**The pragmatics of semantic change: Modeling the progressive-to-imperfective shift**

We implement a computational model of the cyclic progressive-to-imperfective shift, in which languages with a single imperfective marker grammaticize new progressive markers which ultimately broaden in interpretation and displace the older imperfective. While Deo (2015) offers a model of this process within the framework of evolutionary game theory, her model ultimately abandons the semantics she takes such care to construct. Our model offers a more interpretable alternative, operating directly over the semantics. We show how semantic change may be a function of changes in utterance cost—a reflection of morphological complexity or frequency-of-use—as it relates to pragmatic reasoning. Counter to claims that grammaticization is a process of conventionalization of implicature (Traugott & Dasher 2002, Traugott & König 1991), our model holds the denotations of aspectual markers static; what changes is how we reason about their use given their relative cost.

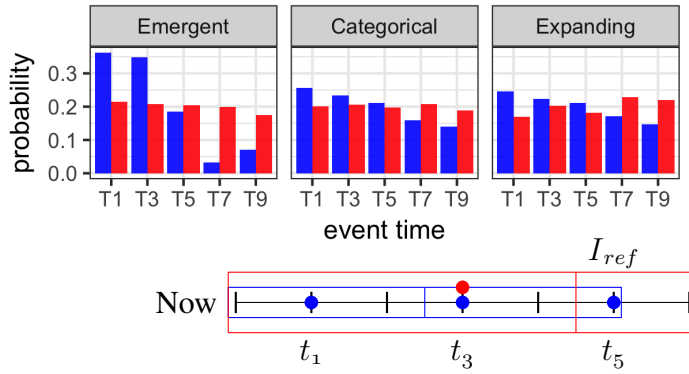
We adopt a version of Deo's (2009, 2015) semantics, wherein both the imperfective and progressive markers feature universal quantification over a partitioned time interval. IMPF and PROG differ in that the partitioned interval over which IMPF quantifies is a superinterval of the reference interval, a contextually-determined interval of time relevant to the evaluation of the utterance; PROG quantifies over subintervals of the reference interval itself. This difference is illustrated in (1), where ${\mathscr{R}_j}^c$ is a regular contextual partition over $j$ and $\text{COIN}(P, k, w)$ holds when $P$ is true at $k$ in $w$. The smaller the partitioned interval, the smaller the partitions are, and the closer each individual partition is to *now*. As a result, events need to be closer to *now* to overlap with each partition cell. The larger the partitioned interval, the further each partition cell is from *now* and the further apart each cell-overlapping event may be. Smaller intervals correspond to event-in-progress readings, as the events are close to *now*; larger intervals correspond to characterizing readings, as events may be far in the future and more sporadically distributed over time. See Fig. 1 (*bottom*) for a visual representation of these intervals and interpretations.

(1) a. $[\![\text{IMPF}]\!] = \lambda P \lambda i \lambda w. \exists j [i \subseteq_{ini} j \wedge \forall k [k \in {\mathscr{R}_j}^c \rightarrow \text{COIN}(P, k, w)]]$

b. $[\![\text{PROG}]\!] = \lambda P \lambda i \lambda w. \forall k [k \in {\mathscr{R}_i}^c \rightarrow \text{COIN}(P, k, w)]]$

In a system where both PROG and IMPF are active, given that PROG describes a narrower set of scenarios than IMPF, the use of PROG entails IMPF. This asymmetry leads to a run-of-the-mill scalar implicature whereby the use of IMPF implies that PROG does not hold (a speaker would use the stronger PROG if it were true), which suggests that $P$ fails to hold for the contextually-salient reference interval. In other words, the use of IMPF rules out the narrower set of interpretations PROG wold have delivered, which reduces event-in-progress reading for IMPF. Conversely, the use of PROG suggests that $P$ does not hold *beyond* the reference interval, ruling out characterizing and habitual readings which extend into the future beyond the reference interval.

We model the pragmatic reasoning that strengthens and refines the interpretations of PROG and IMPF within the Bayesian Rational Speech Act (RSA) framework, where speakers and listeners reason recursively about utterances and the world state those utterances decribe (Frank & Goodman 2012; Goodman & Frank 2016). In our model, a "lifted-variable" RSA variant, listeners also reason about the reference interval ($I_{ref}$) and superinterval ($I_{sup}$) under discussion. Put differently, our model assumes that listeners have uncertainty about the precise reference and superintervals that are relevant when interpreting utterances that appeal to grammatical aspect. Utterance interpretation involves three levels of inference. At the base, the literal listener ($L_0$) interprets utterances according to their literal semantics; $L_0$ updates beliefs about the state of the world $s$ conditioned on the semantics of $u$ and some specified $I_{ref}$ and $I_{sup}$. One level up, the pragmatic speaker ($S_1$)




| stage | PROG | IMPF |
|---|---|---|
| emergent | 100 | 1 |
| categorical | 1 | 1 |
| expanding | 1 | 100 |




Figure 1: *Top left*: model predictions. Probability that the event holds at the specified event time after hearing the PROG or IMPF utterance. *Top right*: relative utterance costs for each stage of aspectual marking. *Bottom*: representation of event-in-progress (blue) and characterizing (red) scenarios, where dots represent events and boxes represent partitioned intervals.

chooses an utterance $u$ to communicate some observed $s$ to $L_0$. $S_1$ makes this choice by maximizing the probability that $u$ would convey $s$ to $L_0$ while minimizing the utterance cost ($C(u)$); the temperature parameter $\alpha$ controls $S_1$'s optimality when maximizing utterance utility. At the top level of inference, the pragmatic listener ($L_1$) observes $u$ and updates beliefs about $s$, together with the likely values for $I_{ref}$ and $I_{sup}$. Thus, $L_1$ uses $u$ to jointly infer the state of the world and the relevant intervals. $L_1$ performs this inference by reasoning about the probability that $S_1$ would have chosen $u$ to communicate about $s$ relative to some specific $I_{ref}$ and $I_{sup}$.

$$
\begin{aligned}
(2) \qquad P_{L_1}(s, I_{ref}, I_{sup}|u) &\propto P_{S_1}(u|s, I_{ref}, I_{sup}) \cdot P(s) \cdot P(I_{ref}) \cdot P(I_{sup}) \\
P_{S_1}(u|s, I_{ref}, I_{sup}) &\propto exp(\alpha \cdot [log(P_{L_0}(s|u, I_{ref}, I_{sup})) - C(u)]) \\
P_{L_0}(s|u, I_{ref}, I_{sup}) &\propto [\![u]\!]^{I_{ref}, I_{sup}} \cdot P(s)
\end{aligned}
$$

To generate model predictions, we fix the free parameters of the model. We model states of the world $s$ as sets of numeric values from 0 to 10, corresponding to time indices when the predicate holds (Fig. 1, *bottom*). The possible utterances are `prog` and `impf`, which return true when $P$ is true of each partition of the relevant interval $I$; our model also features a `null` utterance (i.e., saying nothing at all). By manipulating the relative costs of `prog` and `impf`, we model their prevalence in a language at a particular stage of aspectual marking.

The emergent PROG stage, where PROG constructions are often periphrastic and less frequent, involves PROG constructions with a high utterance cost (i.e., morphologically complex and infrequent) relative to that of IMPF (Bybee et al. 1994, Dahl 1985). At this stage, PROG is used almost exclusively with event-in-progress readings, but its high cost leads to an overall preference for IMPF for both event-in-progress and characterizing scenarios. As PROG's morphology streamlines, its cost lowers and it gets used more frequently. This leads to the categorical PROG stage, where there is a preference for PROG to describe event-in-progress readings while IMPF describes characterizing ones. Here we already see a broadening of PROG's meaning, but the cost symmetry between PROG and IMPF allows each utterance to carve out its own space of usage. In the expanding PROG stage, PROG's meaning broadens yet more, and the relative cost of IMPF leads to an overall preference for PROG for both event-in-progress and characterizing scenarios; PROG has displaced IMPF in its usage. Our model thus captures the progressive-to-imperfective shift while reasoning pragmatically about a stable utterance semantics but changing utterance costs. Note that we have used Deo's semantics for theoretical continuity; any semantics with a similar entailment

asymmetry between PROG and IMPF would deliver the same qualitative pattern of results.