DIACHRONIC PROCESSES IN LANGUAGE AS SIGNALING UNDER CONFLICTING INTERESTS

CHRISTOPHER AHERN, ROBIN CLARK

Department of Linguistics, University of Pennsylvania, 619 Williams Hall, Philadelphia, 19104, United States
cahern@ling.upenn.edu, rclark@ling.upenn.edu

Game-theory has found broad application in modeling pragmatic reasoning in cases of common and conflicting interests. Here we extend these considerations to diachronic patterns of language use. We use evolutionary game theory to characterize the effect of conflicting interests on linguistic signaling in a population over time. We show how sufficiently large conflicts of interest give rise to particular patterns of signaling that can be used to model diachronic processes such as Jespersen’s Cycle.

1. Introduction

In the abstract, both Linguistics and Animal Communication deal with the transmission and interpretation of signals. Yet, the two differ markedly in the use of the term cooperation. Crucially, this difference hinges on the interests of the agents sending and receiving signals. In the Gricean tradition, speakers are taken to be cooperative when they make their contributions relevant to current conversational goals. This use presumes that interlocutors agree to the goals of a given exchange, and benefit from pursuing them. In the biological sense, speakers are cooperative when they forgo some gain to the benefit of others (Nowak, 2006). This use entails that speakers have countervailing goals, but do not pursue them. In what follows, we will use cooperation solely in the biological sense and consider the implications it has for patterns of signaling over time.

Cooperation in the biological sense is in need of evolutionary explanation. If senders have an incentive to pursue some goal, then they should do so. If, in doing so, senders have every incentive to deceive, then receivers should not listen. If receivers do not listen, then senders have no motive to signal in the first place. Crucially, the actions of senders depend on those of receivers and vice versa. Given this interdependence, conflicting interests undermine signaling. Senders will learn or evolve to dissimulate and receivers will learn or evolve to distrust. This process takes on the familiar form of the tragedy of the commons (Hardin, 1968). Individuals will always be tempted to exploit the common resource of credulity, to the collective detriment.
Silence or, at best, meaningless babble is the equilibrium state in the population: neither senders nor receivers have reason to unilaterally change their behavior. So, what keeps signaling in general, and human language in particular, from the downward spiral to silence? The existence and persistence of signaling necessitate evolutionary mechanisms that mitigate conflicts of interest between senders and receivers. Various mechanisms have been proposed to solve this problem, with varying degrees of appropriateness for application to language (Scott-Phillips, 2008).

This leads us to two general questions, which we can consider apart from the details of any particular set of mechanisms. First, how effective are these mechanisms in solving the problem of cooperation in signaling? Given that language exists, such mechanisms are clearly sufficient to stave off total collapse. Given that signals are not always used cooperatively, such mechanisms are not sufficient to ensure the idealized case of Gricean commonality. Second, if language is indeed subject to a host of competing pressures, what impact will this have on how linguistic signals are used over time? This second question will be our chief concern.

Using evolutionary game theory, we consider the range of conflict between senders and receivers, from pure common interest to pure conflicts of interest. We identify the crucial points that determine different regimes of signaling behavior: small conflicts allow for stable signaling, large conflicts lead to a collapse, and in between we observe cyclic signaling. The main contribution of this work is the application of this model to the well-known diachronic process of Jespersen’s Cycle. We show how even a slight tendency to overuse emphatic negation leads to the loss of emphasis, necessitating the introduction of a new emphatic form to take its place. This provides a broader connection between the forces affecting language at both evolutionary and historical time scales.

The rest of the paper is organized as follows. In the next section we provide an overview of the formal framework. We then turn to a particular parametrization of conflicting interests and examine stability. The resulting model is then applied to the cycle. We conclude with a brief discussion of our results and directions for future research.

2. Evolutionary Game Theory

Evolutionary game theory (Maynard Smith, 1982) gives us the formal tools to model interactions between agents with varying degrees of common and conflicting interests, without the too strong assumption of perfect rationality. Here we introduce signaling games (Lewis, 1969) along with the equilibrium concept of an evolutionarily stable strategy as a model of signaling in a population.
2.1. Signaling Games

A signaling game consists of two players: a sender and a receiver. The sender has
some private piece of information, \( t \in T \). The piece of information can be thought
of as some fact about the state of the world. The sender chooses a message,
\( m \in M \), to send to the receiver. A sender strategy \( s \in [T \rightarrow M] \) specifies
what message to send in any given state.\(^4\) The receiver does not know what state
of the world actually holds and must choose an action, \( a \in A \), to interpret the
message sent. A receiver strategy \( r \in [M \rightarrow A] \) specifies what action to take
upon receiving any message. The outcome of the game is determined by the state
of the sender, the message sent, and the action taken by the receiver. The sender
and the receiver have preferences over these outcomes, which are reflected in the
utility functions, \( U_S \) and \( U_R \) respectively, which map outcomes to the set of real
numbers.

The set of possible combinations of sender and receiver strategies are strategy
profiles. That is a sender strategy in the set of possible sender strategies, \( s \in S \),
and a receiver strategy in the set of all possible receiver strategies, \( r \in R \), yield
a strategy profile \( \langle s, r \rangle \). The expected utility for a given strategy profile can be
given as follows, where \( \delta \) specifies a probability distribution over \( T \).

\[
EU_S(s, r) = \sum_t \delta(t) \cdot U_S(t, r(s(t)))
\]

\[
EU_R(s, r) = \sum_t \delta(t) \cdot U_R(t, r(s(t)))
\]

For each possible state the sender and receiver strategies determine an outcome.
\( s(t) \) is the message the sender will employ and \( r(s(t)) \) is the action the receiver
will take given that message. The expected utility is the sum of these outcomes
weighted by the probability of the state that yields them.

2.2. Evolutionarily Stable Strategies

Signaling games provide a description of the interaction we are interested in, but
we require some sort of solution concept to complete our model. An evolution-
arily stable strategy (Maynard Smith, 1982) provides just that. Loosely speaking,
a strategy is evolutionarily stable if, when played by an entire population, it is
resistant to invasion by mutant strategies. For asymmetric games, such as signal-

\(^4\)For simplicity, we only consider pure sender (receiver) strategies, functions from states to mes-
sages (messages to actions). Mixed strategies are a straightforward generalization. A mixed sender
strategy specifies a probability distribution over pure sender strategies, \( \sigma = p_1s_1 + \ldots + p_ks_k \), where
\( \sum_i p_i = 1 \). A mixed receiver strategy specifies a probability distribution over pure receiver strategies,
\( \rho = q_1r_1 + \ldots + q_kr_k \).
ing games, where players have different roles, a strategy profile is evolutionarily stable if and only if it constitutes a *Strict Nash equilibrium* (Selten, 1980).

**Definition 1.** A strategy profile \( (s^*, r^*) \) is a *Strict Nash equilibrium* if and only if:

- For all \( s \in S \), such that \( s \neq s^* \), \( EU_S(s^*, r^*) > EU_S(s, r^*) \)
- For all \( r \in R \), such that \( r \neq r^* \), \( EU_R(s^*, r^*) > EU_R(s^*, r) \)

This simply states that the sender and receiver do worse if they unilaterally deviate from the equilibrium. Thus for a given strategy profile to be an ESS its component strategies must be mutual best responses to each other.

3. **Conflicting Interests**

With the general framework in place, we now turn to a simple way of capturing the degree of conflict between senders and receivers by a single parameter. We then consider different constraints on signaling behavior. From this we determine the existence of evolutionarily stable strategies at various parameter values.

3.1. **Parametrization**

Conflicts of interest between a sender and the receiver are essentially a matter of misaligned preferences. To see this, consider the following game. Let there be two states \( T = \{t_L, t_H\} \), where we treat \( t_L \) as a low value on some scale and \( t_H \) as a high value. For simplicity, \( t_L = 0 \) and \( t_H = 1 \), where both states are equiprobable, \( \delta(t_L) = \delta(t_H) = \frac{1}{2} \). There is some finite set of messages \( M = \{m_1, ..., m_k\} \) available to the sender. Let the set of actions available to the receiver be the interval \( A = [0, 1] \).

Now, we can consider preferences over outcomes. Receivers want to accurately sort the senders by type and act accordingly. For example, suppose there are actions \( a_L \) and \( a_H \) that maximize the receiver’s utility when playing with \( t_L \) and \( t_H \) senders respectively. Following Crawford and Sobel (1982), we define the sender and receiver utility functions as the following, where the constant of conflict, \( b \in [0, 1] \), allows us to express how much some senders gain by exaggeration.

\[
\begin{align*}
U_S(t, a) &= -(a - [t + b])^2 \\
U_R(t, a) &= -(a - t)^2
\end{align*}
\]  

Note that the constant of conflict represents the distance between the interests of the sender and the receiver. When \( b = 0 \) interests are perfectly aligned. Senders wish to reveal their type: a \( t_L \) sender’s payoff is maximized by \( a_L \) and a \( t_H \)
sender’s payoff is maximized by $a_H$. These actions also maximize the receiver’s payoff. In contrast, when $b > 0$, their interests diverge. The sender does best when he convinces the receiver to take an action appropriate for a higher type. For example, if $b = 1$ then senders maximize their utility with the highest possible action. A $t_L$ sender’s payoff is maximized by $a_H$, as is a $t_H$ sender’s payoff. Note that a $t_H$ sender’s payoff is always maximized by $a_H$, regardless of the size $b$.

### 3.2. Signaling

We can now consider the incentives of senders and receivers and define conditions that determine signaling behavior. We begin from the receiver’s perspective before turning to the sender.

The receiver’s best course of action depends on the amount of information available regarding the sender’s type. For example, if senders separate themselves by using distinct signals, then the receiver should make use of this information by taking the unique appropriate action. What should the receiver do if the senders pool together using the same signal? In the absence of any information, the action that maximizes her expected utility is the action that corresponds to the average type of the population, which we will refer to as the pooling action, $a_P$.

$$a_P = \delta(t_L)a_L + \delta(t_H)a_H \quad (3)$$

We can now turn to the sender’s perspective and ask which of the receiver’s actions are preferred. We can state this generally in the following manner. For a sender of type $t$, the expected utility of an action $a'$ exceeds that of an action $a$, where $a' > a$, if the following holds.

$$\frac{1}{2}(U_S(t, a) + U_S(t, a')) < t + b \quad (4)$$

In other words, when the sender’s type plus the constant of conflict exceed the average of the two actions, then the higher action $a'$ is preferred. This gives us a general recipe for determining the sender’s preferences. For any two receiver actions we can find their average. If this is less than the sender’s type plus the conflict, then the sender prefers the alternative higher action.

### 3.3. Stability

We are now in a position to consider the stability of various sorts of signaling behavior for various degrees of conflict. In particular, we will be interested in whether senders separate or pool, and any transitions between the two kinds of behavior. As a point of reference, we will consider a state of affairs where all sender types use a single message $m$, and receivers respond to this message with the pooling action $a_P$. Now, consider the availability of some alternate message.
Consider the case where \( b \) is fairly small. Starting from the pooling state, do senders prefer some alternate action? We consider the types in turn. For \( b \in [0, \frac{1}{4}] \), \( t_L \) senders prefer \( a_L \) to \( a_P \) and \( t_H \) senders prefer \( a_H \) to \( a_P \). That is, both types of senders have an incentive to break from pooling. That is, if receivers respond to \( m' \) with any action other than \( a_P \), then one type will have an incentive to adopt this new signal. Once this takes place the receiver will be able to separate the two types and act accordingly. Moreover, this constitutes an evolutionarily stable separating equilibrium. Neither type wishes to adopt the signal of the other, because it leads to a lower payoff.

If we increase \( b \) we observe different constraints. For \( b \in (\frac{1}{4}, \frac{1}{2}) \), we consider the types in turn. Again, starting from the pooling equilibrium, we know that \( t_L \) senders prefer the pooling action to being identified. But, \( t_H \) senders still wish to be identified. Thus, \( t_H \) senders have an incentive to break from the pooling by adopting a new signal. Receivers will do best by responding to this new signal accordingly and then separating the types. Note that neither sender type has an incentive to pool. Even though \( t_L \) senders prefer the pooling outcome the most, they still prefer being identified to being mistaken for \( t_H \) senders. Thus, separating is still evolutionarily stable.

Increasing \( b \) further, we consider the case where \( b \in (\frac{1}{2}, \frac{3}{4}) \). As before, \( t_H \) senders always have an incentive to adopt a new signal to be identified. Responding to this, receivers will be able to separate the types. However, once the types are separated, \( t_L \) senders will always have an incentive to pool. That is, if receivers respond to \( m' \) with \( a_H \), then \( t_L \) senders will adopt \( m' \). Receivers will then respond with the pooling action and the cycle begins again.

Finally, we consider \( b \in (\frac{3}{4}, 1] \). In this case, both types of senders prefer \( a_H \). Any message that the receiver responds to with \( a_H \) will be adopted by both types. However, if both types adopt the same message, then the receiver does best by responding with the pooling action. Pooling in this case is neologism proof (Farrell, 1993), it cannot be disrupted by the introduction of new signals. Thus, we see that beyond a certain point, conflicts are too large for separating to be sustained, even temporarily.

For sufficiently small conflicts of interest, separating is the only evolutionarily stable strategy. As conflict grows separating is destabilized, but persists in a cyclic pattern. The availability of new signals allows a subset of the types to identify themselves and be separated from the others. Yet, eventually even this escape from pooling is cut off. By considering a range of common and conflicting interests we have identified a range of signaling behaviors, including cyclic behavior where messages are adopted, mimicked, and then abandoned.
4. Diachronic Processes

Jespersen’s Cycle is a broad, cross-linguistic tendency for a change in negation. Starting from an initial point where plain and emphatic negation are two distinct forms, emphatic negation undergoes a kind of rhetorical devaluation and is bleached of its emphasis (Dahl, 2001). Kiparsky and Condoravdi (2006) show the process in Greek repeating itself several times over the centuries; last century’s emphatic negation is bound to be used as this century’s plain negation.

<table>
<thead>
<tr>
<th>PLAIN</th>
<th>EMPHATIC</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>οὐ...τί</td>
<td>οὐ-δε...τί</td>
<td>Ancient Greek</td>
</tr>
<tr>
<td>οὐ...δε...τί</td>
<td>οὐ-δε...τίποτε</td>
<td>Early Medieval Greek</td>
</tr>
<tr>
<td>δεν...τίποτε</td>
<td>δεν... πράμα</td>
<td>Greek Dialects</td>
</tr>
<tr>
<td>δεν...πράμα</td>
<td>δεν... ἀπαντοξη</td>
<td>Modern Cretan</td>
</tr>
</tbody>
</table>

Table 1. Historical Forms of Negation in Greek

The types of senders correspond to contextual standards of use. For example, one might reasonably say, “I didn’t move!” when one has moved an inch, but not “I didn’t move an inch!” In that case, $t_H$ senders represent a stricter standard of use than $t_L$ senders. Where the actions available to receivers are the formation of degrees of belief, receivers would be expected to form a higher degree of belief for any message sent exclusively by $t_H$ senders. If convincing receivers is the goal, regardless of type, we would expect both types to flock to such a message. Receivers would eventually becoming increasingly skeptical, responding more and more with the pooling action. The degree of belief that receivers have in the proposition would decrease over time. In other words, emphatic negation would lose its emphasis.

In light of this, the $t_H$ senders can distinguish themselves by the use of some new signal. Neologisms are abundant, most often formulated by the addition of a minimizer (e.g. “not an inch”) or a generalizer (e.g. “not at all”) (Horn, 1989). Yet, just as $t_H$ senders might distinguish themselves by adopting a new signal, so to can the $t_L$ senders follow. Forms of negation are constantly being devalued and abandoned as interlocutors aim to convince others.

5. Concluding Remarks

Here we have presented a model that allows us to explore various degrees of common and conflicting interests, noting a range of behavior, from separating, to cy--

---

b Schwenter (2006) argues from synchronic data in Romance that the plain/emphatic distinction is not sufficient, and that the different forms of negation are sensitive to information structural constraints. Our analysis could be extended to this account.

c This table is constructed to show the connection between forms of different time periods; see Kiparsky and Condoravdi (2006) for more details and discussion.
cling, to collapse, as conflict increases. We have focused on cyclic signaling, taking Jespersen’s Cycle as a prime example. Such models offer insight into what can be explained by considering signaling under both common and conflicting interests. The existence of neologisms is crucial to the stability of separating and pooling. Future work will involve understanding the interaction between neologism formation and various game dynamics (Hofbauer & Sigmund, 1998).

References