Languages without armies: Dialect alignment in word production

Gary M. Oppenheim*
Manon W. Jones

School of Psychology, Bangor University

Under review: September, 2019

Abstract

More than half of all humans speak multiple languages, and even monoglot speakers tend to encounter multiple regional variations of the same language: *dialects*. Particularly in diverse linguistic communities, successful communication requires the ability to detect and switch between dialects. So how exactly do speakers store and retrieve the elements of a dialect? Are the words of a dialect tagged as such, allowing generalization of a dialect during conversation? In this two-session picture naming study, we consider whether occasional exposure to exemplars of a nonnative dialect are sufficient to trigger a generalized switch in dialect production in immediate retrieval and/or later recall. In the first session, British English speakers named a series of pictures, followed by British, American, or dialect-neutral corrective feedback. Several days later, participants named the pictures again, without feedback. Results show progressive increases in the production of US-dialect-marked words within each session, as well as higher rates of US-dialect-marked word production in the second session that gradually generalized even to those tokens that had been presented with UK feedback. Such generalization, and self-priming in particular, is consistent with the idea that speakers tag the elements of a foreign dialect much as they would the elements of a foreign language, and use these tags to bias lexical retrieval.

*School of Psychology, Bangor University, Bangor, Gwynedd, LL57 2AS, UK, g.m.oppenheim@bangor.ac.uk
Thanks to Sheikh Yaqub, Jamie Wignall, and the rest of the Bangor Language Production Lab for assistance with data collection and transcription.

Introduction

Finding the right word to express an idea is a major challenge that we solve so quickly, easily, and even flexibly, that it usually goes unnoticed. We know vast numbers of words, including not only sets of words with similar meanings, but also words that mean (almost) exactly the same thing. For instance, speakers of American English might use couch (74%, according to Székely et al., 2003) and sofa interchangeably, and even know that an older generation might use davenport or chesterfield to refer to the same object. Similarly, speakers of British English might prefer the term sofa (74%, according to Oppenheim, in prep), but also occasionally use couch (12%, ibid), knowing it as a popular 'Americanism'. Interestingly, in both cases, these English subordinate terms can be identified as 'translations' of a dominant term in a different known dialect, analogous to translations of lexical entries in different languages.

Though the distinction is somewhat murky[1], linguists typically characterize dialects as representing subtler variations than distinct languages (Hazen, 2001; Labov, 1998), with higher mutual intelligibility allowing greater synchronic and diachronic interaction between the forms and their speakers. Because interlocutors typically seek to align their linguistic and conceptual systems for efficient dialogue (e.g. Branigan, Pickering, & Cleland, 2000; Brennan & Hanna, 2009; Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987; Pickering & Garrod, 2004), one key to successful communication in dialect-rich environments may therefore be the ability to detect and switch between dialects, and to store and retrieve context-based dialect information for later reuse (cf Green & Abutalebi, 2013).

However, relatively little is currently known about the linguistic and cognitive architectures supporting dialectal language production, and whether it functions in a manner better resembling bilingual or monolingual language production. In order to speak fluently, bilinguals must be able to activate and produce the words and syntax of one language with minimal interference from the other. Current models of bilingualism therefore posit language control mechanisms that allow speakers to avoid language interference: for example, selection mechanisms might ignore the unintended language (Costa, Miozzo, & Caramazza, 1999; Roelofs, Piai, & Rodriguez, 2011) increase the activation of elements in the target language (de Bot, 1992; Finkbeiner, Gollan, & Caramazza, 2006; La Heij, 2005) or actively inhibit words from the unintended language (Abutalebi & Green, 2007; Green, 1986, 1998). In all scenarios, this separation of languages is thought

[1] As well as rough linguistic criteria such as size and mutual intelligibility, distinctions often incorporate politicized notions like 'prestige' (Hudson, 1996; Wei, 2000), leading to the popular aphorism that "language is a dialect with an army and a navy" (Weinreich, 1945). Thus, geopolitical inconsistencies abound in distinctions between languages and dialects. For instance, Danish, Norwegian and Swedish, are typically considered full, distinct languages, despite their high mutual intelligibility, whereas Chinese authorities have recently declared Cantonese and Mandarin 'dialects', of a single language, despite their low mutual intelligibility.

to be accomplished via some form of "language tagging", that allows the control mechanism to select output from the intended language, thus implementing a language grouping principle on-demand. Such a control mechanism might apply during lexical activation or selection, for example, serving as a source of initial activation or boosting activation of words in the target language (e.g. Runnqvist, Strijkers, Alario, & Costa, 2012).

The idea that bilinguals actively switch between co-activated languages draws on two major empirical effects. First, bilinguals typically exhibit language switch costs in which, for example, production of successive words is hindered by cued switching from one language to another (e.g., see Kiesel et al., 2010 and Declerck & Philipp, 2015, for a review; but cf. Kleinman & Gollan, 2016). These language switching costs are thought to be analogous to those involved in non-linguistic task-switching, possibly even depending on shared cognitive and neural bases (e.g. Anderson, Chung-Fat-Yim, Bellana, Luk, & Bialystok, 2018; Bialystok & Craik, 2010), thus suggesting some sort of similar cognitive control process. Specific evidence that cross-linguistic lexical representations are habitually co-activated is the cognate facilitation effect: Inter-lingual cognates sharing orthographic and/or phonological form tend to be named faster than non-cognates, suggesting that separate lexical entries in each language are jointly activated via overlapping phonological features, boosting retrieval speed (Bernolet, Hartsuiker, & Pickering, 2012; Cai, Pickering, Yan, & Branigan, 2011; Costa, Santesteban, & Caño, 2005). Thus, language separation in bilingual lexical production is thought to promote fluency, minimizing cross-language contamination, and reducing or eliminating possible cross-language competition.

Monolingual language production, in contrast is generally assumed to lack bases for systematic language separation, despite the fact that monolinguals regularly encounter close synonyms and a range of dialects, registers, and accents within a given language (Dylman & Barry, 2018; Foulkes & Hay, 2015). The bilingualism literature typically identifies bidialectals as monolinguals (Marian, Blumfield, & Kaushanskaya, 2007) and indeed, dialects of a given language are often mutually intelligible, in the sense that deviances in dialect only affect isolated lexical utterances (about 10% in Oppenheim's, in prep, comparison of US and UK picture naming norms) and they tend not to be obvious.

A useful analogy for the distinction between languages and dialects might be in terms of that between common taxonomic categories (e.g. kinds of fruit) versus ad hoc or goal-directed categories (e.g., things to keep in a desk drawer; e.g. Barsalou, 1983). Elements of distinct languages—like common categories—can often be distinguished based on surface features (e.g. phonotactic or orthotactic restrictions, distinguishing legal nonwords in one language from another) as well as distributional information (or 'affordances'). Elements of dialects, on the other hand—like ad hoc categories—often differ primarily in the latter[2]. Although people can generate and evaluate exemplars for ad hoc categories, they

[2]Except in the case of e.g. pidgens; note that although accents can provide useful cues to dialect, they color all words, not just dialect-specific elements.

have only weak category-to-exemplar and exemplar-to-category associations, so the process is quite effortful[3]. These associations can, however, be strengthened through experience, as a category gradually loses its ad hoc status, for instance allowing one element of a language or dialect to prime other elements.

The question posed in this paper is whether the selective retrieval mechanism described in relation to bilingualism also characterizes multidialectal production. That is, do speakers selectively activate and retrieve elements of dialects much like elements of distinct languages, via well-established category-exemplar associations, or do they simply store them as declaratively tagged alternatives within a more integrated system, e.g. as simple synonyms or semantic relations that one could retrieve and deliberately test for category membership? The sparse empirical literature on this topic has typically adapted rather artificial experimental paradigms that were derived from research in bilingualism, and has presented conflicting accounts. For instance, in a cued-dialect-switching task, bidialectals' longer naming latencies indicated dialect switch costs reminiscent of the language-switching cost shown in bilinguals, suggesting that bidialectal speakers might utilize similar cognitive control mechanisms (Kirk, Kempe, Scott-Brown, Philipp, & Declerck, 2018). However, in a picture-word interference task, bidialectals' longer naming latencies in a condition with cross-dialect distractors (Melinger, 2018) contrasted with bilinguals' typical facilitation in conditions with cross-language distractors (Costa et al., 1999; Hermans, Bongaerts, de Bot, & Schreuder, 1998). Melinger (2018) speculated that because dialect mixing often remains highly comprehensible to an interlocutor, it may not warrant developing the kind of tag-based control that characterises bilingual language production.

Perhaps a more effective approach to the study of bidialectal language production is to use a paradigm that better reflects the characteristics of the language phenomenon. Given that moving from one dialect to another may not comprise an absolute 'switch' as is often the case in moving from one language to another, it may be more fruitful to consider the proportional increase of dialect use as a function of priming. In fact, although language switching paradigms typically highlight the role of deliberate cognitive control in externally cued changes between languages, automatic language priming may facilitate staying within a language with minimal effort (Léwy & Grosjean, 2000; Li, 1996). One can then ask whether such priming might similarly apply to the elements of dialects: does introducing 'Americanisms' like stroller prime British English speakers to produce other Americanisms like couch instead of the dominant British term, sofa?

The Current Study

The current study is designed to assess whether speakers represent and produce the words of a second dialect in a manner functionally similar to words of a second

[3] For example, Barsalou (1983, pp225-226) posited that people generate exemplars for ad hoc categories via a deliberate "generate-test" loop, serially proposing and evaluating candidates.

language. We focus on British English speakers' production of Americanisms because they tend to have extensive exposure to American English and we have comparable empirical measures for each dialect. In a two-session picture naming/memory experiment, we ask whether a form of lexical priming will generalize to increase the likelihood of selecting other words from the same non-native dialect. In each trial of Session 1, native speakers of British English named a single picture, triggering the appearance of a desired name directly below it (see Fig. 2).[4] In critical trials—which represented only a small proportion of the total—these desired names were implicitly 'dialect-marked', in the sense that the picture's empirically assessed dominant name in British English norms differed from its dominant name in American English norms. Without repeating any pictures, the feedback thus presented names from UK norms for the first 210 pictures (Blocks 1-2), and names from US norms for the remaining 315 (Blocks 3-5), keeping in mind that only a small proportion of these names were in fact dialect-contrastive. The crucial measure is how often UK participants name critical items using the dominant US names, relative to the dominant UK names; the first two blocks establish a baseline against which to assess whether speakers become more likely to choose US-dialect names for new pictures after we introduce US-dialect feedback.

We hypothesized that if the elements of dialects are simply stored as within-language synonyms (e.g. Melinger, 2018), without structural basis for linking dialect members, then priming one Americanism should have no effect on British speakers' production of future Americanisms. If, on the other hand, dialects operate according to the grouping principles of distinct languages, then priming a single-token Americanism (e.g., 'truck$_{\text{US}}$') may increase the likelihood of producing other, unencountered Americanisms ("couch$_{\text{US}}$") in subsequent trials (i.e. increasing the ratio of "couch$_{\text{US}}$" to "sofa$_{\text{UK}}$" responses). In other words, priming the system with a few US-dialect tokens may be sufficient to activate other Americanisms.

In Session 2, we re-tested participants 1-3 days later to assess whether a dialect tag might generalize in storage or retrieval from memory. In other work (Balatsou, Fischer-Baum, & Oppenheim, in revision), we have recently demonstrated that participants' picture name selections are remarkably stable over repeated testing, even after a one-week delay, providing a baseline expectation that, if the feedback in Session 1 has no effect (either due to failures in encoding or retrieval), then speakers should tend to produce the same names in both sessions, and thus the ratio of Americanisms to Britishisms should remain unchanged.

We therefore consider three possibilities:

1. If participants use item-specific feedback, then their feedback-congruent responses should be more likely in Session 2 than Session 1, and this should

[4] Note that Alario et al., 2004, used a very similar procedure for their two-session investigation of picture naming norms, drawing an explicit parallel between their use of feedback in their first session and the common use of familiarization procedures before many picture naming experiments.

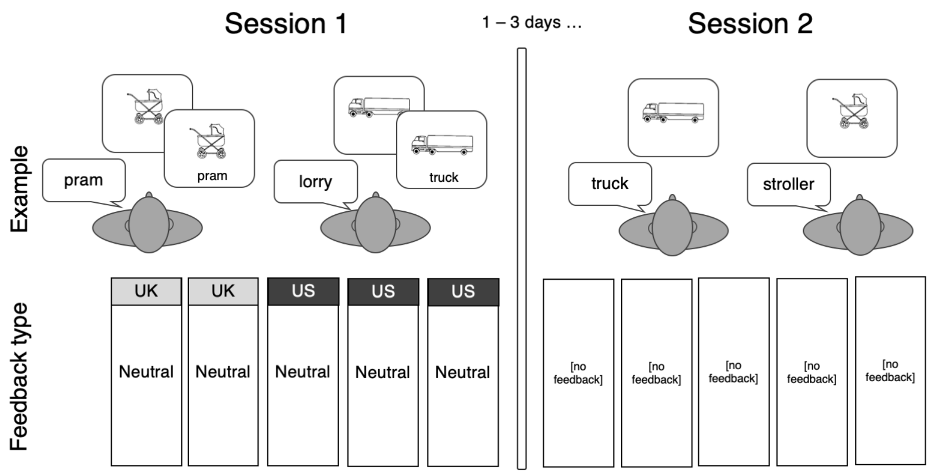

Figure 1: Schematic depicting an example of the experimental procedure (top panel), and experimental blocks according to feedback type (bottom panel). In Session 1, Blocks 1 and 2 presented British dialect feedback, whilst Blocks 3, 4 and 5 presented American dialect feedback on items in which such dialectal divergence was appropriate (~ 10% of trials). In Session 2, no feedback was provided, but participants were instructed to name the pictures using the feedback from Session 1.

hold for the Britishisms presented as feedback, as well as the Americanisms.

2. If participants tag the previous episode with a single dialect (similar to tagging an entire conversation), then they should be more likely to produce US-dialect names in Session 2 than Session 1, regardless of whether they received US- or UK-dialect feedback for a specific item (though this hinges on them being able to guess the desired US name).

3. If dialect generalises during retrieval–for instance by subtly priming a dialect or inducing a controlled switch–then it should increase the production of Americanisms over the course of each Session.

Assuming that speakers already begin Session 2 by attempting to recall dialect-specific feedback as best they can (Prediction 1), these hypothesized increases in their production of Americanisms (Predictions 2 and 3) should be most apparent for those critical items that actually received UK-dialect feedback, analogous to eliciting a false memory. Thus, the re-test manipulation allows us to distinguish relatively ephemeral influences from the contributions of longer-term learning or memory processes, and assess a possible production-based mechanism for dialect generalisation. Thus, our results can speak not only to the structural representation of dialects, but also how they are subject to experience-based adjustments in the utterance-building system, even in adult speakers.

Methods

Participants

Twenty-six monolingual, native-British English Bangor University students (12 female, mean age = 21.24) received course credit for participation. All provided informed consent, and reported normal or corrected-to-normal vision and hearing, and no known language disorders. Additional participants were excluded due to technical difficulties (2), or failure to attend the second session (2).

Stimuli

The 525 black-and-white line drawings of the International Picture Naming Project (IPNP; E. Bates et al., 2003) served as the stimuli for this experiment. Norms for naming these pictures in American English (from Székely et al., 2003) and British English (from Johnston, Dent, Humphreys, & Barry, 2010, and then-preliminary results from Oppenheim, in prep) provided empirical bases for assessing dialect-based differences in their names. We operationalized dialect-based lexical differences as cases where British participants produced the US-dominant name much less frequently than in the US norms (Figure 2), and often produced a different name instead (see Table 1), and thus selected 65 pictures as critical items (see Appendix A for the full list). These items included well-established translation-equivalents (e.g. tap_{UK}/$\text{faucet}_{\text{US}}$), as well as items for which the dialectal terms may also reflect subtle semantic distinctions (e.g. $\text{tortoise}_{\text{UK}}$/$\text{turtle}_{\text{US}}$; note that these alternatives could not be readily distinguished from a simple line drawing). Thus, our operational definition of a 'dialect' included both linguistic and paralinguistic distinctions. The remaining 460 items (87.6%) served as fillers, for which the norm-assessed dominant British and US terms were identical in almost all cases (94.5%; the remaining cases tended to be items with borderline name dominance), with similar dominance in each dialect.

Design & Procedure

The experiment consisted of two 35-45-minute sessions, conducted 1-3 days apart (mean:1.3 days). Each session consisted of a 525-trial picture-naming task, divided into five blocks of 105 pictures and presented on an 85Hz 17" CRT via a PC running PsychoPy (Peirce, 2010) in a sound-attenuated booth. To allow both within-subjects and within-items manipulations, trial sequences were progressively counterbalanced via a generative algorithm that ensured a different order for each session, with each item appearing in each block similarly often across participants. Each trial consisted of a randomly varying 900-1900ms blank screen (inter-trial interval), followed by a 200ms (17 frames) fixation cross, followed by a 506ms (43 frames) blank screen, followed by a single 422px X 422px

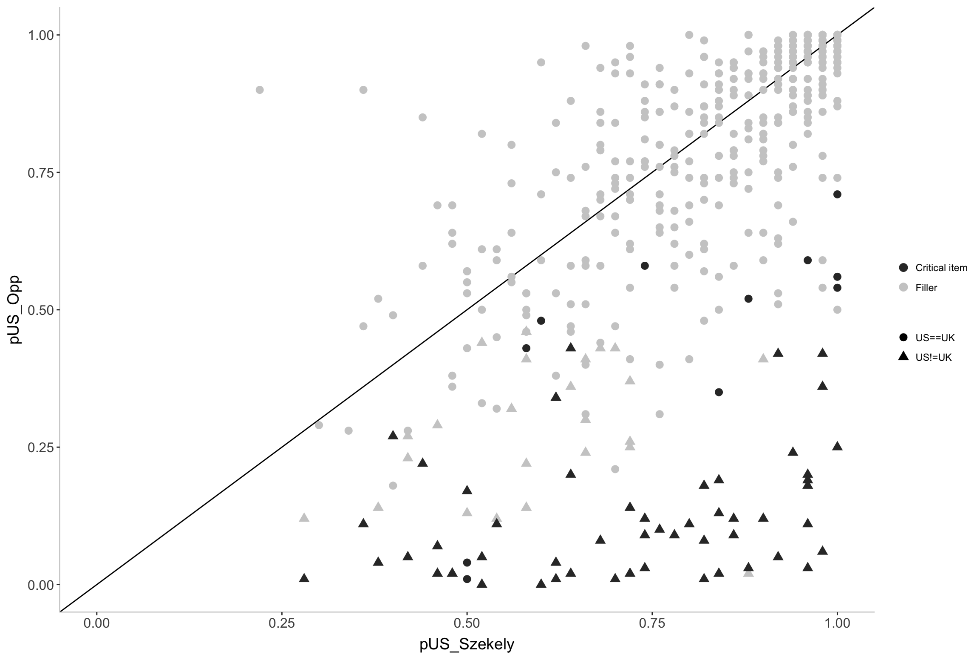

Figure 2: Correspondence between picture name agreement in US norms (from Székely et al., 2003) and the frequencies of the same names for the same pictures in recent UK norms (from Oppenheim, in prep.). British participants tended use the US dominant names much less often for critical items (black) than fillers (grey).

picture that appeared in the middle of the screen for up to 3000ms. Participants were instructed to quickly and accurately name each picture; the vocalization triggered a 50ms delayed-threshold voicekey (Tyler, Tyler, & Burnham, 2005) via a headmounted microphone and was digitally recorded for later transcription.

Each session thus resembled an IPNP picture naming norming experiment (E. Bates et al., 2003), with just two crucial differences. First, in Session 1, participants received *written feedback* after each response: at the end of each trial, the picture's desired name appeared below it for 1250ms, and participants were instructed to remember these names for future use (cf Session 1 of Alario et al., 2004). As mentioned above, this feedback was always one of the picture's dominant names: the UK-dominant name in Blocks 1 and 2, and the US-dominant name in Blocks 3-5. Note that, for most trials (87.6%) the US-dominant and UK-dominant names were identical; they differed only for the 12.4% of trials designated as 'critical'. Second, in Session 2, participants were instructed to name each picture using the names that they had learned for each picture in Session 1. No further feedback was presented in Session 2, however, so it differed from a standard norming procedure only in terms of that instruction. Note that

Table 1: Picture name characteristics for critical items and fillers. US norms come from Szekely et al. (2003; n=50 US English speakers). UK norms come from Oppenheim (in prep, n=100 UK English speakers). Not included: five shape pictures (e.g. square, circle) for which Szekely et al. (2003) did not report norms.

	Frequency of the dominant US name in US norms	Frequency of the dominant US name in UK norms	Proportion of items with the same dominant name in UK and US norms	Frequency of the *specified* US name in UK norms	Frequency of the *specified* UK name in UK norms	Naming latency for the specified US name in UK norms	Naming latency for the specified UK name in UK norms
Critical items (n=65)	0.726	0.179	0.169	0.179	0.559	1028 (11)	1008 (6)
Fillers (n=455)	0.834	0.799	0.945	0.802	0.802	931 (2)	931 (2)

no explicit references to dialects or Americanisms were made in either session.

Analyses

Responses were initially transcribed online and confirmed offline via audio recordings. To avoid ambiguities or post-hoc attributions, we classified each response as an Americanism or Britishism if and only if it exactly matched the picture's predetermined dialect-specific name. For instance, for the $\text{sofa}_{\text{UK}}/\text{couch}_{\text{US}}$ image, a "sofa" response would be classified as a Britishism, a "couch" response would be classified as an Americanism, and "settee", "divan", "davenport", "chesterfield" or any other response would be classified as neither. Trials where the participant did not produce a name within 3000ms were coded as omissions. Any trial that ended prematurely due to a voicekey error was excluded.

Statistical analyses used pre-planned logistic mixed effects regressions via lme4 (D. M. Bates et al., 2019) in R v3.6.1. Unless otherwise specified, all fixed effects are coded as centered continuous predictors, and all models include maximal by-subject and by-item random effects structures, omitting correlations between random effects to facilitate convergence (Barr, Levy, Scheepers, & Tily, 2013). Claims of simple main effects describe results of identical procedures and predictors applied to relevant subsets of the data. Non-directional p-values are estimated via Wald approximation. Although our research questions primarily concern word choices, we also use naming latencies to characterize and constrain possible mechanisms where appropriate.

Results

As a preliminary, we note that, in Session 1, participants produced the expected names for the filler items in proportions that correlated very strongly with their frequencies in our previous UK norms (Pearson's R=0.937, p<.001), and with by-item mean latencies that correlated similarly well (Pearson's R=0.909, p<.001). This validates our use of these norms for predicting participants' other naming behaviours in this experiment.

From this point, we restrict our analyses to the 3380 dialect-critical trials, summarised in Table 2; filler trials are included in the extended table in Appendix B. We first excluded 87 (2.57%) trials that ended prematurely due to lipsmacks, audible hesitations, or equipment errors, because such problems would have allowed feedback to contaminate response selection in Session 1. Each remaining trial was then classified as one of the following:

1. **Americanism**: The participant's response exactly matched the US-dialect name that we had pre-selected for the item and presented as feedback. (22.9% of all responses)
2. **Britishism**: The participant's response exactly matched the UK-dialect name that we had pre-selected for the item and presented as feedback. (54.8% of all responses)
3. Omission: any case where a participant did not produce a codable response within 3000ms. (4.8% of all responses)
4. Any other response: Any other name, regardless of whether it could be subjectively considered characteristic of a particular dialect. (17.5% of all responses)

Focusing our primary analyses on the 2560 (77.7%) responses from the first two categories, Americanisms and Britishisms (depicted in Figure 3), allows us to straightforwardly address our research hypotheses about dialect access. However, we consider omissions and other responses in follow-up tests, where appropriate, as a step toward addressing relative contributions of dialect activation and inhibition. This analytical approach follows the logic of Begg & Gray's (1984) binomial approximation of multinomial regression. Before considering word selections, however, it may useful to contextualise them by considering the response time patterns.

Naming latency patterns within each session, depicted in Figure 4 and considered in further detail in Appendix C, were roughly in line with those previously reported for similar procedures without corrective feedback (e.g. Székely et al., 2003). Briefly, naming latencies in such studies reliably show several basic features that we also find here: First, although naming latencies are affected by many factors, they tend to average around 1000ms in no-feedback versions of this task (cf Table 1); our mean RTs in Session 1 seem close enough to rule out concern that they might reflect qualitatively different cognitive processes.[5] Second, speakers typically produce dominant names faster than non-dominant names; here, such a trend emerges when comparing the Britishisms to the

[5] Note also that the addition of feedback necessarily increases the effective interstimulus interval (ISI), which can itself increase RTs.

Session 1 feedback	Session	Block	Used target UK name	Used target US name	Used any other name	No response	Excluded	US to UK ratio
UK	**1**	**1**	57.5% (188)	18.3% (60)	20.8% (68)	3.4% (11)	(13)	0.319
		2	55.9% (179)	17.8% (57)	21.6% (69)	4.7% (15)	(18)	0.318
		3	-	-	-	-	-	-
		4	-	-	-	-	-	-
		5	-	-	-	-	-	-
	2	**1**	74.2% (95)	4.7% (6)	10.2% (13)	10.9% (14)	(1)	0.063
		2	67.6% (50)	12.2% (9)	14.9% (11)	5.4% (4)	(2)	0.18
		3	77.6% (142)	10.9% (20)	7.1% (13)	4.4% (8)	(2)	0.141
		4	66.7% (78)	17.1% (20)	15.4% (18)	0.9% (1)	(0)	0.256
		5	63.3% (107)	20.7% (35)	14.8% (25)	1.2% (2)	(2)	0.327
US	**1**	**1**	-	-	-	-	-	-
		2	-	-	-	-	-	-
		3	54.5% (177)	17.8% (58)	22.5% (73)	5.2% (17)	(10)	0.328
		4	53.8% (176)	18.3% (60)	19.9% (65)	8.0% (26)	(13)	0.341
		5	50.5% (165)	23.2% (76)	20.2% (66)	6.1% (20)	(10)	0.461
	2	**1**	45.5% (92)	34.7% (70)	18.3% (37)	1.5% (3)	(4)	0.761
		2	41.9% (108)	39.5% (102)	15.9% (41)	2.7% (7)	(5)	0.944
		3	40.9% (61)	38.3% (57)	13.4% (20)	7.4% (11)	(4)	0.934
		4	48.6% (104)	34.1% (73)	13.1% (28)	4.2% (9)	(2)	0.702
		5	48.0% (83)	30.1% (52)	16.2% (28)	5.8% (10)	(1)	0.627

Table 2: By-block outcome summary for the 3380 dialect-critical trials.

less-frequent Americanisms. Third, naming latencies typically increase over the course of a session, a pattern that others have attributed to simple fatigue. This pattern emerges in the current experiment, and while inhibitory mechanisms could contribute to it, its slope is similar to what others have previously reported (cf Figure 2 in Székely et al., 2003). Finally, other studies have previously demonstrated very long-lasting repetition priming in picture naming (e.g. Cave, 1997), and this pattern also holds when comparing Session 1 to Session 2. Thus, considering the prior expectations, derived from similar tasks with the same stimuli, the response times provide no clear evidence that the presence of feedback, occasionally dialect-marked, particularly affected the picture-naming process. We therefore focus our remaining analyses on the outcomes themselves.

Session 1: Immediate dialect priming via token-based feedback

During the two-block baseline in the beginning of Session 1, participants named the critical items using our expected Britishisms in 56.7% of trials, and our expected Americanisms in 18.1%, a ratio of 1:0.319 that closely matches the 1:0.320 ratio (55.9% UK vs 17.9% US) observed in our previous UK norms (given in Table 1, from Oppenheim, in prep.). The three subsequent blocks introduced US-dialect feedback; the question is whether it generalises to increase the likelihood of other Americanisms, despite their lack of conceptual or lexical

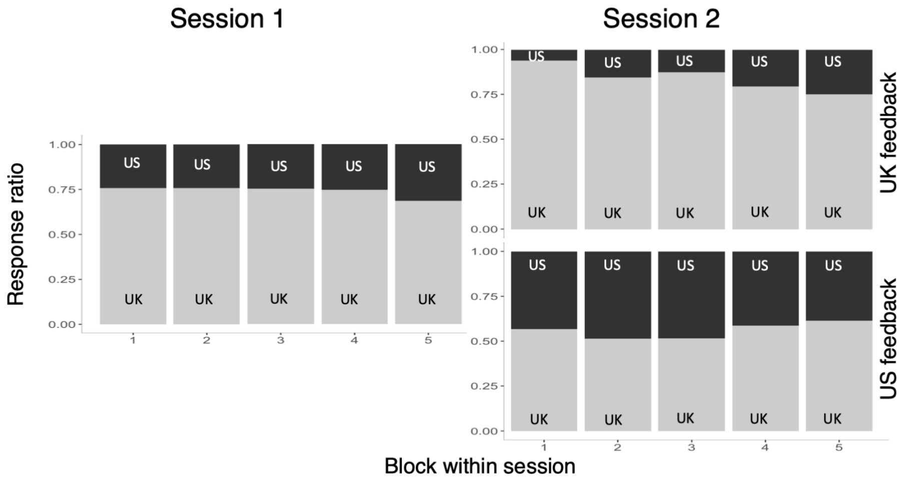

Figure 3: By-block summary of the 2560 dialect-critical trial where participants produced the specified Americanism or Britishism. Responses matching the desired UK-English name are depicted in light grey. Responses matching the desired US-English name are depicted in black.

association. We therefore used a maximal logistic mixed effects regression to assess a link between the introduction of US-dialect feedback in Block 3 and an increase in the log-likelihood of participants volunteering a picture's desired US-dialect name (rather than the UK-dialect name) in Session 1, contrast-coding Block{1:5} as a centered linear effect following a two-block baseline (i.e. contrasts(Block) = {-1.2, -1.2, -0.2, 0.8, 1.8}).[6] Consistent with our dialect-tagging prediction, this analysis showed that participants grew significantly more likely to volunteer Americanisms after the introduction of US-dialect feedback ($\beta_{\text{main effect of Block}}$=0.186, SE=0.0763, p=0.0149, OR=1.200), thus revealing dialect-based generalization to novel pictures/names within the same session.

If switching to a non-native dialect involves inhibiting one's native dialect then we might expect omissions to similarly increase over the course of this session, relative to the same Britishism baseline. Applying the same logistic regression model as above, we see that this ratio of omissions to Britishisms also increases over the course of Session 1 ($\beta_{\text{main effect of Block}}$=0.341, SE=0.134, p=0.0112, OR=1.41), consistent with the inhibition-based prediction.

However, such an increase in omissions could be explained even without dialect-based inhibition, as a corollary of the slowing noted in Szekely et al.'s (2003) monolingual picture-naming norms. If inhibiting one's native-dialect drives the

[6] A similar idea could be statistically framed as a binomial contrast of Feedback Dialect, with similar results, but a continuous coding both minimizes deviance and fits better with our mechanistic assumptions of cumulative priming.

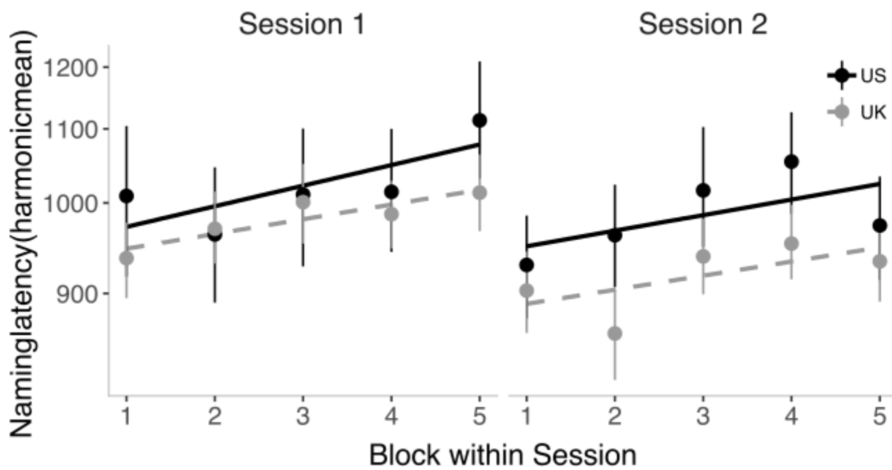

Figure 4: Mean naming latencies for Americanisms (black) and Britishisms (grey), by block within session. Error bars represent bootstrapped 95% confidence intervals; trendlines depict simple linear regression fits and confidence intervals.

production of a non-native dialect, then we might further expect these increases in the rates of Americanisms and omissions to be accompanied by an increase in the rate of other, less frequent responses (e.g. naming 'cot$_{UK}$' as 'cradle', which is not the expected name in either dialect). But, by applying the same logistic regression model to the ratio of these 'other responses' to Britishisms, we see that it remains relatively stable throughout Session 1 ($\beta_{\text{main effect of Block}}$=0.0313, SE=0.079, p=0.692, OR=1.03), a pattern that is inconsistent with the inhibition-based account.

Thus, changes in response distributions over the course of Session 1 demonstrate that introducing non-native-dialect feedback in Block 3 quickly generalized to specifically support the retrieval of other names from the same non-native dialect.

Session 2: Dialect generalisation versus token specificity in memory, and gradual spread

We now consider the second session, which occurred 1-3 days after the first. Participants named all 525 pictures again, in a different pseudorandom order. Recall that this second session omitted the feedback component, and instead simply instructed participants to use the terms that they had learned in the first session. Back-sorting items according to their previously-paired feedback dialect can then provide a basis for assessing dialect-based generalization without immediately available external cues.

The main consideration here is whether a dialect tag might generalize in either storage or retrieval from memory. We envisioned three possibilities. First, if speakers encode and retrieve item-specific lexical information, then they should be more likely to produce Americanisms for items that were previously paired with US-dialect feedback than for items that were previously paired with UK-dialect feedback, and vice versa. Second, if speakers encode and retrieve dialect marking for the previous session as a whole, analogous to tagging a conversation, then they should be more likely to produce Americanisms in Session 2 than in Session 1, regardless of the item-specific feedback. Third, if a speaker's use of a dialect generalizes, or self-primes during retrieval, then this self-priming should gradually increase their production of Americanisms over the course of the second session.

We address the first two predictions by applying a logistic mixed effects regression that considers the ratio of Americanisms to Britishisms in both sessions, as predicted by centered fixed effects of Session{1,2}, Session1FeedbackDialect{UK,US}, and their interaction, plus an uncorrelated maximal random effects structure for subjects and items.[7] As suggested by Figure 3, the analysis confirms that participants were more likely to produce Americanisms for the items that were presented with US-dialect feedback ($\beta_{\text{main effect of Session1FeedbackDialect}}$ =1.33, SE=0.181, p<0.001, OR=3.77), and although they produced somewhat more Americanisms in Session 2 overall ($\beta_{\text{main effect of Session}}$=0.358, SE=0.109, p=0.001, OR=1.43), this increase was strongly moderated by the specific feedback that they had received for each specific picture ($\beta_{\text{Session X Session1FeedbackDialect interaction}}$=1.80, SE=0.252, p<0.001, OR=6.06). Thus, comparing the two sessions provides little or no evidence for the possibility of dialect-tagging interactions on a higher level. Instead, the Session X Feedback interaction suggests that the observed increase in feedback-congruent responses was likely driven by more specific priming or memory of the dialect's individual tokens.[8]

Finally, we can ask whether dialect self-generalizes during the retrieval process itself, as one might expect if the elements of a dialect share some prime-able core representation. To address this question, we apply a logistic mixed effects

[7] To allow model convergence, we excluded one item (traffic light$_{\text{UK}}$ / stoplight$_{\text{US}}$) for which no participant used its US-dialect name in either session.

[8] Regarding the other outcome categories, omissions were more likely for items with US-dialect feedback ($\beta_{\text{main effect of Session1FeedbackDialect}}$ =1.07, SE=0.326, p=0.001, OR=2.90), but did not significantly change with Session ($\beta_{\text{main effect of Session}}$=-0.431, SE=0.355, p=0.225, OR=0.65), nor did Session moderate the feedback dialect effect ($\beta_{\text{Session X Session1FeedbackDialect interaction}}$=0.485, SE=0.465, p=0.297, OR=1.62) . Other names were also more likely for items with US-dialect feedback ($\beta_{\text{main effect of Session1FeedbackDialect}}$ =0.653, SE=0.184, p<0.001, OR=1.92) , and although they became less likely in Session 2 ($\beta_{\text{main effect of Session}}$=-0.657, SE=0.119, p<0.001, OR=0.519), that decrease was driven by those previously presented with UK-dialect feedback 2 (($\beta_{\text{Session X Session1FeedbackDialect interaction}}$=-0.867, SE=0.240, p<0.001, OR=2.38).

regression that considers the ratio of Americanisms to Britishisms in Session 2 alone, as predicted by centered fixed effects of Session1FeedbackDialect{UK,US}, Block{1:5} and their interaction, plus an uncorrelated maximal random effects structure for subjects and items. Consistent with the token-specific priming noted above, this analysis confirms that participants were much more likely to produce Americanisms when naming pictures that they had seen paired with US-dialect feedback in the previous session ($\beta_{\text{main effect of Session1FeedbackDialect}}$ =2.25, SE=0.258, p<0.001, OR=9.53). Remarkably, although their likelihood of producing Americanisms generally increased over the course of this second session ($\beta_{\text{main effect of Block}}$ =1.48, SE=0.0736, p=0.045, OR=1.16), a significant interaction indicates that the increase was driven by the pictures that had previously been paired with UK-dialect feedback ($\beta_{\text{Block X Session1FeedbackDialect interaction}}$=-0.439, SE=0.206, p=0.036, OR=0.645).[9]

Fitting analogous models to subsets of the data, based on their Session 1 feedback dialect, we can confirm that the rate of Americanisms does not significantly change for the items that were previously paired with US-dialect feedback ($\beta_{\text{main effect of Block}}$ =-0.0445, SE=0.105, p=0.671, OR=0.956), but significantly increases for those that were previously paired with UK-dialect feedback ($\beta_{\text{main effect of Block}}$ =0.392, SE=0.148, p=0.008, OR=1.48).

The increase in Americanisms over the course of Session 2, thus internally replicates the gradual generalisation that we previously saw in Session 1, but this time as a purely production-internal phenomenon.

Discussion

Dialects and languages as usage-derived categories

In the category learning literature, researchers have long posited a continuum between the ad hoc categories that people might create on-the-fly in service of particular purposes (e.g. things to save from a burning home, Barsalou, 1983) and more traditional (e.g. taxonomic) categories, based on either readily perceivable 'natural kinds' (animals) or merely theorized differences (e.g. mammals). Ad hoc categories, it has been argued, have the experiential bases to develop into better-established categories—for instance, for a landscape architect, an ad hoc

[9] Regarding the other outcome categories in Session 2, omissions were more likely for items with US-dialect feedback ($\beta_{\text{main effect of Session1FeedbackDialect}}$ =1.05, SE=0.381, p=0.006, OR=2.85), and while they did not significantly change with Block ($\beta_{\text{main effect of Block}}$ =-0.113, SE=0.148, p=0.445, OR=0.893), there was a non-significant trend such that they increased numerically more for items with US-dialect feedback ($\beta_{\text{Block X Session1FeedbackDialect interaction}}$=0.463, SE=0.304, p=0.127, OR=1.59). Other names were also more likely for items with US-dialect feedback ($\beta_{\text{main effect of Session1FeedbackDialect}}$ =1.19, SE=0.257, p<0.001, OR=3.30), and while they did not significantly change with Block ($\beta_{\text{main effect of Block}}$ =0.0495, SE=0.0805, p=0.539, OR=1.05), the feedback-dialect effect marginally decreased as Session 2 progressed ($\beta_{\text{Block X Session1FeedbackDialect interaction}}$=-0.380, SE=0.216, p=0.0789, OR=0.684).

category like "trees that do well in the shade" might develop into a better-established "shade trees" category (Medin, Lynch, Coley, & Atran, 1997)—but have not yet done so. While such nascent categories exhibit some features of better-established categories, such as allowing graded prototypicality judgements, they lack the strong bidirectional category-exemplar associations that would typically support exemplar-to-exemplar priming.

In the Introduction, we asked where dialect knowledge might lie on this continuum. People who speak multiple languages clearly treat them as well-established categories—for instance, they use a representation of the language (a "language tag") as a category to select (or inhibit) its exemplars—but until now it was unclear whether people who use multiple dialects represent them as well-established categories, too. Incorporating a dialect priming manipulation into a simple picture naming task, we therefore asked whether dialect-associated words would be capable of exemplar-to-exemplar priming, such that processing one dialect-associated word (e.g. couch_{US}) would increase the likelihood of producing other words that are associated with the same dialect (e.g. $\text{popsicle}_{\text{US}}$). In each of two sessions, we saw evidence for such priming, and can therefore conclude that people who regularly use alternative dialects (can) represent them as well-established categories, much the same way that they represent distinct languages as such.

Does this mean that speakers *must* represent dialects as distinct, well-established categories? Of course not. An extensive literature demonstrates that individuals learn categories through their own individual goals and experiences. Despite the existence of evidence for particular category distinction out in the world, whether an individual acquires a robust representation of that category will depend on the extent and type of their interactions with it. What is clear here—from our sample of young British English speakers using American English—is that speakers who use multiple dialects can represent and use them in a way that is functionally similar to how they might represent and use distinct languages.

Dialect and generalization in production

But dialects do clearly differ from distinct languages in some important ways. While distinct language pairs *can* show considerable overlap, the overlap between two dialects of a single language is central to their definition as such. Such overlap is clearly evident in picture naming norms, where British and American English speakers prefer exactly the same names for most common experimental stimuli, providing a strong basis for the dialects' mutual intelligibility. That is, when conversing with someone who speaks another dialect of the same language, one can usually expect to both understand and be understood, and this mutual intelligibility has at least two important consequences. First, unlike non-native languages[10], non-native dialects tend to be socially acquired—perhaps more like social registers—rather than taught in formal educational settings, presumably

[10] There are exceptions in both directions, of course; some dialects are formally taught and many multilinguals acquire distinct languages socially.

because the differences are subtle and the consequences of misuse are minor. Second, because a dialect typically remains understandable even when deployed in the wrong context, there is less potential cost for dialect mixing (in most cases it would be unnoticeable), and therefore less need for speakers to inhibit a non-target dialect in the same way that they might inhibit a non-target language.[11]

Our claim, then, is not that languages and dialects are entirely identical. Rather, we claim more specifically that speakers represent and retrieve the elements of dialects via a grouping principle that is functionally similar to 'language tagging'. We see evidence for the use of such 'dialect tagging' in each session of the current experiment. Recall that, in the Introduction, we hypothesized that language-like dialect tagging should make it possible to prime a dialect in a way that generalizes beyond individual tokens. We tested this idea by running British English speakers in a two-session picture naming task that closely resembled tasks that we and others have previously used for picture naming norms: in each session, participants named a long series of pictures, without repetition. In Session 1, feedback predominantly presented each item's dominant UK-dialect name, but starting in Block 3 of 5 occasionally presented a contrasting US-dialect name instead. Consistent with our dialect-tagging prediction, as we introduced US-dialect feedback, our British English speakers increasingly used Americanisms to name the new pictures that they encountered, thus showing token-to-dialect generalization.

This generalization seems to reflect automatic activation of the non-native dialect as a whole, rather than inhibition of the native dialect or a slow, deliberate, retrieval process. If the generalization were driven by either of the latter, then it should be associated with increased naming latencies. But naming latencies in Session 1 were generally commensurate with those from similar procedures without feedback manipulations, and relative to such baselines they showed no additional slowing after the introduction of US-dialect feedback[12]. Similarly, although we saw some increase in omissions over the course of Session 1, it was comparable to effects seen in other studies without feedback[13]. Finally,

[11] That is, if speakers are sensitive in some way to the potential for communication difficulties, and inhibit productions that are likely to give rise to them—functions that could be attributed to either monitoring and teleological reasoning, or simple error-based learning—then they should tend to inhibit productions that would be less contextually appropriate, and this should generally mean inhibiting alternative languages more than alternative dialects. This lack of need to inhibit a non-target dialect may, in turn, provide a clue as to why dialect use may fail to show language-like inhibition effects (e.g. Melinger, 2018): if the lack of semantic inhibition in cross-language picture-word interference is attributed to the habitual engagement of cross-language inhibition, which somehow eliminates competition, then a lack of cross-dialect inhibition could allow interference to proceed normally.

[12] In metanalyses comparing naming latencies for critical items in Session 1 of the current study to those for the same items in Oppenheim's (in prep) single-session norms and Balatsou et al's (in revision) Session 1 norms, the interactions of Block with Experiment do not approach significance (ps $> .6$).

[13] In similar metanalyses comparing omission rates for critical items in Session 1 of the current study to those for the same items in Oppenheim's (in prep) single-session norms and

if participants in Session 1 had taken a strategy of searching for subordinate names for pictures—rather than specifically activating Americanisms, then we might have seen a generalized increase in the use of non-dominant picture names, but our results gave no indication of such an increase. Thus, the dialect-based generalization in Session 1 appears consistent with the engagement of a quick, dialect-specific priming mechanism that links the comprehension of a non-native dialect to increased activation of that dialect in production. As we will discuss below, in normal conversation this mechanism should produce interlocutor dialect alignment.

The second session, days later, provides evidence for dialect-based generalization within the production system itself, again implying the existence and use of a prime-able dialect representation. Here, we had instructed participants to name each picture using the term that they had previously received as feedback, an approach that others have previously validated without dialect manipulations (Alario et al., 2004). When naming pictures that were previously paired with *non-native-dialect feedback*, participants did produce those names more often throughout Session 2, but that does not actually prove dialect-based generalization. Instead, the question of generalization hinges on participants' performance when naming the pictures that we had previously paired with *native-dialect feedback*: because participants had never encountered their *non-native-dialect* names in this experiment, (incorrectly) choosing them would specifically support the claim of dialect-based generalization. And this is precisely what we saw in Session 2: participants gradually, increasingly, offered our expected non-native-dialect names for these pictures, thus supporting the generalization claim.

Importantly, the lack of feedback in Session 2 allows us to attribute that generalization to a mechanism by which retrieving one part of a non-native dialect durably facilitates retrieving other parts of the dialect. That is, we can rule out memory encoding or passive decay as its source. We saw no evidence that participants had tagged the previous session as a whole as involving a distinct dialect: comparing performance at the onset of Session 2 to the Session 1 baseline, although speakers were more likely to use Americanisms when they had received US-dialect feedback for an item, they were also more likely to use Britishisms when they had received UK-dialect feedback instead. Thus, the item-specific lexical labels clearly survived the thirty or so hours that intervened between the first session and the second, and any dialect-tagging of the episode as a whole was, at least, not powerful enough to create false memories (cf Roediger & McDermott, 1995). Moreover, if decay alone had caused a drop in accuracy to 74% after 30 hours, then assuming a power function (e.g. Wixted & Ebbesen, 1991) it would be unlikely to cause a further 11% drop (to 63%) in the next 30 minutes. Thus, the generalization over the course of Session 2 must instead be attributed to an active process, and our current findings indicate that the process of producing a non-native dialect word specifically promotes the retrieval

Balatsou et al's (in revision) Session 1 norms, the interactions of Block with Experiment do not approach significance (ps > .2) of more words from its dialect.

Taken together, the observed dialect-based generalization in the two sessions implies the existence and use of two related, but possibly distinct, mechanisms, that we argue may have parallels in bilingual language use. First, within the production system, speakers generalize, such that producing one word from a non-native dialect increases the accessibility of other words from the same dialect. Such production-based priming would seem to provide a sort of momentum in production, helping speakers remain within a dialect with minimal need for inhibition.[14] But, second, for a speaker to use the appropriate dialect for a given situation, it needs to be primable through their interactions with their environment, implying a role for the comprehension system as well. Here, perceiving a word paired with a picture did not merely prime perceiving that word again, or perceiving that word/picture combination again, but actually 1.) increased that association's accessibility *for production* (i.e. transferring from comprehension to production) and 2.) generalized to prime the production of other terms from the same dialect. This may seem a small point—of course comprehension must somehow influence production—but usefully adds to evidence for tight links between the systems (Dell & Chang, 2014; Pickering & Garrod, 2012). Finally, it is important to note that although our paradigm accomplished this priming through the highly artificial mechanism of computerized lexical feedback, speakers show analogous tendencies to adopt their interlocutor's terminology in spontaneous communication (i.e. lexical alignment). In the sections that follow, we consider a framework for bidialectal lexical-semantic representations. We also suggest how the mechanisms identified here might contribute to both bidialectal and bilingual language production.

Dialect representation

As a framework for understanding how speakers might represent dialect-marked words—and how this might relate to both simple synonyms and cross-language translation equivalents—we can start with the assumption that word representation incorporates at least two kinds of information about any named entity. First, it includes the concept's semantic features or taxonomic information (e.g. a dog is a furry, quadrupedal member of the animal category). Second, it includes fine-grained contextual/sequential knowledge of how a particular word is used—not just grammatical category, but the kinds of lexical cooccurrences and transitional probabilities that can be leveraged as the basis for distributional semantics (e.g. Firth, 1957; Mikolov, Chen, Corrado, & Dean, 2013), and might be represented as connections in a simple recurrent network (e.g., Elman, 1990).[15]

[14] On the assumption that the non-native dialect is more marked than the native dialect, it might follow that such momentum should also asymmetrically help a speaker to switch into it, by allowing one production to prime the next.

[15] Although the methods are better developed for assessing linguistic context, the notion of context can easily incorporate non-linguistic context as well.

Within this framework, we can assume that cross-language translation equivalents should be mapped onto the same semantic feature space. That is, whether 'dog' and 'ci' (i.e. its Welsh translation) share all of their semantic features or merely most of them (Van Hell & de Groot, 1998), they should both be linked to the same *furriness* feature rather than having a separate copy for each language. In terms of distributional semantics, however, cross-language translation equivalents should occupy essentially disjoint spaces.[16] That is, although sentences like "The dog sits on the couch," and, "Mae'r ci yn eistedd ar y gadair," refer to the same entities in the same order, the surface lexical context of 'dog' has no elements in common with that of 'ci'. Within-language synonyms should, by contrast, share not only their featural semantics but occur in essentially the same contexts. In the American English sentence, "The dog sits on the couch," the word 'sofa'[17] can be substituted for 'couch' without changing its meaning or acceptability. Although many synonyms may fall short of this Platonic ideal (e.g. some might argue that couches are less formal than sofas, and the linguistic contexts of 'give' and 'donate' do differ), it is usually what we talk about when we talk about synonymy. Between these two poles lie cross-dialect translation equivalents. Like both cross-language translation equivalents and within-language synonyms, they share most or all of their featural representations. Their distributional semantics, however, will slightly differ: more so than for within-language synonyms, but less so than for cross-language translation equivalents.[18]

On its surface, something like a simple recurrent network (Elman, 1990) should accommodate both aspects of lexical knowledge in production: a meaningword mapping (Oppenheim, Dell, & Schwartz, 2010) would activate all forms (e.g. both $\text{shovel}_{\text{US}}$ and spade_{UK}, as well as $\text{rhaw}_{\text{Welsh}}$), and converging activation from a context layer would bias selection toward a form that better matches the lexical context (for instance the transitional probability of yrhaw is much greater than the transitional probability of yspade). Such an architecture would ease the selection of appropriate forms in continuous discourse, at least for languages or dialects that are more distinctive. But because dialects are defined by the sparsity of their differences (dialect-marked words are often rare in discourse, and in our experiment they occurred only about every ten trials), less is to be gained from tracking such superficial transitions (cf Marcus, 1998). In fact, the particular sequences that participants encountered in our experiment—hundreds of unrelated nouns—were unlikely to resemble any that they would have encountered previously. To accomplish such long-distance dialect priming, and apply it to novel contexts, speakers must represent and use dialect membership

[16] Excepting cognates, of course.

[17] The strongly preferred term in British English, and an equifrequent synonym in American English.

[18] Writing a recipe for ratatouille, for example, one would use most of the same words but deciding to add an $\text{eggplant}_{\text{US}}$ rather than an $\text{aubergine}_{\text{UK}}$, would also commit one to adding $\text{zucchinis}_{\text{US}}$ rather than $\text{courgettes}_{\text{UK}}$, and $\text{mincing}_{\text{US}}$ the garlic rather than $\text{chopping}_{\text{UK}}$ it. Indeed, there is some humorous evidence that dialect users are sensitive to this lexical non-independence (e.g. https://www.theguardian.com/lifeandstyle/wordofmouth/2010/jul/15/how-to-make-perfect-ratatouille#comments).

on a more abstract level, as illustrated in Figure 5, and we suggest that the same functions may also contribute to bilingual language production.

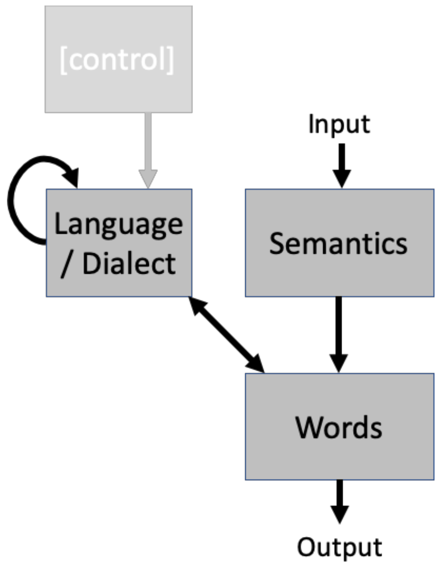

Figure 5: A simple illustration of how dialect use might be integrated into a minimal architecture for word production (e.g. Oppenheim, Dell, & Schwartz, 2010). Word retrieval in production is typically cast as a spread of activation from meaning to words, without language bias, followed by a selection process that might impose coarse constraints on language, dialect, or part of speech (Dell, Oppenheim, & Kittredge, 2008; Runnqvist et al., 2012). Deliberately imposing a language or dialect constraint (e.g. in cued switching) involves routing activation through a language or dialect 'tag'. However, this tag is itself a bottom-up-primable representation, whose activation persists over time in a way that lexical, semantic, or part-of-speech activation does not.

In this model, we assume that word retrieval for production remains largely driven by a spread of semantic activation (based on Oppenheim et al, 2010, as described above), but also incorporates activation from language- or dialect-specific bias nodes. These bias nodes do not require an all-or-nothing switch between dialects (or languages); as a bias node becomes more activated, it simply contributes more to meaning-based word selection. Based on our Session 2 data, which showed priming from participants' own productions, we suggest that these bias nodes are bi-directionally connected to the lexical representations, allowing activation from the lexical level to feed back to the dialect representations, within the production system. Functioning normally as a discourse-level tag, the activation of the bias nodes must be relatively persistent—certainly more persistent than the rapidly dissipating activation patterns that are typically assumed for lexical or syntactic representations (e.g. Bock & Griffin, 2000) or otherwise accomplished via 'check-off' mechanisms (e.g. Dell, 1986). We assume that the connections from both the semantic representations and the bias nodes are learned through experience (e.g. Oppenheim et al., 2010), with the bias nodes themselves emerging through repeated interactions in distinctive contexts (see e.g. Jones, Kuipers, Nugent, Miley, & Oppenheim, 2018, for evidence that people track context in paired associate learning). In addition, although we do not clearly demonstrate it in this study, we assume that these same language or dialect bias nodes can take more deliberate input from a cognitive control process (e.g. the 'booster' in Oppenheim et al, 2010, or the 'syntactic traffic cop' of Dell, Oppenheim, & Kittredge, 2008 and Gordon & Dell, 2003), thereby allowing their engagement by cued switching paradigms (e.g. Kirk et al., 2018). Although aspects of this model are admittedly quite tentative, and the model itself is clearly incomplete, we believe it provides a useful step in describing more explicitly how dialect or language tagging might be represented and how it might contribute to production.

Conclusion

Language rapidly adapts to context. Previous studies show that part of this adaptation involves negotiating specific terms for known entities (lexical alignment). At a higher level, we have shown that speakers can rapidly adjust their word choices to accommodate sparse encounters with a non-native dialect, generalizing to novel exemplars (dialect alignment). They also retain word-specific biases across encounters, and although they do not seem to holistically associate interactions with particular dialects, retrieving these words in the future can induce a generalized dialect switch. We have briefly proposed a model that accounts for this generalization: by representing a dialect tag as a primable representation it persistently supports the retrieval of appropriate forms in production.

References

Abutalebi, J., & Green, D. (2007). *Bilingual language production: The neurocognition of language representation and control. Journal of Neurolinguistics* (Vol. 20). https://doi.org/10.1016/j.jneuroling.2006.10.003

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140–155. https://doi.org/10.3758/BF03195559

Anderson, J. A. E., Chung-Fat-Yim, A., Bellana, B., Luk, G., & Bialystok, E. (2018). Language and cognitive control networks in bilinguals and monolinguals. *Neuropsychologia*, *117*(June), 352–363. https://doi.org/10.1016/j.neuropsychologia.2018.06.023

Balatsou, E., Fischer-Baum, S., & Oppenheim, G. M. (in revision). The psychological reality of picture name agreement.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227.

Bates, D. M., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … Fox, J. (2019). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-21, https://CRAN.R-project.org/package=lme4.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380. https://doi.org/10.3758/BF03196494

Begg, C. B., & Gray, R. (1984). Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions. *Biometrika*, *71*(1), 11. https://doi.org/10.2307/2336391

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2012). Effects of phonological feedback on the selection of syntax: Evidence from between-language syntactic priming. *Bilingualism*, *15*(3), 503–516. https://doi.org/10.1017/S1366728911000162

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, *19*(1), 19–23. https://doi.org/10.1177/0963721409358571

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177–192.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, *75*(2), B13-25. https://doi.org/https://doi.org/10.1016/S0010-0277(99)00081-5

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Cai, Z. G., Pickering, M. J., Yan, H., & Branigan, H. P. (2011). Lexical and syntactic representations in closely related languages: Evidence from Cantonese-Mandarin bilinguals. *Journal of Memory and Language*, *65*(4), 431–445. https://doi.org/10.1016/j.jml.2011.05.003

Cave, C. B. (1997). Very Long-Lasting Priming in Picture Naming. *Psychological Science*, *8*(4), 322–325. https://doi.org/10.1111/j.1467-9280.1997.tb00446.x

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical Selection in Bilinguals: Do Words in the Bilingual's Two Lexicons Compete for Selection? *Journal of Memory and Language*, *41*(3), 365–397. https://doi.org/10.1006/jmla.1999.2651

Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, *94*(1), 94–103. https://doi.org/10.1016/j.bandl.2004.12.002

de Bot, K. (1992). A bilingual production model: Levelt's "speaking" model approach adapted. *Applied Linguistics*, *13*(1), 1–24. https://doi.org/10.1093/applin/13.1.1

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin and Review*, *22*(6), 1630–1645. https://doi.org/10.3758/s13423-015-0836-1

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*.

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*, *23*(4), 583–608. https://doi.org/10.1080/01690960801920735

Dylman, A. S., & Barry, C. (2018). When having two names facilitates lexical selection: Similar results in the picture-word task from translation distractors in bilinguals and synonym distractors in monolinguals. *Cognition*, *171*(June 2016), 151–171. https://doi.org/10.1016/j.cognition.2017.09.014

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Finkbeiner, M., Gollan, T. H., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism*, *9*(2), 153–166. https://doi.org/10.1017/S1366728906002501

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*.

Foulkes, P., & Hay, J. B. (2015). The Emergence of Sociophonetic Structure. In *The handbook of language emergence* (Vol. 87, p. 292). Wiley Online Library.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218. https://doi.org/10.1016/0010-0277(87)90018-7

Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, *27*(1), 1–40.

Green, D. W. (1986). Control, activation, and resource: a framework and a model for the control of speech in bilinguals. *Brain and Language*, *27*(2), 210–223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67–81. https://doi.org/10.1017/S1366728998000133

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(July 2015), 1–16. https://doi.org/10.1080/20445911.2013.796377

Hazen, K. (2001). *Teaching about dialects.* Washington, DC: ERIC Clearinghouse on Languages and Linguistics.

Hermans, D., Bongaerts, T., de Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*(3), 213–229. https://doi.org/10.1017/s1366728998000364

Johnston, R. a, Dent, K., Humphreys, G. W., & Barry, C. (2010). British-English norms and naming times for a set of 539 pictures: the role of age of acquisition. *Behavior Research Methods*, *42*(2), 461–469. https://doi.org/10.3758/BRM.42.2.461

Jones, M. W., Kuipers, J., Nugent, S., Miley, A., & Oppenheim, G. M. (2018). Episodic traces and statistical regularities : Paired associate learning in typical and dyslexic readers. *Cognition*, *177*(April), 214–225. https://doi.org/10.1016/j.cognition.2018.04.010

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching-a review. *Psychological Bulletin*, *136*(5), 849–874. https://doi.org/10.1037/a0019842

Kirk, N. W., Kempe, V., Scott-Brown, K. C., Philipp, A. M., & Declerck, M. (2018). Can monolinguals be like bilinguals? Evidence from dialect switching. *Cognition*, *170*(March 2016), 164–178. https://doi.org/10.1016/j.cognition.2017.10.001

Kleinman, D., & Gollan, T. H. (2016). Speaking Two Languages for the Price of One: Bypassing Language Control Mechanisms via Accessibility-Driven Switches. *Psychological Science*, *27*(5), 700–714. https://doi.org/10.1177/0956797616634633

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. *Handbook of Bilingualism: Psycholinguistic Approaches*, 289–307.

Labov, W. (1998). Co-existent systems in African-American vernacular English. *African-American English: Structure, History and Use*, (August), 110–153.

Léwy, N., & Grosjean, F. (2000). Bilingual model of lexical access: A connectionist model of bilingual spoken word recognition. *Unpublished Manuscript, Laboratoire de Traitement Du Langage et de La Parole, Université de Neuchâtel, Neuchâtel, Switzerland*.

Li, P. (1996). Spoken word recognition of code-switched words by Chinese--English bilinguals. *Journal of Memory and Language*, *35*(6), 757–774.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, *37*(3), 243–282.

Marian, V., Blumfield, H., & Kaushanskaya, M. (2007). Bilinguals and Multilinguals. *Hearing Research*, *50*(August 2007), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: do all roads lead to Rome? *Cognitive Psychology*, *32*(1), 49–96. https://doi.org/10.1006/cogp.1997.0645

Melinger, A. (2018). Distinguishing languages from dialects : A litmus test using the picture-word interference task. *Cognition*, *172*(November), 73–88. https://doi.org/10.1016/j.cognition.2017.12.006

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Nips*, 1–9. https://doi.org/10.1162/jmlr.2003.3.4-5.951

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007

Oppenheim, G. M. (in prep.). On the role of endogenous competition in normal word production: evidence from timed picture naming norms for British English.

Peirce, J. (2010). PsychoPy - Psychology software for Python. *Integration The Vlsi Journal*.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, *27*(2), 169–190; discussion 190-226. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2012). An integrated theory of language production and comprehension. *The Behavioral and Brain Siences*, *36*(c), 329–392. https://doi.org/10.1017/S0140525X12001495

Roediger, H. L., & McDermott, K. B. (1995). Creating False Memories: Remembering Words Not Presented in Lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roelofs, A., Piai, V., & Rodriguez, G. G. (2011). Attentional inhibition in bilingual naming performance: Evidence from delta-plot analyses. *Frontiers in Psychology*, *2*(AUG), 1–10. https://doi.org/10.3389/fpsyg.2011.00184

Runnqvist, E., Strijkers, K., Alario, F.-X., & Costa, A. (2012). Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language*, *66*, 850–869. https://doi.org/10.1016/j.jml.2012.02.007

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K. D., Herron, D., Iyer, G., . . . Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 621–633. https://doi.org/10.3758/BF03195542

Tyler, M. D., Tyler, L., & Burnham, D. K. (2005). The delayed trigger voice key: an improved analogue voice key for psycholinguistic research. *Behavior Research Methods*, *37*(1), 139–147.

Van Hell, J. G., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193–211. https://doi.org/10.1017/s1366728998000352

Wixted, J. T., & Ebbesen, E. B. (1991). On the Form of Forgetting. *Psychological Science*, *2*(6), 409–415. https://doi.org/10.1111/j.1467-9280.1991.tb00175.x

Appendix A. Characteristics of critical items.

Please note that name agreement statistics drawn from Szekely et al. (2003) and Johnston et al. (2010) have been recalculated to include non-responses as negative evidence; in the original papers, non-responses were excluded before calculating name agreement.

IPNP item code	Dominant US name (target Americanism) [a]	Frequency of the dominant US name in US norms [a]	Frequency of the dominant US name in UK norms [b]	Target Britishism	Frequency of the specified UK name in UK norms [b]	Spoken UK norms [b]		Written UK norms [c]		
						Dominant spoken name [b]	Name agreement [b]	Dominant written name [c]	Written name agreement [c]	Spoken name agreement [c]
obj003	airplane	0.70	0.01	aeroplane	0.48	*plane*	0.50	aeroplane	0.52	0.60
obj004	alligator	0.90	0.12	crocodile	0.86	crocodile	0.86	crocodile	0.84	0.80
obj019	stroller	0.46	0.02	pram	0.80	pram	0.80	pram	0.87	0.88
obj020	backpack	1.00	0.71	rucksack	0.12	*backpack*	0.71	rucksack	0.81	0.28
obj027	bandaid	0.92	0.05	plaster	0.93	plaster	0.93	plaster	1.00	0.96
obj033	bathtub	0.78	0.09	bath	0.86	bath	0.86	bath	0.94	1.00
obj039	bug	0.44	0.22	beetle	0.60	beetle	0.60	beetle	0.87	0.68
obj076	cane	0.92	0.42	walking stick	0.43	walking stick	0.43	walking stick	0.81	0.76
obj082	carousel	0.58	0.43	merry-go-round	0.43	*carousel*	0.43	merry go round	0.52	0.52
obj100	clothespin	0.48	0.02	peg	0.78	peg	0.78	peg	0.77	0.84
obj104	dime	0.60	0.00	coin	0.84	coin	0.84	coin	0.97	0.72
obj107	cookie	0.74	0.09	biscuit	0.58	biscuit	0.58	biscuit	0.74	0.36
obj110	corn	1.00	0.56	sweetcorn	0.35	*corn*	0.56	sweetcorn	0.45	0.48
obj115	crib	0.82	0.08	cot	0.49	cot	0.49	cot	0.94	0.64
obj125	diaper	0.46	0.07	nappy	0.30	*pants*	0.34	nappy	1.00	0.48
obj136	dresser	0.48	0.02	drawers	0.54	drawers	0.54	*chest of drawers*	0.61	0.40
obj150	faucet	0.82	0.01	tap	0.95	tap	0.95	tap	1.00	0.96
obj156	firetruck	0.62	0.34	fire engine	0.51	fire engine	0.51	fire engine	0.84	0.80
obj158	fishing pole	0.50	0.04	fishing rod	0.76	fishing rod	0.76	fishing rod	0.74	0.60
obj160	flashlight	0.96	0.18	torch	0.71	torch	0.71	torch	1.00	0.96
obj167	football	1.00	0.25	rugby ball	0.45	rugby ball	0.45	rugby ball	0.65	0.84
obj173	trash	0.42	0.05	rubbish	0.65	rubbish	0.65	rubbish	0.61	0.76
obj174	gas	0.38	0.04	petrol pump	0.23	*petrol*	0.30	petrol pump	0.87	0.52
obj192	hamburger	0.84	0.13	burger	0.84	burger	0.84	burger	0.71	0.72
obj197	hanger	0.88	0.52	coat hanger	0.46	*hanger*	0.52	coathanger	0.61	0.56
obj215	ice cream cone	0.50	0.01	ice cream	0.99	ice cream	0.99	ice cream	0.87	0.92
obj222	puzzle	0.98	0.36	jigsaw	0.59	jigsaw	0.59	jigsaw	0.77	0.60
obj223	jump rope	0.84	0.02	skipping rope	0.66	skipping rope	0.66	skipping rope	0.97	0.88
obj233	ladybug	0.64	0.20	lady bird	0.31	lady bird	0.31	lady bird	0.81	0.76
obj245	light switch	0.64	0.43	switch	0.47	switch	0.47	switch	0.52	0.36
obj255	mailbox	0.84	0.35	post box	0.27	*mailbox*	0.35	post box	0.42	0.40
obj264	mixer	0.36	0.11	whisk	0.78	whisk	0.78	whisk	0.65	0.72
obj265	priest	0.40	0.27	monk	0.52	monk	0.52	monk	0.81	0.44
obj270	mosquito	0.50	0.17	daddy long legs	0.16	*fly*	0.21	daddy long legs	0.58	0.24
obj271	motorcycle	0.96	0.11	motorbike	0.61	motorbike	0.61	motorbike	0.84	0.76
obj291	package	0.94	0.24	parcel	0.62	parcel	0.62	parcel	0.87	0.88
obj298	pants	0.86	0.12	trousers	0.80	trousers	0.80	trousers	0.94	1.00
obj324	pitcher	0.52	0.00	jug	0.89	jug	0.89	jug	1.00	0.96
obj325	pitchfork	0.62	0.34	fork	0.43	fork	0.43	fork	0.48	0.72
obj333	popsicle	0.64	0.02	ice lolly	0.28	ice lolly	0.28	ice lolly	0.32	0.12
obj340	purse	0.98	0.06	bag	0.60	bag	0.60	bag	0.45	0.36
obj352	refrigerator	0.88	0.03	fridge	0.95	fridge	0.95	fridge	0.77	0.88
obj364	rooster	0.54	0.11	cockerel	0.13	*chicken*	0.64	cockerel	0.48	0.20
obj372	sailboat	0.76	0.10	yacht	0.10	*boat*	0.69	yacht	0.39	0.12
obj395	shovel	0.98	0.42	spade	0.51	spade	0.51	spade	0.94	0.92
obj403	sled	0.96	0.20	sledge	0.54	sledge	0.54	sledge	0.84	0.56
obj405	slingshot	0.74	0.58	catapult	0.14	*slingshot*	0.58	catapult	0.48	0.32
obj412	couch	0.74	0.12	sofa	0.75	sofa	0.75	sofa	0.90	0.84
obj426	stove	0.72	0.14	cooker	0.34	*oven*	0.43	cooker	0.58	0.40
obj428	stroller	0.74	0.03	pushchair	0.20	*pram*	0.62	pushchair	0.52	0.44
obj433	sweater	0.52	0.05	jumper	0.87	jumper	0.87	jumper	0.84	0.92
obj436	needle	0.60	0.48	syringe	0.27	*needle*	0.48	syringe	0.68	0.60
obj450	thermos	0.80	0.11	flask	0.67	flask	0.67	flask	0.81	0.76
obj462	top	0.72	0.02	spinning top	0.44	spinning top	0.44	spinning top	0.94	0.68
obj464	railroad tracks	0.28	0.01	railway tracks	0.17	*train tracks*	0.28	railway tracks	0.19	0.16
obj466	stoplight	0.62	0.01	traffic lights	0.88	traffic lights	0.88	traffic lights	0.65	0.76
obj468	trashcan	0.68	0.08	bin	0.70	bin	0.70	*dustbin*	0.65	0.24
obj472	truck	0.96	0.59	lorry	0.37	*truck*	0.59	lorry	0.77	0.84
obj476	turtle	1.00	0.54	tortoise	0.46	*turtle*	0.54	tortoise	0.77	0.72
obj482	vacuum	0.82	0.18	hoover	0.75	hoover	0.75	hoover	0.61	0.64
obj484	vest	0.96	0.19	waistcoat	0.63	waistcoat	0.63	waistcoat	1.00	0.80
obj488	wagon	0.62	0.04	cart	0.24	cart	0.24	cart	0.52	0.12
obj494	closet	0.86	0.09	wardrobe	0.81	wardrobe	0.81	wardrobe	0.94	0.96
obj517	wrench	0.84	0.19	spanner	0.53	spanner	0.53	spanner	0.97	0.92
obj520	zipper	0.96	0.03	zip	0.94	zip	0.94	zip	0.97	0.92

Sources:

[a] Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., … Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 621–633. http://doi.org/10.3758/BF03195542

[b] Oppenheim, G. M. (in prep.). On the role of endogenous competition in normal word production: evidence from timed picture naming norms for British English.

[c] Johnston, R. A., Dent, K., Humphreys, G. W., & Barry, C. (2010). British-English norms and naming times for a set of 539 pictures: the role of age of acquisition. *Behavior Research Methods*, *42*(2), 461–469. http://doi.org/10.3758/BRM.42.2.461

Appendix B. Result summary by block.

By-block outcome and RT summary for all 27,040 trials. Items with ‘Neutral’ Session 1 feedback are the filler trials, reported here as an informal baseline.

Session 1 feedback	Session	Block	Used target UK name %	Used target UK name RT	Used target US name %	Used target US name RT	Used any other name %	Used any other name RT	No response	Excluded	US:UK ratio
Neutral	1	1	80.8% (1860)	956 (8)	-	-	15.6% (358)	1260 (28)	3.6% (84)	2.6% (62)	-
		2	80.7% (1870)	1000 (9)	-	-	15.1% (351)	1294 (30)	4.1% (96)	2.1% (49)	-
		3	79.9% (1844)	1016 (9)	-	-	14.8% (341)	1354 (30)	5.4% (124)	2.5% (60)	-
		4	80.9% (1868)	1028 (9)	-	-	14.1% (326)	1391 (29)	4.9% (114)	2.4% (56)	-
		5	79.6% (1848)	1034 (9)	-	-	14.3% (332)	1411 (30)	6.2% (143)	1.9% (44)	-
	2	1	88.0% (2068)	894 (7)	-	-	7.8% (183)	1110 (32)	4.2% (99)	0.8% (19)	-
		2	90.7% (2126)	915 (7)	-	-	6.6% (154)	1217 (42)	2.7% (64)	0.9% (21)	-
		3	89.6% (2100)	944 (8)	-	-	7.2% (168)	1276 (41)	3.3% (77)	0.9% (21)	-
		4	89.5% (2103)	958 (8)	-	-	7.5% (177)	1268 (45)	3.0% (71)	0.8% (20)	-
		5	90.0% (2106)	954 (8)	-	-	6.3% (148)	1176 (39)	3.7% (86)	0.8% (19)	-
UK	1	1	57.5% (188)	1051 (29)	18.3% (60)	1172 (61)	20.8% (68)	1234 (56)	3.4% (11)	3.8% (13)	0.319
		2	55.9% (179)	1094 (31)	17.8% (57)	1117 (65)	21.6% (69)	1398 (66)	4.7% (15)	5.3% (18)	0.318
		3	-	-	-	-	-	-	-	-	-
		4	-	-	-	-	-	-	-	-	-
		5	-	-	-	-	-	-	-	-	-
	2	1	74.2% (95)	1133 (45)	4.7% (6)	1035 (203)	10.2% (13)	1789 (183)	10.9% (14)	0.8% (1)	0.063
		2	67.6% (50)	1022 (57)	12.2% (9)	904 (107)	14.9% (11)	1290 (166)	5.4% (4)	2.6% (2)	0.18
		3	77.6% (142)	1096 (41)	10.9% (20)	1039 (78)	7.1% (13)	1203 (128)	4.4% (8)	1.1% (2)	0.141
		4	66.7% (78)	964 (35)	17.1% (20)	954 (54)	15.4% (18)	1076 (75)	0.9% (1)	-	0.256
		5	63.3% (107)	935 (30)	20.7% (35)	1012 (52)	14.8% (25)	1187 (115)	1.2% (2)	1.2% (2)	0.327
US	1	1	-	-	-	-	-	-	-	-	-
		2	-	-	-	-	-	-	-	-	-
		3	54.5% (177)	1164 (36)	17.8% (58)	1156 (58)	22.5% (73)	1288 (66)	5.2% (17)	3.0% (10)	0.328
		4	53.8% (176)	1093 (29)	18.3% (60)	1171 (61)	19.9% (65)	1385 (63)	8.0% (26)	3.8% (13)	0.341
		5	50.5% (165)	1161 (37)	23.2% (76)	1297 (61)	20.2% (66)	1333 (65)	6.1% (20)	3.0% (10)	0.461
	2	1	45.5% (92)	906 (34)	34.7% (70)	1000 (34)	18.3% (37)	1161 (73)	1.5% (3)	1.9% (4)	0.761
		2	41.9% (108)	922 (30)	39.5% (102)	1094 (39)	15.9% (41)	1182 (77)	2.7% (7)	1.9% (5)	0.944
		3	40.9% (61)	1087 (52)	38.3% (57)	1185 (56)	13.4% (20)	1209 (109)	7.4% (11)	2.6% (4)	0.934
		4	48.6% (104)	1169 (47)	34.1% (73)	1281 (56)	13.1% (28)	1536 (126)	4.2% (9)	0.9% (2)	0.702
		5	48.0% (83)	1185 (48)	30.1% (52)	1206 (77)	16.2% (28)	1580 (107)	5.8% (10)	0.6% (1)	0.627

Appendix C. Naming latency analyses.

Mean naming latencies in each session were roughly in line with those reported for similar procedures without corrective feedback (e.g. Székely et al., 2003a). Briefly, naming latencies in such studies reliably show three basic features that we would expect here: First, although naming latencies are affected by many factors, they tend to average around 1000ms in no-feedback versions of this task.[19] Second, speakers typically produce dominant names faster than non-dominant names. Third, naming latencies typically increase over the course of a session, a pattern that others have attributed to simple fatigue.

We confirmed these patterns via a maximal linear mixed effects regression of inverse-transformed RTs for target Britishisms and Americanisms, that included centered fixed effects for Dialect{UK,US}, Session{1,2}, Block{1:5} and their interactions, plus uncorrelated random slopes and intercepts for subjects and items. As expected, participants were slower to produce Americanisms than Britishisms ($\beta_{\text{main effect of Dialect}}$=6.24, SE=2.44, t=2.56, p=0.01), slowed over the course of each session ($\beta_{\text{main effect of Block}}$=2.29, SE=0.48, t=4.80, p<0.001), and were faster in the second session than in the first ($\beta_{\text{main effect of Session}}$=-8.09, SE=2.21, t=-3.65, p<0.001). No interactions, however, approached significance (all p>.45). Considering the prior expectations, derived from similar tasks with the same stimuli, the response times provide no clear evidence that the presence of feedback, occasionally dialect-marked, particularly affected the picture-naming process.

[19]Note also that the addition of feedback necessarily increases the effective interstimulus interval (ISI), which can itself increase RTs.