Signalling Games, Sociolinguistic Variation, and the Construction of Style

Heather Burnett (CNRS-Université de Paris 7-Denis Diderot)

1. Introduction. This paper introduces *social meaning* games, which are developed for the analysis of the strategic aspect of sociolinguistic variation (in the sense of Labov 1966 et seq.). While remarks have been made (Clark 2011:88) about the potential usefulness of game theory in the analysis of, for example, variable use of the (ING) suffix (*walk**in'*** vs *walk**ing***), a general framework uniting variationist sociolinguistics with game theoretic pragmatics has yet to be developed. We propose that such a unification is possible through the integration of the *third wave* approach to the meaning of sociolinguistic variables (see Eckert 2012) with *signalling* games (Lewis 1969), which are commonly used in the game-theoretic analysis of Gricean reasoning (Benz et al. 2005, Franke 2009, Clark 2011, Franke & Goodman 2012, among others). We define the games and then we show certain results concerning the framework's predictions for contextually-based *style-shifting*.

2. Game Set-Up. The basic idea is that the sender/speaker, **S**, has a set of (positive) properties that they believe holds of themself (their *type*). S chooses a message m_i out of a set of variants. The receiver, **R**, hears the message and updates the set of properties that they consider to hold of S based on their prior beliefs about S and their beliefs about the kind of person would use m_i versus some other variant m_j. More formally, a **Social Meaning Signalling Game (SMSG)** is a tuple $\langle\{S,R\},\mathbb{P},\mathscr{C},Pr,M,[\cdot],\mathbb{A},U_S,U_r\rangle$, where S and R are the players and $\mathbb{P}=\{p_1,\ldots,p_n\}$ is a finite set of properties. Well-known features of social meaning are that it is highly context-sensitive and flexible (Ochs 1992, Silverstein 2003, Campbell-Kibler 2007, Eckert 2008, a.o.), and that a particular property (such as *articulateness*) can be viewed positively in one context (as eg. *cultured*) but be construed negatively in another (as *pretentious*) (Campbell-Kibler 2007/Eckert 2008). We model the possible positive/negative construals through the introduction of two classes of properties: for all $p_i \in \mathbb{P}$, p_i^+ is p_i's **positive** construal and p_i^- is p_i's **negative** construal. $\mathscr{C}$ notates the family of sets of properties that are **+/- consistent**; that is, sets that do not contain both positive and negative construals of the same property. $Pr \in \Delta(\mathscr{C})$ is a probability distribution over $\mathscr{C}$, which represents R's (prior) uncertainty about which properties hold of S. $M=\{m_1,\ldots,m_n\}$ is the set of messages (i.e. variants) that S can pick from. As in the *third wave* approach to variation, messages are related to **indexical fields** (Eckert 2008) by the indexation function $[\cdot]$. For our purposes, an **indexical field** ($\mathscr{F}_m$) of a message m is a set of ordered pairs $\mathscr{F}_m=\{\langle p_1^+,p_1^-\rangle,\ldots,\langle p_n^+,p_n^-,\rangle\}$. The set of **actions** $\mathbb{A}$ is the set $\mathscr{C}$; that is, R's action is to attribute a set of properties to S.

3. Strategies and Utility. Since S does not know exactly which strategy R will use to interpret the message (and R does not know exactly which strategy S will use to send it), as in Franke 2009, we will use **mixed/behavioural strategies** to describe how the speaker picks m and how R interprets it. In particular, Sender strategies are functions σ from sets of **positive construals** of properties to probability distributions over messages, and Receiver strategies ρ are functions from messages to probability distributions over $\mathscr{C}$ (sets containing both positive and negative properties). We also assume that both S and R share an important bit of knowledge: the indexical field associated with m. Although we do not impose a deterministic link between all of the properties indexed by m and m itself (because of the extreme context-sensitivity of social meaning), we do allow the players to make use some amount of this shared information in interpreting and calculating which message to send. In particular, we propose that Receiver strategies are existentially **anchored** to indexical fields: If $\rho(m)(P)\neq 0$, there is some $p_i^{+/-}$ such that $\langle p_i^+,p_i^-\rangle \in \mathscr{F}_m$. In other words, if R has some non-zero probability of interpreting a message as a set of properties, this is because at least one of

those properties can be indexed by the message. Which strategies will be the most useful depends on the beliefs of S and R concerning the reasoning of their interlocutor. S's beliefs will therefore be represented by a Receiver strategy ρ_S. R's prior beliefs are represented by Pr, and their beliefs about the strategy that S is using is represented as an indexically anchored Sender strategy σ_R. Note that σ_R is a function from any member of $\mathscr{C}$; that is, while S's self evaluation contains only positive properties, R's evaluation of S may contain positive or negative properties. Finally, R's posterior beliefs are derived through Bayesian update (Franke 2009, a.o.).

We view SMSGs as games of cooperation: for S, the optimal situation is one in which R converges with them on the same set of positive properties. For R, such a scenario is also optimal because them not only do they 'get' S, but they have also formed a positive impression of them. The worst outcome for S (and R) is one in which R hears S's message and attributes to S a set of negatively construed properties. An intermediary outcome is one in which S and R do not converge on the same set of properties, but R still forms a positive impression of S. The utility functions for S/R (U_S, U_R) are therefore defined ordinally over these three outcomes and S/R's expected utility and best responses are defined in the standard way (cf. Osbourne 2004).

4. Results. In this paper, we focus on production, so we are interested in S's expected utility for a message m, given their beliefs about R's strategy ρ_S. The formal results will be given in the paper, but for readability, we give them informally here. The first property of the models is that if S has no strong beliefs concerning the strategy that R is using (i.e. if $\rho_S(m)$ is (at least close to) a uniform distribution over $\mathscr{C}$, modulo indexical anchoring), then S's best response is to use the variant that indexes the properties in their type (**Theorem 1**). So, for example, if a speaker wants to construct a persona that incorporates *easy-going-ness*, which is a positive construal of a property in *-in*'s indexical field (Campbell-Kibler 2007), in most cases, their best strategy is to use *-in*. However, in situations in which S has a strong belief concerning how they will be evaluated by their interlocutor (i.e. if $\rho_S(m)$ is highly weighted on certain kinds of sets of properties), we predict possibly different best responses. For example, if S strongly believes that using *-in* will cause R to attribute a negative construal (such as *inarticulateness*) to them rather than *easy-going-ness*, we predict that S's best response will be to use another less risky variant: *-ing*. Likewise, if S strongly believes that using *-ing* will make them sound *pretentious*, then their best response (regardless of their type) is predicted to be *-in*. (**Theorem 2**). These results are consistent with studies such as Labov (2012), which found a large difference between President Obama's use of *-in* (vs *-ing*) in an informal BBQ (72% *-in*) vs a formal speech (3% *-in*). Conversely, if S strongly believes that R will converge on their intended type (eg. {*tough, masculine*}) no matter what, then S's expected utilities for *-in* and *-ing* will approach each other and possibly 'tie' (**Theorem 3**). We argue that this result is consistent with Gratton (2015)'s study of non-binary individuals' use of (ING), which found a large significant difference in an individual who was assigned female at birth's use of *-in* in a café surrounded by strangers (80% *-in*), but found no significant difference in the use of *-in* vs *-ing* when they were in their home with a close friend (56% *-in*).

5. Selected References. 1. Campbell-Kibler, K. (2007). Accent, (ING), and the Social Logic of Listener Perceptions. *American Speech*. 82:32-64. **2.** Clark, R. (2011). *Meaningful Games*. MIT Press. **3.** Eckert, P. (2012). Three waves of variation study. *Ann. Rev. Anthro.* 41:87-100. **4.** Eckert, P. (2008). Variation and the indexical field. *J. of Sociolinguistics*, 12:453-476. **5.** Franke, M. (2009). *Signal to Act: Game Theory in Pragmatics*. PhD, Amsterdam. **6.** Gratton, C. (2015). Indexin' Gender: Variable (ING) and the Creation of Non-Binary Trans Identities. Poster at *NWAV44*. **7.** Labov, W. (2012). *Dialect Diversity in America.* U. Virginia Press.