UNIVERSITY of York

This is a repository copy of *How 'U' are 'U' words? Exploring variation in the usage and perception of class-based lexical shibboleths in British English*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/203582/

Version: Accepted Version

Article:
Sandow, Rhys, Bailey, George orcid.org/0000-0001-5137-8394, Braber, Natalie et al. (1 more author) (2024) How 'U' are 'U' words? Exploring variation in the usage and perception of class-based lexical shibboleths in British English. English World-Wide. A Journal of Varieties of English. pp. 94-124. ISSN 1569-9730

https://doi.org/10.1075/eww.00085.san

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

How 'U' are 'U' words?: Exploring variation in the usage and perception of class-based lexical shibboleths in British English[1]

Running title: How 'U' are 'U' words?

Authors and Affiliations:

Rhys J. Sandow, Queen Mary, University of London

George Bailey, University of York

Natalie Braber, Nottingham Trent University

Eddie O'Hara-Brown, University of Sussex

Abstract

That U (upper-class) and non-U (non-upper class) speakers are identifiable through their vocabulary is an axiom in England. These claims are repeated in books, in print media, on social media, and in conversations regarding social class. However, such claims are seldom investigated empirically. To redress this, we consider the production and perception of U and non-U lexis through two studies. In the first, we identify the sociolinguistic distribution of the usage of three variables which are purported to be indicators of socioeconomic status, namely, LOO, NAPKIN, and SOFA. The second study employs the matched-guise technique to investigate the perception of variants of these three variables. The production results reveal that all three variables exhibit change in apparent-time with limited evidence of class-based variation. In the perception study, we find no systematic class-based indexicalities across the variables. Ultimately, our findings challenge the belief that U words are shibboleths of upper-classness.

Key words: Social Class, U, Lexis, Production, Perception

[1] The authors would like to thank two anonymous reviewers and editors of the journal for their valuable and insightful comments on a previous version of this article. Any remaining shortcomings are the responsibilities of the named authors.

1. Introduction

Social class has always been something of a slippery concept; readily identifiable but, at the same time, somewhat esoteric. Alan S. C. Ross (1954) proposed a system of behaviours which distinguish the 'upper-class' from the 'middle' and 'lower' classes. These socioeconomic indices, which include not playing tennis in braces, an aversion to high tea, as well as use of particular features of language, were purported to differentiate the 'U', that is, the upper-class, from all other socioeconomic groups, the 'non-U' (see Ross, 1954). Ross (1954) identified features of pronunciation and grammar which he perceived to distinguish the upper-class from the rest of the English population. Overwhelmingly, Ross' (1954) article is best remembered for its identification of class-based vocabulary differences which Crystal (2019: 386) has described as generating the 'most famous debate on the English language and social class'.

The categories of U/non-U words have been revised by, among others, Buckle (1978), and Fox (2004), as some of the words identified by Ross (1954) became obsolete and other lexical items developed class-based associations. Often claiming legitimacy from the likes of Ross (1954) and Fox (2004), lists of U/non-U words which make claims such as "if you use these 27 words it proves you're definitely not posh" (Victor, 2017) are commonplace online, as are quizzes of word usage which claim to answer questions such as 'do you talk posh?' (Himefield 2018). This highlights the appetite among the English public to position themselves in a social hierarchy. However, it remains the case that little empirical research has been conducted which tests the veracity of such claims of class-based lexical usage patterns. We seek to redress this by conducting an analysis of the social variation of the usage of lexical items purported to be revealing of social class in Britain. Specifically, we consider the usage and perception of three variables, LOO, NAPKIN, and SOFA, which supposedly have U and non-U realisations. Ultimately, our findings challenge the belief that U words are shibboleths of upper-classness in Britain.

Before we outline our studies, it is important to address exactly what we mean by 'social class' and how this term has been used as the basis for the classification and categorization of speakers in variationist sociolinguistic studies, particularly within Britain.

2. Social class in British sociolinguistics

The British have been described as a notoriously class 'obsessed' people (Fletcher and Dashper 2013). According to Van Doesum et al. (2017:12), social class is a multifaceted construct, as, in addition to economic resources, "class also reflects disparities in cultural capital, status, and power. Ultimately, social class is best captured under the umbrella of societal rank". One's class status is commonly believed to be reflected in one's clothes, belongings, home, and language (Tyler 2008). Savage (2007) showed that concepts of class had changed between 1948 to 1990, being seen as a matter of agency, rather than simply being handed down from generation to generation. In later work, Savage et al. (2013) argued that "social class is a multi-dimensional construct [which] indicates that classes are not merely economic phenomena but are also profoundly concerned with forms of social reproduction and cultural distinction" (Savage et al. 2013:223; see also Bourdieu 1984).

Class-based linguistic variation has interested variationist sociolinguists since the mid-20th century (for review, see Ash 2002; Dodsworth 2009; Kerswill 2009; Rampton 2010 and Block 2015). Labov's (1966) seminal work in New York laid the groundwork for sociolinguistic research into class and its effects on linguistic variation. Since Trudgill's (1972, 1974) work in Norwich, British sociolinguistics has largely concentrated its research on class to the *middle*- and *working*-classes. Trudgill (1972:181) classified speakers' social class using "an index that was developed using income, education, dwelling type, location of dwelling, occupation, and occupation of father as parameters". This index, consisting of the aggregate scores of each of these six indicators of socioeconomic status, was used to classify individuals into five categories, namely, 'middle middle class', 'lower middle class', 'upper working class', 'middle working class', and 'lower working class' (cf. Savage et al. 2013: 230 who discuss seven classes in British society). Trudgill (1972) found sociolinguistic variation to correspond with his classification of social class using the index of six constituent factors. For example, speakers in higher socioeconomic groups, e.g., the middle middle class, were less likely to use sociolinguistic variants which carry negative social evaluation, e.g., the apical variant of (ING), than those in lower socioeconomic groups. This finding, whereby non-standard, or less overtly prestigious, forms being used most frequently by speakers who occupy the lower end of the socioeconomic hierarchy has since been widely attested, such as for h-dropping in Hull, Milton Keynes, and Reading (Williams & Kerswill

1999), the STRUT vowel in Manchester (Turton & Baranowski 2021), and rhoticity in Cornwall, Devon, and Dorset (Malarski 2021).

Despite classifying his participants into five socioeconomic classes, Trudgill's classification reinforces the binary (middle- and working-class) taxonomy of the British class system which can be further divided into more fine-grained classifications, with the upper-class largely disregarded as a small minority. This approach has continued to dominate social class categorizations in British sociolinguistics. For example, Baranowski (2017) uses similar categories to Trudgill (1974) in his study of GOOSE and GOAT vowels in Manchester, even explicitly highlighting the way in which these five classes are understood as subsets of the working and middle classes, as opposed to five whole and distinct classes: "The assignment of informants to either the working or the middle classes reflects the traditional division into blue-collar workers and white-collar labor" (Baranowski 2017:303). The focus on two social classes, the working- and the middle-class is reinforced by a number of studies which consider social class as a binary system e.g., Foulkes and Docherty (2000), Watt (2002), Stuart-Smith, Timmins and Tweedie (2013), Jansen (2017), Sandow (2022). Often, even when more fine-grained classification systems are used (e.g., Robinson 2010; Beal and Burbano-Elizondo 2012), most make use of the over-arching working vs. middle-class typology[2] with the 'upper-class' largely absent from these studies. This is not to say that there are no 'upper-class' individuals in these samples, just that they do not occupy a distinct category in the most common class-based schema used in sociolinguistic studies and they tend to be subsumed under the label of 'middle class'.

While research in sociolinguistics has a tendency to focus on working and middle class speech patterns (for discussion, see Britain 2017), this article contributes to the growing body of research on the usage of features associated with upper-class groups (e.g., Badia Barrera 2015; Alderton 2019; Halfacre and Khattab 2019, see also Fabricius 2002). Also noteworthy here is Shi and Lei's (2021) study which used the *Spoken British National Corpus 2014* to identify features distinctive of the highest socioeconomic groups, focussing on lexical richness, word length, and word class.

[2] Alternatively, some scholars have used numerical classification systems which avoid labelling socioeconomic groups (e.g., Macauley 1977; Dann 2019).

3. Applying sociolinguistic theory to U and non-U vocabulary

In his classification of U and non-U words, Ross (1954) did not use any empirical methodology (Honey 1985), instead relying on his perceptions. Although much has been written about U and non-U words, we know very little about their usage from an empirical, quantitative perspective. That is, it is unclear to what extent U/non-U words are reliable indices of socioeconomic class, particularly in the 21st century. However, by using sociolinguistic theory in the context of the received narrative of U and non-U words, we can make predictions about the social and stylistic variation of U and non-U words, as well as their perception. One important point to note is that at no point did Ross (1954) define U or upper-class. While he did highlight a tripartite system, he did not specify whether the group at the top of the socioeconomic hierarchy, the U, consisted of, for example, the top 33% of society or the nobility only. This vagueness does make any direct comparison of Ross' claims with empirical data challenging.

Due to the prestige claimed to be associated with the U forms, we can predict patterns of social variation based on previous findings in sociolinguistic studies. For example, sociolinguistic variants with overt prestige are typically found to be used the most by higher socioeconomic groups (e.g., Labov 1972; Trudgill 1974; Baranowski 2017). Also, women have typically been found to use higher rates of variants which carry overt prestige than men (Labov 2001). Thus, we can predict that U words that are used more by women and those in higher socioeconomic groups.

Sociolinguistic theory also makes predictions about the linguistic variation observed within individual speakers. Stylistic, or 'intra-speaker', variation emerges in different speech environments in response to contextual and social stimuli. One of the most studied stylistic factors is the degree of attention paid to speech (Labov 1972). Sociolinguistic theory predicts that variants associated with higher socioeconomic groups are used more frequently by all speakers as their attention-to-speech increases (Labov 1990; Snell 2018, although, see Sandow 2022). Thus, as U variants are anticipated to be used the most by speakers in higher socioeconomic groups and by women, we can also predict that these forms exhibit stylistic variation, with U forms occurring most frequently in careful speech styles in our usage-based study. In terms of the perception study, if the U/non-U distinction remains relevant, we would

expect to see U variants evaluated more highly for characteristics related to status, such as educatedness, poshness, and formalness.

4. Data collection and methods

We use two studies to further our understanding of U and non-U lexis. The first study serves to answer the research question 'what, if any, are the social correlates of the usage of 'U' and 'non-U' lexis'? The second study serves to answer the research question 'are there differences in the ways that U and non-U lexis is perceived'? Before outlining these studies more thoroughly, we first discuss the investigated three variables in greater detail.

4.1. The variables and variants

In this article, we focus on three lexical variables which have been identified as having both U and non-U realizations. Specifically, we investigate the usage of NAPKIN, LOO, and SOFA.[3]

Ross (1954) identifies *napkin* as U and *serviette* as non-U (see also Buckle 1978; Fox 2004). Indeed, Ross (1954:46) suggests that these words are "perhaps the best known of all the linguistic class-indicators of English". For the LOO variable, Fox (2004) asserts that *loo* and *lavatory* are U while *toilet* is non-U (see also Buckle 1978) and some U speakers use *bog* ironically (Fox 2004).[4] *Sofa* is considered to be U while the non-U alternatives are *settee* (Buckle 1974; Fox 2004) and *couch* (Fox 2004).

Little research has been conducted on the social variation of the investigated concepts, particularly NAPKIN and LOO. There has been limited research into the lexical variation in the realization of the concept SOFA. In Canada, Boberg (2004) found that the Canadian variant *chesterfield* is recessive, with *couch* being the most common replacement among younger speakers. However, Boberg (2004) did not consider the role of social class or gender in the trajectory of this change. Using the *British National Corpus*, Romaine (2008:103) investigated the socioeconomic variation in the use of *couch*, *sofa*, and *settee* as variants of SOFA in British English. *Settee* was the most common variant for all four socioeconomic groups and was particularly favoured by the lowest socioeconomic group (Romaine 2008:103).[5] *Couch* was seldom used, although it was used to the greatest extent by the

[3] We refer to each variable by its U realization.

[4] Ross (1954) does state that while *toilet paper* is non-U *lavatory paper* is U but does not mention the concept of a LOO specifically.

[5] Although, this is not true of the 2014 *Spoken British National Corpu*s, see Figure 6.

interior social groups, that is, neither the highest nor the lowest. In fact, the highest and lowest social groups did not use *couch* at all. The U variant *sofa* did not exhibit clear social stratification, with the highest and second lowest of four socioeconomic groups using highest frequency *sofa* in relation to the other attested variants. Romaine (2008) did not consider the use of the SOFA variable in relation to other social variables such as age or gender. There is also a regional dimension to this variation. Grieve et al. (2019) observed *sofa* to be most frequent in the South and South-West of England, where our production study was conducted, with *settee* being favoured in the Midlands and the North (see also *Our Dialects* 2022). In addition to the limited research on usage-based variation of these variables, we are not aware of any studies which have investigated the intra-speaker variation or perception of these variables.

4.2. Production study

The first study discussed here investigates lexical usage of the variables LOO, SOFA, and NAPKIN. The study (see Sandow 2021) was chiefly a study of Cornish identity and the Anglo-Cornish dialect (see Sandow 2022, 2023, in press a,b) but also investigated some supra-local features (e.g. Sandow et al. in press).[6] The study took place in the mid-west Cornish towns of Camborne and Redruth and surrounding villages between September 2017 and June 2018. While not rural, Camborne-Redruth is not a metropolitan area and is, thus, a rather different context to the cities in which variationist sociolinguistics studies have typically been conducted (see Britain 2012).

4.2.1. Methods

In order to investigate lexical usage patterns of social and stylistic variation, Sandow (2021) employed two elicitation procedures which vary in the level of attention-to-speech required. Spot-the-difference tasks elicit lexical usage from a relatively casual speech style while naming tasks elicit lexical usage from a relatively careful speech style. These tasks elicit

[6] Data were collected for the three variables investigated here serendipitously, as they were not target variables in the initial study, but distractor variables. This explains why a more comprehensive set of U/non-U words is not investigated.

usage data which enable the analysis of sociolinguistic variation of lexis in the social and stylistic dimensions.

Spot-the-difference tasks were used to elicit lexical usage while the speaker's attention is concentrated on task-completion, not language use. That is, the speaker is focused on spotting the differences, yet, in doing so, they provide lexical items to identify the difference, e.g., 'the sofa is a different colour in that picture' or 'the toilet is green in the left picture and white in the right one'. Thus, it can be said that in the spot-the-difference tasks, participants' attention-to-speech is relatively low. Participants completed five spot-the-difference tasks (an example of which is shown in Figure 1), each of which contained six differences. All of the investigated concepts, NAPKIN, SOFA, and LOO, appeared in two scenes, thus two tokens of each variable were produced per speaker in the relatively casual speech style.

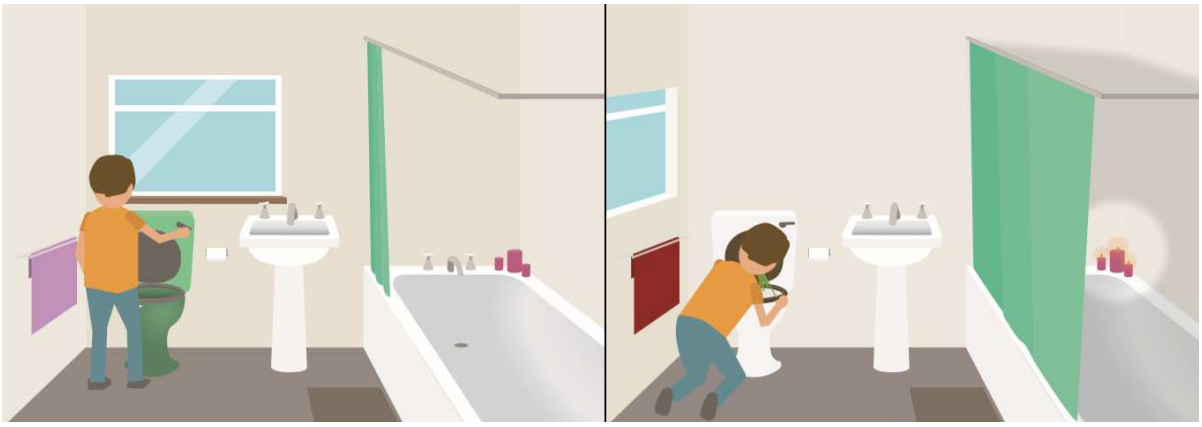


Figure 1: The 'bathroom' spot-the-difference scene

In contrast to the relatively casual style elicited for spot-the-difference tasks, a relatively careful speech style was elicited from naming tasks (see Figure 2). Participants were told that the purpose of the naming tasks was to investigate word usage. Thus, their primary cognitive load was focused on their word choice. As a result, their attention-to-speech is greater than in the task-oriented spot-the-difference games, in which their primary goal was task-completion. In the naming tasks, of which there were twelve, participants were presented with an image and an incomplete sentence and were asked to use the image to complete the sentence, e.g., 'this is a …'. Simple sentence stimuli were used because a more complex or engaging sentence could cause the participant to focus on the phrasing as opposed to on their word

usage. Each participant saw one naming task that related to each investigated variable, meaning that they produced a single token per variable in the relatively careful style.

Figure 2: The LOO naming-task

For each concept, the picture stimuli used in the spot-the-difference tasks and naming-tasks are identical, which minimizes variation in referential meaning between the responses to the different tasks. This enables us to be confident that any observed variation can be attributed to the level of attention paid to the task (e.g., see Sandow 2022, 2023).

4.2.2. The participants

Eighty participants were recruited through the friend-of-a-friend snowball method. The eighty-speaker sample is balanced for the social categories of age, gender, and socioeconomic class. Each of these categories are conceptualized as binary, that is, older and younger, male and female,[7] and working-class and middle-class. Socioeconomic class was calculated using a social-class matrix, consisting of three indicators, namely, education, occupation, and relative deprivation of the speaker's place of domicile (for more detail, see Sandow 2021). In terms of age, speakers were categorized into those older than 40 or younger than 30.[8] The sample of participants was made up of ten participants per cross-section of these social groups (see Table 1), such as ten younger female middle-class speakers.

[7] Speakers were given the option to identify with a non-binary gender, although none elected to do so.

[8] There were no participants in this study who were between the ages of 30 and 40.

Table 1: The socio-demographic profile of speakers used to balance the sample

	18–30 years of age	**40+ years of age**
Working class	Male: N=10 Female: N=10	Male: N=10 Female: N=10
Middle class	Male: N=10 Female: N=10	Male: N=10 Female: N-10

While Table 1 displays the socio-demographic profile of speakers used to balance the sample, we also include further social categories in our analysis which enable us to explore more fine-grained patterns of social variation (see Table 2). Specifically, we conceptualize socioeconomic class as a ternary category, consisting of working-class (n=25), upper-working-class (n=34), and middle-class (n=21) groups. We also considered age as a ternary category, with 40 speakers aged 40 or below, 22 speakers aged 40-55, and 18 aged 56 or older. The indicators of socioeconomic class were also inputted into our statistical models to determine which had the most explanatory power for the different lexical variables under study. In terms of their highest level of education, 30 participants had degrees, 30 had A-levels or equivalents, and 20 received formal education up to 16 or younger. For occupation, 40 speakers were in high-prestige jobs, 22 in mid-prestige roles, and 18 in low-prestige occupations. For place of domicile (see Sandow 2021), 20 speakers experienced low levels of deprivation, 42 mid-levels of deprivation, and 18 experienced high levels of deprivation. More recently, Savage et al. (2013) propose that the six percent of the UK population at the top of the socioeconomic hierarchy, characterized by high levels of economic, social, and cultural capital, are the 'elite' who are over-represented in the alumni of many prestigious universities and in professions such as medicine, law, and at director level in business (see also Cunningham 2019; Hecht et al. 2020).[9] While we could not isolate six percent of participants in this study, we considered those who achieved the lowest possible score on the social class matrix (the lower the score, the higher the social class) to be 'elite'. These individuals typically had high status occupations, such as headteacher, CEO, or author, and

[9] It is difficult to compare directly Ross' (1954) and Savage et al.'s (2013) categories as Ross did not define 'upper-class'.

attended fee-paying schools and elite universities such as Oxford, Cambridge, Imperial College London, and the London School of Economics. Twelve participants (15% of the overall sample) scored the lowest possible score (see Table 2).[10] Our statistical models for our production data considered whether or not these twelve speakers had distinct patterns of lexical usage from the other 68 participants.

Table 2: The non-balanced social categories in the production study

Age_3	>30: N=40	40-55: N=22	>55: N=18
SEC_3	Working Class: N=25	Upper Working Class N=34	Middle Class: N=21
Place of Domicile	High Deprivation: N=18	Mid Deprivation: N=42	Low Deprivation: N=20
Education	Up to 16: N=20	Up to 18: N=30	Degree or higher: N=30
Occupation	Low Prestige: N=18	Mid Prestige: N=22	High Prestige: N=40
Elite	Elite: N=12	Non-elite: N=68	

Thus, we can investigate social class effects using a range of different conceptualisations of social class, including binary (WC and MC), ternary (WC, UWC, MC), and elite vs. non elite, as well as the constituent indices of occupation, education, and place of domicile. While the sample was balanced according to the binary WC and MC distinction, additional and alternative ways of splicing social class enable us to investigate social class effects that are opaque to the initial sampling criteria.

4.3. Perception study

The second study, conducted in the summer of 2022, reported on here involves an adaptation of the matched-guise technique (see Lambert et al. 1960; Campbell-Kibler 2007; Beltrama and Casasantro 2017). Rather than the traditional spoken guises, we use written guises. This

[10] Of these twelve participants, four are 'older' and eight are 'younger', while six are men and six are women.

is so that we can ensure any variation between guises is purely lexical and that phonetic or phonological variations are not confounding variables.

One hundred participants from England were asked to evaluate a number of texts, framed as social media posts, between 1-7, according to the following characteristics; 'educated', 'cool', 'posh', 'young', 'formal', 'friendly', 'honest', and 'attractive.' In addition to analysing the perception of these characteristics independently, we also consider clusters of characteristics which relate to status (educated, posh, formal, and honest) and solidarity (cool, friendly, and attractive) dimensions (see also Brown 1965; Luhman 1990; Dailey-O'Cain 2000; Ng and Diskin-Holdaway 2023).

The study was a between-speakers design with half of participants reading the U variant and half reading the non-U variant within an otherwise identical carrier phrase.[11] These stimuli were adapted from genuine posts on Twitter. An example of the task, administered through *Prolific*, is presented in Figure 3.

[11] For the perception study, we investigated two variants per variable, one U (*loo*, *sofa*, and *napkin*) and one non-U (*toilet*, *settee*, and *serviette*).

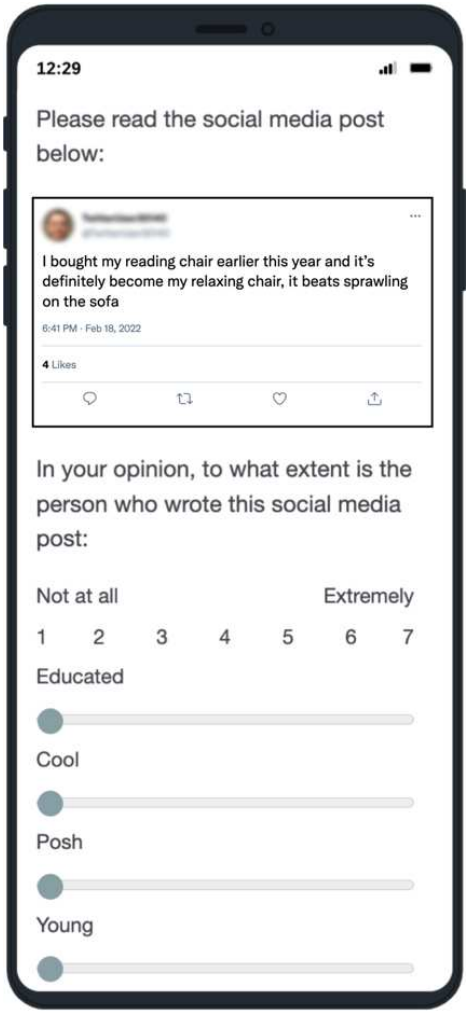


Figure 3: An example page from the online matched-guise experiment, showing the *sofa* stimulus.

Figure 3 displays the stimulus for the SOFA variable and the *sofa* variant, while half of participants saw the same carrier phrase with the word *sofa* replaced with *settee*. For the NAPKIN variable, the carrier phrase was 'My flatmate went to a wedding and I brought takeaway, was almost done eating before I saw something that looks like a fried egg, put it in my mouth and it was a [napkin/serviette]. God why me!?'. 'I woke up and I'd fallen asleep on the [toilet/loo] at work, in that moment I knew I was pregnant. It wasn't my first pregnancy so I knew of the early ridiculous tiredness' was the carrier phrase used for the LOO variable. There were also a number of other stimuli seen by participants which we do not discuss here (e.g., see Sandow et al. in press), including a control guise seen by all.

Participants in this study were asked to self-identify their gender, age, occupation and the region of England they grew up in (see Table 3). For occupation, participants were asked to

state if their job role was best categorized as (1) 'higher managerial and professional' (e.g. lawyer, accountant, business CEO), (2) 'intermediate' (e.g. teacher, manager), (3) ''white-collar' and lower managerial or clerical' (e.g. office worker, secretary, administrative assistant), (4) ''blue-collar' lower supervisory and technical' (e.g. electrician, plumber), or (5) 'semi-routine or routine' (e.g. retail worker, cleaner).

Table 3: The socio-demographic composition of the one hundred participant sample, by gender, age, occupation and region. For region, 'North' refers to the North West, North East and Yorkshire and the Humber, 'Midlands' refers to the East Midlands and West Midlands, and 'South' refers to the South West, South East, London, and the East of England.

Gender	Male: N=30	Female: N=70			
Age	18-29: N=33	30-49: N=52	>49: N=15		
Occupation	1: N=7	2: N=43	3: N=32	4: N=6	5: N=12
Region	North: N=24	Midlands: N=27	South: N=49		

In our analysis, we consider occupation as a binary category, with groups 1 and 2 in Table 3 being categorized as 'higher status' (N=50) and groups 3, 4, and 5 being categorized as 'lower status' (N=50).

5. Results

In this section, for each variable we first present the results from the production study and then the perception study for ease of comparison and to better understand any potential parallels between production and perception. The methods of data analysis naturally differ between the two studies given the different types of measurements involved. At the start of each section, cross-tabulations are provided to offer a general overview of how the U variants are distributed across age, gender and class. The main production results are based on logistic regression models fit to all observations coded in a binary fashion, i.e., tokens are coded as 1 if the U variant was used for a particular variable, and 0 if one of the non-U variants was used instead. In all cases, automatic step-wise regression was used to determine the optimal model structure, starting off from a full model containing all possible independent variables before dropping one at a time in order to reach the best-fitting model as determined by the Akaike Information Criterion (AIC); this is a measure of relative model *quality* that balances a model's explanatory power with the preference for parsimony in order to avoid overly-

complex models. Below, we provide in-text summaries of the main results from each logistic regression model, with the full coefficient tables provided in the Appendix.

We model the perception results using cumulative link mixed models (CLMM), which are more suitable than other techniques such as linear regression when it comes to modelling the type of ordinal data we are analysing in this study. In order to identify statistically significant differences in how the U and non-U variants are evaluated along the different descriptive scales, we conduct pairwise comparisons on the estimated marginal means extracted from the statistical model using the emmeans package in R (Lenth 2021). This technique allows us to determine if a particular U variant received significantly higher or lower ratings on a specific descriptive scale relative to its non-U counterpart, or whether two social groups differ significantly in the ratings given to a particular lexical item.

5.1. LOO

We first address the results of the LOO variable. In terms of the overall patterns in production, the non-U variant *toilet* was used most frequently, by 74 speakers, while *bog* was used by four speakers and the U variants *loo* and *lavatory* were used by ten and one speakers, respectively.[12] The cross-tabulations in Table 4 reveal that these rare uses of *loo/lavatory* are almost entirely restricted to older speakers, particularly those of middle class status, and that women slightly favour the use of these U variants.

Table 4: The use of *loo/lavatory* by age, gender and social class (N=30 in each cell).

	Working Class		**Middle Class**	
	Older	**Younger**	**Older**	**Younger**
Male	3%	0%	23%	0%
Female	20%	0%	30%	10%

[12] It is important to note here that a number of speakers used multiple variants, hence the totals here do not equal the number of speakers, 80.

Modelling the effects of social and stylistic independent variables on the usage of the two U variants *loo* and *lavatory* (vs all non-U variants) confirms that a range of factors influence this variable: the optimal model includes significant effects of age, gender, occupation and domicile deprivation. There is a change in apparent time away from the U variants and towards the non-U *toilet* ($\beta = -3.53$, $p < 0.001$), as well as a significant effect of gender with males disfavouring *loo/lavatory* more so than women ($\beta = -2.01$, $p = 0.003$). In terms of the various indices of socioeconomic status, we find a monotonic function of occupation with the lowest status group most strongly disfavouring *loo/lavatory* ($\beta = -3.01$, $p = 0.02$), and a smaller but non-significant trend of the middle status group also disfavouring it ($\beta = -0.62$, $p = 0.51$) relative to the highest status group. A curvilinear effect of domicile deprivation emerges such that the least *and* most deprived areas favour *loo/lavatory* relative to the middle group ($\beta = 3.49$, $p < 0.001$; $\beta = 4.66$, $p = 0.001$ respectively).

Turning now to the results of the perception study, the overall results aggregated over all respondents actually reveal no significant differences in how *loo* and *toilet* are evaluated for any of the descriptive scales. However, when we fit the same CLMM with the addition of an interaction between variant, scale and social class, a statistically-significant difference does emerge for the 'young' scale: the higher social classes see *loo* as significantly less young than *toilet* ($\beta = 1.52$, $p = 0.03$), which matches the apparent time change found in the usage data we collected independently in the production study. One other significant difference emerges if we flip the direction of analysis and instead look for differences in ratings between social groups for the same variant (rather than differences in ratings within social groups and between the variants). While there are no differences in how the lower and higher status social groups evaluate *toilet* on the 'posh' scale ($\beta = 0.29$, $p = 0.67$), we *do* find that the lower social classes rate *loo* as significantly less posh relative to the higher social classes ($\beta = -2.04$, $p = 0.004$). This of course runs counter to the historical status of the *loo* variant being the U lexical item in this pair.

5.2. NAPKIN

Napkin, *serviette*, and *doily* were used as variants of NAPKIN by participants in this study. The U variant *napkin* was the most frequently used variant (n=72) while the non-U variant *serviette* was used less often (n=18) and the archaic *doily* was used just once by a single participant. Table 5 presents the distribution of *napkin* by binary divisions of age, gender and

social class; the results indicate that young people use *napkin* almost categorically, with most social groups at 100%, compared to the variability exhibited by older speakers.

Table 5: The use of *napkin* by age, gender and social class (N=30 in each cell).

	Working Class		**Middle Class**	
	Older	**Younger**	**Older**	**Younger**
Male	63.3%	90%	56.7%	100%
Female	60%	100%	86.7%	100%

The results of the optimal logistic regression model reveal significant effects of age, gender, and elite status. Younger speakers use *napkin* more ($\beta = 2.91$, $p < 0.001$), suggesting an apparent-time change towards increasing use of this U variant and confirming the trend observed in the cross-tabulation above. Male speakers, however, are less advanced in their use of *napkin* relative to female speakers ($\beta = -0.88$, $p = 0.03$), and the lack of significant interaction with age indicates that this gender difference is consistent across both age groups and therefore not evidence of a female-led change.[13] Finally, we find that speakers categorized as 'elite'—recall that this refers to the group of 12 speakers in this study who scored highly on all three individual measures of social class—are significantly more likely to use the U variant *napkin* relative to the rest of the population sample ($\beta = 3.12$, $p = 0.007$).

We turn our attention now to the perception results. Unlike the LOO variable, the NAPKIN variable seems to elicit significant indexical responses even when aggregating over the whole sample of participants: *serviette* is perceived as significantly more attractive ($\beta = 1.13$, $p = 0.014$) and more posh ($\beta = 0.94$, $p = 0.045$) than *napkin.* There is also a trend towards it being evaluated as more formal than *napkin,* but this does not quite reach the traditional threshold for statistical significance ($\beta = 0.85$, $p = 0.072$). However, when a more complex model is fitted to the data that also includes an interaction with respondent age, we find that these indexicalities are only present for the middle age group of respondents aged 30–49. For this group, *serviette* does significantly index attractiveness ($\beta = 1.38$, $p = 0.037$), poshness ($\beta =$

[13] Table 5 suggests that this effect of gender is largely restricted to older MC speakers. This did not appear in the model as it would have required a three way interaction between age*class*gender, for which we do not have the statistical power to identify.

2.15, $p = 0.003$) and formality ($\beta = 1.77$, $p = 0.013$), but for the other age groups nothing significant emerges for any of the scales.

Grouping participants by class rather than age reveals another interesting finding: the biggest evaluative difference between *napkin* and *serviette* is found on the 'posh' scale and for the higher class participants, who actually rate *serviette* as significantly more posh than *napkin* ($\beta = 1.45$, $p = 0.027$). Just like with the LOO variable, this is unexpected given the historical status of *napkin,* not *serviette,* being the U variant.

5.3. SOFA

For the SOFA variable, the overall patterns of variant usage indicate that the U variant *sofa* is by far the most common variant (n=69), while the non-U variants *settee* (n=19) and *couch* (n=12) were much rarer. In terms of how this patterns by the various independent variables considered in this study, Table 6 presents a cross-tabulation of *sofa* usage by age, gender and social class. The results reveal a general change towards *sofa* in apparent time, a pattern which is particularly strong for working class speakers who move from roughly 47% usage to 83% (among men) and 97% (women).

Table 6: The use of *sofa* by age, gender and social class (N=30 in each cell).

	Working Class		**Middle Class**	
	Older	**Younger**	**Older**	**Younger**
Male	46.7%	83.3%	86.7%	76.7%
Female	46.7%	96.7%	73.3%	90%

We find that the optimal regression model includes an age×gender interaction, as well as the ternary social class predictor and the measure of domicile deprivation. The results confirm that *sofa* is favoured more by younger speakers ($\beta = 2.40$, $p < 0.001$), although a significant age×gender interaction reveals that young men are lagging behind in this change ($\beta = -1.53$, $p = 0.04$) and that this is therefore a female-led change in apparent time (cf. NAPKIN (section 5.3) where the gender effect was independent of age). The effect of deprivation is such that the speakers from the most deprived areas are least likely to use the U variant *sofa* ($\beta = -1.38$, $p = 0.001$). The effect of social class is not statistically significant, but the trend is for the two

working-class groups to use *sofa* less frequently than the middle-class group (β = -0.56, p = 0.33; β = -0.86, p = 0.07).

Similar to the other two lexical variables investigated here, the indexical profile of the SOFA variable is somewhat mixed. When aggregating over the whole population sample, significant differences emerge on the attractive and cool scales, with *sofa* scoring more positively on both dimensions (β = 0.95, p = 0.026; β = 1.03, p = 0.016 respectively). When adding an interaction with social class (based on occupational category) into the statistical model, we see that this effect is driven completely by the higher status participants of the study. As Figure 4 illustrates, the use of *sofa* scores significantly higher on the attractive (β = 1.87, p = 0.002) and cool (β = 1.80, p = 0.003) scales but only among the higher class respondents, with no evaluative differences present among the lower status group. Another noteworthy result evident in Figure 4 is how the two social groups differ in their perception of *sofa* and *settee* in terms of 'poshness': the higher socioeconomic group evaluate *sofa* as more posh than *settee,* which aligns with the respective U and non-U statuses of these variants, but this is flipped for the lower socioeconomic group who actually see *settee* as more posh. This difference in how the two groups evaluate *sofa* on the posh scale is statistically significant (β = 1.30, p = 0.041).

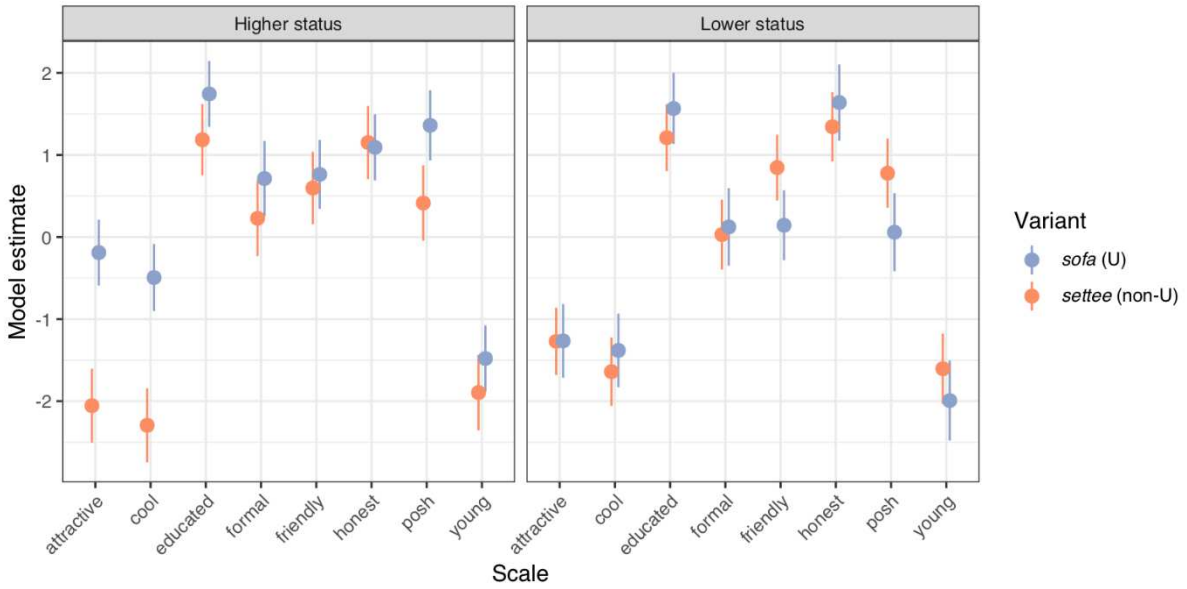


Figure 4: CLMM estimates modelling the interaction between variant, scale and class for the *sofa–settee* variable. Points reflect model estimates, ranges reflect ±1 SE.

Given that the sofa variable shows a strong regional profile (see the contemporary regional distributional here: https://www.ourdialects.uk/maps/furniture/), it is important to consider how these results interact with the regional background of our respondents. We ran an additional model, this time with region as an independent variable using the tripartite split between North/Midlands/South as presented in Table 3.

Significant indexicalities emerge between the two variants for southern listeners who evaluate *sofa* as more attractive (β = 1.54, p = 0.011) and cool (β = 1.44, p = 0.019) than *settee*. This is not surprising, given that *sofa* is not only the U variant but also the most frequently used variant in these southern dialect regions. Descriptively, this effect is weaker in the Midlands (β = 1.05, p = 0.216 for 'attractive'; β = 0.99, p = 0.231 for 'cool') and completely absent in the North (β = 0.23, p = 0.814 for 'attractive'; β = 0.43, p = 0.663 for 'cool'). This monotonic relationship might represent the intermediate status of the Midlands with respect to conflicting norms to both northern and southern dialectal variants (see Figure 5).

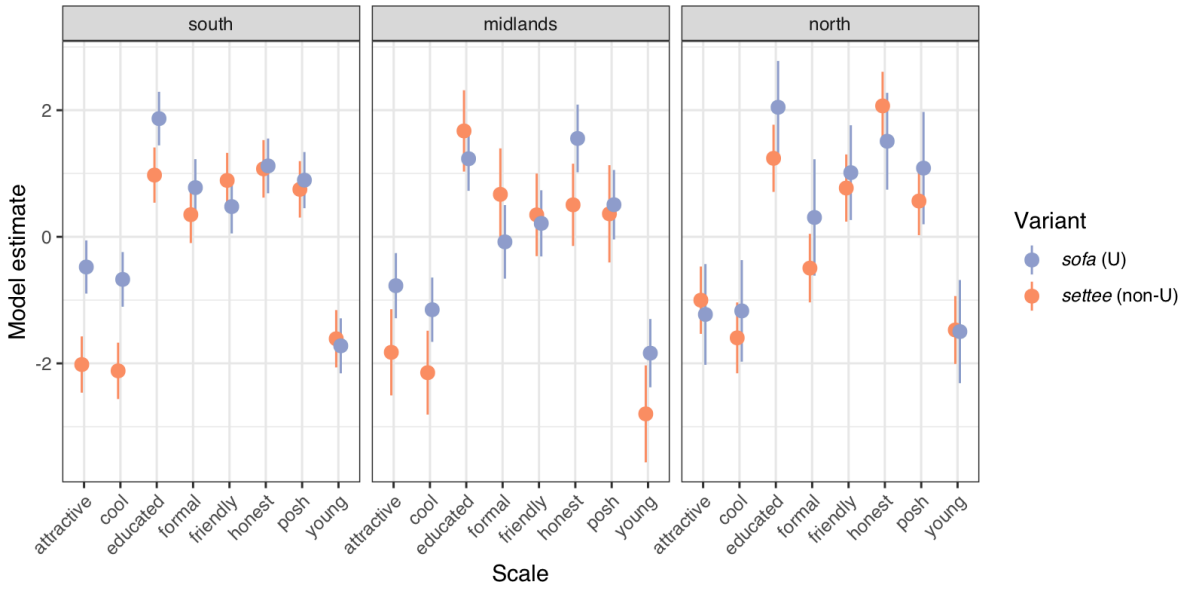


Figure 5: CLMM estimates modelling the interaction between variant, scale and region for the *sofa–settee* variable. Points reflect model estimates, ranges reflect ±1 SE.

6. Discussion

The two studies discussed in this article have served to provide an empirical perspective of the usage and perception of U and non-U words in England. The findings of study 1

challenge the assumption that the usage of U words are shibboleths of upper-classness. The only social variable which exhibits clear and consistent variation in relation to lexical usage is age, the results of which are summarized for all three variables in Figure 6. However, the direction of these apparent-time changes, either towards or away from U forms, is not uniform. While, on the one hand, the U variant *loo* is used more by older speakers, the non-U variants *serviette* and *settee* are also more commonly used by older speakers. This lack of regularity is consistent with other changes in U and non-U words. For example, the U *looking-glass* 'mirror' and *wireless* 'radio' are now largely obsolete while the non-U *mirror* and *radio* are used almost ubiquitously.[14] Alternatively, the non-U *preserve* 'jam' is now seldom used while the U *jam* is commonplace. This suggests that while U lexis is undergoing change, there is not a clear direction of travel.

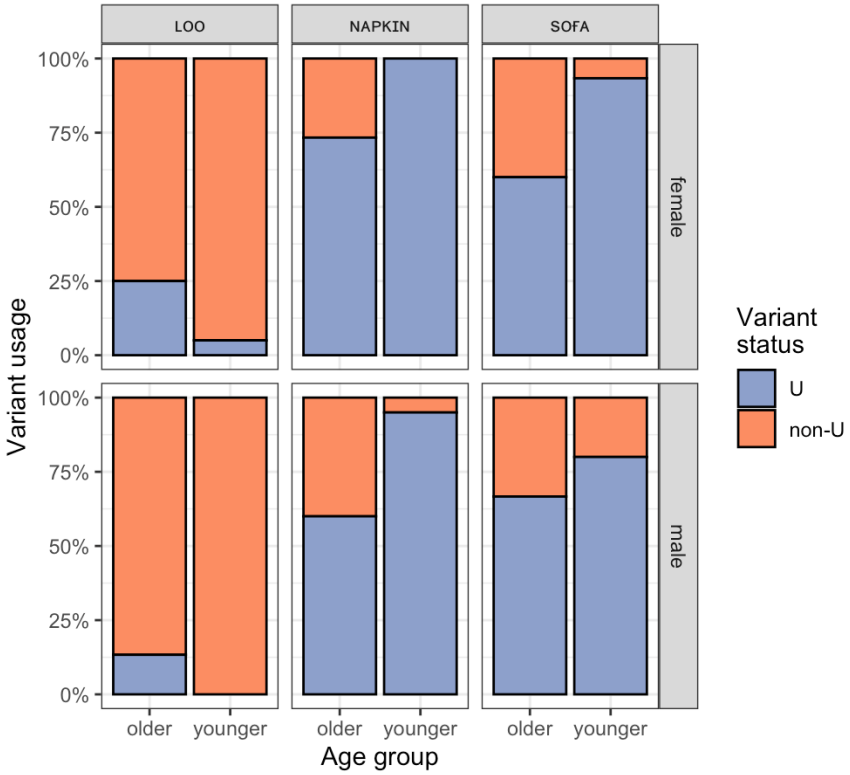


Figure 6: The apparent-time changes in production, split by gender, for all three lexical variables.

The most commonly used variant of each variable in study 1 is not consistently U or non-U. While the U variants *napkin* and *sofa* are the most common variants of the NAPKIN and SOFA variables, the non-U *toilet* is the most common variant of the LOO variable. Thus,

[14] The OED labels *looking-glass* 'mirror' as 'chiefly *archaic* and *historica*l' and states that 'for many users [*wireless* 'radio'] is either *historica*l or somewhat *archaic*'.

for two of the three investigated variables, it is the U variant that is the most common. This challenges Ross' (1954) claims that U lexis distinguishes the 'upper-class' from the rest of society. If most speakers are using these forms, they are not a reliable indicator of upper-classness. That is, they cannot be considered to be distinctive or characteristic of a subset of speakers, if they are used by the majority. In particular, we have demonstrated claims such as if a person uses *sofa* 'they are upper-middle [class] or above' (Fox 2004:25) to be demonstrably false.

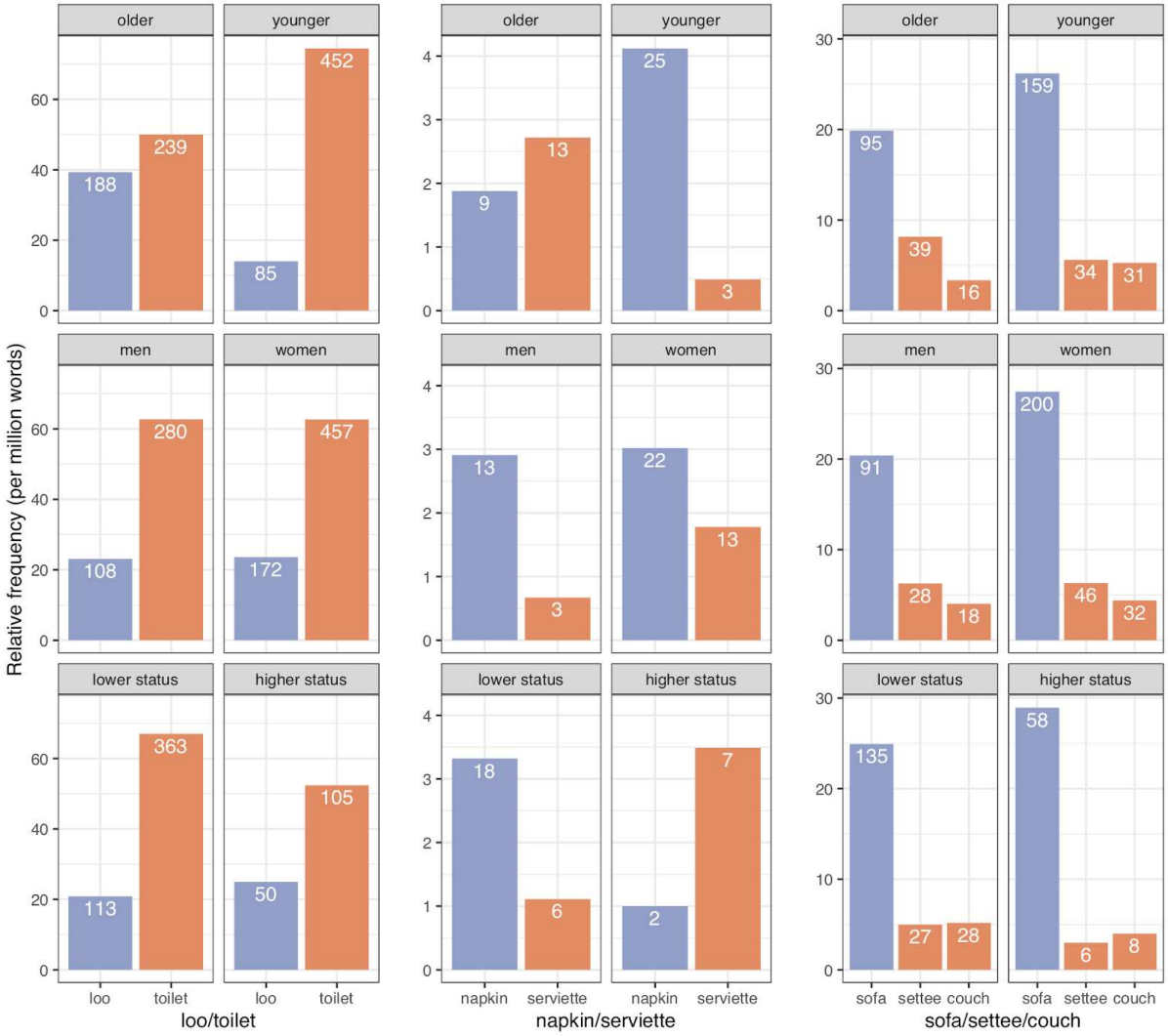
This is not to say that the investigated variables did not exhibit any class-effects in the production study. *Loo* was found to be disfavoured by speakers in the lowest occupational group. The 'elite' group use *napkin* more than 'non-elite' speakers. *Sofa* is used the least by those who experience the highest amounts of deprivation. Thus, while there are no comprehensive and consistent predictors of the usage of the U variant of each of the three variables, they do each show effects of social class in the expected directions.

Sociolinguistic theory predicts that linguistic change is led by women (Labov's Principles 3 and 4 (2001:188, 274)). However, it is not clear how this applies to the level of lexis. While some studies have shown women to lead lexical change (e.g., Johnson 1993; Ito and Tagliamonte 2003; Macauley 2005; Robinson 2010; Sandow et al. in press). Others have found a very limited or complete lack of a gender effect in lexical change (Beeching 2011; Tagliamonte and Brooke 2014; Jankowski and Tagliamonte 2019; Tagliamonte and Pabst 2020; Sandow in press b). Such findings led Tagliamonte and Pabst (2020:13) to question whether or not lexis, particularly lexical change, is subject to the same patterns of social conditioning as other levels of the grammar. While we found all three variables to be undergoing change in apparent-time, only one, the change towards *sofa*, to be female-led. The other two variants increasing in apparent-time, *napkin* and *toilet*, did not exhibit any interaction between age and gender.[15]

One possible interpretation of the lack of class-effects in our production study is that such a finding is specific to Cornwall or of the specific elicitation tasks used, and thus, not representative of broader usage. In order to explore this interpretation, we explore the usage of the three investigated variables in the spoken British National Corpus (BNC 2014). We find that the social distribution of the investigated variables in the spoken BNC data (see

[15] The role of indexicality in the incrementation of lexical change is considered in greater detail in Sandow et al. (in press) and Sandow (in press a). Such a discussion is beyond the scope of the present article.

Figure 7) are remarkably consistent with the social variation of the investigated lexical items in our production study. For example, there are strong effects of age[16] in relation to lexical usage, with older speakers using *loo* and *serviette* more and *sofa* less, in comparison with their younger counterparts. Analogous to our findings, there are limited effects of gender. However, there is a slight tendency in the BNC spoken corpus for women to use *sofa* more than men, which is consistent with our finding of a female-led change towards this variant. There are limited effects of socioeconomic class in the corpus data. An exception is the NAPKIN variable, which shows that the higher status group are less likely to use the U variant *napkin*, contrary to our initial predictions. However, this is consistent with our perception data, which found *napkin* to be perceived as less 'posh' than *serviette*.



[16] In order to make the corpus data more comparable to our production study data, we do not consider speakers aged 30-39.

Figure 7: The social variation of *loo-toilet*, *napkin-serviette*, and *sofa-settee-couch* from the spoken BNC 2014. Relative frequency is presented on the Y axis, raw frequency is presented in the bar labels..[17]

In our production study, none of the variables exhibited any statistically significant style effect. This is consistent with the moderate perceptual differences between U and non-U words in our perception study. While some of the perceptual effects are in the expected direction, many are not. For example, *serviette* is perceived as more posh than the supposedly U *napkin*, which is the inverse of our predictions. However, this is consistent with Fox (2004: 25) who comments that *serviette* is a 'genteelism', perceived particularly by the lower middle-class (many of whom may make up our higher occupational status category) as more refined than napkin, while the upper-class favour the U, *napkin*.

We also identified that the perception of the investigated variants are not uniform across demographic categories. Perceptual differences between U and non-U alternates are greatest among those with higher status. Additionally, it was not always the case that for the higher occupational status groups, the social indices of U/non-U variants are merely stronger. For the SOFA variable, we found that the higher socioeconomic status group perceived *sofa* to be more posh, while the lower socioeconomic group perceived *settee* as more posh. This highlights the social heterogeneity in the perception of U and non-U lexis. These findings are largely consistent with the way that Ross (1954) discussed U and non-U alternates, framing them as markers of in-group membership while the rest of the population is less, or differently, aware of these indexical meanings.

It is important to reflect on a number of limitations of the research reported on in this article. One key limitation is that we have studied three variables, rather than a more exhaustive set of allegedly U words which could alter our conclusions. Additionally, our production study relied on data collected from a study which was framed as a study of regional dialect usage and identity, rather than social class, which could have elicited different results (although these results were consistent with the BNC which is made up of non-elicited data). Another limitation is that our sample of participants, in both studies, incorporates a wide range of the socioeconomic spectrum. It is possible that the very highest socioeconomic groups may use

[17] It is important to note that this corpus analysis is not consistent with Labov's (1972: 72) *principle of accountability*. Nevertheless, we believe that this analysis is both useful and meaningful for our purposes. That is, it is being used to corroborate the findings of a method which does conform to the principle of accountability.

and perceive the investigated lexical items in a distinctive manner which our sampling methods have been unable to attest. While our perception study investigated a range of perceptual categories, these are not comprehensive. Further insights into the indexical fields of the investigated variables could have been uncovered by including other perceptual categories, such as masculinity-femininity. Additionally, the perception study relies on one minimally contextualized example per investigated word. It is possible that different socio-indexical meanings could be conferred by the investigated terms when they are used in other conceptual domains. While the contexts used for the presentation of these terms might themselves bias the evaluative responses (e.g. the mention of a 'reading chair' in the SOFA example might bias participants to give higher ratings of poshness), it is reassuring that the lack of clear class-based indexicalities are consistent across all three variables in our study. It is also important to note that our analysis rests on relative differences within pairs of guises, so any lack of neutrality for the wider context within which the terms are presented does not present a confound.

Despite the limitations outlined above, and the fact that the two studies reported here rely on different samples of participants, the two studies reported in this article are complementary in that they both speak to the finding that "U" words are not shibboleths of upper-classness in 21st century England. The results reported in this article should serve to redirect narratives of class-based social practices in England away from these "U" words. Whether or not such narratives are, in themselves, problematic is a discussion point beyond the scope of the current article.

7. Conclusions

It has been widely posited that use of ostensibly U/non-U words is a good diagnostic of socioeconomic class, specifically U-ness or non-U-ness. However, our data challenge this assumption as we observe limited social class effects. This suggests that U forms are not the shibboleths of socioeconomic status that they have been purported to be. The data from the BNC 2014 (spoken) demonstrates that this finding is not unique to Cornwall but is largely characteristic of lexical usage in Britain. In addition to the lack of class-based variation, the lack of stylistic variation in our Cornish data and the lack of consistent indexicalities, particularly relating to status, in the perception study suggest that U/non-U words may not confer the prestige that they are widely reported to. It is important to acknowledge that our results do not speak to the social distribution of these forms in the past. Especially given the

shifting nature of the socioeconomic landscape in Britain (Savage 2007), it is very possible that while they do not reflect a sociolinguistic reality today, Ross' (1954) claims were accurate at the time they were made. Indeed, Ross (1954: 54) acknowledged the 'ephemeral nature of our present system of linguistic class-indicators' by contrasting it with earlier accounts such as Walker (1791, cited in Ross 1954).

Both production and perception studies show that there is little consistency in the way that each of the investigated variables are used and perceived. This raises questions as to the validity of treating 'U' lexis as a distinct set with a degree of homogeneity. While we argue that claims about U and non-U words, which remain commonplace, are not reflected in the sociolinguistic reality of England in the 21st century, this is not to say that there are no class-based indices in contemporary British English. We suggest that identifying lexical alternates which exhibit strong class-effects in the 21st century would be an interesting topic of future research.

Author contact information:

Rhys Sandow:

r.sandow@qmul.ac.uk

Arts One,

Mile End Road

Bethnal Green

London

England

E1 4PA

George Bailey

george.bailey@york.ac.uk

Vanbrugh College

University of York

York

England

YO10 5DD

Natalie Braber

natalie.braber@ntu.ac.uk

Nottingham Trent University

50 Shakespeare Street

Nottingham

England

NG1 4FQ

Eddie O'Hara-Brown

Meredith.Brown@sussex.ac.uk

Sussex House

Falmer

Brighton

England

BN1 9RH

References

Alderton, Roy. 2022. "T-tapping in Standard Southern British English: An 'Elite' Sociolinguistic Variant?" *Journal of Sociolinguistics* 26: 287-298.

Ash, Sharon. 2002. "Social Class". In Jack K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, eds. *The Handbook of Language Variation and Change*. (pp. 402-422). Oxford: Blackwell.

Badia Barrera, Berta. 2015. *A Sociolinguistic Study of T-glottaling in Young RP: Accent, Class, and Education*. Unpublished thesis. Essex: University of Essex.

Baranowski, Maciej. 2017. "Class Matters: The Sociolinguistics of GOOSE and GOAT in Manchester English". *Language Variation and Change* 29: 301-339.

Beal, Joan., and Lourdes Burbano-Elizondo. 2012. "'All the Lads and Lasses': Lexical Variation in Tyne and Wear: A Discussion of how the Traditional Dialect Terms *Lad* and

Lass are Still Used in the Modern Urban Dialects of Newcastle upon Tyne and Sunderland". *English Today* 28: 10-22.

Beeching, Kate. 2011. "Sociolinguistic Aspects of Lexical Variation in French". In Tim Pooley and Dominique Largorgette, eds. *On Linguistic Change in French: Sociohistorical Approaches*. Chambéry: University of Savoy Press, 37-54.

Beltrama, Andrea., and Laura Staum Casasanto. 2017. "Totally Tall Sounds Totally Younger: Intensification at the Socio-Semantics Interface". *Journal of Sociolinguistics* 21: 154-182.

Block, David. 2015. "Social class in Applied Linguistics". *Annual Review of Linguistics* 35: 1-19.

Boberg, Charles. 2004. "Real and Apparent Time in Language Change: Late Adoption of Changes in Montreal English". *American Speech* 79: 250–269.

Bourdieu, Pierre. 1984. *Distinction*. London: Routledge.

Britain, David. 2012. "Countering the Urbanist Agenda in Variationist Sociolinguistics: Dialect Contact, Demographic Change and the Rural-Urban Dichotomy". In Sandra Hansen, Christian Schwarz, Phillip Stoeckle, and Tobias Streck, eds. *Dialectological and Folk Dialectological Concepts of Space*. Berlin: Mouton, 12-30.

Britain, David. 2017. "Which Way to Look?: Perspectives on "Urban" and "Rural" in Dialectology". In Chris Montgomery and Emma Moore, eds. *Language and a Sense of Place*. Cambridge: Cambridge University Press, 171-188.

British National Corpus 2014 Spoken. [online]. Retrieved from <https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fbnc2014_spoken_tt3>

Brown, Roger. 1965. S*ocial Psychology*. New York: Free Press.

Buckle, Richard. 1978. *U and Non-U Revisited*. London: Debrett's Peerage.

Campbell-Kibler, Kathryn. 2007. "Accent, (ING), and the Social Logic of Listener Perceptions". *American Speech* 82: 32-64.

Crystal, David. 2019. *The Cambridge Encyclopedia of the English Language*, 3rd edition. Cambridge: Cambridge University Press.

Cunningham, Niall. 2019. "Making and Mapping Britain's "New Ordinary" Elite". *Urban Geography* 40: 604-626.

Dailey-O'Cain, Jennifer. 2000. "The Sociolinguistic Distribution of and Attitudes towards Focuser *Like* and Quotative *Like*". *Journal of Sociolinguistics* 4: 60-80.

Dann, Holly. 2019. *Productions and Perceptions of BATH and TRAP Vowels in Cornish English*. Unpublished thesis. Sheffield: University of Sheffield.

Dodsworth, Robin. 2009. "Modelling Socioeconomic Class in Variationist Sociolinguistics". *Language and Linguistics Compass* 3: 1314-1327.

Fabricius, Anne. 2002. "Ongoing Change in Modern RP: Evidence for the Disappearing Stigma of T-Glottaling". *English World-Wide* 23: 115-136.

Fletcher, Thomas. and Katherine Dashper. 2013. "'Bring on the Dancing Horses!': Ambivalence and Class Obsession within British Media Reports of the Dressage at London 2012". *Sociological Research Online* 18: 118-125.

Foulkes, Paul. and Gerard J. Docherty. 2000. "Another Chapter in the Story of /r/: 'Labiodental' Variants in British English". *Journal of Sociolinguistics* 4: 30-59.

Fox, Kate. 2004. *Watching the English: The Hidden Rules of English Behaviour*. St. Ives: Clays.

Grieve, Jack, Chris Montgomery, Andrea Nini, Akira Murakami, and Diansheng Guo. 2019. Mapping Lexical Dialect Variation in British English Using Twitter. *Frontiers in Artificial Intelligence* 2 doi: doi.org/10.3389/frai.2019.00011.

Halfacre, Caitlin., and Ghada Khattab. 2019. "North-South Dividers in Privately Educated Speakers: A Sociolinguistic Study of Received Pronunciation Using the FOOT-STRUT and TRAP-BATH Distinctions in the North East and South East of England". In Sasha Calhoun, Paola Escudero, Marija Tabain, and Pauk Warren, eds. *Proceedings of the 19th International Congress of Phonetic Sciences*, 2665-2669.

Hecht, Katharina, Daniel McArthur, Mike Savage, and Sam Friedman. 2020. *Elites in the UK: Pulling Away? Social Mobility, Geographic Mobility and Elite Occupations*. Research report, The Sutton Trust, London.

Himelfield, Dave. 2018. "Do You Talk Posh? Take our Quiz to See if you Speak with Fancy Upper-Class Words – or those Used by the Rest of us". *The Mirror* [online]. <https://www.mirror.co.uk/news/uk-news/you-talk-posh-take-quiz-12008288> (accessed [June 20th, 2021)

Honey, John. 1985. “Acrolect and Hyperlect: The Redefinition of English RP”. *English Studies* 66: 241-257.

Ito, Rika., and Sali A. Tagliamonte. 2003. “Well Weird, Right Dodgy, Very Strange, Really Cool: Layering and Recycling in English Intensifiers”. *Language in Society* 32: 257-279.

Jankowski, Bridget L., and Sali A. Tagliamonte. 2019. “Supper or Dinner? Sociolinguistic Variation in the Meals of the Day”. *English World-Wide* 40: 169-200.

Jansen, Sandra. 2017. “Change and Stability in GOOSE, GOAT, and FOOT: Back Vowel Dynamics in Carlisle English”. *English Language & Linguistics* 23: 1-29.

Johnson, Ellen. 1993. “The Relationship between Lexical Variation and Lexical Change”. *Language Variation and Change* 5: 285-303.

Kerswill, Paul. 2009. “Language and Social Class”. In Jonathan Culpeper., Francis Katamba., Paul Kerswill., Anthony McEnery., Ruth Wodak., and Tony McEnery, eds. *English Language*. Palgrave Macmillan: Basingstoke, 358-372.

Labov, William. 1966. *The Social Stratification of English in New York*. Washington D.C: Center for Applied Linguistics.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1990. “The Intersection of Sex and Social Class in the Course of Linguistic Change”. *Language Variation and Change* 2: 205-254.

Labov, William. 2001. *Principle of Linguistic Change: Social Factors.* Oxford: Wiley.

Lambert, Wallace E., Richard C. Hodgson., Robert C. Gardner., and Samuel Fillenbaum. 1960. “Evaluational Reactions to Spoken Languages”. *The Journal of Abnormal and Social Psychology* 60: 44-51.

Lenth, Russell. V. 2021. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.6.1. <https://CRAN.R-project.org/package=emmeans> (accessed February 23, 2023)

Luhman, Reid. 1990. “Appalachian English stereotypes: Language attitudes in Kentucky”. *Language in Society* 19: 331-348.

Macaulay, Ronald K. S. 1977. *Language, Social Class, and Education*. Edinburgh: Edinburgh University Press.

Macaulay, Ronald K. S. 2005. *Talk That Counts: Age, Gender, and Social Class Differences in Discourse*. Oxford: Oxford University Press.

Malarski, Kamil. 2021. *Loss of Rhoticity in South-West England*. Unpublished thesis.Poznań : Adam Mickiewicz University

Ng, Jia C. and Chloe Diskin-Holdaway. 2023. "Attitudes Towards English in Contemporary Malaysia". *World Englishes* 42: 562-578.

Our Dialects. 2022. *How do People Refer to the Item of Furniture Pictured Here?* <https://www.ourdialects.uk/maps/furniture/> (accessed May 17, 2022)

Rampton, Ben. 2010. "Social Class and Sociolinguistics". *Applied Linguistics Review* 1: 1-22.

Robinson, Justina.A. 2010. *Semantic Variation and Change in Present-day English*. Unpublished thesis. Sheffield: University of Sheffield.

Romaine, Suzanne. 2008. "Corpus Linguistics and Sociolinguistics". In Anke Lüdeling and Merja Kytö, eds. *Corpus Linguistics: An International Handbook*. Vol. 1. Berlin: Mouton, 96-111

Ross, Alan S. C. 1954. "Linguistic Class-Indicators in Present-Day English". *Neuphilologische Mitteilungen* 55: 113-149.

Sandow, Rhys J. 2020. "'The Anglo-Cornish Dialect is a Performance, a Deliberate Performance': Ideological Orientation and Patterns of Lexical Variation in a Peripheral Dialect". *English Today* 36: 77–84.

Sandow, Rhys J. 2021. *Anglo-Cornish Dialect Lexis: Variation, Change, and Social Meaning*. Unpublished PhD thesis, University of Sussex.

Sandow, Rhys J. 2022. "Attention, Identity and Linguistic Capital. Inverted Style-Shifting in Anglo-Cornish Dialect Lexis". *English Language & Linguistics* 26: 677–95.

Sandow, Rhys J. 2023. "*Maid* in Cornwall: Social, Stylistic, and Cognitive Factors in Lexical Levelling". *English World-Wide* 44: 157–83.

Sandow, Rhys J. in press a. “An Emmet’s Tale: The Duality of Lexical and Social Change”. To appear in *Language & Communication*.

Sandow, Rhys J. in press b. “Interpreting Identity Effects in Sociolinguistic Variation and Change: The Role of Identity in the Usage and Recognition of Anglo-Cornish Dialect Lexis”. To appear in the *Journal of English Linguistics*.

Sandow, Rhys J., and Justyna A. Robinson. 2018. “‘Doing Cornishness’ in the English Periphery: Embodying Ideology through Anglo-Cornish Dialect Lexis”. In Natalie Braber & Sandra Jansen, eds. *Sociolinguistics in England*. Palgrave Macmillan: London. 333-61.

Sandow, Rhys J., George Bailey, and Natalie Braber. in press. “Language Change is *Wicked*: Semantic and Social Meaning of a Polysemous Adjective”. To appear in *English Language & Linguistics*.

Savage, Mike. 2007. “Changing Social Class Identities in Post-War Britain: Perspectives from Mass-Observation”. *Sociological Research Online* 12, doi: 10.5153/sro.1459.

Savage, Mike, Fiona Devine, Niall Cunningham, Mark Taylor, Yaojun Li, Johs Hjellbrekke, Brigitte Le Roux, Sam Friedman, and Andrew Miles. 2013. “A New Model of Social Class? Findings from the BBC’s Great British Class Survey Experiment”. *Sociology* 47: 219-250.

Shi, Yaqian, and Lei Lei. 2021. “Lexical Use and Social Class: A Study of Lexical Richness, Word Length, and Word Class in Spoken English”. *Lingua* 262, doi: 10.1016/j.lingua.2021.103155.

Snell, Julia. 2018. “Solidarity, Stance, and Class Identities”. *Language in Society* 47: 665-691.

Stuart-Smith, Jane, Claire Timmins, and Fiona Tweedie. 2013. “‘Talkin’ Jockney’? Variation and Change in Glaswegian Accent”. *Journal of Sociolinguistics* 11: 221-260.

Tagliamonte, Sali A., and Julian Brooke. 2014. “A Weird (Language) Tale: Variation and Change in the Adjectives of Strangeness”. *American Speech* 89: 4-41.

Tagliamonte, Sali A., and Katharina Pabst. 2020. “A *Cool* Comparison: Adjectives of Positive Evaluation in Toronto, Canada and York, England”. *Journal of English Linguistics* 48: 3-30.

Trudgill, Peter. 1972. “Sex, Covert Prestige and Linguistic Change in the Urban British English of Norwich”. *Language in Society* 1: 179-195.

Trudgill, Peter. 1974. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Tyler, Imogen. 2008. ““Chav Mum Chav Scum”: Class Disgust in Contemporary Britain”. *Feminist Media Studies* 8: 17-34.

Van Doesum, Niels J., Joshua M. Tybur, and Paul A.M. Van Lange. 2017. “Class Impressions: Higher Social Class Elicits Lower Prosociality”. *Journal of Experimental Psychology* 68: 11-20.

Victor, Tom. 2017. “If you Use these 27 Words it Proves that you’re Definitely not Posh”. *ShortList*. <https://www.shortlist.com/news/posh-words-upper-class-etiquette-u-non-u> (accessed September 21, 2020).

Watt, Dominic. 2002. “‘I don’t Speak with a Geordie Accent, I Speak, like, the Northern Accent’: Contact-Induced Levelling in the Tyneside Vowel System”. *Journal of Sociolinguistics* 6: 44–63.

Appendices

Here we report the optimal model structure for the production data of all three lexical variables, as well as the full table of coefficients for each:

loo/toilet

```
loo/lavatory ~ age2 + gender + class2 + occupation + area_deprivation
```

Factor/level		**Estimate**	**Std. Error**	**z-value**	***p*-value**	
(Intercept)		-1.9702	0.6195	-3.180	0.00147	**
Age	*older*	(reference level)				
	younger	-3.5286	0.8010	-4.405	< 0.001	***
Gender	*female*	(reference level)				
	male	-2.0058	0.6741	-2.975	0.00293	**
Social class (binary)	*middle*	(reference level)				
	working	-1.8835	1.3706	-1.374	0.16938	
Occupation	*highest*	(reference level)				
	lowest	-3.0144	1.3299	-2.267	0.02341	*
	middle	-0.6163	0.9404	-0.655	0.51219	
Deprivation	*middle*	(reference level)				
	least	3.4888	0.8351	4.177	< 0.001	***
	most	4.6582	1.4544	3.203	0.00136	**

napkin/serviette

```
napkin ~ age2 + gender + area_deprivation + elite
```

Factor/level		Estimate	Std. Error	z-value	*p*-value	
(Intercept)		4.0885	1.2253	3.337	< 0.001	***
Age	*older*	(reference level)				
	younger	2.9103	0.6308	4.614	< 0.001	***
Gender	*female*	(reference level)				
	male	-0.8771	0.4034	-2.174	0.02969	*
Deprivation	*middle*	(reference level)				
	least	-1.8835	1.3706	-1.374	0.16938	
	most	1.1410	0.5979	1.908	0.05637	
Elite	*elite*	(reference level)				
	non-elite	-3.1237	1.1494	-2.718	0.00658	**

sofa/settee

```
sofa ~ age2*gender + class3 + area_deprivation
```

Factor/level		Estimate	Std. Error	z-value	*p*-value	
(Intercept)		1.46873	0.52844	2.779	0.00545	**
Age	*older*	(reference level)				
	younger	2.39860	0.60755	3.948	< 0.001	***
Gender	*female*	(reference level)				
	male	0.09203	0.40363	0.228	0.81965	
Social class (ternary)	*middle*	(reference level)				
	lower working	-0.55769	0.56841	-0.981	0.32652	

	upper working	-0.85738	0.47807	-1.793	0.07291	
Deprivation	*middle*	(reference level)				
	least	-0.37926	0.44675	-0.849	0.39592	
	most	-1.37796	0.43222	-3.188	0.00143	**
Age × gender	*older female*	(reference level)				
	younger male	-1.53246	0.74422	-2.059	0.03948	*