1 A Formal Semantics for Social Meaning

This book presents a new framework for studying the relation between language, ideologies and the social world. The framework combines two main ideas. The first idea is that tools from formal semantics can be used to formalize theories from sociolinguistics, linguistic anthropology and feminist/gender studies. Formal semantics is the domain of linguistics that uses logic and mathematics to study linguistic meaning (see Dowty et al., 1981; Gamut and Gamut, 1991; Heim and Kratzer, 1998, Chierchia and McConnell-Ginet, 2000, for textbook overviews). The second idea is that tools from epistemic game theory can be used to bring those theories in closer line with empirical studies of sociolinguistic variation and identity construction through language. Game theory is a mathematical framework for studying interaction between agents (see Osborne et al. (2004) for an introduction), and its epistemic branch focuses on how the reasoning of sentient agents, like humans, affects how their interactions unfold (see Perea, 2012; Pacuit and Roy, 2017 for introductions). I argue that a game-theoretic framework, elaborated using formal semantics and informed by sociolinguistic theory, can make significant contributions to our knowledge of how speakers use their linguistic resources to construct their identities and the role that language and identity play in gender-based inequalities and discrimination.

This introductory chapter presents an overview of the main empirical phenomenon studied in this work: *social meaning*. It then presents the main lines of the framework that I will develop in subsequent chapters, and the main mathematical tools that I will use to develop this framework: decision theory and game theory.

1.1 Social Meaning: An Overview

Although the meaning of the term *social meaning* varies widely across the humanities, following Podesva (2011) I will use this term to refer to how information encoded in pronunciations, morphemes, words or constructions expresses aspects of speakers' properties, attitudes and identities. I will use the phrase *identity construction through language* to refer to how we use socially meaningful language to build, establish and reinforce our place(s) in our communities of practice.

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

The empirical domain of linguistic social meaning is very large and includes linguistic phenomena that have been studied in semantics and pragmatics, philosophy of language, sociolinguistics and linguistic anthropology. In semantics and philosophy, the most frequently studied social meaning phenomena are *expressives*: items that express a speaker's attitude towards an individual or an event (Potts, 2005). For example, swear words, such as *fucking* (1-a), usually express a speaker's heightened attitude towards an individual or event. As discussed by McCready (2012), the English word itself does not specify whether the attitude is positive (1-a) or negative (1-b), and the listener must reason about aspects of the utterance, the context and the speaker's identity to correctly identify it. This reasoning process will be one of the primary focuses of this book.

(1) Mike Tyson won another fight. (McCready, 2012)
 a. **Fucking** Mike Tyson won another fight. He's wonderful!
 b. **Fucking** Mike Tyson got arrested again for domestic violence.
 # He's wonderful!

Another class of expressions that have been analysed as expressives are *slurs*, like *dyke* (2-b) (Kaplan, 1999; Kratzer, 1999; Potts, 2005; Jeshion, 2013a, among others). Slurs often appear to express the speaker's negative attitude towards their referent, in a way that more 'neutral' identity terms, like *lesbian* (2-a), do not. The attitudes expressed by slurs, and the pragmatic and social functions of these elements, will be the topic of Chapter 4.

(2) Slurs and identity categories (Kaplan, 1999; Potts, 2005)
 a. Heather is a **lesbian**.
 b. Heather is a **dyke**.

In addition to attitudes, socially meaningful language can also express aspects of a speaker's place in their communities in terms of their relationship to other people. For example, depending on the community and the context, using the French second person pronoun *tu* (3-b) signals that the speaker views themself as having a closer or more intimate relationship with the addressee than if they used the pronoun *vous* (3-a). Similarly, *honorific* morphemes, like Japanese *o-* (4-b), communicate that the speaker honours the subject of the utterance, and *terms of address*, such as *dude* (5-b) or *sweetheart* (5-c) also communicate the speaker's beliefs about their relationship with the addressee: that it is one of cool non-sexual solidarity (see Kiesling, 2004) or sexist condescension (see Shear, 2010; Cameron, 2019).

(3) a. Je peux **vous** aider?
 b. Je peux **t'**aider?
 'Can I help you?' (Brown et al., 1960)

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

(4) a. Sam-ga warat-ta.
Sam-NOM laugh-PAST
'Sam laughed.'
b. Sam-ga **o**-warai-**ninat**-ta.
Sam-NOM **subj.hon**-Iaugh-**subj.hon**-PAST
'Sam laughed.' (Potts and Kawahara, 2004, 253)

(5) a. What are we doing tonight?
b. **Dude**, what are we doing tonight? (Kiesling, 2004)
c. **Sweetheart**, what are we doing tonight?

Most of the examples cited so far concern expressions whose sole (or at least primary) function is to help the speaker communicate information about their attitudes and place in society, including aspects of their age, generation, nationality, etc. However, social meaning can be associated with almost any expressions in a language. For example, as discussed in Acton (2014, 2019), a speaker can use the English definite determiner *the* in utterances like (6-b) to distance themselves from the group denoted by the noun phrase (in this case *Americans*); a speaker can use a *precise* number (7-b) rather than a *round* number (7-a) to construct themselves as knowledgeable and confident, although if they're not careful they may come off as arrogant and pedantic (Beltrama, 2019); and in languages that mark grammatical gender, like French, the use of masculine grammatical gender to refer to a woman (8-b) can signal the speaker's advanced age or socially conservative views (Abbou, 2011b; Burnett and Bonami, 2019b, among others), much in the same way that insisting on using an explicitly masculine marked noun (*chairman* instead of *chair* or *chairperson*) might do so in English.

(6) a. Americans love fast cars.
b. **The** Americans love fast cars. (Acton, 2014, 2019)

(7) a. The package was delivered at **9** pm.
b. The package was delivered at **9:03** pm. (Beltrama, 2019)

(8) a. **La** professeur**e** a oublié son livre.
b. **Le** professeur a oublié son livre.
'The (female) professor forgot her book.'

Much of the literature on social meaning and identity construction in sociolinguistics has been focused on *sociophonetic variants*: different pronunciations of the same word. Two examples of sociophonetic variants that appear in most dialects of English are (ING) (9) (i.e. having either a velar or alveolar pronunciation of the final consonant in a word like *working* (Labov, 1966; Hazen, 2006; Tamminga, 2014)) and *t-release* (having more or less aspiration on the final consonant in a word like *meet* (Bucholtz, 1999; Bunin Benor, 2001; Podesva et al., 2015)).

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

(9) a. I'm work**ing** on my paper. [iŋ]
b. I'm work**in'** on my paper. [in]

(10) a. We should mee[**t^{h}**]. released 't'
b. We should mee[**t**]. unreleased 't'

Unlike alternations involving whole words or phrases, the informational differences communicated by sociophonetic variants are so subtle that we must often use a methodology more sophisticated than introspection to observe them. One of the main ways in which social meaning differences between sociophonetic variants can be diagnosed, which has been commonly used in social psychology and variationist sociolinguistics, is through an experimental paradigm known as the *matched guise technique* (MGT) (Lambert, 1967). 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 under investigation. 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 far as possible, modulo the forms under study, so that any observed differences in inferences that participants draw from different guises can be attributable to the variable under study, not to some other aspect of the voice of the speaker or of the content of their discourse. To give an example: Campbell-Kibler (2007) performed an MGT study with American college students investigating how the use of the variable (ING) influences listener beliefs and perceptions. This study yielded a variety of complex patterns; however, one of her main results was 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 a similar vein, Podesva et al. (2015) investigated the social meaning of the t-release variable (10) through an MGT study with American participants using stimuli formed from political speeches of six American politicians (Barak Obama, John Edwards, Nancy Pelosi, George W. Bush, Hilary Clinton, and Condoleezza Rice). As in Campbell-Kibler's study, the t-release study yielded a number of results concerning associations with released vs unreleased/flapped /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 wat*[ɾ]er*.). On the other hand, Nancy Pelosi was rated as significantly less friendly and less sincere when she used released /t/, and Barak Obama was rated as significantly more passionate in his flapped guise than in his released /t/ guise. Thus, this methodology allows us to assess how socially meaningful language affects listeners' subtle beliefs about speaker identity.

Although the MGT is particularly useful for studying sociophonetic variables, it has also been used to study the social meaning of variants beyond the domain of sounds. For example, using the MGT paradigm, Maddeaux and Dinkin (2017) show that speakers using the discourse particle *like* to modify a noun phrase (11-b) sound less articulate and intelligent to listeners in Toronto, Canada than speakers who do not use *like* (11-a).

(11) a. This speech she had to give about herself . . .
b. This, **like**, speech she had to give about herself . . .

Likewise, Beltrama and Staum Casasanto (2017) show that speakers using the English intensifier *totally* to modify a relative gradable adjective (12-c) are rated as more friendly, outgoing and cool than those using variants *really* (12-b) and *very* (12-a). At the same time, speakers using (12-c) are rated as significantly less intelligent, mature and articulate than those using another intensifier.

(12) a. John is **very** tall.
b. John is **really** tall.
c. John is **totally** tall.

Finally, Beltrama (2019) shows that speakers using precise numerical expressions (such as (7-b), repeated below as (13-b)) are rated as more articulate, intelligent and educated than those using less precise expressions (13-b); however, they are also perceived as more annoying, obsessive, pedantic and uptight than authors of more approximate statements.

(13) a. The package was delivered at **9** pm.
b. The package was delivered at **9:03** pm. (Beltrama, 2019)

The MGT has also been instrumental in diagnosing social meaning differences not only between elements of a single language, but also between languages themselves; indeed, the MGT was actually first developed to investigate the different properties attributed to speakers of French vs English in Montréal, Québec (Lambert et al., 1960; Lambert, 1967). For a more detailed example of social meaning differences between languages, we can consider Kit Woolard's work on the social meaning of Castilian vs Catalan in Barcelona (Woolard, 1989, 2009, 2016; Woolard and Gahng, 1990). Woolard began her investigations of this topic in 1980, shortly after Catalonia became an autonomous region. Under the Franco regime, the Catalan language was repressed in Catalonia, with Castilian being the sole language of government and education. Furthermore, large numbers of Castilian speakers from southern Spain in the 1960s, although by the 1970s, this immigration had stagnated. After Franco's death, Catalonia became autonomous and became officially bilingual in 1979. Starting in the 1980s, the Catalan government enacted aggressive

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

policies making Catalan the language of government and education. One of the results of this complex social situation is that, starting in the 1980s, large portions of the population of Barcelona became bilingual to some degree, having at least a passive understanding of both Catalan and Castilian (Woolard, 1991, 2009).

Woolard wanted to investigate the social meanings of the two languages in Barcelona in the context of this situation of bilingualism. She therefore performed an MGT experiment in 1980 (Woolard, 1989; Woolard and Gahng, 1990) with young native Catalan and Castilian listeners, where the guises consisted in the same speakers speaking in Catalan and then in Castilian. She found a complicated pattern of social interpretations, one which partially broke down along ethnic lines. Firstly, she found that Catalan guises were rated significantly higher on what Woolard calls *status* properties: 'intelligent' (*intelligent*), 'cultured' (*persona culta*), 'hardworking' (*persona treballadora*), and to a lesser extent in those of 'self-confident' (*té confiança en ella mateixa*) and 'worthy of confidence' (*digna de confiança*). This was regardless of whether the speaker in the Catalan guise was a native Catalan or native Castilian. Woolard suggests that this social meaning derives from social differences between native Catalans and native Castilians which are observed all across society: 'in the workplace, where Catalans are more often found in managerial positions and Castilian-speaking immigrants in manual labor; in residential neighborhoods, where Catalans tend to occupy prime locations and Castilian immigrants the high-rises of the periphery; in private shops and services, where Catalans are more often owners, particularly in the more desirable areas, and Castilian speakers more often clients.' (Woolard, 1985, 742)

Secondly, she found differences between native Castilian and native Catalan listeners with respect to what Woolard calls *solidarity* (or *likeability*) properties: 'likeable' (*simpàtica*), 'amusing' (*divertida*), 'has a sense of humor' (*té sentit de l'humor*), 'open' (*oberta*), 'attractive' (*atractiva*) and 'generous' (*generosa*). Native Catalan listeners gave native Catalan speakers higher solidarity ratings in Catalan guises, and lower solidarity ratings to native Catalans speaking Castilian. Likewise, native Castilian speakers gave native Castilian speakers higher solidarity ratings when they spoke Castilian, penalizing them on solidarity when they spoke Catalan. Both Catalan and Castilian speakers gave neutral solidarity ratings to members of the opposite ethnolinguistic group when they spoke their own language. In other words, 'listeners rewarded linguistically identifiable co-members of their ethnolinguistic group for using their own language, and penalized them with significantly lower solidarity ratings when they used the out-group language' (Woolard, 2009, 133).

In addition to showing that even the language of communication itself can have social meaning, Woolard's work, particularly Woolard (1991), also builds an important link between social meaning, as observed through listener-internal judgements about the properties of the speaker, and external aspects of speaker behaviour such as language use. She says (p. 64),

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

Table 1.1 *Linguistic profile of adults in Barcelona urban area in 1983, based on (Woolard, 1991, 64)*

Age group	Catalonia-born %	Speak Catalan frequently %
15-20	87	43
21-30	68	49
31-40	48	44
41-50	45	46

> These positive sanctions for the maintenance of Catalan by native speakers and negative sanctions against its use by Castilian speakers helped explain patterns of language proficiency and use. A survey in 1983 (Direcció General de Política Lingüística 1984) found that, of those born in Catalonia of parents born in Catalonia, 93% claimed Catalan as their principal language. This shows remarkably minimal attrition of the Catalan language group. However, the demographic structure of Catalonia has changed significantly over the twentieth century, from a largely native-born to a massively immigrant population by the 1960's, and then with economic stagnation in the 1970's, returning to an increasingly native-born population. Immigration has virtually ceased, and among the 15-20 year olds in the DGPL sample, 87% were born in Catalonia, while over half of some older age brackets were immigrants. . . . Table [1.1] shows that while Catalonia is again becoming much more native and less immigrant in character, its native-born are much less likely to be Catalan-speaking.

Table 1.1 shows that, in the early 1980s, a greater proportion of young people were born in Catalonia than in the previous generation; however, the rate of frequent use of Catalan remains the same across age groups. In other words, young Castilian speakers born in Barcelona are not switching to Catalan. According to Woolard, this fact about language use is understandable based on the social meaning of Catalan for Castilian speakers: she says, 'The matched guise test showed that Castilian speakers had little to gain in cementing relations with Catalans by attempting to speak Catalan, while they had much to lose in solidarity and support from co-members of their own native ethnolinguistic group' (Woolard, 1991, 64).

In 2007, Woolard did a follow up MGT study (2009)[1] in Barcelona with participants of a similar demographic profile as in her 1980 study. She found that, as in the 1980s, Catalan was still associated with higher status ratings than Castilian for all listeners; however, thirty years later, solidarity ratings had changed drastically. She explains that 'in the experiment with the new case study group in 2007, there was no statistical difference between Catalan and Castilian guises in the Solidarity ratings. The general likeability of a speaker was not affected by the language she used; in contrast to earlier

[1] Woolard did another MGT study in 1987, which showed an intermediary pattern between the ones I describe here.

years, ratings neither rose nor fell with a speaker's use of Catalan or Castilian' (Woolard, 2009, 134). In other words, the ethnolinguistic boundary observed in the early 1980s appeared to have been broken down in the mid-2000s.[2] The change in the social meanings of Catalan for native Castilian listeners appears to also be correlated with a change in language use for these individuals: in the mid-2000s, when they are not penalized on the solidarity dimension, native Castilians are much more likely to speak Catalan. This is the case with the participants of Woolard's (2009) MGT study (a class of Barcelona high school students), where 'students' accounts of family history in interviews showed a clear trend toward Catalan across the generations', and 'any language change between parent and child or between home and habitual language was toward Catalan' (Woolard, 2009, 130). The pattern of change in language use is also observed in the broader population, as shown by the Government of Catalonia's report showing a rise in habitual use of Catalan by native-born Catalans (Generalitat de Catalunya, 2013).

Having observed this case of socio-semantic change, and a parallel change in language use, we would, of course, like to know what caused the solidarity-related social meanings of Catalan and Castilian to change from the 1980s to the 2000s. Woolard (1991, 2009) argues that the change in the social interpretations of Castilian/Catalan for speakers in Barcelona is a result of a more general process of ideological change in the ethnolinguistic categories of *Catalan* vs *Castilian*. In the 1980s, these social categories were conceptualized as being rooted in the circumstances of one's birth or family; however, 'basic terms of social identity have moved from an essentialist treatment of Castilian linguistic origins or habits as defining Castilian identity in contrast to Catalan twenty years ago, to a voluntarist conceptualization of *espanyol* vs. Catalan identity as a matter of politics and style' (Woolard, 2009, 145).[3] It is likely that the aggressive pro-Catalan language policies instituted in government and education after Franco's death played a role in this ideological change, though, as Woolard notes, 'because of a myriad other changes – political, social, economic, demographic, cultural – over the same period, we cannot know how directly these developments in young people's linguistic consciousness can be attributed to educational linguistic policy' (Woolard, 2009, 147) .

In summary, in this section we have seen that a wide variety of linguistic features, ranging from phonetic to morpho-syntactic to linguistic code, can change the way that listeners perceive the identity of the speaker. Furthermore, studies such as Woolard's (among many others) have observed connections between the social meanings of linguistic elements and speaker/listener ideologies, on the one hand, and speaker/listener behaviour on the other. These

[2] This was already starting to be the case in Woolard's 1987 follow up (Woolard and Gahng, 1990; Woolard, 2009).

[3] See Heller (2003, 2011) for somewhat similar ideological changes in French Canada.

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

studies have also pointed to the role of the social world, its material properties and structures, in shaping ideologies, that constrain which social meanings can be associated with which linguistic forms.

This book presents a formal model of how these different components (social structure, ideologies, social meanings and language use/interpretation) interact. Broadly speaking, the framework I will develop can be schematized as in Figure 1.1, where solid arrows represents connections that will be studied in detail in the book, while dashed arrows represent connections whose detailed characterization is left to future work. The social world consists of non-linguistic actions, social institutions, (non)social facts, among other things.[4]

Individuals' ideologies are shaped through their interactions with the social world in a number of ways, either through their direct observations or through their exposure to and subsequent integration of *discourses*: ways of talking about or representing objects which, simultaneously, serve to define and create them (Foucault, 1969, 1976). The question of how discourses shape ideologies relevant to socially meaningful language will be discussed in Chapters 4 and 5.

Ideologies play an important role in the model because they constitute what formal semanticists call the *domain of interpretation* of socially meaningful language. As such, ideologies provide properties, social categories and identities that can be associated with language, and they impose constraints on what the social meanings of linguistic elements can be. How ideologies constrain meaning is one of the major topics of Part II of the book (Chapters 4 and 5). The actual mappings between ideological objects and linguistic forms can be established in a number of different ways. For example, they can be established though individuals' direct observations about the kind of people who use a particular sociolinguistic variant (Labov, 1972; Trudgill, 1986; Kerswill and Williams, 2002; Preston, 2011 among very many others) or through meta-linguistic discourses invoking language ideologies (see Silverstein, 1979; Irvine and Gal, 2000; Cameron, 2012). Although the formation of both sociolinguistic variables and ideologies about language are currently important topics in sociolinguistics, I will have nothing new to contribute to these interesting debates in this work. What will occupy the bulk of this book is the connection between social meanings and language use and interpretation, in other words, the socio-semantic system. The fine-grained properties of the socio-semantic system will be the focus of Chapters 2 and 3. I will also briefly discuss the effects of socially meaningful language on the world, and its role in strategic discourse, in Chapters 2 and 4; however, my

[4] Both language use/interpretation and ideologies are also part of the social world; however, I distinguish these two components in Figure 1.1 because they will be given a more sophisticated treatment in this work than other aspects of the world.

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

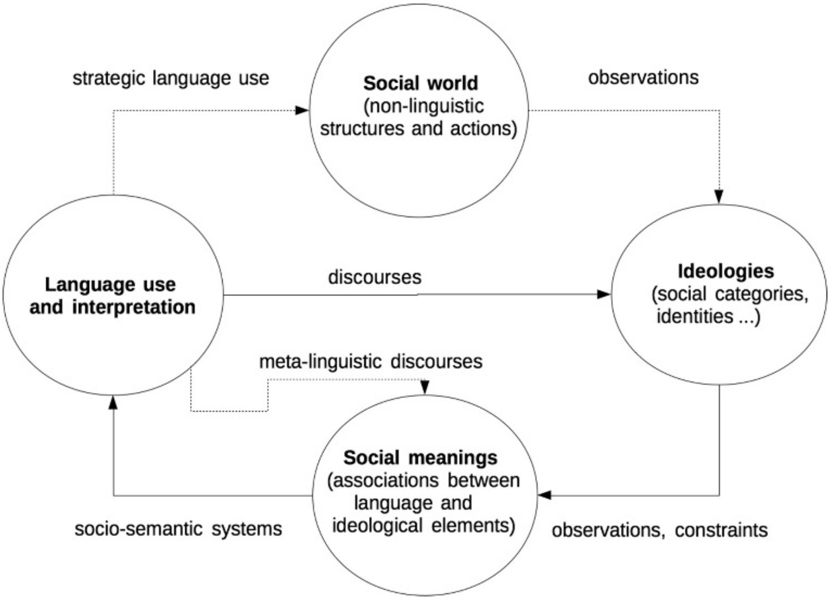
strategic language use
Social world
(non-linguistic structures and actions)
observations
Language use and interpretation
discourses
Ideologies
(social categories, identities ...)
meta-linguistic discourses
Social meanings
(associations between language and ideological elements)
socio-semantic systems
observations, constraints

Figure 1.1 The relationship between language and the social world (as Heather Burnett sees it)

remarks will by no means do justice to this fascinating and important topic whose exploration is left to future research.

As mentioned above, a key aspect of my framework is that it is formal: it is constructed using objects and definitions from mathematics, particularly formal logic and game theory. This is unusual; these tools are rarely used in sociolinguistic studies. Therefore, before proceeding, it is worthwhile discussing why (on earth) one would (ever) want to have a formal theory of social meaning and identity construction through language.

1.2 Why Should We Formalize?

Social meaning and identity construction has been extensively studied in linguistic anthropology and in sociocultural linguistics more generally (see Bucholtz and Hall, 2005, 2008, for reviews). They have likewise been studied in variationist (quantitative) sociolinguistics (Labov, 1963; Weinreich et al., 1968), especially within the *Third Wave* (TW) approach (Eckert, 2000, 2008, 2012, 2018), which will be described below. There has historically been less interest in treating these phenomena within formal linguistics, especially in semantics and pragmatics. A notable exception is the work of Sally

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

McConnell-Ginet, who has explored how notions and theories from semantics and analytical philosophy can help us better understand complicated socially meaningful phenomena such as identity terms, pronouns, gender marking and even sociophonetic variation since the early 1980s. Readers are encouraged to see McConnell-Ginet (2011) for a collection of her papers spanning three decades, as well as McConnell-Ginet (2013, 2020) for more recent work. In the past fifteen years, however, there has been an increased interest in analysing social meaning within formal semantics and pragmatics (Kaplan, 1999; Van Rooij, 2003; Potts, 2005; McCready, 2008, 2012, 2015; Smith et al., 2010; Acton, 2014, 2019; Beltrama, 2016, 2019; Beltrama and Staum Casasanto, 2017; Jeong, 2018, among others), and this book is yet another contribution to this growing research domain.

As I see it, there are (at least) three ways in which having a formal framework which can explore questions related to social meaning, identity construction and social structure is (or at least has the potential to be) beneficial. They are laid out in (14).

(14) Why have a formal theory of social meaning and identity construction?

(i) Formalization allows us to better test the predictions of our sociolinguistic theories.
(ii) Formalization opens up new ways of looking at sociolinguistic phenomena which allows us to make new empirical discoveries.
(iii) Formalization facilitates interactions between sociolinguistics and the cognitive and information sciences.

I will go into these three points in more detail below, but first I will say a bit more about what I understand the project of formalization to be, particularly with respect to theories from sociolinguistics and gender studies. At its heart, I view formalization as a process which takes a theory that has been described only in words, identifies what its main lines (i.e. its essential ‘moving parts’) are, identifies what the main empirical generalizations it aims to explain are, and then uses mathematics to characterize these moving parts/generalizations such that it becomes possible to relate the two by means of a proof (in the case of a logical formalization) or a simulation (in the case of a computational formalization). Of course, people may (and will) differ in both what they take to be the ‘essential moving parts’ of an informal theory and what they take to be the explananda of the theory. Furthermore, it can often be the case that our chosen mathematical tools are adapted to formalizing only a subpart of the informal theory, so the formalization that we propose will not do justice to all the ideas that we aim to capture. It is for these reasons that the main general proposal in this book is a framework for formalizing sociolinguistic theories, not simply, for example, a particular formalization of Penelope Eckert’s Third

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

Wave approach to the meaning of variation. Therefore, we will speak in this work of ***a** formalization* of theory X and not ***the** formalization* of X, even if it happens that the one proposed is the only one currently on the market. In fact, Chapter 2 alone will contain multiple formalizations of ideas from the Third Wave, and I will compare these formalizations to each other in the light of empirical data.

The models outlined in this book are similar in spirit to TV or film adaptations of novels. Like formalizations in linguistics, TV and film adaptations allow the ideas in the source material to reach a new audience and can explore themes from the original in new ways. This being said, because of the format (full-length novel vs two-hour movie) and the visual medium, adaptations can often treat only a subset of the original material. Because of this constraint, different adapters may have different parts that they wish to prioritize. And of course every adaptation is also an instance of interpretation: the adaptation is filtered through the adapter's vision when it is presented to the public. For example, if we consider Stanley Kubrick's 1980 film adaptation of Stephen King's 1977 novel *The Shining*, we see that, in addition to making cuts necessary in order to fit a the novel into a 2.5 hour movie, Kubrick made a number of interpretative choices: (MINOR SPOILER ALERT) he downplayed the supernatural element of the novel and backgrounded themes like the disintegration of the family and the dangers of alcoholism that were featured in the original. Kubrick did this in order to focus on the creation of the chilling mood that his movie is now famous for, something the film medium allowed him to do (Miller, 2013). Although the reception of Kubrick's *The Shining* was initially mixed, and King himself was unhappy with some of Kubrick's interpretations of his work,[5] the 1980 film is now considered one of the greatest horror films of all time and was selected for preservation by the United States National Registry by the Library of Congress in 2018. Seventeen years later, in 1997, Mark Garris and King adapted *The Shining* to create a TV mini-series, which was much longer (six hours) but remained much more faithful to the book. Reception of this adaptation was, and continues to be, very mixed, with some audiences appreciating the aspects of the original which were incorporated into TV adaptation, and others comparing it unfavourably to Kubrick's film.[6] In other words, both the film and the TV mini-series adaptations of King's novel incorporated and emphasized different aspects of the original, and were appreciated by different people for these different aspects, even though clearly one (the film version) has emerged as much more culturally significant than the other. In an ideal case, then, a successful formalization is like Kubrick's adaptation of *The Shining*: it captures key insights of the original, but its new form allows for these insights to be explored in a new way in new domains.

[5] https://scrapsfromtheloft.com/2018/03/08/stephen-king-playboy-interview-1983/.

[6] See https://en.wikipedia.org/wiki/The_Shining_(miniseries) for an overview.

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

All this is on the understanding that people will have different ideas about how well the formalization/adaptation captures the key insights of the source and how insightful its application to the new domains is. On the other hand, an unsuccessful formalization/adaptation is one that, rather than using the power of the new medium to accomplish new things, uses it as a box into which the adapter/formalist crams the exciting, complex and illuminating original, and in doing so loses the source's insights, stripping it of everything that everyone loved about it. Ultimately, then, one hopes that, when one undertakes to formalize complex theories from sociolinguistics, what one ends up with is closer to *The Shining, Carrie* (1976) and *Misery* (1990) than to the film adaptations of *The Hitchhiker's Guide to the Galaxy* (2005), *Dune* (1984) and (keeping with the Stephen King theme) *The Dark Tower* (2017).

Suppose we arrive at a formalization of a theory of social meaning and identity construction that we consider successful in the sense I described above. What would be the benefits? Firstly, I argue that such a formal theory can help us refine our sociolinguistic theories. Linguistic variation and identity construction are extremely complex cognitive and social processes, and a lot of open issues in the study of language, variation and identity are very subtle. Formalization can be a powerful tool for carefully distinguishing different aspects of theoretical proposals and for precisely identifying empirical predictions made by competing analyses. For example, in Eckert's Third Wave approach (which will be outlined in greater detail in Chapter 2), linguistic variation is proposed to arise from a combination of variants' social meanings and how different speakers use these meanings in different contexts to construct personae (identities/social types). According to Eckert (2008), a variant's social meaning is its indexical field: a set of properties or stances that members of a community of practice associate with a variant. As an illustration, following Campbell-Kibler (2007), Eckert proposes that the variants of variable (ING) have the indexical fields in (15).

(15) a. **Indexical field of *-ing*:** {articulate/pretentious, effortful, educated, formal}
b. **Indexical field of *-in'*:** {inarticulate/unpretentious, easygoing/lazy, uneducated, relaxed}

Researchers working in the Third Wave commonly analyse the social meaning of a variant through giving a list of the properties/stances in its indexical field similar to (15) (Campbell-Kibler, 2008; Moore and Podesva, 2009a; Walker et al., 2014; Beaton and Washington, 2015; Podesva et al., 2015; Tyler, 2015, among many others). Although the indexical fields featured in these works often make intuitive sense, these proposals have not yet been accompanied by an explicit, precise theory of how exactly the fields determine the patterns of sociolinguistic variation and interpretation they are supposed to

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

be analysing. Because of this, we currently have no principled way of arguing that one indexical field is a better analysis of the social meaning of a variant than another possibly very similar one. In other words, we are currently missing a linking theory (in the sense of Marr, 1982) between abstract ideological structures, like indexical fields, and linguistic behaviour, like variation and interpretation. My book proposes that formal semantics and pragmatics, combined with game-theoretic tools used in computational psycholinguistics, can give us this precise theory. Chapter 3 shows how we can test different hypotheses about the structure of the indexical fields associated with sociolinguistic variants in Montréal French.

Secondly, having a formalization of an informal theory can often draw attention to new aspects of that theory which were not clear before the formalization. The formal models that we will develop will have their own properties, and studying them as formal systems can reveal unexpected predictions that the systems make. Thus, foregrounding aspects of the theory that were backgrounded in the literature on the original can result in new empirical discoveries. We will see an example of such a new result in Chapter 2, where properties of the formal system that I propose will cause us to look more closely at the relationship between propositional/truth-conditional meaning and social meaning, and this will give rise to a new empirical generalization about the difference between the two.

Thirdly, having formal adaptations of informal theories from the humanities provides an important step towards allowing insights from these theories to be incorporated into theories of language from cognitive science, which are themselves often formalized. The mathematical/computational modelling of language variation and change is a vibrant research area (Clark and Roberts, 1993; Niyogi and Berwick, 1997; Yang, 2000; Yang, 2002; Adger and Smith, 2005; Adger, 2006; Kauhanen and Walkden, 2018, among many others). Since the late 1990s, formal models have yielded enormous advances in our theories of the cognitive and linguistic factors underlying variable language use. For example, we now have a clearer understanding of how parsing ambiguous utterances can trigger language change (Clark and Roberts, 1993; Yang, 2000) and how production biases can affect the shape of that change (Kauhanen and Walkden, 2018). However, although many (if not most) patterns of linguistic variation are socially conditioned, mathematical models have been almost exclusively focused on the grammatical and/or psychological aspects of change, neglecting its social aspects. Likewise, the computational modelling of human cognition is a vast area, which has yielded important results concerning how humans process language. There has been some recent interest in incorporating social factors in computational cognitive models, such as Jaeger and Weatherholtz (2016); Kleinschmidt et al. (2018); however, the sociolinguistic theory underlying these approaches is minimal. A formalization of a sociolinguistic theory could serve as a bridge across the humanities and

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

cognitive science, which would allow for identity-oriented theoretical insights from sociolinguistics and anthropology to inform the formal modelling of language variation, change and processing. Chapter 5 specifically addresses the question of the modelling of morpho-syntactic change within a formal system driven by social meaning.

All this being said, for a mathematical approach to sociolinguistic variation and identity construction to be helpful, we need to use a formalism that is appropriate for the data that we want to model. And it turns out that this is not a trivial matter. In fact, as discussed by Recanati (2004), many mathematical approaches to meaning, such as classic formal semantics (Montague, 1970; Heim and Kratzer, 1998), allow contextual factors to play only a restricted role, primarily in the evaluation of indexical expressions such as *I* or *you*, or quantifier domain restriction, such as saying *Everyone brought their book* to mean *Everyone in this class brought their book* (Stanley, 2000). Furthermore, these approaches focus primarily on truth-conditional interpretation, so they tend to study the behaviour of the listener only, and not that of the speaker. Studies of interaction in sociolinguistics and linguistic anthropology have stressed the extreme context-sensitivity and speaker-dependent nature of social meaning, which suggests that many classic formal frameworks are ill-equipped to analyse sociolinguistic variation and identity construction through language.

Developing appropriate, mathematically precise frameworks for capturing the relation between language, meaning and use is a long-standing problem in linguistics. Already in her ([2011] 1985) paper ‘Feminism in Linguistics’, Sally McConnell-Ginet reflects on the supposed ‘trade-off’ between formal rigour and interactivity as follows (2011, 64):

> Many critics would say that rigor in linguistics has been achieved at the price of rigor mortis. The radical operation required to ‘isolate’ the language system has killed it: formal rules and representations provide no insight into language as a human activity. The defense against this malpractice charge, of course, is to develop an account of the relation between abstract linguistic systems and the mental states and processes, social actions and cultural values, that infuse them with life.

The guiding idea of the book is that game theory gives us a way to answer McConnell-Ginet’s challenge of providing ‘vibrant’ formal theories of linguistic communication.

Indeed, the idea that language can be conceptualized as a game dates back at least to Wittgenstein (1953), and the proposal that game theory could be useful for analysing sociolinguistic interaction and its relation to identity has been explored by a number of scholars in anthropology, sociology and philosophy, including Goffman (1970) Bourdieu (1977) and Gumperz (1982). The initial interest in marrying mathematical and ethnographic studies of linguistic interaction dates back to the 1970s, when game-theoretic methods started to

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

become widely used outside economics (Osborne et al., 2004). However, this interest never developed into a full-fledged research programme.

I believe there were two reasons for this: firstly, the game-theoretic models accessible to scholars in the humanities at the time were not adapted for modelling language use and interpretation. For example, understanding an utterance in its social context almost always involves reasoning under uncertainty: when they hear a speaker use the *-in'* variant, the listener must decide which subset of (15-b) to attribute to this particular speaker based on this particular utterance. However, epistemic models, which take into account agents' reasoning, only became popular in the 1980s (Perea, 2012). Fortunately, in the past fifteen years, significant advances have been made in the application of epistemic game theory to linguistic communication (see Benz et al., 2005; Franke, 2009, 2017; Frank and Goodman, 2012; Franke and Jäger, 2016; Goodman and Frank, 2016, among many others), and the work presented in this book builds on these advances. Secondly, the way that we think about identity is very different now from how it was conceived in the 1970s and 1980s. Since the mid-1990s, the dominant view in the humanities and (some) social sciences has been that aspects of our personal and social identities (including gender, race, ethnicity and age) are not fixed; rather, they are constructed through a combination of our own actions and 'and the interpretations and reactions to them of others'? The social constructionist view of identity has its roots in studies of knowledge and power by Foucault (1969, 1976), and was most famously developed for gender identity by Butler (1991, 1993), although similar arguments can be made for other aspects of identity. As Eckert and McConnell–Ginet (2013, 40–41) say,

> This new focus on the role of invisible power in knowledge construction led to a recognition that social structures, including the nature of the categories *male* and *female*, are the outcome of interested historical forces (hence thinkers such as Foucault are often referred to as *poststructuralist*) . . . Gender, then, surfaced not as given, but as emergent; not as natural or essential, but as socially constructed. It went from something that people 'have' to something that people 'do'. In this view, gender doesn't just exist, but is continually produced, reproduced, and indeed changed through people's performance of gendered acts, as they project their own claimed gender identities, ratify or challenge others' identities, and in various ways support or challenge systems of gender relations or privilege and the ideologies in which they figure.

This new focus on how the actions of individuals construct their identities leads naturally to an analysis of identity construction as an interactive phenomenon. And, as discussed above, explicitly analysing interactive phenomena is what game theory does best. Therefore, I suggest that development of poststructuralist 'performative' theories of identity make it appropriate to treat this complex social and psychological phenomenon using game theory. I will return

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

to this point in Chapter 2 when I discuss an early attempt at relating language and identity using game theory: Erving Goffman's *Expression Games* (1970).

1.3 Decision Theory and Game Theory: An Overview

This section presents a brief overview of some of the formal tools that I will use in the analysis of social meaning and sociolinguistic variation: *decision theory* and *game theory*. These two mathematical frameworks are closely related: they both involve studying decisions and how decision makers (called *agents*) make them under different conditions.[7] They can both be interpreted *prescriptively*, i.e. as studying how idealized agents should make their decisions, or *descriptively*, i.e. as studying how human agents do make their decisions. Since we are using game and decision theory as tools for analysing actual linguistic data, this book adopts exclusively the descriptive perspective. The main difference between decision theory and game theory is that decision theory studies the reasoning underlying an agent's choices in situations where their decisions do not depend on the behaviour of other agents with whom they are interacting. Game theory studies the reasoning underlying agents' choices in situations in which their decisions are interdependent. Since decision theory focuses on a single agent, the decision maker, it is a good place to start to outline some formal tools which will be expanded on in the next chapter.

Decision theory is based on the idea that decision makers are rational: they adopt the actions that will have the best chance of achieving their goals.

(16) **Theory of Rational Choice:** an agent chooses the best action according to their preferences, among all the actions available to them.

Note that this technical sense of the word *rational/rationality* is different from what this word means in our everyday language, where it usually means something like 'sensible, reasonable' or 'guided by objective logic'. If someone wants to do something that seems strange to us (and not in their best interests), such as burn all their money, some people might call them irrational; however, if they perform actions that will get them to their goal (going to the bank, taking out all their money in cash, dowsing the pile of money in gasoline, etc.), then they will be considered rational in the sense in (16).

A basic model has three components:

[7] For a general introduction to decision theory, see Resnik (1987) and Peterson (2017); for an introduction to its use in pragmatics, see Merin (1999) and Benz et al. (2005). For a general introduction to game theory, see Osborne and Rubinstein (1994) and Osborne et al. (2004); for introductions to its use in formal linguistics, see Benz et al. (2005), Jäger (2011), and Franke (2017); and for an introduction to game theory in light of sociolinguistic data, see Dror et al. (2013, 2014).

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

(17) Basic decision-theoretic model

(i) An agent (decision maker).
(ii) A set A consisting of all the actions that, under some circumstances, are available to the decision maker.
(iii) A specification of the agent's preferences.
- Preferences are represented by a *utility function u* from actions to $\mathbb{N}$ such that $u(a) > u(b)$ iff the decision maker prefers a to b.

Decision-theoretic models can have different kinds of utility functions. Ordinal utility functions are those in which the precise values assigned to the actions don't matter; only the > relation between them does. Cardinal utility functions are those in which the precise values assigned to the actions encode degree of preference. So suppose we have two models, one with utility function u_1 and one with utility function u_2. u_1 assigns the values 42 and 5 to actions a and b respectively ($u_1(a) = 42$; $u_1(b) = 5$), and u_2 assigns the values 6 and 5 to a and b: $u_1(a) = 6$ and $u_2(b) = 5$. If we are treating the utility functions as ordinal, then the two models are equivalent. If the utility functions are viewed as cardinal, then the two models are not equivalent, since the decision maker prefers a to b much more in the first model than in the second model (over eight times more, in fact).

Even a simple decision-theoretic model can explicitly capture the link between social meaning and language use in Woolard's studies of Catalan vs Castilian language choice in Barcelona. We will first consider the situation in Barcelona in the early 1980s. Suppose you are Castilian and you meet another Castilian. You have to choose which language to use speak to them. We can represent this choice as choosing elements from a set of actions $A = \{$Speak Castilian (CAST), Speak Catalan (CAT)$\}$. Because of the social meanings of the different languages, choosing CAST or CAT will have different outcomes. Recall from section 1.1 that for Castilian speakers interacting with Castilian listeners in the 1980s, speaking Catalan made one sound more intelligent but less likeable; whereas speaking Castilian made one sound more likeable but less intelligent. Thus, in this time period Castilians need to consult their preferences: would they prefer to come across as intelligent or likeable? We can make models of two different types of Castilians: one type who values status over solidarity, and therefore who would value speaking Catalan over Castilian (18), and another who values solidarity over status (19).

(18) Status-oriented Castilian (speaking to Castilian) model (1980)

(i) A Castilian decision maker
(ii) $A = \{\text{CAST}, \text{CAT}\}$
(iii) $u(\text{CAT}) > u(\text{CAST})$ Ordinal utility function

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

(19) Solidarity-oriented Castilian (speaking to Castilian) model (1980)

(i) A Castilian decision maker
(ii) $A = \{\textsc{cast}, \textsc{cat}\}$
(iii) $u(\textsc{Cast}) > u(\textsc{Cat})$ Ordinal utility function

The Theory of Rational Choice (16) gives us a way of linking abstract preference structures like utility functions and language use. Recall that it states that an agent chooses the best action according to their preferences, so the models in (18)-(19) predict that the status oriented Castilian should speak in Catalan, and the solidarity oriented Castilian should speak in Castilian.

(18) and (19) provide models of native Castilian agents with different ideologies (systems of attitudes and values (Maio et al., 2006)): one values solidarity over status, and the other has the opposite values. An important discovery from sociology is that ideologies are not equally distributed across society. In particular, there are often correlations between ideologies surrounding status and solidarity and other aspects of social structure, particularly social class. In Europe and North America, (upper) middle-class individuals tend to value status properties more highly; whereas working-class individuals tend to put greater value on solidarity properties (Bourdieu, 1977, 1979; Lamont et al., 1992; Lamont, 2009).[8] As mentioned above, the native Castilian/Catalan ethnolinguistic distinction largely coincides with a social class distinction: even during the Franco regime, 'Catalans continue to dominate the internal economic structure of Catalonia (which contributes significantly to that of Spain). Although the Castilian language was successfully imposed by state institutions as the means of access to functionary positions, Catalans continue to be predominant in ownership and management of the private sector, which is still characterized by small and mid-sized industries' (Woolard, 1985, 742). Therefore it is reasonable to think that, in Barcelona society in the 1980s, there are far more native Castilians whose ideologies are better represented by the model in (19) than in (18). If we take the distribution of ideologies across society into account, our model predicts that, in the early 1980s, there will be far more Castilians who stick to their native tongue than those who switch to Catalan.

I have given ordinal utility functions in the models (18)–(19): the decision maker's preferences are simply represented with the *greater than* relation (>). With this kind of utility function, the models make qualitative predictions: (18) predicts that Catalan will be used more and (19) predicts that Castilian will be used more. We can have the models make quantitative predictions for language use if we change the utility function to a cardinal one, where there are particular numerical values associated with each language (20). If we were doing a new empirical study, we could ask participants to give a measure of *how much* more

[8] See also Trudgill's (1972) notion of the *covert prestige* (solidarity signalling) of variants favoured by working-class men in the UK.

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

they value solidarity (or status) at the time of the MGT experiment, but, for illustration, I have just set the preferred language to be twice as preferred as the dispreferred language.

(20) Solidarity-oriented Castilian (speaking to Castilian) model (1980)

(i) A Castilian decision maker
(ii) $A = \{\text{CAST}, \text{CAT}\}$
(iii) $u(\text{CAST}) = 2, u(\text{CAT}) = 1$ Cardinal utility function

For quantitative predictions, we need to weaken the Theory of Rational Choice and say that speakers and listeners are instead *approximately rational*: they are *rational* in the sense that they are trying to maximize their utility (as in (16); 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.), and there may simply be a certain amount of inherent variability in the system (Weinreich et al., 1968). Therefore, in order to account for possible variability in action selection, we will assume that, rather than just picking the action with the highest utility, the decision maker chooses the best option given a noise-perturbed assessment of utilities. One such weaker choice rule, called the *Soft-Max choice rule* (Luce, 1959; Bridle, 1990; Sutton and Barto, 1998), is widely used in both reinforcement/machine learning and game-theoretic approaches to a variety of pragmatic phenomena (Frank and Goodman, 2012; Degen et al., 2013; Bergen et al., 2016; Lassiter and Goodman, 2015; Franke and Jäger, 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

The Soft-Max choice rule is given in (21), where a is an action and λ is a real number.

(21) For a parameter $\lambda \in \mathbb{R}$ and an action $a \in A$,

$$P(a) = \frac{exp(\lambda \times u(a))}{\sum_{a' \in A} exp(\lambda \times u(a'))}$$

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

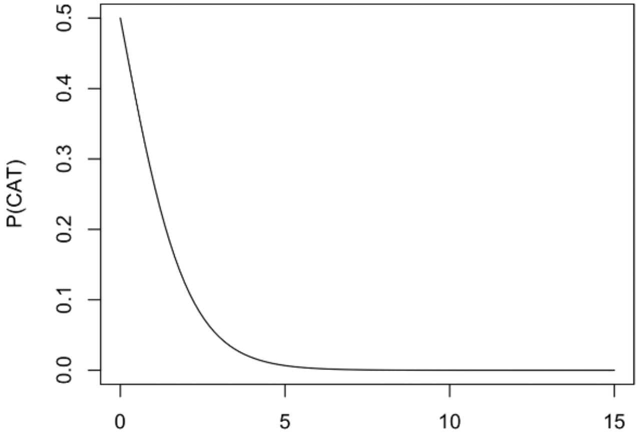

Figure 1.2 Predicted probability of using Catalan by solidarity-oriented Castilians in 1983 by values for λ

The parameter λ in (21), called the *reverse temperature*,[9] introduces some non-determinacy into the model. Setting $\lambda = \infty$ recovers the rational choice rule (16), but any value for lambda that is less than ∞ will result in variation in most models.

When modelling actual quantitative studies, the value for λ that best fits the observed data can be estimated (as in Goodman and Stuhlmüller, 2013; Franke and Degen, 2016; Scontras et al., 2018); however, as an illustration, the probability of using Catalan for solidarity-oriented native Castilians interacting with Castilians in 1980 by values for λ is displayed in Figure 1.2: the probability of using Catalan is never predicted to be higher than 0.5, and it decreases as λ increases.

Our model for native Castilians in the early 1980s predicts that they should largely avoid Catalan, but, as Woolard shows, by 2007, the social meaning of Catalan and Castilian for native Castilians, has changed: Catalan still makes Castilian speakers sounds more intelligent, but now speaking in Castilian to other Castilians doesn’t make one sound more likeable. So there is no longer any ‘trade-off’ between sounding nice and sounding smart, and there is an advantage to choosing Catalan over Castilian, even for solidarity-oriented Castilians (22).

[9] The *temperature* metaphor comes from physics where the hotter molecules get, the more they move around. λ in this work is a ‘reverse’ temperature because the higher λ is, the less variable the system is.

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

(22) Native Castilian model (2007)

(i) A Castilian decision maker
(ii) $A = \{\textsc{cast}, \textsc{cat}\}$
(iii) $u(\textsc{Cat}) = 2, u(\textsc{Cast}) = 1$ Cardinal utility function

Of course, for native Catalans, there was never any advantage to speaking Castilian: speaking Catalan always gives one a status bonus, and, at least in 1980, Catalans receive a penalty on solidarity when they don't speak their native language to other Catalans. Therefore the structure of the decision problem for native Catalans speaking to both Catalans and Castilians in both the 1980s and 2000s can be captured by the model in (23).

(23) Native Catalan model (1980, 2007)

(i) A Catalan decision maker
(ii) $A = \{\textsc{cast}, \textsc{cat}\}$
(iii) $u(\textsc{Cat}) > u(\textsc{Cast}) = 1$ Ordinal utility function

The little decision-theoretic models that I have just laid out provide the skeleton of a formal model of how social change can be related to language change: social changes (economic changes, education changes, language policy change, etc.) change ideologies (in this case social meanings of linguistic expressions), which then change how useful the expressions are to speakers in the construction of their identities. This is a good start, but clearly much more remains to be developed.

For instance, I have grouped the decision maker's reasoning about the social meanings of Catalan vs Castilian together with preferences over the action of speaking Catalan vs Castilian. Actual social meanings, what they look like and how exactly they enter into reasoning, are nowhere in the models above. Additionally, I have treated the problem of choosing to speak in Catalan vs Castilian as a decision problem: I assumed that the choice of CAST vs CAT depends only on how the individual wishes to present themself. In the models in this chapter, the interlocutor plays no role in this choice, beyond their ethnicity, which partially determines the social meanings of Catalan/Castilian. But we might think that actions of the person that we are communicating with influence which language we choose to speak. [10] In my own experience of bilingualism, in most situations, there is pressure for speakers to coordinate on the language of communication, even when both are perfectly bilingual. So a speaker's choice of language may depend not only on their ideologies and the

[10] See Myers-Scotton (2000); Myers-Scotton and Bolonyai (2001) for a model of code-switching based on the Theory of Rational Choice.

languages' social meanings, but also on the language choice of their interlocutor. In order to model interdependencies such as these, we need to use game theory rather than decision theory. This is what we will do in the next chapter.

1.4 Plan of the Book

This book is composed of a mixture of published work, new unpublished work and everything in between (presentations, posters, etc.).

In Chapter 2, I extend the skeleton of a model that was outlined above into a full game-theoretic system which will provide a 'formal semantics' for sociophonetic variation. I will do this by giving a formalization of some aspects of Eckert's Third Wave approach to the meaning of variation. The resulting framework is called Social Meaning Games (SMGs). The first part of Chapter 2 is a synthesis of two published papers on the topic: 'Sociolinguistic Interaction and Identity Construction: The View from Game-Theoretic Pragmatics' (Burnett, 2017) and Signalling Games, Sociolinguistic Variation and the Construction of Style (Burnett, 2019). The second part of Chapter 2 evaluates the models presented in these works in light of a number of empirical and theoretical considerations, and presents some directions in which they could be refined. One of the proposed refinements is based on joint work with Eric Acton. (Eric Acton) presented as a poster 'Markedness, Rationality and Social Meaning' at the 2019 meeting of the Linguistic Society of America (Acton and Burnett, 2019). The final refinement in Chapter 2 involves embedding the linguistic game of social signalling modelled by SMGs within a larger game of non-linguistic interaction, and then looking at the effects of socially meaningful language on players' strategies in the non-linguistic game. Consequently, we build a model of the link between sociolinguistic variants and the non-linguistic outcomes they contribute to creating for the speaker. This larger game-theoretic model reflects the idea that identity construction is, at the end of the day, in the service of the non-linguistic interaction: what people are trying to do with their words. This framework also allows us to provide a new characterization of the meanings of sociolinguistic variants, one that relates the variant not to truth conditions or internal mental representations, but to the material conditions that the variant contributes to creating for the speaker. By *material conditions*, I mean not only economic conditions, but also social, cultural and health-related conditions that language has a hand in producing (Jackson, 2001). I call this way of analysing linguistic social meaning a *materialist semantics*, since it is focused on behaviour and outcomes in addition to mental representations. I argue that this perspective has the potential to yield a unified theory of the social meanings of the range of linguistic expressions discussed at the beginning of this chapter, and, eventually, for better studying the link between linguistic and non-linguistic behaviour.

In Chapter 3, I show how the SMG framework can be used to test different hypotheses concerning the social meanings associated with sociolinguistic variants. Concretely, I present a new empirical study of socially conditioned variation in the use of negative polarity items *du tout* and *pantoute* 'at all' in the Montréal 84 corpus of spoken Montréal French (Thibault and Vincent, 1990) and show how the SMGs, combined with the corpus data, will allow us to arbitrate between different possible analyses of the social meanings of these lexical items. This chapter has never been published before but has been presented at the 2017 Integrating Approaches to Social Meaning conference in Toulouse, the 2018 meeting of the Association for French Language Studies in Toronto, and at the University of Chicago as a colloquium.

Chapter 4 extends the materialist semantics proposed in Chapter 2 to provide an analysis of the meaning and use of (some) slurs. Slurs are linguistic expressions used to denigrate individuals based on some aspect of their identity. I focus on one slur in particular: *dyke*, which is generally considered to be a derogatory term for lesbians. I argue that previous research on this word in formal semantics and analytical philosophy has been limited in (at least) two ways: firstly, I argue that not enough attention has been paid to the use of *dyke* by members of the target group, who can often use it in a non-insulting manner; secondly, I argue that not enough attention has been paid to the use of the 'neutral' form *lesbian*, which is generally treated as having a simple, clear meaning, such as 'engage[s] in same-sex sex' (Jeshion, 2013a, 312). Following McConnell-Ginet (2002), I argue that the semantics of *lesbian* is actually quite complex, and that taking into account all the uses of both *dyke* and *lesbian* requires a new semantics and pragmatics for both terms. More specifically, I propose that *dyke* and *lesbian* are associated with different sets of *personae*: abstract identities or stereotypes. *Dyke* is associated with an *anti-mainstream persona*, which the vast majority of speakers views negatively; whereas *lesbian* is associated with at least one *mainstream persona*, which many speakers view favourably. To make this proposal explicit, I present a formal framework for capturing the link between ideological structure and language use/interpretation: Gärdenfors' (2000, 2014) *Conceptual Spaces*, and I set my analysis of the meaning of *dyke* and *lesbian* in this framework. I show how the semantic puzzles associated with *dyke* and *lesbian* can be resolved through the combination of a theory of these personae in Conceptual Spaces and a game-theoretic theory of how listeners' beliefs about their interlocutors' ideologies affect utterance interpretation. Again, I set the game of linguistic communication within a larger non-linguistic game and show how the use of a slur like *dyke* or an identity term like *lesbian* can change the outcomes for the target in the non-linguistic interaction. Much of Chapter 4 comes from 'A Persona-Based Semantics for Slurs' (Burnett, 2020).

Chapters 5 deals with formalizing ideologies about social gender and how these ideologies relate to the social meaning of grammatical gender in French. This presents a corpus study of variation and change in the use of grammatical gender in the speech of politicians in the Assemblée Nationale (the

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press

French House of Representatives). In 1986, Prime Minister Fabius legislated the use of feminine grammatical gender in the Assemblée Nationale and similar government institutions; however, as Burnett and Bonami (2019b) show, this prescription had little to no effect on the speech of the politicians at the time. Then, in 1998, Prime Minister Jospin issued a statement reiterating Fabius' policy. Unlike twelve years earlier, the feminine form successfully replaces the masculine form within the space of a year. This chapter argues that changes in the use of feminine grammatical gender and differences in the effectiveness of Fabius/Jospin's language policy are the result of changes in gender ideologies in France between the mid-1980s and mid-1990s. We argue that the mid-1990s saw the emergence of a new persona for female politicians, which only feminine grammatical gender can construct. We hypothesize that Jospin's reinforcement of Fabius' policy in 1998 was successful because it strengthened an existing association between feminine grammatical gender and a female political persona; whereas Fabius' original policy was unsuccessful because it tried to build on ideological structure that was not shared by a large portion of the Assemblée Nationale. In this chapter I show how the framework developed in Chapters 2 and 4 can be used to formalize change in gender ideologies between 1986 and 1998, and how Jospin's success and Fabius' failure can be predicted (in the form of theorems) in the resulting system. The bulk of Chapter 5 is a synthesis of 'Linguistic prescription, ideological structure, and the actuation of linguistic changes: Grammatical gender in French parliamentary debates' (Burnett and Bonami, 2019) and 'A Conceptual Spaces model of socially conditioned language change' (Burnett and Bonami, 2019).

Chapter 6 summarizes the main results and themes of the book, and then discusses directions for improvement and extension in the future.

https://doi.org/10.1017/9781108894524.003 Published online by Cambridge University Press