponsors of NLP are complicated, recent research has shown that bias can be reliably identified in training data and models can be cleansed of it. That leaves room for hope.

References:

[1] Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K. W., Wang, W. Y. (2019). *Mitigating gender bias in natural language processing: Literature review*.
[2] Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y., & Denuyl, S. (2020). *Social biases in NLP models as barriers for persons with disabilities*.
[3] A group of YouTubers is claiming the site systematically demonetizes queer content. (2021). https://www.vox.com/culture/2019/10/10/20893258/youtube-lgbtq-censorship-demonetization-nerd-city-algorithm-report
[4] The algorithms that detect hate speech online are biased against black people. (2021). https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter
[5] Chang, K. W. (2019). *Bias and Fairness in NLP*. http://web.cs.ucla.edu/~kwchang/documents/slides/emnlp19-fairNLP-part1.pdf
[6] Conviction in the Philippines Reveals Facebook's Dangers. (2020). https://www.nytimes.com/2020/06/16/technology/facebook-philippines.html
[7] Johnson, K. (2021). MIT takes down 80 Million Tiny Images data set due to racist and offensive content. https://venturebeat.com/2020/07/01/mit-takes-down-80-million-tiny-images-data-set-due-to-racist-and-offensive-content/
[8] Bias in Natural Language Processing (NLP): A Dangerous But Fixable Problem. (2021). https://towardsdatascience.com/bias-in-natural-language-processing-nlp-a-dangerous-but-fixable-problem-7d01a12cf0f7
[9] Hao, K. (2020). We read the paper that forced Timnit Gebru out of Google. Here's what it says. https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-gout-timnit-gebru/
[10] Apte, P. (2021). How Hugging Face is tackling bias in NLP. https://venturebeat.com/2021/08/25/how-hugging-face-is-tackling-bias-in-nlp/
[11] Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6, 587-604.

Wordslut:

A Feminist Reinvention of the English Language

Beatrix Livesey-Stephens is a final-year Language and Linguistics student at the University of Aberdeen.

Wordslut[1] was the book I needed as a fresh-faced Linguistics student who didn't yet have any idea about just how much there was to say about gender and language. Here, Amanda Montell presents an innately accessible account of the relationship between the two from a number of fascinating angles. As the title suggests, Montell doesn't hesitate to dive into taboo aspects of language and gender, such as curse words and genitalia, right from the get-go. This sets the reader up to quickly understand that the language of femininity, gender, insults, and more, are all inextricably connected. Readers of *Wordslut* won't just have a new understanding of how language and gender affect each other, but will undoubtedly have numerous thoughts of their own about why and how this happens, long after closing the book. Above everything else, *Wordslut* is hugely thought-provoking, and will inspire many more linguists in years to come.

Montell gives considerable thought to how women's speech styles differ from men's, while acknowledging that hardly any research has been done on those who belong to neither group. Women tend to have collaborative speech styles which attempt to encourage conversation from multiple members in the group, while men are less collaborative, and more competitive. What particularly struck me was the dichotomy between gossip, a term reserved for women, and *(locker-room) banter*, which Montell calls "a manlier-sounding synonym for gossip". I had never considered this before, and was suddenly reminded of that time a few years ago where everyone around me referred to any sort of lighthearted chatter not as *gossip*, but as *banter*. When banter reached its peak at the end of 2014, its most common occurrence was undoubtedly *banter with lads* — essentially, "gossip with men". Comparing *banter* and *lads* in Google gives you remarkably similar trends[2]. Montell asserts that "the word *gossip* and its trivial implications have been pegged a feminine thing", and she's absolutely right.

Montell covers a wide range of topics in *Wordslut*, there's something for everyone. She devotes chapters to how women speak to each other when men aren't around, differences in speech between genders, the double standards of how women are spoken about, such as Hillary Clinton and Scarlett Johansson, and more.

I guarantee that whatever gender you are, this book will encourage you to reevaluate your perceptions of gendered language. *Wordslut* has helped me embrace the way I speak as a woman, and I've begun to challenge my own sexist thinking in regard to criticising how other women speak. I can't count the number of times I read one of Montell's sentences and thought "I couldn't put this into words, but now it's on a page and I understand the specifics of my idiolect so much better!" Did you know that hedging is in fact most often used as a tool to be tactful and negotiate sensitive topics, rather than indicating indecision? Women will typically use terms like "maybe", "possibly" or "kind of" much more than men, in order to avoid sounding rude, bossy, or controlling. I've spent my whole life hedging, and didn't quite understand why until reading *Wordslut*. I've tried to stop hedging as much as I used to, and can already feel myself becoming more confident.

The accessible nature of *Wordslut* hinges on Montell's brilliant and chatty style. It really brings the book full circle, and implicitly challenges readers' perceptions and assumptions about how women "should" speak. Whoever Montell had in mind when writing, I think she's found a voice that will delight anyone who's interested in finding out more about anything language and gender related, from catcalling to cursing. To say anything else would be robbing readers of *U-Lingua* of the delight of discovering *Wordslut* for themselves. I'm sure every reader of *Wordslut* will get something different out of Montell's work, and I know I'll be referencing it for a long time.

References:

[1] Montell, A., 2019. *Wordslut*. 1st ed. New York, N.Y: Harper Wave.
[2] https://trends.google.com/trends/explore?date=2010-10-08%202021-11-08&geo=GB&q=banter,lads

Book Review

28
Inventions and Reinventions

THROUGH THE AGES

Alice Eddyshaw, University of Manchester

Helicopter comes from the Greek words *helix* (the root *helik-* meaning "spiral") and *pteron* (the root *pter-* meaning "wing") so *helicopter* literally means "spiral wing". It is also interesting to note that because English doesn't allow the cluster /pt/ at the start of a syllable, speakers resyllabify *helicopter* as /hɛ.lɪ.kɔp.təɹ/, even though that doesn't align with the boundary between the two root words.

helicopter. (2021). In *Oxford English Dictionary*. Oxford: Oxford University Press.

Anon., Bangor University

Plumber comes from Latin *plumbum*, meaning "lead" (the metal), and *plumbarius*, meaning "someone who works with lead".

plumber. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/plumber

Jake O'Keeffe, Durham University

Algebra comes from the Arabic *al jebr* which means "reunion of broken parts". *Jebr* means a splint or a cast, which is used to reset bones in the right way. In fact, when *algebra* was first loaned into English in the 15th-16th centuries it also meant "bone-setting".

algebra. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/algebra

Anon., Bangor University

Salary comes from the Latin word *sal* meaning salt, because Roman soldiers were paid in salt.

salary. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/salary

Anon., Bangor University

In *nightmare*, *mare* originally referred to a demon from Germanic folklore that brings people bad dreams, specifically of suffocation. Originally *nightmare* referred to the suffocating sensation the mare brings, but in the 19th century came to refer to bad dreams in general.

nightmare. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/nightmare

Anon., Bangor University

Martelé is a bowing technique for string instruments that produces an emphatic sound. It comes from the French verb *marteler* meaning "to hit with a hammer" as it is quite an aggressive playing technique.

Bachmann, W. et al. (2001). Bow. *Grove Music Online*. https://doi.org/10.1093/gmo/9781561592630.article.03753

Anon., Bangor University

Skunk comes from the word *squunck* from the Algonquian language Massachusett. The Proto-Algonquian root is */šeka:kwa/*, a combination of */šek-/* ("to urinate") and */-a:kw/* ("fox").

skunk. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/skunk

Will Rimer, University of Cambridge

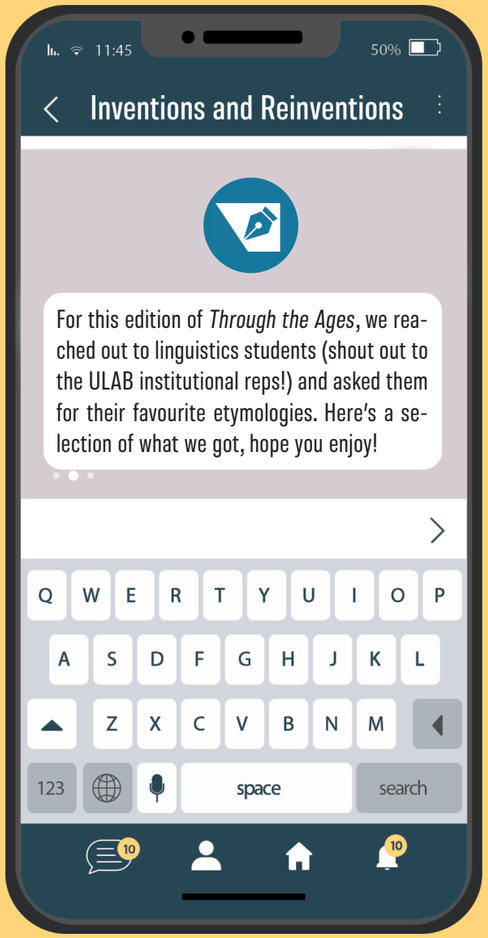
The Gaulish word *braga* was borrowed into Old French as *brague*, meaning *trousers*. Then the diminutive form *braguette* was taken to mean *codpiece*. At some point, someone decided that the right-angled supports which hold up a roof, for example, looked like Henry VIII's particularly fancy armoured codpiece, and so began using the newly-borrowed-into-English word bracket to refer to them. Then the reference of the word was extended to two right-angles put together, i.e. [ ]. Finally, this was generalised to all parentheses (at least for people who are less terminologically precise than the Americans), so ( ) came to be called *brackets*.

bracket. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/bracket

Romany Amber, University of Cambridge

Avocado comes from Spanish. Its original form was *aguacate*, but its form changed due to a folk-etymological link with the unrelated Spanish word for "lawyer", *avocado*. *Aguacate* derives from the Nahuatl (Aztecan) *āhuacatl*, which is also the Nahuatl word for 'testicle' – based, apparently, on similarity in shape and surface texture. There doesn't seem to be consensus on which word was named after the other, begging the age-old question: What came first, the avocado or the testicle? So modern *avocado* either inspired, or is derived from, an ancient word for testicle.

avocado. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/avocado



James McLean, University of Manchester

The suffix *-core* (as in *cottagecore*, *grandmacore*) is a back-formation from *hardcore*. The noun phrase *hard core* originally referred to a solid core foundation in buildings, but became an adjective in the 1930s (e.g. *hard-core unemployment*) and was then picked up by the porn industry in the 1950s to describe intense films, i.e. *hardcore porn*. It eventually came to describe particularly fast, loud rock music (e.g. *hardcore punk*), so *soft-core* arose in contrast to it. The suffix has since then become productive, and is used to describe a style associated with a particular aesthetic, such as *cottagecore*!

Why do we use the phrase "hard core"?. (n.d.). Retrieved from Merriam-Webster website: https://www.merriam-webster.com/words-at-play/hard-core-meaning-origin

Ruth Chapman, University of Cambridge

Black originally meant both "black" and "white" because in Middle English people couldn't agree if *blak* meant the ashes after a fire or the bright light during it. Both meanings come from the same root **bhleg-* meaning "to burn" in PIE, but the forms were different in Old English, with *blæc* meaning "ashes" and *blac* meaning "light". The loss of the letter <æ> in Middle English meant that the written forms merged.

black. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/black

Anon., Bangor University

The word *crystal* and the surname *Crystal* (as in David) are false cognates! *Crystal* the rock comes via Old French and Latin from the Greek *krystallos*, from *kryos* meaning "frost". The surname *Crystal*, and its alternate forms such as *Christal*, *Christole*, and *McCrystal*, come from a diminutive spelling of *Christopher*, which means "Christ bearer", from the Greek *Christos* ("Christ") and *pherein* ("to carry").

crystal. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/crystal
Christopher. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/christopher
Crystal Last Name Origin. (n.d.). Retrieved November 4, 2021, from The Internet Surname Database website: https://www.surnamedb.com/Surname/Crystal

Anon., Bangor University

The adjective *jumbo* with its present meaning "unusually large" emerged in the late 19th century, when in 1882 an unusually large elephant named Jumbo was sold form London Zoo to American circus showman P T Barnum. *Jumbo* meant a "clumsy, unwieldy fellow" in 19th century English slang, is of uncertain origin, but the *OED* proposes that it comes from *mumbo-jumbo*, a borrowing/corruption of *maamajomboo* in Mandinka, a Mende language spoken in Senegal, referring to a masked dancer participating in religious ceremonies.

jumbo. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/jumbo
mumbo-jumbo. (2021). In *Oxford English Dictionary*. Oxford: Oxford University Press.

Anon., Bangor University

Assassin comes into English via medieval French and Italian, but ultimately comes from the 12th century Arabic word *hashīshīn*. That's also the source of *hashish*, a word for cannabis. *Hashīshīn* was a nickname for that 12th-13th century Europeans gave to Muslim assassins because believed the assassins would would eat lots of hashish before attempting assassinations.

assassin. (n.d.). In *Online Etymology Dictionary*. Retrieved from https://www.etymonline.com/word/assassin

A HANDS-ON APPROACH

How To... Rebuild A Language

In this instalment, Cara McSherry gives us ideas about what we can do to revitalise a language, taking Scots as a working example.

Scots gets it tough. Its status as a language is hotly contested — more often than not it gets labelled everything from "dialect" to "slang" to "bad English". We've all heard the overplayed (and frankly offensive) line on Scottish people needing subtitles to be understood. But there's a very vocal group of online activists who are trying to shift people's perceptions.

1. Write poems.

At the forefront of these efforts is poet Len Pennie, who's combatting the anti-Scots feeling, quite literally, one word at a time. Aside from her fantastically fiery poems, such as *I'm no havin' children*, Len has amassed over 100,000 followers on Twitter for her Scots word of the day videos, where she celebrates the language in a fun, accessible, and very digestible way. Her mission is bolstered by a plethora of others who feel similarly passionate about their *leid* (language).

2. Sing renditions of hit songs.

Iona Fyfe is another big name — a singer from Aberdeenshire known for singing in Scots, with a particularly iconic Scots rendition of Olivia Rodrigo's *Driver's License*.

3. Share on social media.

Dr Michael Dempster is a similarly enthusiastic proponent — the Director of the Scots Language Centre and Creative Scotland's Scots Scriever, he brings both academic expertise and a touch of humour to Scots Twitter, through his *accidental Scots* posts highlighting the many things he comes across that have a dual meaning for Scots speakers.

These individuals are some of many: There is now a very significant community of Scots speakers who connect online to speak in, celebrate and promote their language. And such advocacy is necessary — after all, Scots is one of Scotland's minority languages and, in many contexts, finds itself overshadowed by its close neighbour — English.

Scots and English developed distinctly over many centuries but have always existed in close quarters and this has, unsurprisingly, led to them growing broadly similar. But it is this similarity which causes a lot of grief for Scots — it is seen as inferior, either a relic of times-gone-by or the slang flung about by "wee neds". And it is undeniable that the borders are blurry.

Take, for example, the events of August 2021. It emerged that, over the course of many years, one young enthusiast had taken it upon themselves to write thousands of articles for the Scots language version of Wikipedia. The catch? This individual was based in America and had zero connection to the Scots language — they had simply been writing "English with an accent". This caused quite a stir among online Scots activists, with many pointing to the damage this individual had done to the status of an already minoritized language. It was clear that this individual hadn't gotten Scots *right*. And yet, few could point to *why* this individual was so wrong. Scots lacks any standardised orthography — its speakers generally use Standard English for writing, although activists have been constructing an *ad lib* orthography by using mediums like Twitter to share Scots content. And that intertwined relationship with English made it difficult for activists to meaningfully criticise what they perceived as an attack. Scots' connection to English also proves a sticking point for activists, who receive regular backlash for simply expressing themselves in Scots online. Critics deride their use of Scots, goading them to "speak properly" and suggesting that Scots is nothing more than a cheap knock-off of English — the *real* language. This abuse has grown so bad that Len Pennie has given her account to a friend in order to avoid the trolls.

Clearly, it's not easy to resurrect a language. Activists for Scots face resistance at every turn and have to fight hard to enshrine their language's status, and indeed its existence. The case of Scots highlights just how political language can be — the borders between one language and another are rarely just linguistic, but rather come with myriad social and ideological complexities that trace divides between communities of speakers. And that's not something that can be changed overnight, as Scots activists know all too well. The best you can do is to stick up for your own tongue and, when the inevitable backlash descends, try your hardest to keep the *heid*.

Your Turn: Translating Poetry

Why not try your hand at translating some Georgian modern poetry? Excavate all the sources and dictionaries you need, and send your translations in to ulinguamagazine@gmail.com — we'll print them in the next Issue!
Bonus tip: Can the English translation match the Georgian metrical structure?

აუტიზმი. ამეტყველება (Autism: Beginning to Speak)

Diana Anphimiadi[1]

როგორ სათითაოდ ჩამოხოცა
მავთულზე ჩამომსხდარი სიმღერები
ყინვამ.
ყურებზე ხელი ავიფარე
ვისმენ – სიტყვა სად იბადება,
ვინ ვარ.
დავდივარ თვალებით – ღია გალიებით –
რომ შიგნით მოვიმწყვდიო
ჭრელი თუთიყუში-ქვეყანა.
ჩემ წინ და ჩემ უკან სამიოდ ნაბიჯი
და მერე ხელახლა
და მერე თავიდან
მსოფლიოს გარშემო რკინიგზა გამეყვანა
მსურდა.
და ვტრიალებ, ვტრიალებ
ვტრიალებ
ვიწევი. ვმუსუბუქდები, ვმჩატდები, ვვარდები . . .
ღრუბლების ხაო და ნოტიო შეხება
დედამ ჩამოკიდა ახალი ფარდები.
პირში მიწის ოდნავ მომჟავო გემო აქვს
ბიჭს, ჩემში რომ უდგას საწოლთან ტორშერი
და თბილი ხელების დათრგუნველ შეხებას
ძვირფასი თვალების ალესილ შეხედვას
კაშკაშა პროჟექტორებს თოფივით მოშვერილს
იშორებს . . .
ცაში კი სიტყვები ჩხავიან
როგორც თოფის სროლით დამფრთხალი ყვავები
გაჩუმდი, გეყოფათ!
ერთს მოვდე ანკესი –
და თავზე
და თავში
მთელი ცა მემხობა!
ვიწყებ –
დავაწერე ყვავების გუნდები
სანოტო ხაზებს, ტელეფონის მავთულებს –
პირში მიჩხავიან მარტივი სიტყვები –
რომელთაც თან და ქვე ვაწყობ
და
ვართულებ.

[1] Anphimiadi, D. (2018). *Beginning to Speak*. The Poetry Translation Centre.

Words, Words, Words

The Weary Self-Awareness of Contemporary Poetry

Literature columnist Olivia Szczerbakiewicz (University of Edinburgh) delves into what makes a poem a poem in the contemporary world.

What do we understand as poetry? The answer to this question always seems to have blurry boundaries: a nebulous, whimsical thing marked with emotion, evading definitions. Something that *isn't what it is*, dealing with tools which, by definition, point to or hide other things, like metonymy and metaphor. Yet, if we go far back enough to the tradition of *Beowulf* or Shakespeare's sonnets, poetry becomes something easily measurable and — provided we stick to the rule book — decipherable. A sonnet, after all, has a set number of lines, a set rhyme scheme, a pre-ordained limit its poetics are constrained within.

Contemporary poetry, however, offers us no such thing. Shying away from traditional rhyme or structure, it tends towards free verse and seems to employ a wholly different set of tools to elicit an emotional response in the reader.

Let us consider the opening of Richard Siken's *Litany in Which Certain Things Are Crossed Out*:

A thing that strikes us almost instantly after reading it is the poem's own self-referentiality. The text is self-aware — *metatextual* — and hence recognises itself as what it is, pointing out its own shortcomings and predictability.

From the first verse, Siken juxtaposes the "prosaic" — trivial and belonging to the real world — with the literary. The anaphoric *every morning* highlights the mundane recurrence of both the described world (maple leaves) and the act of writing itself. *The Hero's Journey* remains the same trope across the ages, and the same "big and little words" —

> Every morning the maple leaves.
> Every morning another chapter where the hero shifts
> from one foot to the other. Every morning the same big
> and little words all spelling out desire, all spelling out
> You will be alone always and then you will die.
> So maybe I wanted to give you something more than a catalog
> of non-definitive acts,
> something other than the desperation[1].

the very metaphors and figurative language we established lie at the core of poetry — remain the signifiers of the same things they always have been.

The act of writing doesn't change, and poetry grows old and predictable even to the poet. What comes next?

Something quite cynical, Siken notes, with his cutting assertion of *You will be alone always and then you will die*.

The lyrical expectation of what comes out of the creative process is subverted, but it does invoke emotion — doubly so. On one level, we are faced with the despair of someone doubting his own condition and ability to love. On the other, we are faced with a lapse in the poet's trust in the function of poetry. His definition of it is not beautiful but hopeless — and it doesn't seem to be easily contained. Instead, it escapes him.

And so, the opening section thus ends with an admission: *Maybe* — says the lyrical *I* — *I wanted to give you* — both the lyrical object and the reader, intertwined in the double-level of the poem's analysis — *something more*.

The *catalog of non-definitive acts* can be read as metaphor itself, standing in for the "back-log" of pre-existing, well-worn literary tradition. The speaker is tired of repeating something which to him feels hollow — and of the "desperation" with which he sought to escape what he perceives as the cliché of his feelings and his own lyrical expression of them.

Different poets take this notion even further — parallel to Siken's Litany, let us consider a fragment of Hera Lindsay Bird's poem *Monica*, titled after one of the characters of the well-known sitcom *F.R.I.E.N.D.S.*.

> Often I look at the world
> And I am dumbfounded that anyone can function at all
> Given the kinds of violence that
> So many people have inherited from the past
> But that's still no excuse to throw
> A dinner plate at your friends, during a quiet game of Pictionary
> And even if that was an isolated incident
> And she was able to move on from it
> It still doesn't make me want to watch her on TV
> I am falling in love and I don't know what to do about it
> Throw me in a haunted wheelbarrow and set me on fire
> And don't even get me started on Ross[2]

A refreshing pulse of audacity runs through the structure of this poem: There is something surprising and almost a little disturbing about it. Involving a degree of intertextuality, the poem expects the reader to reconcile their pop-cultural knowledge of *F.R.I.E.N.D.S.* — something, arguably, quite *un-poetic* — with the expressed fears and anxieties of the speaker.

Poignant questions such as being able to "function at all" under the weight of "inherited" (and, in a sense, traditional violence and horrors of the past are contrasted with a popular story. The speaker contrasts her frustration with the expectations on how to deal with trauma with her frustration with Monica — the ill-adjusted, in her opinion, and jarring representation of what humanity has created as a response. She doesn't want to identify with Monica or repeat the same well-worn response to the old well-worn problems of humanity.

And yet she is "falling in love" and asks to be "thrown in a haunted wheelbarrow and set on fire" — which is neither subjective nor unexpected. It is, however, both quite ironic and touching to the reader in its anxious familiarity.

The melancholy and frustration linger — much like Siken, Bird is conscious of her own allegiance to the literary tradition, understood also as a *pattern*: broadly, a pattern of creativity, and more specifically, of poetry. Likewise, Bird employs a startling set of tools to disrupt the "old order" as much as possible. She weaves a highly self-aware web of intertextuality, cutting language and irony, names it after a character in a sitcom and daringly tells us — *this is a poem*.

And this, perhaps, is the conclusion we can draw about what a contemporary poet does with the language at their disposal: much like a linguist, they step back and deconstruct it even while using it. Emotion remains at the core of the poem, but the mechanisms of evoking it change and shift, often calling into question the old patterns. Self-awareness becomes of the driving forces of modern poetry, alongside the questioning of the expected in order to seek "something more."

References:

[1] Siken, R. (2005) *Crush*. New Haven, Ct: Yale University Press.
[2] Bird, H. L. (2017). *Hera Lindsay Bird*. Penguin UK.

Curiouser and Curiouser

Getting into Postgraduate Study

Thinking of pursuing a Masters or PhD? Unsure how to go about it? In this issue we look at practical steps you can take to prepare for the next big step of your academic life. Columnist Caitlin Wilson shares advice from recent graduates to help you navigate postgraduate applications.

The start of the new year brings new beginnings and new opportunities. For students in their final year, this time of the year can bring about a lot of questions about the future. *What am I going to do after graduation? Do I want to do another degree? Should I be applying for a PhD? What about the graduate job market?* This issue, I hope to answer a few of these questions and provide some information on one of the post-graduation possibilities: Applying for a postgraduate degree. I've asked a few postgrads at various stages in their degrees to tell me about how they found the right programme, what the application process was like, and more.

01 Choosing the right programme

Postgraduate degrees are not one-size-fits-all. There are many to choose from and it can be hard to know which is the right path to take. The general options are MA or MSc if you've just finished your undergraduate, or in some cases you might be able to jump straight into a PhD. Certain universities, such as Oxbridge, also offer MPhils which are more research intensive. There are of course other options such as PGCerts although these are less common for theoretical subjects like Linguistics. The main thing you need to think about, says Steph (MPhil), is to decide whether you'd rather be taught or do research. Thinking about this will help you decide what type of programme to apply to, but also which institution is best suited to you, as some universities place more importance on teaching compared to independent research and vice versa. If you've already finished a masters level course, you might be thinking of moving on to doctoral level study, in which case you'll be committing to three or more years of research along with the possibility of doing some teaching yourself.

02 Picking a topic

Once you've narrowed down the type of degree you want to do and a university you're interested in, you need to think about what it is you want to research and who will be your supervisor (No need to worry about this if you're applying to a taught masters!). Eloise (PhD) recommends having a fairly stable research topic in mind that isn't too vague before you start your applications. Universities want to know what you'll be bringing to the department with your study, so pick something specific enough that it won't have been done before. Some general advice is to read lots! Maggie (PhD) found her dissertation topic by reading a particularly interesting article on vowel perception and then realised that there was lots more potential research to be done in this area. Tom (MPhil) seconds this idea. He recommends scrutinizing your past research as well as others': "Find the gaps and figure out if there's more to be said". If you're still not 100% certain on a topic, don't worry! Most people I've spoken to agree that there is a certain amount of leeway surrounding your topic. Once you've been accepted onto a programme, you'll probably be able to change directions further down the line.

03 Picking a supervisor

Once you've determined what it is you want to research, you'll need to find someone to supervise you. This can be a very simple process. For instance, if you want to stay at your undergraduate university, you'll already know a lot of staff and might have already worked with some on previous research. However, if you're starting fresh in a new institution, you'll need to figure out if anyone there will be able to help you. Eloise recommends starting with the university website to figure which staff members work in the same field as you and if they're accepting new supervisees. Steph recalls sending lots of emails. "Don't be shy to reach out to as many professors as possible", she says. You might have a specific researcher in mind when starting your application but don't bank on them being able to accept you. Tom reminds us not to take it personally if someone rejects your request: "If they say no, it's not a personal attack. It's all about interests. If their research interests don't align with yours, it was never going to work."

04 Practicalities of applying

So you've decided on a degree, narrowed down a topic, and spoken to a few potential supervisors. Time to bite the bullet and start the application. The actual practicalities of the application process will vary from university to university but in general you'll have to fill out some forms, write a personal statement/cover letter, research proposal, and provide a few references. If you're applying to a PhD programme, the research proposal will take up the bulk of this work and will be far more extensive than for a research masters. Certain programmes also interview potential applicants. If you've got one coming up Tom suggests being overprepared: "Think of all the possible questions they could ask and prepare for them."

05 Last pieces of advice

I'd like to finish this column with a final piece of advice from the postgrads I spoke to.

Steph: "The most important step is deciding that you do actually want to stay in education for the next (few) year(s)."

Maggie: "Postgrad personal statements tend to be more academic focused than undergrad personal statements: universities want to know what you have studied so far and how you want to build on your current knowledge, rather than about extracurriculars and non-academic work."

Tom: "Strive to be innovative and passionate at all stages of the application process. Don't be afraid of rejection because it will come, you'll get there eventually."

Eloise: "My main advice is to actually seek advice! It can feel like a lonely process but there are plenty of people around you who can help."

Good luck, you can do this!

Anatomy of a Linguist

What Keeps Us Up at Night

The Philosophical Value of Big Data: A Plea for Intention Tracking

In this edition of Anatomy of a Linguist, columnist T. R. Williamson tries to shed new light on the use of big data in linguistic research. Understanding the philosophical potential of corpus linguistics may lead to a wealth of opportunities, academic and beyond.

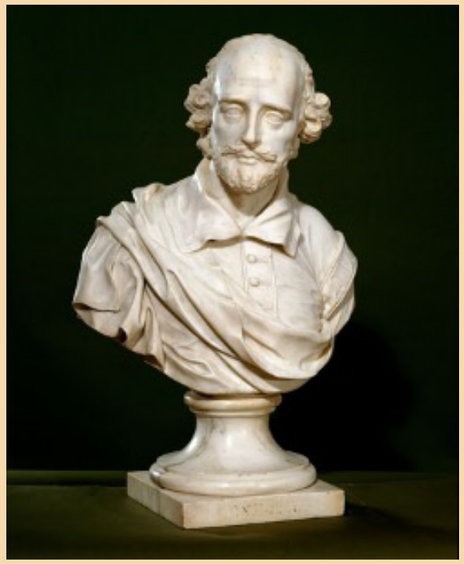
Linguistics departments across the world have a secret. Somewhere in their buildings – perhaps in a basement, a disused teaching room, or an old office – they keep confidential equipment. You may have seen such areas yourself. Perhaps you noticed a suspicious *DO NOT ENTER* sign, spotted boarded-up windows, or even walked by as a whisp of ghostly, cold vapour seeped from underneath a door. What are these departments doing? What are they hiding?

Had you followed this eerie, gaseous substance and entered where you were not permitted, you'd have uncovered their covert operations. For decades, linguists have been conspiring with government agents to cryogenically freeze language and store it away from prying, undergraduate eyes. They hold words and sentences captive in massive vats, frozen with nitrous oxide, and run dangerous tests on them. From checking parts of speech to counting bigram frequency and everything in between, these evil language scientists poke and prod without their specimens' consent. They must be stopped.

And what is the name of this malevolent discipline? Corpus linguistics.

This is, of course, nonsense: corpus linguists are lovely, they do not conspire with the government, and words have no feelings. Yet, certain parts of this depiction seem to present interesting parallels with the work corpus linguistics actually involves. It is the field that concerns itself with the collection, organisation, and presentation of instances of language use for further analysis[1]. A corpus linguist might digitalise old documents or record audio from real-world conversations with the aim of collecting that data and presenting it in a searchable format.

Want to know how many times the term *fake news* was said by Donald Trump in tweets from 2016-2020? You can make a Twitter corpus and find out[2]. Want to know how many times the word *bastard* was used informationally or offensively throughout the entirety of Shakespeare's works? You needn't make your own corpus; it's already been done[3]!



In this way, parallels arise. Corpus linguists take samples of language, frozen in time, and test them to find out the answers to further questions. But there is a deeper potential to corpora, one for which we do not (and perhaps even should not) use them but nevertheless one which they may yet still fulfil.

Famous cognitive linguist Charles Fillmore once drew quite a stark comparison between what he called "armchair" (i.e., generative, Chomskyan, nativist) and "corpus" linguists. The former type, he suggests, is often caricatured as utilising methods akin to sitting in an armchair and thinking up sentences that seem to present novel facts about the structure and organisation of language in the mind (i.e., Universal Grammar)[4]. This is not true, of course – plenty of fieldwork and other typological efforts inform generative linguistics. In any case, such sentences might, for instance, be grammatical in surprising ways or demonstrate a surprisingly felicitous usage of a word with seemingly inflexible semantics.

However these linguists get their examples, one thing remains clear. As objects of analysis, they are taken as isolated, decontextualised linguistic expressions with morphosyntax that can shed light on the structure of Universal Grammar via deduction. The meanings of words in these sentences are interesting only insofar as they constitute the conditions under which those sentences would be true (e.g., nouns and verbs referring to phenomena in the world).

The latter, "corpus", linguists take a different approach. With the linguistic examples presented to them by corpora, Fillmore suggests they are often caricatured as running unenlightening statistical analyses on phenomena that have no utility for describing the ultimate nature of language. Rather than attempting to explain morphosyntax or semantics, they look to observe and describe language use, in-context, with any and all other factors considered.

> "These two don't speak to each other very often, but when they do, the corpus linguist says to the armchair linguist, 'Why should I think that what you tell me is true?', and the armchair linguist says to the corpus linguist, 'Why should I think that what you tell me is interesting?'"[4]

Words like *statistical* and *describe* don't seem to have much philosophical relevance. When you think of philosophy, in many ways, perhaps even the concept of an "armchair" linguist comes to mind: One might picture an old man stroking a bushy white beard atop an ivory tower making grand generalisations about the nature of the universe. The exclusion of corpora from philosophical considerations is a mistake, I conjecture, and one that is derived from a lack of appreciation for a third crucial term in corpus linguistics: *Context*.

One of the corpus linguist's most vital assets is that their examples come packaged into the linguistic environment of their original usage. For a given search item, which might be a word, phrase, or anything else, one can observe the linguistic context in which it arose; what came before, what came after, what it tends to occur around, where in the sentence it tends to appear, and so on. But moreover, the corpus analyst also gains privileged access to a number of important features that give them a feel for wider, extra-linguistic clues. In the Spoken British National Corpus (BNC) from 2014, for example, one can search by number of interlocutors, region of the UK, and even social class[5].

Your query "door" returned 2,134 matches in 632 different texts (in 11,422,617 words [1,251 texts]; frequency: 186.822 instances per million words) 0.078 seconds – retrieved from cache

|< << >> >| Show Page: 1 Line view Show in random order Choose action...

Solution 1 to 50 Page 1 / 43

No	Text		
1	S1JA 2796	many things **S0094:** oh er **S0095:** and you can kinda compare it to the	deer down the stairs which has been oiled **S0021:** okay **S0095:** >>cos that 's oak
2	S24E 338	things like that **S0520:** mm **S0521:** but the Norwegians they were out of the	deer at four o'clock whatever happened **S0520:** yeah no chance **S0521:** it was very much
3	S263 380	rug **S0616:** so he does n't ye- I mean I notice like the	deer 's been open sometimes some morning and I presume the cat comes
4	S263 401	think not cat **S0590:** mm (.) and then he stands and he wants the	deer open and you say **S0616:** >>yeah **S0590:** you 've got a you 've got
5	S263 405	boy **S0616:** >>that 's it yeah off you go **S0590:** >>so I want the	deer open **S0616:** mm (...) mm (...) **S0590:** apparently --ANONnameM I was speaking to --ANONnameM this morning
6	S263 1487	er I mean th- she 's just moved into the house next	deer to ours erm **S0616:** mm **S0590:** or next door but one to our house
7	S263 1489	moved into the house next door to ours erm **S0616:** mm **S0590:** or next	deer but one to our house in in --ANONplace I think retired **S0616:** >>mm
8	S263 1591	good representation **S0616:** mm **S0590:** er **S0616:** mm (...) **S0590:** well I think I shall go next	deer **S0616:** yeah I 'll I 'll **S0590:** >>yeah **S0616:** >>I 'll speak to --ANONnameF and
9	S263 1635	yeah **S0588:** that sort of thing **S0616:** that 's true (...) **S0589:** could you shut the	deer ? I know you like to leave it open dear but it
10	S263 1767	go out do you --UNCLEARWORD ? **S0616:** (...) but to we (.) decided the snug	deer is a standard size ? **S0589:** it must be **S0616:** yeah **S0590:** the snug door
11	S263 1770	door is a standard size ? **S0589:** it must be **S0616:** yeah **S0590:** the snug	deer is what --ANONnameM ? **S0616:** a standard size **S0589:** >>standard size but it 's
12	S263 1775	it 's **S0589:** >>it 's about five mils smaller or wider than that	deer there **S0616:** but the one into --ANONnameF 's studio **S0589:** is weird **S0616:** >>is different
13	S263 1784	if it 's a standard size I might I might have a	deer you know and and I (.) I know that I have n't got
14	S263 2629	it 's no it 's actually it 's that length less the	deer **S0616:** yeah **S0590:** that 's actually what it is **S0616:** yeah **S0590:** because that door oh
15	S263 2633	the door **S0616:** yeah **S0590:** that 's actually what it is **S0616:** yeah **S0590:** because that	deer oh no it 's a little bit in **S0616:** yeah **S0589:** no it 's
16	S263 3106	like McDonald 's to him **S0590:** yeah **S0616:** so what does er --ANONnameM next	deer feed his cats ? **S0588:** I think it 's Felix **S0589:** it 's Felix
17	S26N 389	yesterday and **S0012:** oh yeah **S0013:** and **S0012:** we 'd only then walked in the	deer **S0013:** our sister-in-law **S0012:** just finished off a cup of tea were n't we
18	S26N 868	went to the lighthouse and we we said we were just next	deer come from just next door and --UNCLEARWORD asked her **S0012:** oh yeah **S0013:** and
19	S26N 868	we we said we were just next door come from just next	deer and --UNCLEARWORD asked her **S0012:** oh yeah **S0013:** and said **S0012:** oh you bought a
20	S27D 491	you know **S0024:** (.) yeah (.) sitting with the I- the student loans man next	deer **S0024:** in his mansion (.) they probably all go round to each other 's
21	S2AJ 91	sits in the lawn and **S0439:** so what happened what happened with next	deer ? **S0441:** oh yeah so he was fine he text me sort of
22	S2AJ 106	back but I did n't get it so he knocked on the	deer and was like can I borrow the machine ? and I was
23	S2AJ 110	grateful again now **S0441:** then he came back and he knocked on the	deer and I had it ready for him just like --UNCLEARWORD he was
24	S2AJ 112	he brought that round (.) and then (.) this was like standing outside his	deer like and I saw him like fill the kettle up and put
25	S2AJ 234	because I he caught me by surprise when he came through the	deer first of all I was just wearing like a little summer dress
26	S2AJ 235	be in your entire life and the window cleaner came to the	deer and I was just like hello **S0439:** and you know when you just
27	S3B5 58	so as well **S0024:** quite rightly you do n't want that on your	deer step **S0144:** >>yeah (.) but it was I used to hate taking that dog
28	S2C9 454	train ? **S0336:** yeah **S0362:** I think I was asleep right next to the	deer I was just knackered **S0336:** but yeah erm (.) we do n't have like
29	S2C9 1307	**S0336:** >>no on their like you know as you walk through the	deer on the right-hand side ? **S0362:** >>yeah yeah **S0336:** like I think they **S0362:** >>that
30	S2C9 1400	that erm one of --ANONnameM 's friends is buying the house next	deer ? **S0336:** oh I did hear that **S0336:** that 's my idea of a
31	S2C9 1403	love you guys but like I would n't wan na live next	deer to you like (.) I think I 'd I 'd start to really
32	S2C9 1403	whatever like is n't that awkward to not like knock on their	deer and be like oh **S0362:** too much **S0336:** too much do you want a
33	S2C9 1430	'd think well that 's a bit weird that --ANONnameF 's next	deer and we should probably invite her **S0336:** I know **S0336:** you and it 's
34	S2C9 1441	'd be almost pissed off if someone was like knocking on the	deer every so often going hi like completely uninvited cos you 're not

Figure 1: Search for "door" in the Spoken BNC2014 via CQPweb[6]

These linguistic and extra-linguistic clues help paint a picture of an utterance's context. We can know who said what, where, when, to whom, about what subject matter, within what conversation, and more. And from all this surface-level, description-informing, statistically-analysable information emerges the more philosophical; words' *pragmatics*, not just their *semantics*, are discernible.

This is by no means novel; information of this kind can help to inform further inquiry into cases of semantic ambiguity, for example, to discern a word's intended meaning[7]. Corpus data is comprised by authentic language use, where words' meanings may often change from instance to instance, so an understanding of this seems necessary almost as a prerequisite.

Instead, philosophical profundity arises from the implications of this realisation. If we can work out from corpus data what the words were intended to mean (pragmatics) beyond just what they mean out-of-context (semantics), what's stopping us from inferring speakers' actual intentions behind different uses of the same linguistic expressions? There is some debate on whether inference to people's mental linguistic structure is permissible from corpus data, though this often concerns whether frequency statistics are reliable indicators of entrenchment (strength of a mental representation) in a mental grammar[8]. Discussion also exists on the utility of corpora for researching speech acts (which occur when we perform an action with an utterance; e.g., "I do" in a wedding ceremony)[9]. The entire field of discourse analysis rests on the premise that speakers' motivations can be reliably grafted from a text[10]. And yet, this deeper philosophical question remains. Can we come to an understanding of how intentions differ and change for identical expressions to better understand them and why they're used?

Little has been said on such a subject. Perhaps this is for good reason. Pragmatics is notoriously difficult to determine from written language alone. We need facial cues, prosody, and even gesture to fully understand what someone means by what they say (if someone says "yes" whilst shrugging, what meaning do they actually communicate?)[9]. This also may not seem pertinent for corpus linguists to investigate insofar as such proposed analyses seem ultimately to rely upon qualitative methods, if we take corpus data to best lend itself to quantitative measures (which is not definitely true). So, perhaps this is a normative issue; should we even try to come to this more holistic understanding of the intentions behind expressions' use if such attempts may not be reliable?

Well, why should we? If it's not futile, is it even useful? I think it is. Developing a sophisticated understanding of the intentions that commonly underpin expressions' usage has enormous practical value. Knowing why *person type x* often says *expression type y* in *context type z* could offer a wealth of opportunities for better understanding

how people communicate. There is as much potential for academia as there is for industry.

This, of course, raises a number of practical and ethical issues. How can a researcher (e.g., a corpus annotator) know the actual intentions that underpin someone's utterances? How will findings have to be handled across academia and business to ensure people's communicative tendencies aren't exploited? Would people start purposefully speaking differently to avoid having regularities picked up in another attempt to stay off the big-data, big-tech radar?

The first of these is most answerable: With some difficulty, I would imagine. With all the linguistic context and background information (e.g., who, what, where, when), a human could take a stab at working out the whys of people's discourses in many cases. If a machine could learn to do this, though, that would be helpful (and terrifying). An annotator would have to be prepared for a lot of uncertainty, it seems. Whether this is a set-back, though, is unclear: I can envision an annotation method where uncertainty over ambiguous (though, not vague) intentions could be represented as a set of probabilities for the annotator's level of confidence based on the helpfulness of the linguistic context with the number and quality of sources of background information computed. Quantifying all that, though, I'll leave to the computational linguists.

The second and third issues require more speculation. The exploitation of people's personal information has been a hot topic in current affairs for the last few years, and it doesn't look like a resolution is in sight. To some, the prospect of linguists working out *why* they say things could seem like too much of an invasion of privacy. I'm not convinced it is: the intentions behind what you say and do are being collated and evaluated all the time by everyone you speak to (if they're actually listening to you). People in the public eye, especially politicians, live and die by the fine-grained distinctions in interpretations of what they say. One has to be wary of the unintended consequence, to which we might have chalk up the third issue, but there are many possible benefits (assuming faultless function): Academically, better theories of linguistic interaction, a better understanding of behavioural psychology, and a step closer to solving the mystery of language; more widely, there are applications in the legal system to aid court cases and in the world of marketing to conduct better consumer research.

So, I think we could give it a shot. But who's to say? It seems to me that this is where we introduce philosophers and computer scientists alike into the discussion. They should be the ones to work out how the tools corpora offer can be best and most creatively leveraged to track the ever-changing nature of intention, each expression's usage at a time. That's important; the intentions underpinning expressions' pragmatics will not only vary synchronically but also diachronically.

Thus, I conclude with a plea: Arise, armchair linguists — swap the cushions for a keyboard, your mind's eye for a monitor, and plunge into the potential (intentional) profundity of corpora.

References:

[1] McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

[2] Davies, I. (2021). Building Your Own Twitter Corpus: A Step-By-Step Guide. *U-Lingua Magazine*, 1(4), 18-19.

[3] Culpeper, J., & Archer, D. (2020). Shakespeare's language: Styles and meanings via the computer. *Lang. Lit.*, 29(3), 191-202.

[4] Fillmore, C. J. (1992). „Corpus linguistics" or „Computer-aided armchair linguistics". Paper presented at the *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82*, Stockholm, 4-8 August 1991. Berlin, New York: Mouton De Gryer.

[5] Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *Int. J. Corpus Linguist.*, 22(3), 319-344.

[6] Hardie, A. (2012). CQPweb: combining power, flexibility and usability in a corpus analysis tool. *Int. J. Corpus Linguist.*, 17(3), 380-409.

[7] Williamson, T. R. (2021). The Graded Co-Salience Hypothesis for Polysemous Ambiguity. *JoULAB*, 1(1), 57-77.

[8] Arppe, A., Gilquin, G., Glynn, D., Hilpert, M., & Zeschel, A. (2010). Cognitive Corpus Linguistics: five points of debate on current theory and methodology. *Corpora*, 5(1), 1-27.

[9] Garcia McAllister, P. (2014). Corpora and speech acts. In C. Rühlemann & K. Aijmer (Eds.), *Corpus Pragmatics: A Handbook* (pp. 27-114). Cambridge: Cambridge University Press.

[10] Mullet, D. R. (2018). A General Critical Discourse Analysis Framework for Educational Research. *J. Adv. Acad.*, 29(2), 116-142.

ULAB Undergraduate Essay Competition 2021-2022

We are inviting undergraduate students to write us an essay in response to one of the three questions below. Essays should be 2500 words long (+/- 10%), and referenced using a referencing style of your choice. For each of the questions, we have provided a list of recommended reading to help you get started – if you are unable to find any of the sources online, please contact us (via social media or email) and we will help as best as we can. These references are, of course, just a starting point; you are welcome to bring in other sources!

Essays will be marked in accordance with criteria established internally to ULAB, but the best submissions will take a novel approach to a question, will think creatively to pose original challenges to existing literature, and, most importantly, will be exciting to read! We look forward to reading your entries – good luck!

Submission

Please send your essay in PDF format to ulablinguistics@gmail.com by 23:59 GMT on Monday 13th January 2022. Submitted files should be fully anonymised — please ensure that neither your name nor any other personal information is visible in the document.

Eligibility

All undergraduate students across all disciplines and countries are welcome to submit! This includes people who started their undergraduate degree in September, as well as those who have graduated from an undergraduate degree in 2021. If you are a member of a ULAB Subcommittee, please contact your Subcommittee Chair to check your eligibility.

Prize

The winner of this year's essay competition will receive £50 in prize money, a free ticket to ULAB's 2022 conference, and the opportunity to work with the *U-Lingua* team to perfect their article and publish it in a future issue. There will also be monetary prizes for runners up.

Questions

Question 1: Questioning Grammaticality

The concept of grammaticality is one that does not have a precise definition, and a sentence may be differentially grammatical across multiple dialects of the same language family. As new coinages come into the lexicon, more novel sentences get generated by speakers, and more languages come into contact, more data is available to be analysed for grammaticality.

Write an essay addressing any of the following issues:

- What is grammaticality and is it reducible to anything else?
- If grammaticality can be dialect-specific, why can it not be idiolect-specific, and does this make the concept itself redundant?
- Will increased language use and contact trend towards all sentences being grammatical?

Question 2: Linguistic Universals

Languages differ in many ways, including in their word order, sound inventories, lexicons and morphological processes. However, languages are similar in many ways too, and a number of linguistic universals exist. Linguistic universals can be absolute (all languages sharing a property or pattern of properties) or statistical (the vast majority of languages sharing a property or pattern of properties).

Making sure to address both absolute and statistical universals, write an essay exploring the possible reasons that linguistic universals arise.

Question 3: Critical Discourse Analysis in Society

Critical discourse analysis (CDA) is a multidisciplinary methodology that allows researchers to explore and uncover ideologies and examine how they affect language use. They do this by denaturalising ideologies that are often hidden and assumed to be "common sense". As Gomez-Jimenez (2018: 101) explains, "CDA addresses social problems and so it can clarify our understanding of forms of social inequality".

Discuss the value of critical discourse analysis in the context of one of the following topics (or an intersection of multiple):

- Environment (ecolinguistics)
- Gender & sexuality (feminist & queer linguistics)
- Disability
- Race, ethnicity & nationality
- Religion
- Social class

Recommended reading for each question available on our website: https://www.ulab.org.uk/essay-competition

Editorial Team
Editor-in-Chief Stephanie Jat
Section Editor for *Hot off the Press* Elif Yildiz
Section Editor for *Behind the Bookshelves* Alessandra Terranova
Section Editor for *Through the Ages* Kitty Liu
Columnist for *A Hands-On Approach* Cara McSherry
Columnist for *Words, Words, Words* Olivia Szczerbakiewicz
Columnist for *Curiouser and Curiouser* Caitlin Wilson
Columnist for *Anatomy of a Linguist* Tom Williamson
Editorial Designer Xinmei Sun

Authors
"Secret" and Constructed Languages: How We Change Our Language
Gabriel Bonté
Emergence of a Dialect: The Origin of African American English
Andrew Tobin
Toxic Data: Biases in Natural Language Processing
Nicholas Daines
***Wordslut*: A Feminist Reinvention of the English language**
Beatrice Livesey-Stephens
Inventions and Reinventions
edited by Kitty Liu

Photo Credits
Illustration of printers upklyak on Freepik
Section Cover *Hot off the Press* sigmund on Unsplash
Illustration of whispering stories on Freepik
Illustration of books Freepik on Freepik
Illustration of people in conversation pikisuperstar on Freepik
Section Cover *Behind the Bookshelves* visualsbybrad on Unsplash
Illustration of Natural Language Processing vectorjuice on Freepik
***Wordslut* cover** HarperCollins
Section Cover *Through the Ages* jcomp on Freepik

Illustration of text boxes Freepik on Freepik
Illustration of cell phone dancounsel on Freepik
Illustration of writing stories on Freepik
Illustration of music and singing pchvector on Freepik
Illustration of social media stories on Freepik
Illustration of typewriter ivector on Adobe Stock
Illustration of book Freepik on Freepik
Illustration of feather Freepik on Freepik
***Curiouser and Curiouser* column background** Freepik on Freepik
Sculpture of Shakespeare Birmingham Museums Trust on Unsplash
Screenshot of CQPweb Tom Williamson
Photo of man in amazement Usman Yousaf on Unsplash
Photo of laptop dancounsel on Unsplash

Facebook
Undergraduate Linguistics Association of Britain
Twitter
@ULAB_Ling, @ULAB_Magazine
Instagram
@ulab_ling @ulinguamagazine
LinkedIn
Undergraduate Linguistics Association of Britain
Discord
www.ulab.org.uk/discord
Strava
www.strava.com/clubs/ulab