

Signalling Games, Sociolinguistic Variation and the Construction of Style*

Heather Burnett
Laboratoire de Linguistique Formelle
CNRS-Université de Paris-Diderot

February 11, 2017

Abstract

This paper develops a formal model of the subtle meaning differences that exist between grammatical alternatives in socially conditioned variation (called *variants*) and how these variants can be used by speakers as resources for constructing personal linguistic styles. More specifically, this paper introduces a new formal system, called *social meaning games* (SMGs), which allows for the unification of variationist sociolinguistics and game-theoretic pragmatics, two fields that have had very little interaction in the past. Although remarks have been made concerning the possible usefulness of game-theoretic tools in the analysis of certain kinds of socially conditioned linguistic phenomena (Goffman, 1961, 1967, 1970; Bourdieu, 1977; Dror et al., 2013, 2014; Clark, 2014, among others), a general framework uniting game-theoretic pragmatics and quantitative sociolinguistics has yet to be developed. This paper constructs such a framework through giving a formalization of the *Third Wave* approach to the meaning of variation (see Eckert (2012) for an overview) using (modified) *signalling games* (Lewis, 1969) and a probabilistic approach to speaker/listener beliefs of the kind commonly used in the Bayesian game-theoretic pragmatics framework (see Goodman and Lassiter, 2014; Franke and Jäger, 2016, for recent overviews).

*This research has been partially supported by the program “Investissements d’Avenir” overseen by the French National Research Agency, ANR-10-LABX-0083 (Labex EFL), and a fellowship from the Center for the Study of Language and Information at Stanford University. I thank Leon Bergen, Adrian Brasoveanu, Judith Degen, Chantal Gratton, Erez Levon, Eric McCreedy, Jessica Rett, Devyani Sharma, Elizabeth Smith, Sali Tagliamonte, Meredith Tamminga, audiences at UCL, Institut Jean Nicod, LLF Paris-Diderot, Stanford, UCLA, UCSC and *NWAV45*, and especially Penny Eckert and Dan Lassiter for very helpful comments and discussions. All errors are my own.

1 Introduction

This paper develops a formal model of the subtle meaning differences that exist between grammatical alternatives in socially conditioned variation (called *variants*) and how these variants can be used by speakers as resources for constructing personal linguistic styles. The range of empirical phenomena that the proposed model aims to capture is exemplified through the grammatical alternations in (1)-(3). In terminology commonly used in the field of *variationist sociolinguistics* (Labov, 1963, 1966; Weinreich et al., 1968, et seq.), alternations such as those shown below are called sociolinguistic *variables*, and they can be, among other things, phonetic in nature (like (1)-(2)) or morpho-syntactic/discursive (as in (3)).

- (1) (ING)
- a. I'm working on my paper. [ɪŋ]
 - b. I'm workin' on my paper. [ɪn]
- (2) /t/ release
- a. I want a glass of wa[t^h]er. released /t/
 - b. I want a glass of watblue[r]er. flap
- (3) Discourse *like*
- a. What are we doing tonight? bare
 - b. **Like**, what are we doing tonight? like

More specifically, this paper introduces a new formal framework, called *social meaning games* (SMGs), which allows for the unification of variationist sociolinguistics and game-theoretic pragmatics (see Benz et al., 2004; Jäger, 2011, for overviews), two fields that have had very little interaction in the past¹. Although remarks have been made concerning the possible usefulness of game-theoretic tools in the analysis of certain kinds of socially conditioned linguistic phenomena (Goffman, 1961, 1967, 1970; Bourdieu, 1977; Dror et al., 2013, 2014; Clark, 2014, among others), a general formal framework uniting game-theoretic pragmatics and quantitative sociolinguistics has yet to be developed. This paper constructs such a framework through giving a formalization of the *Third Wave* approach to the

¹Note that I am speaking here of the field of variationist (quantitative) sociolinguistics. There is a rich (and developing) tradition of work within the game-theoretic paradigm on some other sociolinguistic/pragmatic topics such as the formal modelization of politeness, swear words and social networks (Van Rooy, 2003; Mühlenbernd and Franke, 2012; Mühlenbernd, 2013; Quinley and Mühlenbernd, 2012; McCready, 2012; McCready et al., 2012, among others). Additionally, there is already some work aiming at integrating formal semantics/pragmatics and variationist sociolinguistics using non-game-theoretic methods (Lassiter, 2008; Smith et al., 2010; Acton, 2014; Acton and Potts, 2014; Acton, 2016; Beltrama, 2016, for example). However, (to my knowledge) there is no account within the game-theoretic paradigm of the kinds of phenomena that have been the focus of empirical work within the variationist tradition (to be described below).

meaning of variation (see Eckert (2012) for an overview) using (modified) *signalling games* (Lewis, 1969) and a probabilistic approach to speaker/listener beliefs of the kind commonly used in the Bayesian game-theoretic pragmatics framework (see Goodman and Lassiter, 2014; Franke and Jäger, 2016, for recent overviews).

The paper is laid out as follows: in section 2, based on results from sociolinguistic perception studies, I observe (following others) that the use of one grammatical variant versus another can induce inferences on the part of the listener about the kinds of properties that characterize the speaker. I further argue that, in addition to being triggered by particular linguistic forms, these inferences share certain non-trivial properties with other kinds of inferences commonly studied in formal pragmatics. I therefore propose that social meaning of the kind studied in this paper should be viewed as an instance of pragmatic enrichment, and that, as a consequence, a unified framework that can treat both social meaning and other kinds of meaning in context should be developed. In section 3, I consider what the properties of such a framework should be. In particular, based on the results of sociolinguistic production studies, I argue that the social aspects² of linguistic variation should be analyzed as instances of interactive rational language use. This is most obviously seen through studies of intra-speaker variation (also known as *style shifting*); however, following previous research, I argue that there are reasons to think that inter-speaker variation (a.k.a. social stratification) should also be analyzed as the result of interactive rational language use. I then give a brief description of one influential theory within sociolinguistics which aims to derive both style shifting and social stratification from (informal) principles governing rational use: *Third Wave variation* theory (TW). Based on conclusions from TW that both interactivity and rationality characterize all the social aspects of variation, I propose that a game-theoretic approach can be useful in modelling this kind of linguistic communication.

This being said, game-theoretic tools are extremely general and have been used in the analysis of a wide range of economic, social and biological phenomena³. Thus, for my proposal to have any substantive content, I must be more precise concerning the definition of the games (the players, the architecture, utility functions etc.) and what their solution concepts are. In section 4, I give a concrete proposal for how to integrate sociolinguistic variation into the broader framework of game-theoretic pragmatics: social meaning games. I first define the games, and then I give some illustrations of the kinds of predictions that this framework makes for quantitative patterns of sociolinguistic variation, on the one hand, and the options for and constraints on the construction of personal linguistic styles, on

²The main proposals in this paper are limited to modelling the aspects of linguistic variation that are determined by what sociolinguists call *social*, *external* or *non-linguistic* factors. Patterns of linguistic variation are also determined by other factors which are not social/strategic in nature (ex. general cognitive factors associated with linguistic production and comprehension, as well as grammatical factors (what Labov calls *internal* factors). I will make some remarks concerning how the analysis of social factors given here could be integrated into a broader theory of linguistic variation and change; however, I will not discuss non-social factors in great detail.

³See Osborne and Rubinstein (1994) for an introduction to this vast field.

the other. I further show how this framework can capture both similarities and differences between social meaning and other kinds of pragmatic enrichment discussed in section 2. Section 5 provides some concluding remarks and explores how the proposals made in this paper for social meaning and the construction of linguistic style could be extended to other aspects of stylistic performance.

2 Social Meaning as Pragmatic Enrichment

Suppose we are having a conversation and the person that we are talking to says (1-a) (repeated as (4-a)). What do we understand from this utterance?

- (4) a. I'm **working** on my paper. [ɪŋ]
b. I'm work**bluein'** on my paper. [ɪn]

From hearing (4-a), we can certainly conclude that the speaker is working on their paper. Intuitively, it also seems as if we might be able infer some additional thing(s) from (4-a), possibly something about the properties of the speaker, of the working event, or maybe even of both. Likewise, if our interlocutor says (1-b) (repeated as (4-b)), we will definitely understand from this sentence that they are working on their paper. But again, it seems as if we might want to infer something extra from this utterance, crucially something that is different from what we inferred from (4-a).

One of the most common ways in which the properties of these extra inferences have been investigated in both social psychology and variationist sociolinguistics is through the use of an experimental paradigm known as the *matched guise technique* (MGT) (Lambert et al., 1960). In a MGT experiment, participants listen to samples of recorded speech that have been designed to differ in very specific and controlled ways. Participants hear one of two recordings (called *guises*) which differ only in the alternation studied. After hearing a recording, participants' beliefs and attitudes towards the recorded speaker are assessed in some way, most often via focus group and/or questionnaire. All efforts are made to ensure that the two recordings match as possible, modulo the forms under study. Indeed, many recent studies (such as the ones described below) use digital manipulation of naturalistic speech recordings to ensure that any observed differences in inferences that participants draw in different guises are directly attributable to the variable under study, not to some other aspect of the voice of the speaker or of the content of their discourse.

In her 2006 dissertation and subsequent work, Campbell-Kibler (2006, 2007, 2008) performed an MGT study with 124 American college students using stimuli formed from the speech of 8 different speakers investigating how the use of the variable (ING) influences listener beliefs and perceptions. This study yielded a variety of complex patterns (which will be further discussed below), but her results show that there exist certain consistent associations between linguistic forms (*-ing* vs *-in'*) and property attributions for the listeners who

participated in the experiment. For example, all speakers were rated as significantly more educated and more articulate in their *-ing* guises than in their *-in'* guises. In other words, we see the existence of relationships between linguistic variants and cognitive representations associated with education and eloquence, at least for the participants of Campbell-Kibler's study.

Other studies on different variables have yielded the same kinds of results. For example, in order to investigate the social meaning of the /t/ release variable (2), Podesva et al. (2015) performed a MGT study with 70 American participants (the majority in their early 20s) using stimuli formed from political speeches of 6 American politicians (Barack Obama, John Edwards, Nancy Pelosi, George W. Bush, Hillary Clinton, and Condoleezza Rice). As in Campbell-Kibler's study, the /t/ release study yielded a number of results concerning associations with released vs flapped/unreleased /t/: for example, John Edwards and Condoleezza Rice were rated as significantly more articulate in their released /t/ guises than in their flapped guise (i.e. when they say things like wa[t^h]er, rather than wa[r]er⁴). On the other hand, Nancy Pelosi was rated as significantly less friendly and less sincere when she used released /t/, and Barack Obama was rated as significantly more passionate in his flapped guise than in his released /t/ guise.

The results concerning Pelosi and Obama in the /t/ release study serve to highlight an important feature of social meaning: depending on a variety of factors (to be further discussed below), it may be the case that use of a reduced or 'non-standard' variant triggers property attributions on the part of the listener that the speaker could find desirable (see also Trudgill, 1972; Rickford and Closs Traugott, 1985, among many others). In other words, even though a speaker who uses a flap may risk being perceived as less articulate than if they had used a released /t/, they also have a better chance of coming across as friendly, sincere or passionate with the non-standard variant. Therefore, depending on the persona that they are trying to construct in the context, it may be worth the speaker's while to risk being perceived as inarticulate in favour of being considered more authentic and solidary with their interlocutors.

Finally, we note that similar patterns can be seen beyond the domain of phonetic variation. For example, in another MGT experiment investigating the social meaning of discourse like (3), Maddeaux and Dinkin (2015a) show that speakers using like in a sentence like (5-a) are perceived to be significantly less articulate and less intelligent by a group of 69 Canadian undergraduates than in their guise without *like* (5-b).

- (5) a. I couldn't get, **like**, a peaceful sleep. (Toronto English corpus (2/r/f11)
Cited from (D'Arcy, 2005, 21))
b. I couldn't get a peaceful sleep.

⁴These results are unsurprising given that articulateness has been associated with released /t/ in many other ethnographically-based studies, such as Bunin Benor (2001); Bucholtz (1996); Podesva (2006); Eckert (2008).

In sum, I suggest that we can conclude from these studies (and the many others like them) that, in addition to extra information derived through pragmatic processes that are more familiar to researchers in formal pragmatics, listeners derive extra information from an utterance concerning the properties that hold of the speaker, and these inferences are based on the particular linguistic forms that the speaker has chosen to employ. In other words, I suggest that the inferences triggered by socially meaningful variants are kinds of implicatures, similar (although not identical) to scalar implicatures (6-a) or implicatures generated by expressions with expressive content (6-b) (see also McConnell-Ginet, 2011; Acton, 2014; Smith et al., 2010, for additional support for versions of this claim).

- (6) a. Mary ate **some** of the cookies.
Extra inference: *Mary did not eat all of the cookies.*
- b. That **bastard** Kaplan got promoted! (Kaplan, 1999, 9)
Extra inference: *The speaker does not like Kaplan.*

For some variables (such as (ING)) all or most listeners draw the same robust inferences no matter who the speakers are. However, in many cases, which property attributions a particular variant will trigger will depend greatly on which other properties are believed to hold of the speaker. This feature can already been seen in the discussion of Podesva et al.'s /t/ release study above. In particular, while these researchers found significant relationships between articulateness and released /t/ with Edwards and Rice, these results were found only with these two speakers. Likewise, in this experiment, flapping made only Nancy Pelosi sound more friendly and sincere; no significant effect of friendliness or authenticity was found with the other politicians. As (Podesva et al., 2015, 79) say,

Listeners appear to be less likely to associate released /t/s with competence-based meanings in Pelosi's speech—such as articulateness, intelligence or authoritativeness—as illustrated by the quotation in [(7)] from one of our focus group respondents [...]. The focus group participant's comment acknowledges that release can be used to sound more authoritative, but asserts that in Pelosi's voice the authoritativeness is not genuine.

- (7) **Participant 1:** Um, I don't want to sound, say, fake, but she just sounded-like she was-it was just very-it contrasted from her other speech, so I-um, she sounded like she was trying to be more authoritative in her speech, perhaps.

Thus, because (as Podesva et al. hypothesize) speakers find Pelosi's use of released /t/ to be pretentious and fake, using the flap makes her sound more sincere and creates a positive evaluation. In other words, these results show that social enrichment is dependent on speaker identity, but also (more importantly) on listeners' **interpretations** of speakers' linguistic performances. I will call this property of social meaning **LISTENER ORIENTATION**.

That social enrichment is listener oriented can also be seen in Campbell-Kibler’s study of (ING). In addition to a fair number of interpretative effects being limited to the speech of a single speaker or to a proper subset of speakers⁵, Campbell-Kibler (2006, 2008) finds that listeners can assign different interpretations to the exact same linguistic performances of a single speaker. For example, after hearing recordings by the speaker (pseudo)named Elizabeth in which she discusses groups of people to which she herself does not belong, listeners in this MGT study were divided on which properties to attribute to this speaker. As shown in the table reproduced (from Campbell-Kibler, 2008, 646) as Table 1, after hearing the *-in’* guise, 30% of speakers thought Elizabeth was compassionate; however, another 17.5% of speakers thought that she sounded condescending in her *-in’* guise. Unsurprisingly, the selections for compassionate and condescending were virtually disjoint, with only a single speaker selecting both of these properties (Campbell-Kibler, 2008, 645). On the other hand, when speakers heard Elizabeth say *-ing* (which some focus group members describe what they think that she would “naturally” say (Campbell-Kibler, 2008, 648)), no speakers think that she sounds condescending and only 7.4% of speakers selected *compassionate*.

| Checkbox label | % listeners selecting checkbox | | |
|----------------|--------------------------------|------|-------|
| | -in | -ing | sig. |
| compassionate | 30.0 | 7.4 | 0.022 |
| condescending | 17.5 | 0.0 | 0.005 |

Table 1 – Compassionate/condescending selections (Elizabeth’s “other” recordings)

These results (and others like these) showing listener orientation highlight the differences between other kinds of enrichment and social enrichment of the kind discussed here: as Campbell-Kibler says (p.639),

The audiences for whom we perform on a day-to day basis are not obligated to accept our accounts of ourselves, even if they share a common ground with us regarding the basic meaning of our semiotic choices. As a result the process of constructing linguistic (and other social) performances is not like encoding a secret message, where we can trust that the recipient is seeking to uncover exactly the message we intend to send. Instead, social performance is more like choosing a name for a child: we may study name books and quiz friends about childhood memories of insulting nicknames, but once the name is chosen, we ultimately have no control over what someone gets called on the playground—that is, what interpretations others assign to our chosen resources.

Enrichment of the kind that we find with scalar implicatures (6-a), which proceeds by strengthening the truth conditional information of an utterance **is** often considered to occur as a kind of ‘secret message’ transmission, where the hearer tries to reason cooperatively to try to figure out what the speaker meant in the context (Lewis, 1969; Grice, 1975). Likewise, enrichment of the kind that we find with expressives (6-b) (ex. *bastard*) has

⁵ See (Campbell-Kibler, 2006, chapter 4) for a summary.

been argued to be *speaker oriented* (Kaplan, 1999; Potts, 2007; McCready, 2012, among others); i.e. the listener’s action is to simply attribute a subjective belief to the speaker; they certainly don’t get to decide what belief is. For example, McCready (2012) observes that the swear word *fucking* can be used in English to express either the speaker’s positive attitude towards a situation or their negative attitude. These two uses can be brought out in the mini-discourses in (8).

- (8) (McCready, 2012, 245)
- | | | |
|----|---------------------------------------------------------------------------------------|-----------------|
| a. | Fucking Mike Tyson won another fight. He’s wonderful. | Positive |
| b. | Fucking Mike Tyson got arrested again for domestic violence. # He’s wonderful. | Negative |

However, as McCready shows, with an expressive like *fucking*, the speaker is perfectly licensed to correct the listener on their interpretation if it does not line up with what the speaker wanted, as shown in (9).

- (9) **A:** Somebody stole my **fucking** car.
B: That sucks.
A: No, that’s not what I meant—I hated that car, and now I get the insurance.
(McCready, 2012, 258)

As McCready says (p.259),

Here, B’s attempt to commiserate with A fails. A’s use of *fucking* is meant to convey a positive attitude. This attempt fails; [...] This example, then, indicates that genuine coordination is not at issue, unlike the case of more standard contextual parameters. Rather, communication requires the proper recovery of the speaker’s intention.

Thus, social enrichment displays an importantly different pattern from expressives (at least those like *fucking*).

Finally, I note that social enrichment shares another notable property with (some) other kinds of pragmatic enrichment: it shows (what I will call) GRAMMATICAL DEPENDENCE. More specifically, as has been observed since the beginnings of variationist sociolinguistics (ex. Labov, 1966), the environments in which social enrichment is possible, and what the nature of that enrichment is, are defined by the phonological or morpho-syntactic grammar. To continue with the examples discussed already in this paper: the conditioning of the variable (ING) has been shown to be affected by both grammatical category (Labov, 1966, and many others) and the abstract morphological structure of the variant (Houston, 1985; Tamminga, 2014). Additionally, Podesva et al. (2015) shows that social enrichment for /t/ release is easier when the pertinent variant occupies a word-medial position (10-a) rather than final position (10-b).

- (10) a. I want a glass of wa/**t**/er.
b. We should mee/**t**/.

Likewise for discourse *like*: Maddeaux and Dinkin (2015b) show that while noun-phrase initial *like* (5-a) can trigger inferences of unintelligence, the use of discourse *like* in other syntactic positions, such as verb-phrase initial *like* (11), is not socially meaningful in the same way.

- (11) I'm not sure if my eight year old **like** understands that.
(N/X/m/46), cited in (D'Arcy, 2005, 172)

A more striking example of grammatical dependence of social meaning comes from variation in the future tense in dialects of French. In some varieties of Canadian and European French, the semantic difference between the periphrastic future (12-a) and (12-b) has been neutralized, and these forms are in variation in the dialects in question. This variation conditioned primarily by properties such as age and social class, with (depending on the study) older and/or better educated speakers favouring the more 'standard' simple future form (12-b) compared to younger and/or less educated speakers (see Deshaies and Laforge, 1981; Poplack and Turpin, 1999; Wagner and Sankoff, 2011; Sankoff et al., 2012, among many others).

- (12) a. Je **vais** manger.
b. Je manger**ai**.
'I will eat.'

Interestingly, in quantitative studies of this variable, such as (Wagner and Sankoff, 2011; Sankoff et al., 2012, for Montréal French), social stratification is found **only** in affirmative sentences. If future tense is interpreted within the scope of a semantically decreasing operator, such as negation (13), then stratification disappears and all members of the community strongly prefer the simple future.

- (13) Je vais **pas** manger. \approx Je mangerai **pas**.

Although this phenomenon would need to be further investigated (using, for example, experimental methodology), these preliminary results suggest that social enrichment of the future variants, which creates the stratificational patterns observed in the studies cited above, appears to be blocked in decreasing contexts. It is well-known that pragmatic processes like scalar enrichment are grammatically dependent (Levinson, 2000; Récanati, 2003; Chierchia et al., 2008; Chemla and Spector, 2011; Potts et al., 2015), and these inferences are particularly sensitive to decreasing operators (Ducrot, 1969; Cohen, 1971, among many others). Thus, I tentatively suggest that social enrichment has a very close relationship with

the structure building operations of the grammar, possibly similar to that of scalar implicatures⁶; however, I leave the very complex (but empirically rich) project of fleshing out the grammatical dependencies that social enrichment is subject to to future research.

In summary, following previous research into sociolinguistic perception, I have argued that social meaning should be viewed as an implicature that is triggered by the use of particular variants. As such, I propose that social meaning should be integrated into a broader theory of formal pragmatics, while still accounting for its listener oriented quality. Of course, there are very many pragmatic frameworks with very many different properties available in the literature that we might choose from for this integration. In the next section, I argue that there is one framework in particular that looks especially promising: game-theoretic pragmatics.

3 Sociolinguistic Variation as Rational Language Use

This section argues, following previous work on sociolinguistic production, that speakers have implicit knowledge of the kinds of inferences that listeners draw based on their linguistic usage patterns, and that they exploit this knowledge in order to influence which properties their interlocutor will attribute to them. In other words, in this section, I will argue that linguistic variation is a social phenomenon that is both interactive, in the sense that speakers and listeners make hypotheses concerning their interlocutors' beliefs and interpretation strategies⁷, and (approximately) **rational**, in the sense that speaker/listener behaviours are (loosely) optimized to some criteria (Anderson, 1991). I first demonstrate these proposals with reference to intra-speaker variation (*style shifting*) and then make similar observations with respect to inter-speaker variation based on the perspective developed in the *Third Wave* approach to variation. The general conclusion to be drawn from this section is that a formal model of social meaning and its relation to socially conditioned patterns of linguistic variation should be able to capture both the interactive and rational aspect of the phenomena under study. Since interactivity and rationality are built into the architectures of game-theoretic approaches to meaning in context, I suggest that game theoretic tools are particularly well-suited to modelling this kind of communication.

⁶This being said, expressives have been argued not to have so many interactions with the grammar, i.e. inferences associated with expressive content tend to be preserved in embedded environments and remain within the scope of decreasing operators (Kaplan, 1999; Potts, 2007, among others). Therefore there is work to be done establishing exactly whether/how social meaning is grammatically constructed.

⁷This was already demonstrated for listeners in the previous section and within the works cited. So this section concerns speakers.

3.1 Style-shifting as rational language use

A particularly clear example of linguistic variation as rational language use comes from existence of contextually-based intra-speaker variation, i.e. *style shifting*. This is a robust, well-documented phenomenon, and we can give a first rather simple illustration of it from Labov (2012)'s study of President Obama's use of (ING). (Labov, 2012, 22) finds significant differences in Obama's use of (ING) across three different recordings taken in three different contexts: (what Labov calls) *casual*, *careful* and *formal*. The first recording that Labov studied was one of Obama barbecuing at a Father's Day barbecue on the Whitehouse lawn: a 'casual' context. Labov finds that Obama uses *-in'* 72% of the time in this context, i.e. he is doing a lot of *grillin'*, *eatin'* and *drinkin'* at the barbecue. Then the barbecue finishes, and Obama moves to answer political questions from reporters on the Whitehouse lawn. In this 'careful' context, his rate of *-in'* drops to 33%. Finally, Labov studied Obama's use of (ING) in a scripted acceptance speech at the Democratic National Convention (a 'formal' context). He finds that, in this recording, the President uses *-in'* only 3% of the time. Obama's use of (ING) across three different contexts is summarized in Figure 1, reproduced from (Labov, 2012, 22).

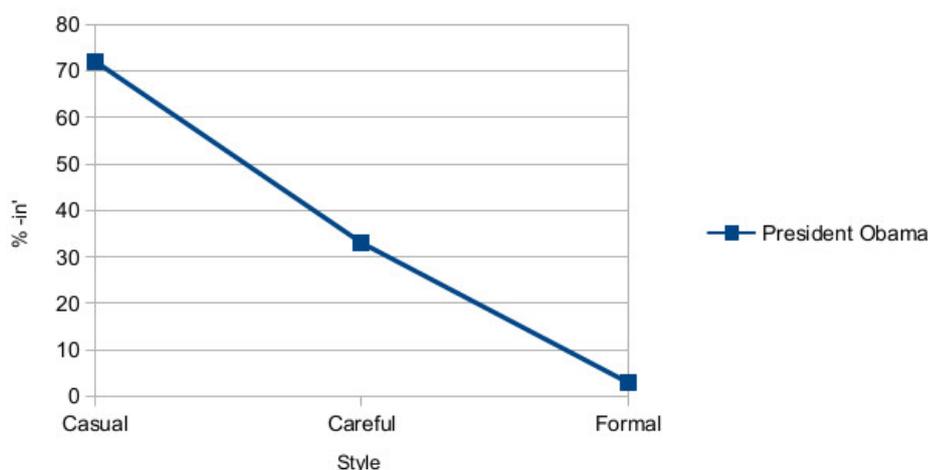


Figure 1 – President Obama's use of (ING) across three contexts

Why do we find this pattern? According to Labov, it arises because we have conventionally associated meanings with *-in'* and *-ing*, which allow us to communicate extra information to each other through phonetic variation. He describes what he calls our *hidden consensus* as follows (p.22),

This consensus is publicly available and in one sense, understood by all. In the classroom, or on the pulpit, people will attribute the use of the **-in'** form to laziness, ignorance, or just plain rascality. Yet the high value we put on

the **-in'** norm in other contexts is not hidden from public view. When we see the large illuminated sign, DUNKIN' DONUTS, we recognize the claim that **dunkin'** doughnuts taste better than **dunking** doughnuts. . . A Philadelphia travel agency is named with an electric sign spelling out CRUSIN'. We understand this as an advertisement that we will have a better time **cruisin'** than we would **cruising**.

I have chosen to give a first illustration of style shifting using President Obama, and (for concreteness) we will continue to study this example throughout the paper. However, it is important to stress that style-shifting is not a phenomenon that is uniquely associated with public figures, although these are the kinds of individuals for whom we tend to have the most available data. For example, Podesva (2004) (cited in Eckert (2005)) finds significant differences in a medical student's use of /t/ release in a clinic setting and when he is at a barbecue. Indeed, recent studies of style shifting of private citizens have shown that intra-speaker variation is widespread, with people significantly changing their use of variants across contexts (Cheshire, 1982; Kiesling, 1998; Podesva, 2007; Gratton, 2016; van Hofwegen, 2017, among many others) and even across sections of discourse (see Kiesling, 2009; Calder, 2016, for example).

It is also important to note that, while, at first glance, it may be tempting to reduce the use of variables like (ING) or /t/ release to a question of formality or 'register', with the unreduced more 'standard' variant being limited to formal contexts and the reduced 'vernacular' variant being limited to informal contexts, there are reasons to think that such a theory is too simplistic and that (ING) can be used to communicate properties that are not directly related to formality. One study that shows this particularly well is Gratton (2016)'s investigation of the link between (ING) and gender presentation (see also Kiesling, 1998). In this work, Gratton conducted group interviews with *non-binary* individuals, that is, individuals whose gender identity does not respect the male/female binary. She focussed on two people: Flynn, who was assigned female at birth, and Casey, who was assigned male at birth, and she conducted two sets of interviews with these consultants: the first one being in a queer-friendly environment (their home and a queer café, respectively) and the second one taking place in a public café. As shown in Figure 2, reproduced from Gratton (2016), neither Flynn nor Casey show a significant difference in their use of *-in'* vs *-ing* in the queer-friendly environments. However, in the public café, things are very different: Flynn, who was assigned female at birth, uses significantly more *-in'* (80%), while Casey, who was assigned male at birth, uses significantly more *-ing* (89%).

A naive theory in which the use of *-in'* vs *-ing* was uniquely determined by formality or 'register' cannot account for this pattern, since it is not obvious that the two interview contexts differ in formality. Furthermore, even if we suppose (for example) that the queer-friendly café is less formal than the public café, our 'formality' analysis could only explain Casey's behaviour; Flynn's pattern, where they increase the use of *-in'* in the public café, is unexpected. Instead of formality, what Gratton suggests is going on in Figure 2 (based on analysis of ethnographic interviews) is that "consultants vary their use of (ING) between

| | | Home/Queer café | | Popular café | |
|-------|---------|-----------------|----|--------------|----|
| | Variant | N | % | N | % |
| Flynn | -ing | 18 | 44 | 13 | 20 |
| | -in' | 23 | 56 | 52 | 80 |
| | Total N | 41 | | 65 | |
| Casey | -ing | 42 | 58 | 86 | 89 |
| | -in' | 30 | 42 | 11 | 11 |
| | Total N | 72 | | 97 | |

Table 2 – (ING) in queer-friendly vs public settings, based on Gratton (2016).

situations based on their perceived level of security in presenting their gender identity”. Indeed, in studies of demographically balanced corpora of naturalistic speech, it is common to find a stratificational pattern in which men favour *-in'* and women favour *-ing* (see Hazen, 2006, for a review). Thus, there appears to be some kind of (possibly indirect) association between *-in'* and masculinity⁸.

In summary, the studies discussed in this section show that speakers assess how their speech will be evaluated by their interlocutors in a particular discourse context, i.e. which properties that they think their interlocutors will attribute to them. In other words, sociolinguistic variation is an *interactive* phenomenon. Then, after this evaluation, speakers choose the form that (they think) will be the most successful to construct their desired persona. In other words, there is an aspect of *optimization* or *rationality* as well. Since interactivity and rationality form an important part of the architecture of game-theoretic frameworks, I propose that such approaches are particularly well adapted to modelling this kind of linguistic communication.

This being said, style-shifting is only one of the focusses of variationist sociolinguistics, with the majority of the most famous results from this field being associated with patterns of *social stratification*, i.e. inter-speaker differences in the use of variants. In the next section, (following others) I will argue that cases of social stratification should also be analyzed as interactive language use, and thus the game-theoretic model that will be presented in section 4 aims to model the full range of social aspects of linguistic variation.

3.2 Social stratification as rational language use

Since the beginning of the quantitative study of sociolinguistic variation, there has been an interest in developing a unified theory of style shifting and the kind of variation that has been the principal empirical focus of variationist sociolinguistics: social stratification. An obvious motivation for such a theory comes from the observation that the exact same linguistic variables are used in both the intra-speaker and inter-speaker dimensions of

⁸Eckert (2005) and Campbell-Kibler (2006) suggest that this relationship is mediated by *casualness*.

variation. For example, consider the graph in Figure 2, taken from Labov (1966)’s famous study of language use in New York City. As we saw with Obama in the previous section, the use of the *-in’* variant decreases with a more formal style⁹, but also with a rise in social class. In this way, (in the words of Labov) it becomes difficult to distinguish “a casual salesman from a careful pipefitter” (Labov, 1972, 240).

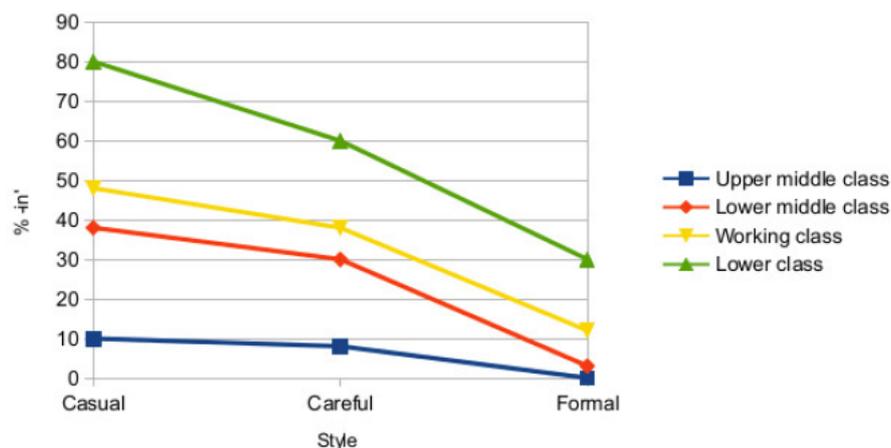


Figure 2 – Labov (1966): (ING) by social class and style (casual, careful, formal) in NYC

Likewise, in developing his influential *Audience Design* theory of style shift, Bell (1984) says that the relation between style shifting and social stratification is “more than an interrelation. It is a *derivation* [Bell’s emphasis]”; in particular, he says that “variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the “social” dimension.” (Bell, 1984, 151). Following this line of research, in this section, I argue that stratificational patterns are derived from the same basic principles as those that underly the creation of patterns of style shifting: the principles governing rational language use.

As Eckert (2008) notes, this point can already be made (for at least some cases of social stratification) based on Labov (1963)’s study of, among other variables, the centralization of diphthongs /ay/ and /aw/ in Martha’s Vineyard, an island south of Cape Cod in Massachusetts. Based on data from 69 sociolinguistic interviews, Labov found that, rather than being conditioned by gender, age or other demographic categories, centralization on Martha’s Vineyard was best predicted through looking at speakers’ stances towards the changing demographics of the island. During the period of Labov’s investigation, the main industries on the Vineyard were in the process of moving from whaling and fishing to tourism, creating significant hardships for islanders who had built their lives around the

⁹Note that in the 1966 study, *casual*, *careful* and *formal* styles correspond to interview speech, reading passages and word lists, respectively. This is different from categories in the 2012 Obama study; however, the overall point remains the same here.

fishing industry. As such, the participants in Labov's study were divided with respect to how they viewed these changes, having reactions "varying from a fiercely defensive contempt for outsiders to enthusiastic plans for furthering the tourist economy" (Labov, 1963, 28). Furthermore, as Labov observes, speakers' orientations towards or away from the island and the old way of life makes an enormous difference with respect to which variant they prefer to use, with more locally oriented individuals showing much higher degrees of centralization (Labov, 1963, 30). This contrast is particularly clear among younger speakers, who must decide whether or not to leave the island for good when they finish high school. Labov points out (p.32) that "comparatively few of the sons of the English-descent group will be earning their living on the Vineyard in the next 20 years" and he says that "a marked contrast was observed between those who plan to leave the island and those who do not. The latter show strong centralization, while the former show little, if any."

Another particularly important study in this vein is Eckert (1989, 2000)'s studies of linguistic variation in Belten High, a high school in a middle class suburb of Detroit. In these works, Eckert studies the linguistic and social conditioning of a variety of sociophonetic variables and one syntactic variable (negative concord). Eckert observes that within the high school there was a high degree of inter-speaker variation, polarized largely around two social groups in the school: the *jocks* and the *burnouts*. As Eckert (1989) shows, members of these two groups distinguish themselves through their actions, their attitudes, their ways of dressing, and also their use of the variables under study. Although the jocks and the burnouts were the most 'extreme' users of the different variants, Eckert (2000) found that, in the whole population of Belten High, use of variants was best predicted by students' urban/suburban orientation and/or institutional engagement, where urban orientation was measured by activities like *cruising*¹⁰, as shown in Table 3 (for 5 variables), and their institutional engagement was measured by participation in (school's) extra-curricular activities¹¹.

In summary, studies such as those conducted by Labov and Eckert show that (at least in the cases that they studied) inter-speaker differences break down along lines associated with other social practices, such as moving, cruising or participating in extra-curricular activities. As Eckert (2000) observes, these correlations can be immediately explained if the use of one variant over another is itself also a social practice, saying (p.169),

To the extent that we find linguistic differences correlating across the population in such a way that social category recedes into the background, one might ask whether the linguistic behaviour of the categories themselves is not simply derivative.

Thus, the formal theory that we are developing should also be able to capture the rationality

¹⁰Cruising involves going joy-riding around the edge of the urban centre of Detroit.

¹¹Note that (Eckert, 2000, 170) found that involvement in school organized extra-curricular activities was an even better predictor of linguistic variation than academic achievement, so her results show that students' use of variables is determined by factors other than simple mastery of the 'standard language' taught in class.

| Variable | | +Cruising | -Cruising | Input | Sig. |
|-----------------------|-------|-----------|-----------|-------|------|
| (∧) backing | Girls | .563 | .458 | .422 | .000 |
| | Boys | .530 | .460 | .447 | .014 |
| (e) backing | Girls | .544 | .464 | .331 | .029 |
| | Boys | .557 | .437 | .368 | .001 |
| (ay) raising | Girls | .765 | .381 | .011 | .000 |
| | Boys | .636 | .295 | .009 | .004 |
| (ay) monophongization | Girls | .634 | .405 | .036 | .000 |
| negative concord | Girls | .777 | .294 | .106 | .000 |
| | Boys | .637 | .338 | .241 | .000 |

Table 3 – Factor weights for cruising at Belten High (Eckert, 2000, 151)

underlying social stratification.

3.3 The Third Wave approach to variation

Above, I argued that, given its empirical properties, it is natural to attempt to model sociolinguistic variation within a game-theoretic framework. This being said, since game theoretic tools are so general, without saying anything else, we are still very far away from a full theory of social meaning and variation¹². Fortunately, there is already a well-articulated and influential theory in the sociolinguistics literature that can provide the basis for our formal implementation. As mentioned above, the Third Wave approach to variation pursues a unified analysis of style shifting and social stratification as rational language use, and (in a nutshell) the account is as follows¹³: The use of a variant in context is related to (or in this discipline’s terminology *indexes*) sets of properties, stances or other concepts/ideas that are to be attributed to the speaker (Ochs, 1992, 1993; Silverstein, 1979, 2003; Eckert, 2008, and others)). Speakers use these linguistic resources to (attempt to) construct the persona that will be the most useful to them in their context-specific goals. Here we take the notion of *goals* to be very general, encompassing concrete aims such as getting a job in a flower shop, but also more abstract things such as making friends or even communicating one’s ‘true’ self to their interlocutor. By virtue of what speakers’ goals and desires are, and by virtue of what resources they have to use, different variants will be more useful to different speakers in different contexts. Thus, in the same way that the properties indexed by *-ing* are more useful to someone like Obama in a formal setting than in an informal

¹²See also the discussion in Dror et al. (2013).

¹³Properly speaking, what I present here is just a small portion of the full TW theory. In particular, my model does not cover the parts concerning how variants come to index particular properties and their implications for a theory of language change. In the model that will be presented in the next section, variants come with associated sets of properties (indexical fields) and the model does not have any evolutionary or large-scale diachronic component. However, I believe that it would be of great interest to extend my proposal to capture these other aspects of TW in the future.

setting, the properties indexed by *-ing* are more useful to upper middle class speakers (in the context of being interviewed by a researcher) than to working class speakers¹⁴.

As described above, TW clearly makes reference to notions like interactivity and rationality; however, it is not a mathematical theory. Thus, although its insights concerning social meaning, sociolinguistic variation and identity/persona construction are clear, by virtue of its form, such a theory cannot be directly incorporated into a broader formal theory of pragmatics. I argue that such an incorporation is desirable for a number of reasons. Firstly, I argued above (following others) that inferences triggered by sociolinguistic variants should be analyzed as instances of pragmatic enrichment; thus, formal theories of language use and interpretation will be incomplete if they cannot be applied to this empirical domain. Secondly, their lack of formalization has isolated the insights of influential sociolinguistic theories such as TW inaccessible to have been isolated from work in cognitive science, computer science and artificial intelligence, to the detriment of these fields¹⁵. Finally, beyond the field of linguistic pragmatics, social meaning and persona/identity construction are fundamental theoretical notions in many disciplines of the humanities and social sciences. They play a key role in our understanding of linguistic and non-linguistic phenomena studied in anthropology, sociology, philosophy, psychology and gender studies (see Cerulo, 1997; Hacking, 1999, for overviews). Moreover, methods based on identity construction through language are widely employed outside academia, for example in education (eg. Roberts, 1991; Varelas, 2012; Kelly, 2014, among others)), management (DeRue and Ashford, 2010; Dutton et al., 2010, among others), social work (Miehls and Moffatt, 2000, among others), digital communication (Zhao et al., 2008), and even social justice (Taylor, 2000; Charmaz, 2011, among others). It is therefore crucial that our understanding of the relationship between language, meaning, identity and variation be as detailed, as explicit and as well-founded as possible. Formalization is a powerful tool that we can use to carefully distinguish between different aspects of theoretical proposals made in sociolinguistics and for precisely identifying empirical predictions made by competing analyses.

With these considerations in mind, in the second half of the paper I propose to formalize the Third Wave approach to variation using signalling games and a Bayesian approach to speaker/listener uncertainty. In doing so, I hope to bring social meaning into the domain of game-theoretic pragmatics and, more generally, identity construction through language into the domain of formal pragmatics.

¹⁴See Bourdieu and Passeron (1970); Bourdieu (1979, 1980); Lamont (2009) for detailed studies of the relationship between social class and values, goals and desires. This point will be further elaborated in section 4.2.2

¹⁵See Cameron (2016) and Hardaker (2016) for recent discussion of the importance of incorporating more sophisticated information about social meaning and its relation to identity into online gender-based harassment prevention tools. See Nguyen et al. (2016) for a comprehensive overview of the difficulties of integrating informal insights from sociolinguistics into NLP.

4 Social Meaning Games

This section presents the *social meaning game* (SMG) framework. In what follows, I will define the games in detail and illustrate the different aspects of the definitions with a small example. For convenience, the full framework is more succinctly laid out in the appendix.

As mentioned above, the framework combines a modification of Lewis (1969)’s signalling games with a probabilistic/Bayesian approach to speaker/listener beliefs and uncertainty (see Tenenbaum et al., 2011, for an overview). In a nutshell, a *signalling game* is a game of coordination between two agents, *S* (the *speaker/sender*) and *L* (the listener/receiver). *S* has a piece of information that they wish to communicate to *L* (their *type*). *S*’s action is to choose a message *m* to send *L*, and *L*’s action is to assign an interpretation to *m*, and, in doing so, update their prior beliefs about the world using the information communicated by *m* (Stalnaker, 1978; Lewis, 1979; Heim, 1982, among others). In signalling games, *S* and *L*’s payoffs are calculated based on coordination: (broadly speaking) both players win if *L* correctly interprets *S*’s message, updating their beliefs accordingly, and they both lose if *S*’s type and *L*’s interpretation do not converge, and *L* comes to believe something different about the world than that which *S* intended.

Social meaning games will have a similar structure: they are games of interaction between two agents: *S* (speaker/sender) and *L* (listener/receiver). *S* has a set of properties characterizing themselves they wish to communicate to *L* (their *type*). *S*’s action is to choose a message *m* to send *L*¹⁶, and *L*’s action is to attribute a set of properties to *S* based on *m* and their prior beliefs about *S*, and, in doing so, update their beliefs.

S and *L*’s payoffs and, consequently, the **solution concept** (or rule that determines how the game is played) will be very similar to what is found in *Iterated Best Response* (IBR) or *Rational Speech Act* (RSA) models, which have been widely used in formal approaches to Gricean pragmatics (Franke, 2009; Frank and Goodman, 2012; Lassiter and Goodman, 2013; Degen and Franke, 2012; Franke and Jäger, 2016; Degen and Tanenhaus, 2015; Bergen et al., 2016, among others). However, contrary to classic IBR/RSA models, something else will play an important a role in calculating agent’s utility (and subsequent actions): *S* and *L*’s personal preferences in the context, which we will call their *values*.

¹⁶Note that the framework does not at all assume that all or even most aspects of message/interpretation selection or utility calculation are conscious or intentional. We know from the psychological literature (for example, the literature on motion planning (Rosenbaum et al., 2007, 2012)) that we make enormously many subconscious choices and calculations when we engage in daily cognitive activities (see Dennett, 1993; Graziano, 2013, among others). Furthermore, work in the field of *evolutionary game theory* (see Gintis, 2000) has shown that even agents not possessing consciousness (like single-celled organisms) appear to engage in the kinds of utility maximizing calculations that are presented in this paper. Thus, the present proposal has nothing to say about the role/limitations of consciousness in sociolinguistic variation and interpretation. I thank Dan Lassiter and Eric McCready for discussions of these points.

4.1 Basic Setup

The definition of a SMG is laid out more formally in Def. 4.1. Some of the lines of Def. 4.1 will doubtlessly be opaque to the reader; however, they will be further elaborated in the rest of this section.

Definition 4.1. A *Social Meaning Game* is a tuple $\langle \{S, L\}, \langle \mathbb{P}, > \rangle, M, C, [\cdot], \mu, Pr \rangle$ where:

1. S and L are the players.
2. $\langle \mathbb{P}, > \rangle$ is the **universe** (a relational structure), where
 - $\mathbb{P} = \{p_1, \dots, p_n\}$ is a finite set of properties.
 - $>$ is a relation on \mathbb{P} that is irreflexive.
3. M is a finite set of **messages**.
4. C is a measure function on M describing the **cost** of each message.
5. $[\cdot]$ is the **indexation** relation (to be described below).
6. μ is a measure function on sets of properties describing S 's **values** in the context.
7. Pr is a probability distribution over sets of properties describing L 's **prior beliefs** about S .

As shown above, the basic domain of interpretation is \mathbb{P} , a set of properties. In this paper, we will have the relation $>$ encode relationships between properties, namely incompatibility; that is, $P_1 > P_2$ just in case P_1 and P_2 are contraries: they cannot both be true of an individual at the same time. This will be the extent of the structure that we will impose on the universe; however, in future extensions of the model it may be desirable to enrich the universe with scales, antonymy relations or other more complicated structures.

As a concrete example, let us consider a universe specified as shown in (14), where it is impossible to be both competent and incompetent at the same time, and it is impossible to be both friendly (where we should understand *friendliness* as also regrouping properties such solidarity and authenticity) and aloof (where we should understand *aloofness* as regrouping properties such as pretension, exclusion and snobbishness).

- (14) $\mathbb{P} = \{\text{competent, incompetent, friendly, aloof}\}$
- a. competent $>$ incompetent
 - b. friendly $>$ aloof

In addition to the attribution of individual properties and the meaning of individual variables, TW also focuses on how those variables combine together into *styles*, which are

both related to and construct particular social types called *personae* (see Podesva, 2004; Zhang, 2008; Eckert, 2008, among many others). In this paper, we take personae to be particular collections of properties that ‘go together’. Thus, the set of possible personae are the maximally consistent sets of properties, as shown in Def. 4.2.

Definition 4.2. *P is possible persona* ($P \in \text{PERS}$) iff

1. $P \subseteq \mathbb{P}$ and there are no $p_1, p_2 \in P$ such that $p_1 > p_2$. *Consistency*
2. There is no $P' \in \text{PERS}$ such that $P \subset P'$. *Maximality*

In our simple example, then, by Def. 4.2, the possible personae in the universe in (14) are shown in (15): we have the set {competent, friendly}, (what we might think of as) the ‘cool guy’ type; {competent, aloof}, the ‘stern leader’ type; {incompetent, friendly}, the ‘doofus’ type; and {incompetent, aloof}, the ‘arrogant asshole’ type.

$$(15) \quad \text{PERS} = \{\{\text{competent, friendly}\}, \{\text{competent, aloof}\}, \{\text{incompetent, friendly}\}, \{\text{incompetent, aloof}\}\}$$

As in classic signalling games, we have a set of messages, which come with a set of costs.

(16) **Messages and Costs.**

- a. $M = \{m_1, \dots, m_n\}$ is the set of messages (i.e. variants) that S can pick from.
- b. C is a function from M to the real numbers that assigns a cost to each message.

In order to show how SMGs work, we will start by showing how to model Labov (2012)’s study of President Obama’s use of (ING) across three contexts. Thus, in the game, we will have two messages (17).

| MESSAGE | COST |
|-------------|------|
| <i>-ing</i> | 0 |
| <i>-in'</i> | 0 |

How should we interpret the costs associated with variants? One idea might be to identify the cost of a message with the comfort or ease (or lack thereof) that a speaker has with manipulating it. For example, if m is a prestige or standard form which requires a certain amount of exposure/engagement with educational institutions in order to manipulate properly, then, for speakers who have not had such exposure, m would be more costly to use than a more vernacular message m' (see Bourdieu and Boltanski, 1975; Bourdieu, 1980). Parallely, if m is a highly vernacular form that the speaker is not familiar with or does not form part of the speaker’s ‘native’ dialect (as in cases of *language crossing* (Rampton, 1995; Bucholtz, 1999, 2010, among others), the same principle may apply. Since both variants of (ING) are used by members of all educational levels (Hazen, 2006) and the use of *-in'* is

not particularly stigmatized¹⁷, we will assume for our example that there are no differences in the cost of using *-in'* than in the cost of using *-ing*.

This being said, having articulated costs may become important when it comes time to integrate SMGs within a broader model of linguistic production and interpretation. As mentioned in the introduction, SMGs aim only to model the **social** or **strategic** aspect of linguistic variation. When we go to speak, which form we end up picking depends on a wide range of factors, only a subset of which depend on social meaning and persona construction. In addition to social (or what Labov calls *external*) factors, physiological or psycholinguistic factors such as ease of articulation, frequency, priming or other processing factors may play a role in favouring the use of one variant over another. Likewise, grammatical factors of the type discussed in section 2, which Labov calls *internal* factors, may induce a bias in favour of one variant over another. For example, it has been shown that (ING) is conditioned by grammatical category and other abstract properties of morphological structure (Labov, 1966; Houston, 1985; Tamminga, 2014), so it seems reasonable to capture the generalization that *-ing* is disfavoured in some grammatical environment compared to *-in'* by assigning an occurrence of *-ing* in that environment a higher cost than is assigned to *-in'* in the same context, possibly through the use of a *harmonic grammar* (Legendre et al., 1990; Smolensky and Legendre, 2006)¹⁸. Of course, adding internal/grammatical conditioning factors to the model would require much more elaborate message representations which would complicate the exposition here, so we will assign a cost of zero to both variants of (ING) in (17).

As mentioned in section 3, in Third Wave, individual variants have meaning that goes beyond their truth conditional meaning. More precisely, variants are proposed to index sets of properties, called their *indexical field* (Eckert, 2008). In SMGs, messages are proposed to be related to their field via the *indexation* relation, as shown in (18).

- (18) **Indexation relation** ([·]).
For all messages $m \in M$, $[m] \subseteq \mathbb{P}$.

Much current work within TW is devoted to studying the structure of indexical fields, investigating whether there are different orders of indexicality within a field (Silverstein, 2003), whether there are meaningful relations between the properties that make up a variant's field (Eckert, 2008), and, if so, whether there exists some kind of algorithm that can extract these relations automatically (Oushiro, 2015). In this paper, I will keep things as simple as possible and not impose any structure on these sets, but, again, the structure of the fields could easily be enriched, should we find empirical arguments in favour of doing so.

In today's example, following (simplified) Eckert (2008); Campbell-Kibler (2009), we will

¹⁷That is, (ING) is a *marker* rather than a *stereotype* in the sense of Labov (1966).

¹⁸Indeed, there exist straightforward mathematical connections between game-theoretic syntax/semantics and optimality-based syntax-semantics (Dekker and Van Rooy, 2000).

assume that the variants of (ING) are associated with the sets shown in (19), which I will call *Eckert fields*.

- (19) Eckert fields associated with (ING)
- a. [-ing] = {competent, aloof}
 - b. [-in'] = {incompetent, friendly}

The Eckert fields shown in (19) correspond to the standard representation within sociolinguistics (see the representations proposed by Campbell-Kibler, 2008, 2009; Eckert, 2008; Moore and Podesva, 2009; Walker et al., 2014; Beaton and Washington, 2015; Tyler, 2015; Drager, 2015; Oushiro, 2015); however, it turns out that the objects in (19) are not exactly of the right type to incorporate into an IBR/RSA-style model. Therefore, we will take advantage of Richard Montague’s important observation that (formally speaking) we often have multiple ways of looking at an object: either we can look at it directly, or, equivalently, we can look at it as its set of characterizing properties. Thus, in the spirit of Montague (1973), we will look at Eckert indexical fields equivalently through the personae that they have the potential to construct¹⁹; in other words, the Eckert-Montague field on {competent, friendly} will consist of all the personae that are either competent or friendly. We will call these type-lifted fields *Eckert-Montague fields*, shown in Table 4. As shown in this table, there is some overlap in the Eckert-Montague fields of (ING) but crucially only *-ing* can be used to construct the {competent, aloof} (stern leader) persona; whereas, only *-in'* can be used to construct the {incompetent, friendly} persona (the doofus).

| Variant | Eckert field | Eckert-Montague field |
|---------|-------------------------|-----------------------------------------------------------------------|
| -ing | {competent, aloof} | { comp. , aloof }, {comp., friend.}, {incomp., aloof} |
| -in' | {incompetent, friendly} | { incomp. , friend }, {comp., friend}, {incomp., aloof} |

Table 4 – Messages in Obama example

We are now ready to describe how the game is played; that is, how the moves of the speaker and the listener are chosen.

As in IBR/RSA models, the model of the speaker and the listener are parallel: the listener (L) has prior beliefs about the properties of the speaker before they speak. These beliefs can be specific (i.e. *George Bush is like X*) or general (i.e. *Americans are like X*). Likewise, if we are modelling the behaviour of the speaker, we propose that S makes a hypothesis about L’s beliefs concerning which persona S instantiates. These beliefs are represented as a probability distribution (Pr) over personae.

Returning to Labov’s Obama example: consider the *casual context* in which Obama is at the barbecue. Suppose, in this context, Obama is worried about coming off as too aloof,

¹⁹In the terminology of *Generalized Quantifier Theory* (Barwise and Cooper, 1981; Keenan and Stavi, 1986), the Eckert-Montague field on {competent, friendly} is derived through taking the Montagovian Individual on *competent or friendly*.

since he is the president. Indeed, one of the main reasons that politicians do such ‘meet and greet’ events is to try to counteract this impression. The way that we encode this worry in the model is by putting more probability mass on the personae that are aloof (the stern leader and the asshole) than on the personae that are friendly (the cool guy and the doofus) in Obama’s hypothesis about his listener’s prior beliefs, as shown in Table 5.

| Persona | stern leader | cool guy | asshole | doofus |
|---------|---------------|----------------|-----------------|------------------|
| P | {comp, aloof} | {comp, friend} | {incomp, aloof} | {incomp, friend} |
| Pr(P) | 0.30 | 0.20 | 0.30 | 0.20 |

Table 5 – Obama worries about seeming **aloof**.

As in IBR/RSA models, the social interpretation process proceeds in a couple of steps: when they hear a variant, the listener first focuses their attention on the personae in the variant’s Eckert-Montague field, discarding the other personae and updating their beliefs accordingly. More technically, L conditions their beliefs on the meaning of the message (20): they intersect each persona with the variant’s indexical field and then normalize the measure.

$$(20) \quad Pr(P|m) = \frac{Pr(\{P\} \cap [m])}{Pr([m])}$$

Conditionalization

The results of conditionalization for both variants of (ING) on listener priors at the barbecue are shown in Table 6. If the listener hears *-ing*, they are certain that their interlocutor is not a doofus, but they remain uncertain about the remaining possibilities. Likewise, if they hear *-in’*, they are certain that S is not the stern leader (because a stern leader would not say *-in’*); so they assign this persona a zero probability and normalize over the cool guy, asshole and doofus.

| | stern leader | cool guy | asshole | doofus |
|------|---------------|---------------|-----------------|-----------------|
| m | {comp, aloof} | {comp, frien} | {incomp, aloof} | {incomp, frien} |
| -ing | 0.375 | 0.25 | 0.375 | 0 |
| -in’ | 0 | 0.286 | 0.428 | 0.286 |

Table 6 – L’s beliefs immediately after hearing m at the BBQ (Pr(P|m)).

IBR/RSA models aim to formalize certain aspects of Gricean pragmatics, most notably, the role that informativity plays in successful communication. According to this framework, coordination, and therefore communication, occurs because speakers try to make the most informative statement possible, and listeners know this. Thus, informativity serves as an external convergence point for both speaker and listener. As such, message informativity is encoded as part of S’s utility function (U_S): the speaker’s utility of a message, given that they wish to construct a particular persona, is measured as the informativity of the message (given the desired persona), minus whatever costs are associated with the message

(21). Following Frank and Goodman (2012), who follow Shannon (1948), we measure the informativity of a message m for a persona P as the natural log (\ln) of the prior probability of P conditioned on the meaning of m .

$$(21) \quad U_S(m, P) = \ln(\Pr(P|m)) - C(m) \quad \text{RSA-style utility function}$$

In SMGs, we will also adopt the proposal that persona/identity construction through language is driven (in part) by informativity²⁰: the speaker is trying to give the listener the most information possible about their desired persona, and the listener assumes that the speaker is giving them (intentionally or not) the most information about the kind of person that they are. Note that the joint assumption of informativity does not require that the listener is positively disposed or actively trying to coordinate with their interlocutor; rather, they are simply trying to extract the most information possible out of their interlocutor’s linguistic offering.

If we were to adopt exactly the utility function in (21), SMGs would be almost exactly equivalent to the RSA framework, which, as mentioned above, was designed to give a Bayesian game-theoretic treatment of a variety of pragmatic phenomena involving propositional communication (scalar/specificity implicatures, resolution of the reference of context-sensitive expressions etc.). However, as discussed in Eckert (2016a), we have reason to believe that propositional communication and persona/identity construction are different²¹. For one thing, propositional communication of the kind modelled by classic signalling games is *reportative*: S observes a fact about the world (their type) and then tries to report it to L. As such, S’s type exists independently from both S’s linguistic action reporting it and S’s feelings towards it. Suppose I observe that it’s 10:30am. Whether or not I choose to tell you that it’s 10:30am and whether or not I would prefer it to be some other time are of no consequence for whether or not it is, in fact, 10:30am. Identity construction, on the other hand, is generally proposed to be dependent both on speakers’ preferences and their linguistic actions. In particular, this phenomenon has been argued to be more like promising or threatening (Butler, 1990, 1993; Bucholtz and Hall, 2005, among others): suppose I say *I promise to be there by 10:30am*. Before my linguistic act of promising, there was no promise that existed in the world; rather, my promise was brought into existence in part **through** my utterance (Austin, 1962). Because of this *performative* aspect of promising, I have a certain amount of agency in creating the promise: if I decide not to say *I promise to be there by 10:30am*, then the promise doesn’t come into existence.

And the same holds for identity. One of the most powerful ideas of the past quarter century is that personal identity is not “a stable structure located primarily in the individual psyche or in fixed social categories”, but rather is “a relational and socio-cultural phenomenon that emerges and circulates in local discourse contexts of interaction” (Bucholtz and Hall, 2005, 586). It turns out that the set of personae/identities that have been found to be

²⁰See Burnett (2017b) for arguments that informativity also plays a role in sociolinguistic production.

²¹I thank Penny Eckert for discussion of these points.

the most useful in predicting patterns of linguistic variation lies within the set **idealized social types** rather than in collections of (possibly) independently verifiable demographic properties. Personae such as *doofuses*, *arrogant assholes*, *alley saunterers* (Zhang, 2008), *gay divas* (Podesva, 2007), *sufer dudes* (Eckert, 2016b) or *valley girls* (D’Onofrio, 2015) clearly do not exist outside our social construction of them; therefore, the proposition that *The speaker is a gay diva*. is not the appropriate kind of information to constitute the type of a classical signalling game. Furthermore, even if we were to take the *social address* view of identity, in which personae are constructed from demographic properties (such as gender, age, sexual orientation, ethnicity, social class etc.), as shown by research such as Butler (1990, 1993) for gender/sex, Cameron and Kulick (2003) for sexual orientation, Alim et al. (2016) for ethnicity (among others), there are good reasons to think that a proposition like *The speaker is a middle-aged male gay WASP* is just as socially constructed through repeated interaction as a proposition like *The speaker is a gay diva*. As (Butler, 1990, 34) says (with respect to gender identity), “there is no gender identity behind the expressions of gender; that identity is performatively constituted by the very “expressions” that are said to be its results.” Thus, unlike the RSA-style set up described above (and in signalling games more generally), a model for identity/persona construction through language should capture the **performative** aspect of this linguistic phenomenon and, correspondingly, the role that speaker **agency** plays in the process.

We therefore add to the model a measure of speaker preferences over how their interlocutor will perceive them in the context²². We call these preferential relations, S’s *values* (22).

- (22) **Speaker Values.** Let μ be a measure function from personae to real numbers that encodes S’s preferences over personae the context.

For example, suppose that at the barbecue Obama prefers to come off as the cool guy. Furthermore, while the stern leader may not be his top choice at the barbecue (because it is informal), suppose that he prefers this more distinguished persona over both the doofus and the asshole (how unpresidential!). We can therefore represent Obama’s values in the barbecue context as in Table 7.

| Persona | P | $\mu(P)$ |
|------------------|-------------------------|----------|
| cool guy | {competent, friendly} | 1 |
| stern leader | {competent, aloof} | 0.5 |
| doofus | {incompetent, friendly} | 0 |
| arrogant asshole | {incompetent, aloof} | 0 |

Table 7 – Obama’s values at the BBQ

²²Again, this preferential relation is not necessarily consciously available to the speaker. In fact, I conceive of it as being determined by something along the lines of Bourdieu (1979, 1980)’s *habitus*: our habits and dispositions which have been deeply ingrained in our unconscious from our experience.

In order to capture the performative nature of identity construction, I propose to slightly modify the speaker’s action: rather than being given a type and then choose a message to report that type, speakers in SMGs will do two things simultaneously:

1. Choose a persona to construct.
2. Choose a message to construct the chosen persona.

Thus, we can represent S’s action as a choice of a particular $\langle \text{persona}, \text{message} \rangle$ pair $(\langle P, m \rangle)$, and speaker utility will be defined with reference to the informativity of m to signal P , the cost associated with m , and, inspired by Yoon et al. (2016), the value that S associates with P , as shown in (23).

$$(23) \quad \begin{array}{l} \text{For all } \langle P, m \rangle \in \text{PERS} \times M, \\ U_S(\langle P, m \rangle) = \ln(\text{Pr}(P|m)) - C(m) + \mu(P) \end{array} \quad \text{SMG utility function}$$

Plugging the values from Tables 6 and 7 into the utility function in (23) assigns the following utilities to pairs of personae and variants in Table 8.

| Persona | P | m | $U_S(\langle P, m \rangle)$ |
|------------------|-------------------------|------|-----------------------------|
| Cool guy | {competent, friendly} | -ing | -0.386 |
| | | -in' | -0.253 |
| Stern leader | {competent, aloof} | -ing | -0.481 |
| | | -in' | $-\infty$ |
| Doofus | {incompetent, friendly} | -ing | $-\infty$ |
| | | -in' | -1.253 |
| Arrogant asshole | {incompetent, aloof} | -ing | -0.981 |
| | | -in' | -0.847 |

Table 8 – Obama’s utility for trying to construct P with m at the BBQ

The most useful actions for Obama are predicted to be to try to construct the cool guy using *-in'*, followed by constructing the cool guy using *-ing* and constructing the stern leader using *-ing*.

4.2 Predicting Speaker Behaviour

A major assumption underlying game-theoretic treatments of language use and understanding is that speakers and listeners are (at least) *approximately rational*. We assume that they are *rational* in the sense that they are trying to maximize their utility; however, we assume that they are only approximately so, meaning that they may **not** in fact always pick the optimal action. It is well-known that mental computation can be impeded by a variety of things (tiredness, attention deficits etc.). Therefore, in order to account for possible variability in action selection, we assume that, rather than just picking the variant with the

highest utility, S chooses the best option given a noise-perturbed assessment of utilities. One such weaker choice rule, called the *Soft-Max Choice Rule* (Luce, 1959; Sutton and Barto, 1998), is widely used in both reinforcement learning and Bayesian game-theoretic approaches to a variety of pragmatic phenomena (Frank and Goodman, 2012; Degen et al., 2013; Lassiter and Goodman, 2015; Franke and Jäger, 2016; Bergen et al., 2016, among others). For example, in their accounts of both vague adjectives and scalar implicatures, (Lassiter and Goodman, 2015, 9) “employ a relaxed version of this model according to which agents choose stochastically, i.e., that speakers sample actions with the probability of making a choice increasing monotonically with its utility. . . Apparently sub-optimal choice rules of this type have considerable psychological motivation. They can also be rationalized in terms of optimal behavior for an agent whose computational abilities are bounded by time and resource constraints, but who can efficiently approximate optimal choices by sampling from a probability distribution”.

Set in the SMG framework, the Soft-Max choice rule looks as in (24), where $P_S(\langle P, m \rangle)$ notates the probability of S choosing to try to construct a persona P using m . The constant α is called the *temperature*, which represents how much indeterminacy the model allows. Setting α to ∞ recovers deterministic choice; whereas, setting it to a low value allows more variation.

$$(24) \quad P_S(\langle P, m \rangle) = \frac{\exp(\alpha \times U_S(\langle P, m \rangle))}{\sum_{\langle P', m' \rangle} \exp(\alpha \times U_S(\langle P', m' \rangle))} \quad \text{Soft max choice rule}$$

Using the choice rule in (24), we can calculate a probability distribution over the speaker’s choices of ⟨persona, message⟩ pairs (Table 9), and, from this table, we can recover a predicted probability distribution over variants: with the temperature set at $\alpha = 8$, we predict that Obama will use *-in'* around **67%** of the time (and *-ing* around 23% of the time), which is close to what Labov found.

| Persona | P | m | $P_S(\langle P, m \rangle)$ |
|------------------|-------------------------|------|-----------------------------|
| Cool guy | {competent, friendly} | -ing | 0.23 |
| | | -in' | 0.66 |
| Stern leader | {competent, aloof} | -ing | 0.11 |
| | | -in' | 0 |
| Doofus | {incompetent, friendly} | -ing | 0 |
| | | -in' | 0 |
| Arrogant asshole | {incompetent, aloof} | -ing | 0 |
| | | -in' | 0.01 |

Table 9 – Predicted probability of Obama trying to construct P with m at the BBQ ($\alpha = 8$)

4.2.1 Style shifting as variation in priors and/or values

After the barbecue finishes, Obama moves to take questions from reporters on the White House lawn. In this situation, we might imagine that although he is still in a relatively informal context (and so most highly values the cool guy persona), he may be worried about appearing incompetent in front of antagonistic journalists. Again, we represent this worry in Obama’s hypothesis about his interlocutor’s priors: in Table 10, personae that are incompetent are weighted higher than competent personae.

| Persona | stern leader | cool guy | asshole | doofus |
|---------|---------------|----------------|-----------------|------------------|
| P | {comp, aloof} | {comp, friend} | {incomp, aloof} | {incomp, friend} |
| Pr(P) | 0.20 | 0.20 | 0.30 | 0.30 |

Table 10 – Obama worries about seeming **incompetent**.

This change (from Table 5 to Table 10) has an important effect on Obama’s predicted linguistic choices since changing prior beliefs will cause the informativity of the messages to change, which, in turn, will cause the speaker utility of a ⟨persona, message⟩ pair to change, resulting in the speaker probability distribution shown in Table 11. Correspondingly, the predicted use of *-in’* drops to around **25%**, which again is similar to Labov’s observation.

| Persona | P | m | $P_S(\langle P, m \rangle)$ |
|------------------|-------------------------|------|-----------------------------|
| Cool guy | {competent, friendly} | -ing | 0.73 |
| | | -in’ | 0.25 |
| Stern leader | {competent, aloof} | -ing | 0.01 |
| | | -in’ | 0 |
| Doofus | {incompetent, friendly} | -ing | 0 |
| | | -in’ | 0 |
| Arrogant asshole | {incompetent, aloof} | -ing | 0.01 |
| | | -in’ | 0 |

Table 11 – Probability of Obama trying to construct P with m after the BBQ ($\alpha = 8$).

Finally, when Obama makes a speech at the DNC, he is in a very formal situation. In such contexts, it is generally not very useful to appear particularly friendly; rather, what is valued in very formal contexts is the aloofness of the stern leader.

| Persona | P | $\mu(P)$ |
|------------------|-------------------------|----------|
| cool guy | {competent, friendly} | 0.5 |
| stern leader | {competent, aloof} | 1 |
| doofus | {incompetent, friendly} | 0 |
| arrogant asshole | {incompetent, aloof} | 0 |

Table 12 – Obama’s values in front of the DNC.

If we substitute the speaker values in Table 7 with those in Table 12, we generate the probability distribution in Table 13, and correspondingly predict that Obama will use *-in'* only **1%** of the time.

| Persona | P | m | $P_S(\langle P, m \rangle)$ |
|------------------|-------------------------|------|-----------------------------|
| Cool guy | {competent, friendly} | -ing | 0.02 |
| | | -in' | 0.01 |
| Stern leader | {competent, aloof} | -ing | 0.97 |
| | | -in' | 0 |
| Doofus | {incompetent, friendly} | -ing | 0 |
| | | -in' | 0 |
| Arrogant asshole | {incompetent, aloof} | -ing | 0 |
| | | -in' | 0 |

Table 13 – Probability of Obama trying to construct P with m at the DNC ($\alpha = 8$)

Thus, the complex patterns observed in Labov’s sociolinguistic study of style shifting can be captured in the SMG framework through the interplay between speakers’ (context-sensitive) values and their hypotheses concerning how they are viewed by their interlocutor.

4.2.2 Social Stratification as variation in values

The SMG framework also provides a way to formally realize the idea inherent in TW and in many other sociolinguistic theories (discussed in section 3) that both intra-speaker (style shifting) and inter-speaker (stratification) variation can be derived from the same basic principles. In what follows, I give a quick exemplification of how stratificational patterns can be modelled in this framework. For a more detailed exposition of how SMGs can be useful in the study of the relationship between social meaning and stratificational patterns in large sociolinguistic corpora, see Burnett (2017a).

Consider again Labov (1966)’s result concerning class-based stratification of (ING) in New York City, particularly what we find in the sociolinguistic interview portion of the study reproduced in Table 14. How can a pattern such as this arise from the combination of the social meanings of the variants, speaker values and conjectures about listener prior beliefs?

| SOCIAL CLASS | APPROX. % <i>-in'</i> |
|--------------------|-----------------------|
| Upper middle class | 10 |
| Lower middle class | 40 |
| Working class | 50 |
| Lower class | 80 |

Table 14 – (ING) by social class (casual style) in NYC.

Following the line of explanation in Eckert (2000, 2012), I suggest that the key to a social meaning-based account of stratificational patterns lies in the idea that speakers of different social classes differ with respect to their preferences over personae, i.e. their value functions (μ). An abundance of work in sociology (such as Bourdieu and Passeron, 1970; Bourdieu, 1979; Gans, 1974; Lamont, 1992, 2009, among others) has detailed how individuals of different education levels and occupations value different kinds of properties in themselves and in others. For example, in a study of symbolic boundaries in working and lower-middle class culture in the United States and in France, Lamont (2009) shows that, in semi-structured interviews, working class participants expressed admiration of properties such as *interpersonal altruism*, *generosity* and *collective solidarity*, much more frequently than the lower middle class participants (see Lamont, 2009, Table on p.21). Furthermore, building on Bourdieu (1979)’s pioneering work on the relationship between social class and taste, Lamont (1992) shows that upper-middle class participants (especially those with an elite university education) are more likely to value properties such as *intelligence* and *sophistication*²³ than individuals without a university education.

With these observations in mind, we might distinguish between three classes of speakers with three different value functions: a *working class* value function (μ_{WC}), which values friendly personae over non-friendly ones; a *upper-middle class* value function (μ_{UMC}), which values competence/educated personae over uneducated personae; and, a *lower middle class* value function (μ_{LMC}), which has no great preference for friendliness or competence/education. These value functions are shown in Table 15.

| PERSONA | μ_{WC} | μ_{LMC} | μ_{UMC} |
|-------------------------|------------|-------------|-------------|
| {comptent, friendly} | 1 | 0 | 1 |
| {competent, aloof} | 0 | 0 | 1 |
| {incompetent, friendly} | 1 | 0 | 0 |
| {incompetent, aloof} | 0 | 0 | 0 |

Table 15 – Class-differentiated value functions (inspired by Bourdieu and Lamont)

One of the properties of the classic Labovian sociolinguistic interview is that the interviewer is not typically close with the person being interviewed. In fact, in many studies in this tradition, participants are interviewed by complete strangers (Tagliamonte, 2006). Thus, when modelling speaker behaviour in a stratified sociolinguistic corpus, we might assume that speakers hypothesize that their interlocutors have no particularly strong prior beliefs about them and treat the function Pr as uniform over personae²⁴.

Since we have both the functions Pr and μ for the different categories of speakers, we can

²³Although there are important differences between how exactly the French and American elites cash out these terms in Lamont (1992).

²⁴This is undoubtedly an idealization, since studies have also shown that listeners bring their ideological baggage even to the interpretation of strangers’ linguistic performances (see Campbell-Kibler, 2010; Levon, 2007, 2014, among many others); however, the context of the Labovian sociolinguistic interview is probably one in which listeners’ priors would be as close to uniform as we would ever find in real communication.

now calculate the predictions that the SMG models make for the distribution of (ING). These are found in Table 16, and they reproduce the main lines of the pattern found in Labov (1966).

| SOCIAL CLASS | APPROX. % -in' |
|--------------------|----------------|
| Upper middle class | 33 |
| Lower middle class | 50 |
| Working class | 66 |

Table 16 – SMG model prediction (uniform Pr ; values from Table 15).

Thus, I suggest that, in combination with detailed sociological and anthropological work on ideologies within a speech community, SMG models can make a contribution to the study of macro-patterns of sociolinguistic variation.

4.2.3 Predicting listener behaviour

When we modelled the speaker’s behaviour, we proposed that the speaker knew their own values and made a hypothesis concerning their listener’s prior beliefs about their identity. Correspondingly, when we look from the listener’s perspective, we assume that L knows their own prior beliefs and makes a hypothesis about their interlocutor’s values. Following work in Bayesian game-theoretic pragmatics, I propose that sociolinguistic interpretation proceeds (roughly) along the same lines as interpretation of other kinds: Bayesian-style inference²⁵. The interpretation of a socially meaningful expression is calculated by taking into account both L’s prior beliefs about S and what L thinks the likelihood of S constructing a particular persona given the message that they heard, which is determined as outlined in the previous section. More specifically, the probability that the listener assigns a particular persona P to the speaker, having heard a message m is proportional to the likelihood that S decided to choose P as the first coordinate of their pair $\langle P, m \rangle$, as shown in (25).

$$\begin{aligned}
 (25) \quad & \text{a. } P_L(P|m) \propto Pr(P) \times P_S(\langle P, m \rangle) \\
 & \text{b. } P_L(P|m) = \frac{Pr(P) \times P_S(\langle P, m \rangle)}{\sum_{P'} Pr(P') \times P_S(\langle P', m \rangle)} \qquad \text{Bayesian-style inference}
 \end{aligned}$$

In order to exemplify how the framework works, we will model a subset of Podesva et al. (2015)’s results on the interpretation of (un)released /t/ by American politicians. Following (simplified) Eckert (2008), we assume an Eckert field for released and flapped/unreleased

²⁵Because of the social constructionist aspect of identity construction, what is written in (25) is not exactly Bayesian inference since the speaker is also choosing the persona (not observing an event in the world). I thank Leon Bergen for discussion of this point.

/t/ as shown in (26) which is similar to the fields of (ING), only that we replace *(in)competence* with *(in)articulateness*²⁶.

- (26) Eckert fields associated with /t/
- a. $[t^h] = \{\text{articulate, aloof}\}$
 - b. $[r/t] = \{\text{inarticulate, friendly}\}$

For the illustration, we will consider interpretations related to articulateness/authoritativeness and friendliness assigned to three different politicians: Condoleeza Rice, Nancy Pelosi, and George W. Bush. Since the recordings that were used were of politicians' speeches, we might assume that the listener hypothesizes that the speakers are most valuing the {articulate, aloof} persona, a kind of *stern leader* social type. Thus, the μ function used by L is shown in Table 17.

| Persona | P | $\mu(P)$ |
|------------------|--------------------------|----------|
| cool guy | {articulate, friendly} | 0 |
| stern leader | {articulate, aloof} | 1 |
| doofus | {inarticulate, friendly} | 0 |
| arrogant asshole | {inarticulate, aloof} | 0 |

Table 17 – L's hypothesis concerning politicians' values

I propose that the different interpretations of (non)released /t/ assigned to different politicians arise as the result of L having different kinds of prior beliefs about each politician. Condoleeza Rice is not a particularly (in)famous politician, so we might assume that listeners' in Podesva et al.'s study do not have particularly strong beliefs about her articulateness or friendliness. We will represent this belief as a uniform prior distribution over personae (Table 18). Thus, before hearing Rice say anything, the listener has a 0.5 belief that she is articulate and a 0.5 belief that she is friendly.

| | | | | |
|-------|-----------------|------------------|-------------------|--------------------|
| P | {artic., aloof} | {artic., friend} | {inartic., aloof} | {inartic., friend} |
| Pr(P) | 0.25 | 0.25 | 0.25 | 0.25 |

Table 18 – Uniform prior beliefs about Rice.

We obtain the likelihood of Rice trying to construct a persona P with a variant m in the way described above, and the results are shown in Table 19.

And using the weights in Table 18, the weights in Table 19, and our interpretation rule (25), we obtain a measure of the listener's beliefs concerning Rice's persona after hearing

²⁶Indeed, Eckert (following others) proposes that the indexical fields of $[t^h]$ and *-ing* share some properties, but also differ on other properties (that are not be relevant in this paper). The same goes for the fields of $[r/t]$ and *-in'*.

| P | m | $P_S(\langle P, m \rangle)$ |
|--------------------------|-------|-----------------------------|
| {articulate, friendly} | t^h | 0.02 |
| | r | 0.02 |
| {articulate, aloof} | t^h | 0.90 |
| | r | 0 |
| {inarticulate, friendly} | t^h | 0 |
| | r | 0.02 |
| {inarticulate, aloof} | t^h | 0.02 |
| | r | 0.02 |

Table 19 – Probability of Rice trying to construct P with m at the DNC ($\alpha = 4$)

a variant. As shown in Table 20, although L’s posterior belief in Rice’s friendliness is a bit higher after she flaps than if she releases ($0.\overline{33}$ vs 0.02), by far the greatest difference in interpretation lies in articulateness: after hearing a release /t/, L’s posterior belief that Rice instantiates an articulate persona is 0.98; whereas, it is only $0.\overline{33}$ if she flaps²⁷.

| m | {artic., aloof} | {artic., frien} | {inartic., aloof} | {inartic., frien} |
|-------|-----------------|-----------------|-------------------|-------------------|
| t^h | 0.96 | 0.02 | 0.02 | 0 |
| r | 0 | 0.33 | 0.33 | 0.33 |

Table 20 – L’s predicted interpretation of P given m for Rice.

On the other hand, the SMG framework predicts very different results if listeners have different prior beliefs. Consider the case of Nancy Pelosi: in Podesva et al.’s study, “listeners appear to be less likely to associate released /t/ with competence-based meanings in Pelosi’s speech—such as articulateness, intelligence, or authoritative-ness” (Podesva et al., 2015, 79). According to Podesva et al., this is because there is a ‘clash’ between speakers’ beliefs about her (that she is not genuinely authoritative/articulate/intelligent) and what they perceive her as trying to do: sound authoritative/articulate/intelligent (recall the quotation in (7)). In the SMG framework, we can represent this situation as one in which L’s hypothesis about Pelosi’s values (Table 17) clash with their prior beliefs about her (Table 21).

| P | {artic., aloof} | {artic., friend} | {inartic., aloof} | {inartic., friend} |
|-------|-----------------|------------------|-------------------|--------------------|
| Pr(P) | 0.05 | 0.05 | 0.45 | 0.45 |

Table 21 – L thinks that Pelosi is not articulate.

The predictions for listener interpretations of Pelosi’s linguistic performances, given Pr in Table 21 and μ in Table 17, are shown in Table 22. There is no predicted difference

²⁷Note that given the simple indexical fields and personae presented here, I do not have an immediate account for why Podesva et al. find the *opposite* pattern with Obama than with Rice and Edwards. A promising line of analysis would be to go deeper into the different classes of personae that Obama, Edwards and Rice are likely to instantiate (see Alim and Smitherman, 2012) and to complicate the meanings of /t/; however, I leave working this out in detail to future work.

in articulateness between hearing $[t^h]$ and $[r]$ (both variants trigger almost exclusively inarticulate personae); however, there is predicted to be a difference in terms of friendliness: $[t^h]$ results in an almost certain attribution of the {inarticulate, aloof} persona; whereas, $[r]$ results in an almost certain attribution of the {inarticulate, friendly} persona.

| m | {artic., aloof} | {artic., frien} | {inartic., aloof} | {inartic., frien} |
|-------|-----------------|-----------------|-------------------|-------------------|
| t^h | 0.01 | 0 | 0.99 | 0 |
| r | 0 | 0 | 0.1 | 0.9 |

Table 22 – L’s predicted interpretation of P given m for Pelosi.

Finally, we can consider the cases of George W. Bush and Hilary Clinton. For these politicians, Podesva et al. say (p.80),

The speech of two remaining politicians—George W. Bush and Hillary Clinton—was not judged to sound different in the released and unreleased guises for any of the nine attributes, either word-medially or word-finally. We argue that listeners hold particularly strong views for these politicians, to the point that slight modifications in their speaking styles produce no effect on listener ratings. [...] Bush was rated as the least articulate, intelligent, authoritative, and sincere, often by a wide margin, and he was also rated as the second-to-least passionate and spontaneous. Focus group commentary was equally unfavorable. In a similar vein, Clinton’s speech was evaluated by focus group participants as sounding, above else, clear-irrespective of the realization of /t/. In other words, extreme ideological stances taken toward speakers may generate floor or ceiling effects in the evaluation of their speech.

The ‘bulletproofing’ effect that strong prior beliefs can have on sociolinguistic interpretation is predicted by the model. In line with the Podesva et al. quotation above, suppose, that L has an extreme belief that Bush is an *arrogant asshole*-type: {inarticulate, aloof}, shown in Table 23.

| P | {artic., aloof} | {artic., friend} | {inartic., aloof} | {inartic., friend} |
|-------|-----------------|------------------|-------------------|--------------------|
| Pr(P) | 0.01 | 0.01 | 0.97 | 0.01 |

Table 23 – L is almost sure that Bush is an asshole.

As shown in Table 24, it is predicted that using different variants will have no effect on L’s beliefs about Bush. In other words, Bush can try as much as he likes to use his linguistic resources to construct a different identity, he will be ‘stuck’ being viewed as {inarticulate, aloof}.

In sum, although the context-independent meanings of variants are fixed and shared across speakers, in the SMG framework, different prior listener beliefs can create different interpretations of those variants with different speakers in context. This is my account of the

| m | {artic., aloof} | {artic., frien} | {inartic., aloof} | {inartic., frien} |
|----------------|-----------------|-----------------|-------------------|-------------------|
| t ^h | 0 | 0 | 1 | 0 |
| r | 0 | 0 | 1 | 0 |

Table 24 – L’s predicted interpretation of P given m for Bush.

listener orientation property discussed in section 2. Furthermore, note that the framework also has a straightforward account of the phenomenon of language crossing (Rampton, 1995) mentioned in section 4.1, in which members of one dialect group use a variant commonly associated with another dialect group, and this use is interpreted as only indicating a (usually desirable) subset of the (un)desirable properties usually attributed to members of the second group²⁸.

Given my account of listener orientation (as the result of the interaction between listener prior beliefs and the optimization process inherent to the SMG architecture), one might wonder how different my analysis of social enrichment actually is from other analyses of other kinds of pragmatic enrichment within the Bayesian game-theoretic framework. Priors that are heavily weighted on certain objects and not others have been proposed to give rise to particular and sometimes strange kinds of interpretations associated with other kinds of pragmatic phenomena, such as vagueness (Lassiter and Goodman, 2013, 2015) and even scalar implicatures (Degen et al., 2015). Thus, it would be worth investigating in the future whether or not the difference between listener oriented social enrichment and ‘objective’ scalar enrichment can be reduced to the degree to which listeners’ prior beliefs impact the pragmatic computation.

5 Conclusion

In this paper I presented a new formal model of the social/strategic aspects of sociolinguistic variation, one that analyzes social meaning as a kind of pragmatic enrichment. In order to give an analysis of the fine-grained differences in social meaning between variants and how speakers can use them to construct identities and linguistic styles, I introduced the social meaning games (SMG) framework, which is a formalization of the Third Wave approach to variation as a kind of signalling game with a Bayesian-style approach to interpretation. This system thus represents the coming together of influential movements in quantitative sociolinguistics and formal pragmatics. I showed how the SMG framework can be used to predict patterns of language use based on speakers’ values, i.e. the personae they are trying to construct in the context and their beliefs about how their interlocutors will interpret their linguistic offerings. I also showed how the system can be used to predict patterns of linguistic interpretation based on listeners’ prior beliefs about the speaker and their

²⁸See, for example, Bucholtz (1999, 2010) for the properties of crossing and the role of white American boys’ use of African American variants in the construction of white masculinity in the US.

hypotheses about the speaker's values.

As stated, the SMG framework opens up a number of clear paths for future research into social meaning and language variation and change, the most pressing ones having to do with the relationship between social meaning and the grammar. I have said very little in this article about how social meaning is grammatically constructed and where the indexical fields and persona-based inferences 'live' in relation to other kinds of non-at-issue content such as presuppositions, scalar implicatures and expressive content. This was principally because I argued in section 2 that there are reasons to think that at least some kinds of social enrichment interact in non-trivial ways with grammatical structure. As such, the natural next step would be to integrate the SMG approach into a broader model of grammar and language processing. Additionally, in this paper, I have only treated individual variants (eg. (ING) and /t/ release), so a next related step would be to investigate extending the current static system that treats single messages (variants) to a dynamic system treating sequences of messages (styles). Finally, since this paper has dealt only with individual variants in isolation, in reality there is nothing in what I have said that limits the proposals in this paper to the meaning sociophonetic or even morphosyntactic variation. Since the game-theoretic tools used here are so general, the system extends directly to the meaning of non-linguistic variation such as systems of makeup, dress or other kinds of social signalling systems. Indeed, in the SMG approach, the differences between linguistic variation and sartorial variation would simply boil down to the inventory of messages, the kinds of meanings that can be associated with them, and the particular 'grammatical' rules that can be used to combine messages together in a more or less compositional way. In conclusion, then, I suggest that the new formal tools developed in this paper have rich applications within linguistics and semiotics more generally, and that they have potential to yield new theoretical and empirical insights into the relationship between form, meaning, identity and stylistic performance.

References

- Acton, E. (2016). Beyond grice: A socio-pragmatic framework for non-entailed meaning. *Linguistic Society of America Annual Meeting*.
- Acton, E. K. (2014). *Pragmatics and the social meaning of determiners*. PhD thesis, Stanford University.
- Acton, E. K. and Potts, C. (2014). That straight talk: Sarah palin and the sociolinguistics of demonstratives. *Journal of Sociolinguistics*, 18(1):3–31.
- Alim, H. S., Rickford, J. R., and Ball, A. F. (2016). *Raciolinguistics: How language shapes our ideas about race*. Oxford University Press.
- Alim, H. S. and Smitherman, G. (2012). *Articulate while Black: Barack Obama, language, and race in the US*. Oxford University Press.

- Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14(03):471–485.
- Austin, J. (1962). How to do things with words.
- Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. In *Philosophy, Language, and Artificial Intelligence*, pages 241–301. Springer.
- Beaton, M. E. and Washington, H. B. (2015). Slurs and the indexical field: the pejoration and reclaiming of favelado ?slum-dweller? *Language Sciences*, 52:12–21.
- Bell, A. (1984). Language style as audience design. *Language in society*, 13(02):145–204.
- Bell, A. (1997). Language style as audience design. In *Sociolinguistics*, pages 240–250. Springer.
- Beltrama, A. (2016). *Bridging the gap. Intensifiers between semantic and social meaning*. PhD thesis, University of Chicago.
- Benz, A., Jäger, G., Van Rooij, R., and Van Rooij, R. (2004). *Game theory and pragmatics*. Springer.
- Bergen, L., Levy, R., and Goodman, N. D. (2016). Pragmatic reasoning through semantic inference. *Semantics & Pragmatics*.
- Bourdieu, P. (1977). The economics of linguistic exchanges. *Social science information*, 16(6):645–668.
- Bourdieu, P. (1979). *La distinction: critique sociale du jugement*. Les éditions de minuit, Paris.
- Bourdieu, P. (1980). *Le sens pratique*. Les éditions de minuit, Paris.
- Bourdieu, P. and Boltanski, L. (1975). Le fétichisme de la langue. *Actes de la recherche en sciences sociales*, 1:2–32.
- Bourdieu, P. and Passeron, J.-C. (1970). *La reproduction: Éléments pour une théorie du système d'enseignement*. Les éditions de minuit, Paris.
- Bucholtz, M. (1996). Geek the girl: Language, femininity, and female nerds. *Gender and belief systems*, pages 119–132.
- Bucholtz, M. (1999). You da man: Narrating the racial other in the production of white masculinity. *Journal of sociolinguistics*, 3(4):443–460.
- Bucholtz, M. (2010). *White kids: Language, race, and styles of youth identity*. Cambridge University Press.
- Bucholtz, M. and Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse studies*, 7(4-5):585–614.

- Bunin Benor, S. (2001). The learned /t/: phonological variation in orthodox jewish english. *University of Pennsylvania Working Papers in Linguistics*, 7(3):2.
- Burnett, H. (2017a). A meaning-driven approach to stratification in large sociolinguistic corpora. *under review*.
- Burnett, H. (2017b). Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of sociolinguistics*, accepted.
- Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.
- Butler, J. (1993). *Bodies that matter: On the discursive limitations of sex*. Routledge, New York.
- Calder, J. (2016). Hand/s/ome women: the role of /s/ in multi-modal gender performances among SoMa drag queens. In *New Ways of Analyzing Variation*, volume 45.
- Cameron, D. (2016). Misogyny by the numbers.
- Cameron, D. and Kulick, D. (2003). *Language and sexuality*. Cambridge University Press.
- Campbell-Kibler, K. (2006). *Listener perceptions of sociolinguistic variables: The case of (ING)*. PhD thesis, Stanford University.
- Campbell-Kibler, K. (2007). Accent,(ing), and the social logic of listener perceptions. *American speech*, 82(1):32–64.
- Campbell-Kibler, K. (2008). I'll be the judge of that: Diversity in social perceptions of (ing). *Language in Society*, 37(05):637–659.
- Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change*, 21(1):135–156.
- Campbell-Kibler, K. (2010). Sociolinguistics and perception. *Language and Linguistics Compass*, 4(6):377–389.
- Cerulo, K. A. (1997). Identity construction: New issues, new directions. *Annual review of Sociology*, 23(1):385–409.
- Charmaz, K. (2011). Grounded theory methods in social justice research. *The Sage handbook of qualitative research*, 4:359–380.
- Chemla, E. and Spector, B. (2011). Experimental evidence for embedded scalar implicatures. *Journal of semantics*, pages 359–400.
- Cheshire, J. (1982). Variation in an english dialect: A sociolinguistic study. *Cambridge Studies in Linguistics London*, 37.
- Chierchia, G., Fox, D., and Spector, B. (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In *Semantics: An International Handbook of Natural Language Meaning*. Mouton de Gruyter, Berlin.

- Clark, R. L. (2014). *Meaningful games: Exploring language with game theory*. MIT Press.
- Cohen, L. J. (1971). Some remarks on grice's views about the logical particles of natural language. In *Pragmatics of natural languages*, pages 50–68. Springer.
- D'Arcy, A. F. (2005). *Like: Syntax and development*. PhD thesis, University of Toronto.
- Degen, J. and Franke, M. (2012). Optimal reasoning about referential expressions. *Proceedings of SemDIAL*.
- Degen, J., Franke, M., and Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. In *Proceedings of the 35th annual conference of the cognitive science society*, pages 376–381.
- Degen, J. and Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science*, 39(4):667–710.
- Degen, J., Tessler, M. H., and Goodman, N. D. (2015). Wonky worlds: Listeners revise world knowledge when utterances are odd. In *CogSci*.
- Dekker, P. and Van Rooy, R. (2000). Bi-directional optimality theory: An application of game theory. *Journal of Semantics*, 17(3):217–242.
- Dennett, D. C. (1993). *Consciousness explained*. Penguin UK.
- DeRue, D. S. and Ashford, S. J. (2010). Who will lead and who will follow? a social process of leadership identity construction in organizations. *Academy of Management Review*, 35(4):627–647.
- Deshaises, D. and Laforge, E. (1981). Le futur simple et le futur proche dans le français parlé dans la ville de québec. *Langues et Linguistique*, 7:21–37.
- D'Onofrio, A. (2015). Persona-based information shapes linguistic perception: Valley girls and california vowels. *Journal of Sociolinguistics*, 19(2):241–256.
- Drager, K. K. (2015). *Linguistic variation, identity construction and cognition*. Language Science Press.
- Dror, M., Granot, D., and Yaeger-Dror, M. (2013). Speech variation, utility, and game theory. *Language and Linguistics Compass*, 7(11):561–579.
- Dror, M., Granot, D., and Yaeger-Dror, M. (2014). Teaching & learning guide for speech variation, utility, and game theory. *Language and Linguistics Compass*, 8(6):230–242.
- Ducrot, O. (1969). Présupposés et sous-entendus. *Langue française*, 4:30–43.
- Dutton, J. E., Roberts, L. M., and Bednar, J. (2010). Pathways for positive identity construction at work: Four types of positive identity and the building of social resources. *Academy of Management Review*, 35(2):265–293.

- Eckert, P. (1989). *Jocks and burnouts: Social categories and identity in the high school*. Teachers College Press.
- Eckert, P. (2000). *Language variation as social practice: The linguistic construction of identity in Belten High*. Wiley-Blackwell.
- Eckert, P. (2005). Variation, convention, and social meaning. In *Annual Meeting of the Linguistic Society of America*. Oakland CA, volume 7.
- Eckert, P. (2008). Variation and the indexical field. *Journal of sociolinguistics*, 12(4):453–476.
- Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology*, 41:87–100.
- Eckert, P. (2016a). The meaning of meaning in variation. *New Ways of Analyzing Variation*, 45.
- Eckert, P. (2016b). Variation, meaning and social change. *Sociolinguistics: Theoretical Debates*, page 68.
- Frank, M. C. and Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.
- Franke, M. (2009). *Signal to act: Game theory in pragmatics*. PhD thesis, Institute for Logic, Language and Computation.
- Franke, M. and Jäger, G. (2016). Probabilistic pragmatics, or why bayes’ rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, 35:3–44.
- Gans, H. (1974). *High culture and popular culture: an analysis and evaluation of taste*. Nova York, Basic Books.
- Gintis, H. (2000). *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton university press.
- Goffman, E. (1961). *Encounters: Two studies in the sociology of interaction*. Bobbs-Merrill.
- Goffman, E. (1967). *Interaction ritual: essays on face-to-face interaction*. Aldine.
- Goffman, E. (1970). *Strategic interaction*, volume 1. University of Pennsylvania Press.
- Goffman, E. (1977). The arrangement between the sexes. *Theory and society*, 4(3):301–331.
- Goodman, N. D. and Lassiter, D. (2014). Probabilistic semantics and pragmatics: Uncertainty in language and thought. *Handbook of Contemporary Semantic Theory*. Wiley-Blackwell.
- Gratton, C. (2016). Resisting the gender binary: The use of (ING) in the construction of non-binary transgender identities. *Penn working papers in linguistics*, 22.
- Graziano, M. S. (2013). *Consciousness and the social brain*. Oxford University Press.

- Grice, P. (1975). Logic and conversation. *Syntax and Semantics*, 3:41–58.
- Hacking, I. (1999). *The social construction of what?* Harvard university press.
- Hardaker, C. (2016). Misogyny machines, and the media: or how science should not be reported.
- Hazen, K. (2006). In/ing variable. In Brown, K., editor, *Encyclopedia of language and linguistics*, pages 580–581. Elsevier.
- Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. PhD thesis, University of Massachusetts, Amherst.
- Houston, A. (1985). *Continuity and change in English morphology: The variable (ING)*. PhD thesis, University of Pennsylvania.
- Jäger, G. (2011). Game-theoretical pragmatics. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 467–491. Elsevier, Amsterdam.
- Kaplan, D. (1999). The meaning of ouch and oops: Explorations in the theory of meaning as use. *Manuscript, UCLA*.
- Keenan, E. L. and Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and philosophy*, 9(3):253–326.
- Kelly, G. J. (2014). Discourse practices in science learning and teaching. *Handbook of research on science education*, 2:321–336.
- Kessler, S. J. and McKenna, W. (1978). *Gender: An ethnomethodological approach*. University of Chicago Press.
- Kiesling, S. F. (1998). Men’s identities and sociolinguistic variation: The case of fraternity men. *Journal of Sociolinguistics*, 2(1):69–99.
- Kiesling, S. F. (2009). Style as stance. *Stance: sociolinguistic perspectives*, 171.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3):273–309.
- Labov, W. (1966). *The social stratification of English in New York city*. Center for Applied Linguistics.
- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Labov, W. (2012). *Dialect diversity in America: The politics of language change*. University of Virginia Press.
- Lambert, W. E., Hodgson, R. C., Gardner, R. C., and Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, 60(1):44.
- Lamont, M. (1992). *Money, morals, and manners: The culture of the French and the American upper-middle class*. University of Chicago Press.

- Lamont, M. (2009). *The dignity of working men: Morality and the boundaries of race, class, and immigration*. Harvard University Press.
- Lassiter, D. (2008). Semantic externalism, language variation, and sociolinguistic accommodation. *Mind & Language*, 23(5):607–633.
- Lassiter, D. and Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Semantics and Linguistic Theory*, volume 23, pages 587–610.
- Lassiter, D. and Goodman, N. D. (2015). Adjectival vagueness in a bayesian model of interpretation. *Synthese*, pages 1–36.
- Legendre, G., Miyata, Y., and Smolensky, P. (1990). *Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations*. Citeseer.
- Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.
- Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society*, 36(04):533–554.
- Levon, E. (2014). Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society*, 43(05):539–566.
- Lewis, D. (1969). *Convention*. Harvard UP, Cambridge.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of philosophical logic*, 8(1):339–359.
- Luce, R. D. (1959). On the possible psychophysical laws. *Psychological review*, 66(2):81.
- Maddeaux, R. and Dinkin, A. (2015a). Is like like like? evaluating the same variant across multiple variables. *New Ways of Analyzing Variation*, 44.
- Maddeaux, R. and Dinkin, A. (2015b). Is like like like? evaluating the same variant across multiple variables. *New Ways of Analyzing Variation*, 44.
- McConnell-Ginet, S. (2011). *Gender, sexuality, and meaning: Linguistic practice and politics*. Oxford University Press.
- McCready, E. (2012). Emotive equilibria. *Linguistics and Philosophy*, 35(3):243–283.
- McCready, E., Asher, N., and Paul, S. (2012). Winning strategies in politeness. In *JSAI International Symposium on Artificial Intelligence*, pages 87–95. Springer.
- Miehls, D. and Moffatt, K. (2000). Constructing social work identity based on the reflexive self. *British Journal of Social Work*, 30(3):339–348.
- Montague, R. (1973). The proper treatment of quantification in ordinary english. In *Approaches to natural language*, pages 221–242. Springer.

- Moore, E. and Podesva, R. (2009). Style, indexicality, and the social meaning of tag questions. *Language in Society*, 38(04):447–485.
- Mühlenbernd, R. (2013). *Signals and the Structure of Societies*. PhD thesis, Universität Tübingen.
- Mühlenbernd, R. and Franke, M. (2012). Signaling conventions: who learns what where and when in a social network. *Proceedings of EvoLang IX*, pages 242–249.
- Nguyen, D., Doğruöz, A. S., Rosé, C. P., and de Jong, F. (2016). Computational sociolinguistics: A survey. *Computational Linguistics*.
- Ochs, E. (1992). Indexing gender. *Rethinking context: Language as an interactive phenomenon*, 11:335.
- Ochs, E. (1993). Constructing social identity: a language socialization perspective. *Research on Language and Social Interaction*, 26:287–306.
- Osborne, M. J. and Rubinstein, A. (1994). *A course in game theory*. MIT press.
- Oushiro, L. (2015). Social meanings of (-r) in sao paulo: a computational approach for modeling the indexical field. *NWAV 44*.
- Podesva, R. (2004). On constructing social meaning with stop release bursts. In *Sociolinguistics Symposium*, volume 15.
- Podesva, R. (2006). *Phonetic detail in sociolinguistic variation: Its linguistic significance and role in the construction of social meaning*. PhD thesis, Stanford University.
- Podesva, R. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of sociolinguistics*, 11(4):478–504.
- Podesva, R. J., Reynolds, J., Callier, P., and Baptiste, J. (2015). Constraints on the social meaning of released/t: A production and perception study of us politicians. *Language Variation and Change*, 27(01):59–87.
- Poplack, S. and Turpin, D. (1999). Does the FUTUR have a future in (canadian) french? *Probus*, 11:133–64.
- Potts, C. (2007). The expressive dimension. *Theoretical linguistics*, 33(2):165–198.
- Potts, C., Lassiter, D., Levy, R., and Frank, M. C. (2015). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, page ffv012.
- Quinley, J. and Mühlenbernd, R. (2012). Conquest, contact, and convention: simulating the norman invasion’s impact on linguistic usage. *Proceedings of BRIMS 2012*, pages 113–118.
- Rampton, B. (1995). Language crossing and the problematisation of ethnicity and socialisation. *Pragmatics*, 5(4).

- Récanati, F. (2003). Embedded implicatures. *Philosophical perspectives*, 17(1):299–332.
- Rickford, J. and Closs Traugott, E. (1985). Symbol of powerlessness and degeneracy, or symbol of solidarity and truth? paradoxical attitudes towards pidgins and creoles. In Greenbaum, S., editor, *The English Language Today*, pages 252–261. Pergamon, Oxford.
- Roberts, B. (1991). Music teacher education as identity construction. *International Journal of Music Education*, 18(1):30–39.
- Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., and van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychological bulletin*, 138(5):924.
- Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., and Van Der Wel, R. (2007). The problem of serial order in behavior: Lashley’s legacy. *Human movement science*, 26(4):525–554.
- Sankoff, G., Wagner, S. E., and Jensen, L. (2012). The long tail of language change: Québécois french futures in real time. *University of Pennsylvania Working Papers in Linguistics*, 18(2):13.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.
- Silverstein, M. (1979). Language structure and linguistic ideology. *The elements: A parasesion on linguistic units and levels*, pages 193–247.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3):193–229.
- Smith, E., Hall, K. C., and Munson, B. (2010). Bringing semantics to sociophonetics: Social variables and secondary entailments. *Laboratory Phonology*, 1(1):121–155.
- Smolensky, P. and Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar (Cognitive architecture)*, Vol. 1. MIT Press.
- Stalnaker, R. (1978). Assertion. In Cole, P., editor, *Syntax and semantics*, volume 9, pages 315–332. Academic Press, New York.
- Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT press.
- Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge University Press.
- Tamminga, M. (2014). *Persistence in the production of linguistic variation*. PhD thesis, University of Pennsylvania.
- Taylor, D. E. (2000). The rise of the environmental justice paradigm: Injustice framing and the social construction of environmental discourses. *American behavioral scientist*, 43(4):508–580.

- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285.
- Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban british english of norwich. *Language in society*, 1(02):179–195.
- Tyler, J. C. (2015). Expanding and mapping the indexical field: Rising pitch, the uptalk stereotype, and perceptual variation. *Journal of English Linguistics*, 43(4):284–310.
- van Hofwegen, J. (2017). *Everyday styles: Investigating the full scope of variation in the life of an individual speaker*. PhD thesis, Stanford University.
- Van Rooy, R. (2003). Being polite is a handicap: Towards a game theoretical analysis of polite linguistic behavior. In *Proceedings of the 9th conference on Theoretical aspects of rationality and knowledge*, pages 45–58. ACM.
- Varelas, M. (2012). *Identity construction and science education research: Learning, teaching, and being in multiple contexts*, volume 35. Springer Science & Business Media.
- Wagner, S. E. and Sankoff, G. (2011). Age grading in the montréal french inflected future. *Language Variation and Change*, 23(3):275.
- Walker, A., García, C., Cortés, Y., and Campbell-Kibler, K. (2014). Comparing social meanings across listener and speaker groups: The indexical field of spanish/s. *Language Variation and Change*, 26(02):169–189.
- Weinreich, U., Labov, W., and Herzog, M. (1968). Empirical foundations for a theory of language change. In Lehmann, W. P., editor, *Directions for historical linguistics: A symposium*, pages 95–195. University of Texas Press, Austin.
- Yoon, E., Tessler, M., Goodman, N., and Frank, M. (2016). Talking with tact: Polite language as a balance between kindness and informativity. *Proceedings of the 38th annual meeting of the Cognitive science society*.
- Zhang, Q. (2008). Rhotacization and the ‘beijing smooth operator’: the social meaning of a linguistic variable. *Journal of Sociolinguistics*, 12(2):201–222.
- Zhao, S., Grasmuck, S., and Martin, J. (2008). Identity construction on facebook: Digital empowerment in anchored relationships. *Computers in human behavior*, 24(5):1816–1836.