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.

1

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)

(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)).

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

(10) a. We should mee[**tʰ**].   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[tʰ]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

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),

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 nega-
tive 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 signifi-
cantly 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 cement-
ing 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.

studies have also pointed to the role of the social world, its material proper-
ties 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 stud-
ied 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 ideolo-
gies 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 iden-
tities that can be associated with language, and they impose constraints on
what the social meanings of linguistic elements can be. How ideologies con-
strain 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 for-
mation of both sociolinguistic variables and ideologies about language are
currently important topics in sociolinguistics, I will have nothing new to con-
tribute 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
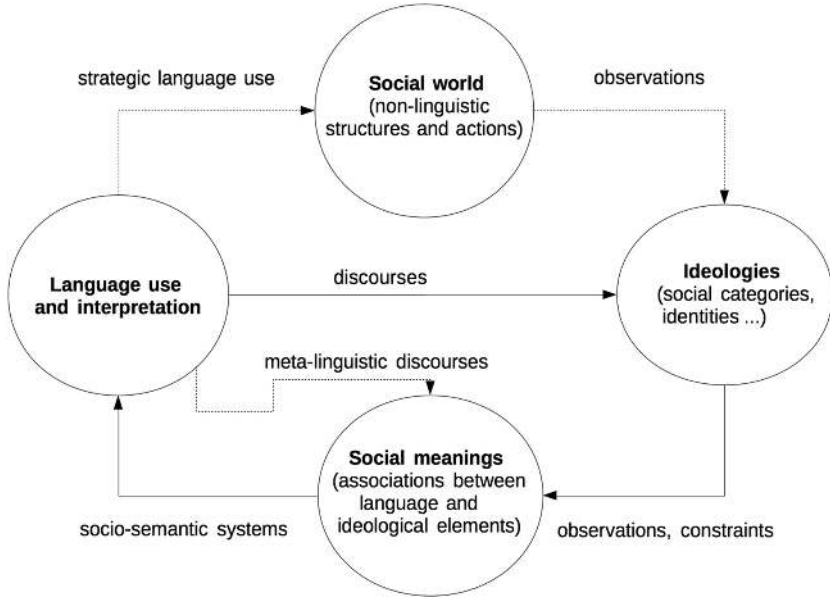treatment in this work than other aspects of the world.

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