

# Egocentric and allocentric learning of social-indexical meaning in American English, Datooga, and Murrinhpatha

**Péter Rác** ([peter.racz@bristol.ac.uk](mailto:peter.racz@bristol.ac.uk))

Department of Anthropology and Archaeology  
University of Bristol

**Alice Mitchell** ([alice.mitchell@bristol.ac.uk](mailto:alice.mitchell@bristol.ac.uk))

Department of Anthropology and Archaeology  
University of Bristol

**Joe Blythe** ([joe.blythe@mq.edu.au](mailto:joe.blythe@mq.edu.au))

Department of Linguistics  
Macquarie University

## Abstract

We address competing perspectives on how social-indexical meaning is learned in language, using data from artificial language learning experiments and two studies in small-scale societies. Our results indicate that learning social-indexical meaning is primarily allocentric as opposed to egocentric: speaker success in learning a social-indexical meaning pattern depends on overall exposure to the pattern more than the pattern's relative importance to the speaker. We base these claims on data from American English-speaking adults, Datooga-speaking children, as well as adults and children speaking Murrinhpatha. The results highlight the importance of widening the sample of methods and data sources in studying how variation in language is learned and maintained.

**Keywords:** language learning, variation, American English, Datooga, Murrinhpatha

## Learning social-indexical meaning

Social-indexical meaning in language refers to the non-linguistic context, such as the addressee or the setting. Such meanings together constitute the sociolinguistic norms of a community (Eckert, 2008).

In this paper, we compare *egocentric* and *allocentric* perspectives on learning social-indexical meaning in language. The distinction comes from the spatial memory literature, indicating (to put it simply) directions that are relative or absolute to the individual (e.g. 'right/left' versus 'North/South') (Burgess, 2006). In terms of social-indexical meaning, an egocentric learner attributes more importance to patterns that are more relevant to them. In this case, individuals who rely more on social-indexical variation will show more sensitivity to it. In contrast, allocentric learning is unweighted statistical learning, which means that learning success is determined by exposure to the pattern. This, in turn, will depend on the pattern's prevalence in the community – not its relevance to the speaker.

This distinction comes to the fore in classic variationist sociolinguistic discussions of language change above and below the level of consciousness (Labov, 2001), and the robust (though, here, simplified) observation that women are more likely to adopt both innovative variants 'from below' and overtly standard social-indexical variants 'from above' – the so-called 'gender paradox'.

The allocentric explanation of this pattern is that women in the West are more likely to be well-connected and upwardly mobile, and thus more exposed to sociolinguistic variation in a wider range of fine-grained interactions. As a result, they are more flexible in using social-indexical meaning in language (Milroy, 1980).

The egocentric explanation is that women are both more likely to be policed for their language use and rely more on social capital (Lakoff, 1973). As a result, they are more sensitive to social-indexical meaning. The difference between the two explanations lies in the emphasis on exposure versus sensitivity.

The problem is that, in observational data, exposure and sensitivity can be conflated, and so these explanations are difficult to distinguish from each other. Pressures on sociolinguistic variation are easier to capture during language learning.

However, the learning of social-indexical meaning in language is understudied (Foulkes, 2010). We know children are sensitive to it from an early age (Smith, Durham, & Richards, 2013; Samara, Smith, Brown, & Wonnacott, 2017), but the main body of sociolinguistic work comes from surveys of established conventions used by adults in large-scale Western societies (Labov, 2001).

At the same time, different large-scale societies, such as Arabic-speaking countries, show a different picture of gendered linguistic variation (Bakir, 1986). What is more, small-scale societies function on different scales and result in very different cognitive profiles than large-scale societies (Henrich, Heine, & Norenzayan, 2010). This means that a comprehensive study of sociolinguistic variation will benefit from a broader range of sources and methods.

Here, we rely on psycholinguistic and anthropological methods to argue for the primacy of allocentricity in learning social-indexical meaning in language. We do so by using limiting cases from different settings: American English speakers exposed to a controlled artificial language, Datooga, a language with categorical, gendered speaker variation, and Murrinhpatha, a language with categorical, context-specific

word variation.

## Aims

We draw on three case studies to look at egocentric versus allocentric learning of social-indexical meaning. Our aims are to investigate how social-indexical meaning is learned and generalised in three different settings and to show that *allocentric learning takes primacy*: learner accuracy reflects the overall prevalence of a social-indexical pattern as opposed to its relevance to the individual (or the individual's subgroup).

## Artificial language learning with American English-speaking adults

**Design** We used a series of short games, played on Amazon Mechanical Turk, to train American English-speaking adults on a simple suffixation pattern in an artificial language (AL). In the game, the participant guides a bird through a series of rooftops to its nest. Each rooftop has a riddle that the bird has to answer in order to proceed.

In training, participants see a nonce word and two possible suffixed variants, as well as a 'conversation partner'. They have to choose one suffixed variant. The nonce words name objects and the suffixed forms denote the diminutive variant of the object. The correct choice depends on the conversation partner.

Participants see six words with four conversation partners. They receive feedback on whether their choice of a suffix was correct. Training is followed by test. Here, in addition to the previously seen words and partners, they see six additional words and four additional partners<sup>1</sup>. The layout of a training and a test trial can be seen in Figure 1. The order of images is randomised for each participant, and the words and suffixes are randomly selected from a set of nonce syllables.

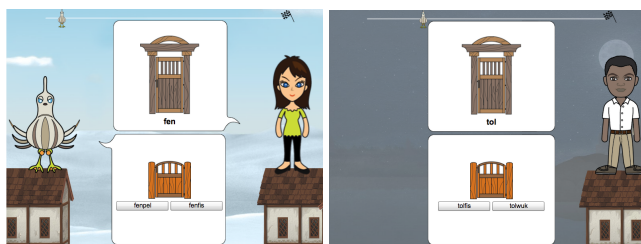


Figure 1: Layout of the task: training is daytime (left), test is nighttime (right). Conversation partner on the right (AL experiment).

Conversation partners across tasks can be seen in Figure 2. A participant sees two simplified distinctions out of four: *gender* (female/male), *age* (adult/child), *ethnicity* (darker/lighter), or *view* (*spatial orientation*) (front/side). Participants have to focus on one distinction as the main cue of the suffixation and ignore a different distinction, which is

<sup>1</sup>For details, see (Rácz, Hay, & Pierrehumbert, 2017, 2018)

the competitor cue. Any distinction can serve as main cue or competitor cue for a given participant.

So, for instance, in training, the participant has to learn that the suffix is '-pel' for a woman and '-fis' for a man, irrespective of their age. The main cue here is gender, the competitor cue is age. In test, they have to generalise this pattern to female and male conversation partners of a different skin tone.



Figure 2: Conversation partners across tasks (AL experiment)

369 participants completed the tasks successfully (mean age = 33, 201 women), with 27-45 participants in a given across-participants condition.

**Hypotheses** The allocentric expectation is that learning success mirrors exposure to the pattern in the real world: The gender distinction will be the easiest to learn as a main contextual cue in the artificial language. The spatial orientation distinction will be the hardest. The age and ethnicity distinctions will be in between. Gender is learned early by infants (Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991) and is a robust sociolinguistic marker (Labov, 2001). This is likely because, while gender is not binary, differences are perceptively more bimodal than in the case of either age or ethnicity.

The egocentric expectation ties back to the *gender paradox* outlined in the first section. If women rely more on social-indexical knowledge, female participants should show higher learning success in the task.

**Results** We used mixed-effects logistic regression to analyse test results, considering random slopes for robust predictors. The outcome is correct answer in the test trials. The predictors are participant age and gender, whether the participant has seen the trial target item / conversation partner in training, and the main cue and the competitor cue type in training.

Participant test accuracy hinges on the main cue. It is lower with the age and ethnicity cues than with the gender cue. However, this is mediated by familiarity with the conversation partner: test accuracy is similar across the three main cues if the conversation partner is familiar from training, but the age cue and the ethnicity cue fare worse than the gender cue with new conversation partners ( $p < 0.05$ ). This can be seen in Fig 3, showing the raw data. While participant distribution is bimodal, the means are indicative of the overall

trend. The view (spatial orientation) cue is worse overall.

Fig 3 shows how participants generalise the different cues. For the gender cue, accuracy is the same with previously seen and new partners. The pattern is generalised easily. For the view cue, accuracy is around chance in both cases. For the two intermediate cues, however, participants are more accurate when they see a conversation partner who is familiar from training<sup>2</sup>.

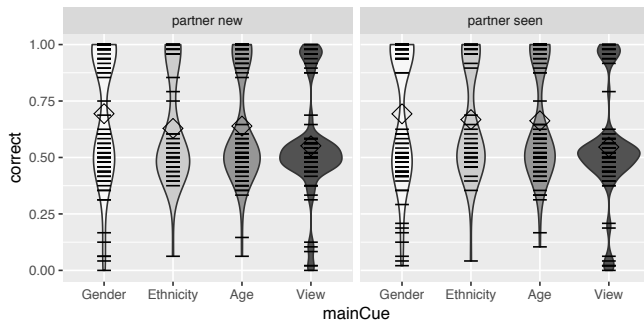


Figure 3: Response accuracy across main cue type and generalisation (AL experiment)

Older participants have higher test accuracy ( $p < 0.05$ ), but self-reported participant gender makes no difference.

**Discussion** The results speak for the prevalence of allocentric learning. Participant behaviour reflects knowledge of sociolinguistic variation. Gender, which is robust, and learned early by infants, is relatively easy to learn as a non-linguistic contextual cue of suffixation. Spatial orientation is very difficult. Distinctions that are also robust but more finely grained in the ambient language, age and ethnicity, are more restricted in scope to individual conversation partners and are harder to generalise to new partners.

Participant accuracy does not depend on participant gender. While the task does not explicitly test weighted learning, this suggests that even if the gender paradox can be explained via the heightened sensitivity of American English-speaking women, this sensitivity is not utilised in this sample.

This simple study offers a controlled way of testing predictions on learning social-indexical meaning. However, sociolinguistic variation in American English is typically diffuse, interlaid, and stochastic. For instance, while gendered language use is attested, it is not based on categorical distinctions in the speech of sharply distinct social groups (Eckert & McConnell-Ginet, 1992). We address this by drawing on field data from two communities and two social-indexical meaning patterns that offer limiting cases. Our first case is a linguistic pattern that is explicitly negotiated and draws sharp bound-

<sup>2</sup>We calculated the Minkowski distance between image pairs to assess how distinct these are visually. We found no effect of the size of Minkowski distance between a pair on how easy it was to learn to contrast the pair.

aries with respect to speaker gender: Datooga name avoidance.

## Name avoidance in Datooga

Datooga is a Nilotic language spoken by a traditionally pastoralist ethnic group living in northern Tanzania. They practice name avoidance: married women do not use the names of many of their husband’s ancestors. These women also avoid words that sound similar to these names, and instead rely on a conventionalised avoidance register that is a mixture of Datooga circumlocutions (‘donkey’ = ‘thing that is loaded’), phonological mutations, and borrowings from Swahili, the *lingua franca* of Tanzania. The register is not used by men, though they have to understand it to communicate with women (Mitchell, 2016).

**Design** The data presented here come from a name avoidance questionnaire completed by 30 Datooga-speaking children in five locations in Mbulu district of Manyara region in northern Tanzania<sup>3</sup>. The children’s estimated age is between 7-11; 17 are girls.

The questionnaire is read out to each participant by the fieldworker. It contains a comprehension and a production part. In comprehension, children hear 10 avoidance terms and have to give the ordinary term. In production, children hear 20 ordinary Datooga terms and have to give the avoidance term.

**Questions** The allocentric expectation is that children who live in a more mono-ethnic and more mono-lingual environment will be more immersed in avoidance practice and, as a result, give more accurate responses. The task warrants a strong egocentric expectation: since only women use the register, we expect girls to be more accurate than boys, especially in the production task.

**Results** We used mixed-effects logistic regression to model the data. The outcome is a correct or incorrect form provided by the participant, the predictors are participant age and gender, whether the participant goes to school, and task type (comprehension versus production). Task type is a within-participant condition. We used a participant random intercept, grouped under a location random intercept.

Overall, accuracy of both girls and boys in both task types is low. The best participant has a mean accuracy of 0.5 (15 correct answers), the worst has 0.07 (2 correct answers). Girls are slightly more accurate than boys ( $p < 0.05$ ), which either shows a genuine advantage in the register, or an overall advantage in fluency. (Young girls tend to be more fluent speakers than young boys, and our measures of controlling for this

<sup>3</sup>The design was pre-registered by the Open Science Foundation (<https://osf.io/xFjgq/>), the data were collected by Alice Mitchell during her field trip between March-October 2017. The data have not been published elsewhere.

difference in the field have proven ineffective.) There is no difference in accuracy across task type (Figure 4).

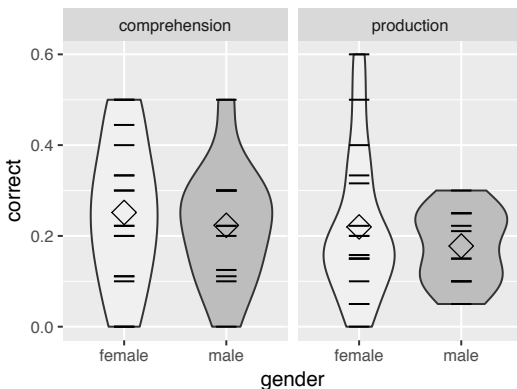


Figure 4: Response accuracy across gender and task type: all locations (Name avoidance task)

Response accuracy varies strongly across locations of data collection. Some locations are more rural and mono-ethnic, have no access to a local school, and are less influenced by Christianity than others. Children in these locations show higher overall accuracy. We speculate that the traditional norm is enforced more strongly in these locations, so that they run counter to the more general trend that name avoidance is falling out of use across the Datooga-speaking regions of Tanzania (Mitchell, 2016)<sup>4</sup>.

**Discussion** The artificial language (AL) task tests a suffixation pattern in a nonce language. Participants are adult speakers of American English. The name avoidance task has a much smaller sample size and focusses on children, testing their knowledge of a Datooga lexical stratum. The AL task references social-indexical patterns that are diffuse and stochastic in American English. The name avoidance task draws from a register that shows almost no within-speaker variation and categorical across-speaker variation, that is socially demarcated, involves explicit instruction, and is associated with traditional gender roles in Datooga.

Still, the Datooga data further support the primacy of allo-centric learning. Despite the fact that name avoidance has a radically different role in the life of women and men, we do not see a substantial gendered difference in children’s knowledge of the avoidance register. At the same time, the rate of exposure seems important: children who grow up more immersed in the register show more knowledge. This can be compared to the success of American English-speaking adults who are trained with more robust cues in the AL task.

<sup>4</sup>We collected data from children whose families avoid names for up to 2 generations of the male lineage. We had to remove an outlier, a boy whose family avoids names for up to 5 generations. Unsurprisingly, this boy was far more accurate than the rest of the cohort.

Name avoidance can be seen as a limiting case of gender-based sociolinguistic variation: married women have to produce avoidance terms but men only have to interpret them. It is very likely that a sample of teens and young adults would show a different picture, one that is marked by stronger gendered differentiation. Still, the children in our sample do have some knowledge of name avoidance, without showing robust gendered variation.

A different limiting case on sociolinguistic variation is one in which a pattern shows speaker-level variation – it is used by both men and women (as in the case of variation in American English) but context-specific word variation is categorical, as social-indexical meaning has a categorical mapping to the social environment. This is typified in sibling marking in the verbal morphology in Murrinhpatha.

### Sibling morphology in Murrinhpatha

Murrinhpatha is an Australian Aboriginal language spoken as a regional *lingua franca* predominantly at Wadeye, the former Port Keats mission in Northern Australia<sup>5</sup>. Its speakers are largely monolingual. Its kinship system is classificatory. Every member of the community can be classified using a kin term. For example, parallel cousins are classificatory siblings, whereas cross-cousins are regarded as cousins.

Kinship has infused the Murrinhpatha verbal paradigm. As in many Indo-European languages, subject and object gender and number are marked on the verb. If the subject or object number is either dual (‘two’) or paucal (‘a few’, cf. plural, ‘many’), the shared kinship status of the subjects / objects – whether they are siblings or not – is also marked. That is, a Murrinhpatha sentence informs the listener via verbal inflection on whether the referents are siblings.

Sibling marking likely evolved in a complex way to allow more discrimination in reference without the use of proper names, as Murrinhpatha practices the avoidance of the names of the deceased (Blythe, 2013). In any case, the correct use of Murrinhpatha requires the speaker to know about the detailed family relations of everybody that they want to talk about.

**Design** The data presented here come from a comprehension task completed by 39 Murrinhpatha speakers in Wadeye (age range: 5-40, mean age 14; 16 females). The task is auditory and visual, and was carried out using a tablet. In the task, each trial presents an activity (fighting / laughing / pointing / walking / waving) and an audio clip of a Murrinhpatha sentence is played. The sentence depicts the activity with a set subject number, gender, and sibling marking (indicating presence or absence of a sibling relation).

Then, two pictures are presented, both showing members of the local community (coming from one large and well known extended family). Participants have to choose the picture that matches the sentence. This is either a match of the

<sup>5</sup>The data were collected by Joe Blythe and Jeremiah Tunmuck in 2015. The data have not been published elsewhere.

verb's gender (all males versus at least one female), the number (dual or paucal, i.e. more than two) and sibling marking (siblings or not siblings). The sibling dimension is made more complex as some of the pairs depict biological siblings versus non-siblings; others depict parallel cousins versus cross-cousins. (Parallel cousins are labelled as siblings in the Murrinhpatha kinship system.) Figure 5 shows an example of a trial picture pair, depicting fighting, and two pairs of men in the community (one pair are brothers, the other pair are not).

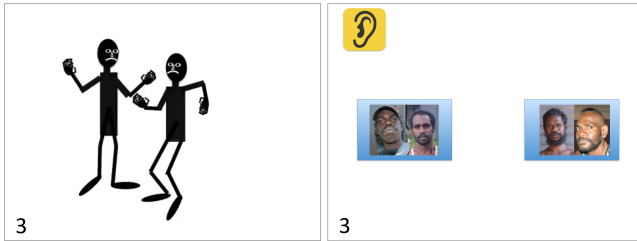


Figure 5: Task trial example (sibling task)

Matching the correct picture with the sentence requires prior knowledge of the world. If the match is based on gender or number, the required knowledge is relatively simple. If it is based on a sibling / non-sibling contrast, this knowledge is more complex (entailing which pairs of humans are likely to be siblings). If it is based on a cross- / parallel-cousin contrast, it is even more complex (entailing specific knowledge of the kin relations of the community).

**Questions** Murrinhpatha sibling-marking offers a different angle on allocentric / egocentric learning, focussing on the word-level, rather than the speaker-level. The allocentric expectation is that distinctions in the verbal morphology are learned together, irrespective of the amount of social information they require – if this social information were not naturally available, the language would be very hard to learn. In contrast, the egocentric expectation is that socially conditioned verbal morphology is harder to learn than morphology based on number and gender. We expect older participants to be more accurate in the task in either case, but the egocentric expectation is that the additional social knowledge gives older participants further advantage.

**Results** We used mixed-effects logistic regression to model the responses. The outcome is whether the correct picture is chosen to pair with the sentence that the participant has heard (see Figure 5). The predictors are participant gender and age, as well as type of verb agreement (gender / number / sibling / classificatory sibling). Verb agreement is a within-participant condition. We tested for the interaction of participant age and verb agreement.

Response accuracy is high. Participant age is a strong predictor of response accuracy in the task ( $p < 0.001$ ). Participant gender makes no difference (Figure 6).

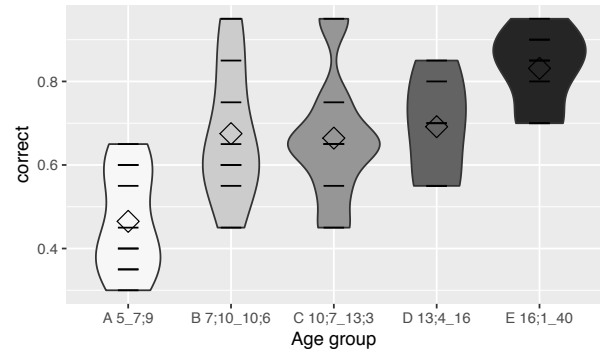


Figure 6: Response accuracy across age groups (sibling task).

In terms of the tested distinction types in verbal morphology, there is no discernible difference between morphological gender, number, and sibling agreement, indicating that by the time children learn verb agreement, they already have enough genealogical information to extend agreement to sibling marking. In contrast, the identification of parallel-cousins (but not cross-cousins) as siblings is slightly harder ( $p < 0.1$ , Figure 7). That is, from a genealogical point of view, generalisation proceeds from specific close-kin relationship to broader classification. From a developmental point of view, however, by the time children have mastery of the morphological template for sibling marking, they also have most of the genealogical information – we see no interaction between participant age and the relative differences in task type accuracy.

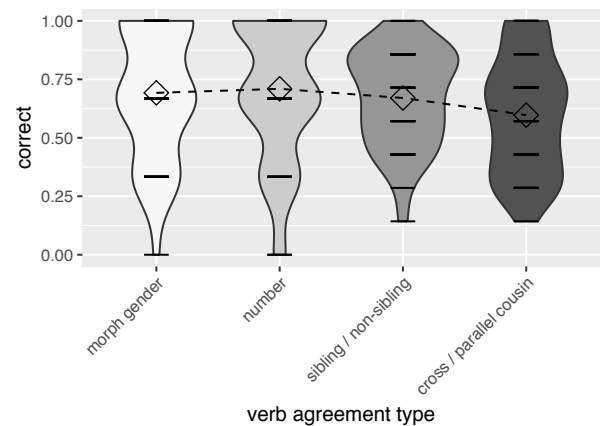


Figure 7: Response accuracy across distinction types (sibling task).

**Discussion** Siblinghood in Murrinhpatha is determined by close relations (having the same biological parents) and more complex, socially defined relations (making a distinction between parallel- and cross-cousins). This socially negotiated distinction is unambiguously coded in the verbal morphol-

ogy. Unlike in Datooga, the linguistic pattern is anchored firmly in the non-linguistic context (a context of siblings carries sibling-agreement, and so on), and it is used by both female and male speakers of the language. As in our Datooga and American English case studies, we see no evidence that female speakers would enjoy an advantage over male speakers.

Murrinhpatha supports the allocentric perspective: the similar rate of learning for morphological gender and number and siblinghood indicate that social meaning permeates language acquisition. However, we see a hint of egocentric learning: more close-knit biological siblings are easier to recognise than broader, socially constructed sibling relations.

### General Discussion

Allocentric and egocentric biases overlap in processing sociolinguistic variation. A speaker's social position might entail that they are both more exposed to a social-indexical pattern and are more sensitive to it. In this paper, we approach exposure *versus* sensitivity from different angles. Our artificial language learning experiments reveal that participants bring their prior knowledge with them into the task, that the overall robustness of the underlying pattern is more important in predicting participant success, and that women enjoy no advantage over men in learning. This advocates for an allocentric reading, with the addendum that the stochastic nature of social-indexical meaning in American English renders these readings difficult to disentangle, even in a controlled experimental task.

The allocentric reading is supported by data on patterns that are categorical, rather than stochastic, encodings of social-indexical meaning (in both the Datooga and the Murrinhpatha case) and the use of which is overtly and strongly gendered (in the Datooga case). While the Datooga example applies to a different linguistic domain (vocabulary, rather than morphology), it offers a more clear-cut case of gendered language than other examples from anthropology (Trechter, 1995). The Murrinhpatha case offers a specific example of learning language alongside indexed social information: for Murrinhpatha speakers, as for all of us, proficiency in language is necessarily combined with understanding the social world.

The overall picture that emerges from these specific studies is that our knowledge of social-indexical meaning is built largely on exposure, not filtered by sensitivity, and that learning social language use is, despite our conscious awareness of social structure, largely automatic. What is more, this is not restricted to a sample from a large-scale Western speaker group.

Needless to say, the egocentric and allocentric perspectives are not incompatible. In addition, these represent only a small part of the problem space of learning and maintaining sociolinguistic variation. The types of variation we draw on here also differ from each other in key aspects of indexicality, which definitely merits further discussion. The results re-

ported here demonstrate the usefulness of a broad toolkit and an inclusive use of evidence in studying focussed problems in social language learning, problems that can be generalised to other aspects of learning theory.

### References

- Bakir, M. (1986). Sex differences in the approximation to Standard Arabic: a case study. *Anthropological linguistics*, 28(1), 3–9.
- Blythe, J. (2013). Preference organization driving structuration: Evidence from Australian Aboriginal interaction for pragmatically motivated grammaticalization. *Language*, 89(4), 883–919.
- Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *TICS*, 10(12), 551–557.
- Eckert, P. (2008). Variation and the indexical field. *Journal of sociolinguistics*, 12(4), 453–476.
- Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual review of anthropology*, 21(1), 461–488.
- Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology*, 1(1), 5–39.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not weird. *Nature*, 466(7302), 29–29.
- Labov, W. (2001). *Principles of linguistic change volume 2: Social factors*. Wiley-Blackwell, Hoboken, NJ.
- Lakoff, R. (1973). Language and Woman's Place. *Language in society*, 2(1), 45–79.
- Milroy, L. (1980). *Language and social networks*. Basil Blackwell.
- Mitchell, A. (2016). Words that smell like father-in-law: A linguistic description of the Datooga avoidance register. *Anthropological Linguistics*, 57(2), 195–217.
- Rác, P., Hay, J. B., & Pierrehumbert, J. B. (2017). Social salience discriminates learnability of contextual cues in an artificial language. *Frontiers in psychology*, 8.
- Rác, P., Hay, J. B., & Pierrehumbert, J. B. (2018). *Learning sociolinguistic variation: an experimental study*. (submitted)
- Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive psychology*, 94, 85–114.
- Smith, J., Durham, M., & Richards, H. (2013). The social and linguistic in the acquisition of sociolinguistic norms: caregivers, children, and variation. *Linguistics*, 51(2), 285–324. doi: 10.1515/ling-2013-0012
- Trechter, S. (1995). Categorical gender myths in Native America: Gender deictics in Lakota. *Issues in Applied Linguistics*, 6(1), 5–22.
- Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, 3(2), 55–75.