The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift *

Ashwini Deo
Yale University

Abstract This paper offers an analysis of a robustly attested semantic change in which progressive markers "spontaneously" emerge in languages, become entrenched in the grammatical system, and diachronically generalize by turning into imperfective markers. The pattern is cyclic in that the generalization is often followed by a re-emergence of new progressive markers. The analysis has a semantic component that characterizes the relation between the progressive and imperfective operators as a privative semantic contrast. Its dynamic component rests on the proposal that imperfective and progressive sentences crucially distinguish between two kinds of inquiries: phenomenal and structural inquiries (Goldsmith & Woisetschlaeger 1982). The trajectory — consisting of the recruitment of a progressive form, its categorical use in phenomenal inquiries, and its generalization to imperfective meaning — is modeled within the framework of Evolutionary Game Theory. The diachronic path is reconstructed as a cyclic pattern in which alternative communication strategies rise and fall in dominance over time due to the differences in communicative success and learnability associated with them.

1 Introduction

(1) a. Jane is sorting the mail.

    b. Jane sorts the mail.

I hear an utterance such as (1a) and understand that the referent is engaged in a particular event that is in progress as I process the utterance. I hear (1b) and understand there is some principled link that connects the referent with the sorting of the mail — perhaps a habit or an assignment of responsibilities. On the face of it, the meaning conveyed by the tense+participle construction in (1a) has little in common with the meaning conveyed by the simple present tense verb in (1b).[1] However, it is a crosslinguistically attested fact that pairs of meanings like those conveyed by (1) seem to be related to each other in a *diachronic* way. In several languages, expressions that are primarily employed in describing events in progress at one temporal stage extend to being used in describing habits and principled generalizations at a later stage. Conversely, languages in which the same expression can be used to describe both events in progress and principled generalizations, may spontaneously innovate new expressions to describe events in progress. The immediate goal of this paper is to work towards an understanding of these robustly attested observations.

The progressive≫imperfective path is only one kind of crosslinguistically attested systematic change. Research on the typology and grammaticalization of categories like tense/aspect, modality, and possession has uncovered several systematic patterns in the links between form and meaning over time. These patterns take the form of unidirectional diachronic trajectories — recurring regularities in the ways that grammatical forms change in meaning. A few examples are in (2).

---

1 I am ignoring, for the moment, the obvious fact that both sentences have divisive reference along the temporal dimension, or in other words, the subinterval property.

(2) a. Resultative markers generalize to markers of perfect aspect and past tense. (Bybee, Perkins & Pagliuca 1994, Dahl 1985, 2000).

b. Expressions encoding location evolve into expressions encoding alienable or inalienable possession. (Clark 1978, Aristar 1996, Heine 1997, Stassen 2009).

c. Expressions of deontic modality diachronically acquire epistemic uses, but not vice versa. (Traugott 1989, Traugott & Dasher 2002).

The broad goal of this paper is to attempt to answer two general questions that emerge from the kinds of empirical observations in (2): Why and how do the meanings of functional expressions change over time? Why do we see cross-linguistic similarities in patterns of semantic change?

Deo (2014, 2015) proposes that any adequate explanation for paths of semantic change that involve functional expressions must contain two components — a structural one, and a dynamic one. The structural component involves giving a precise characterization of the logical relation between the meanings of related functional categories. Following Weinreich, Labov & Herzog (1968), we can view this as the *constraints* problem for semantics. The dynamic component draws from theories of language use and evolution in order to provide a plausible account for the morphosyntactic emergence of a new category and its subsequent generalization to a broader meaning, under normal conditions of usage and transmission. This relates to the *actuation* and the *transition* problems of Weinreich, Labov & Herzog 1968. The account proposed in this paper for the progressive≫imperfective path illustrates the workings of these basic components in a theory of semantic change.

Here is the plan for this paper. In Section 2, I briefly report on crosslinguistic evidence for the progressive–imperfective connection, and in particular, the diachronic path that the two categories are implicated in. Section 3 describes a semantic analysis that treats the progressive–imperfective contrast as involving a privative opposition, and derives the division of labor between them as a pragmatic outcome. Section 4 introduces the idea that the realization of both imperfective and progressive markers in a language allows for formal disambiguation between phenomenal and structural inquiries. It also outlines the three transitions in the progressive≫imperfective grammaticalization path — recruitment, categoricalization, and generalization. Section 5 offers an evolutionary game-theoretic analysis of the path. The *Imperfective Game* described in this section shows how a set of concrete assumptions about imperfective and progressive meaning, the communicative success

associated with distinct strategies of expressing these meanings, and acquisitional asymmetries in learning formally distinct strategies (represented as mutations) allow for a plausible modeling of the cyclic trajectory observed in this semantic domain. In Section 6, I submit that the structure of the analysis and the game-theoretic model are both extendable to a wider range of semantic domains in which trajectorial changes have been empirically observed.

2 The progressive–imperfective connection

Crosslinguistically, imperfective marking is associated with at least three distinct readings:

   i. the progressive or *event-in-progress* reading;

  ii. the habitual or generic *characterizing* reading; and

 iii. the *continuous* reading with lexically stative predicates.

The three readings are illustrated in (3) with examples from Gujarati, an Indo-Aryan language with imperfective marking.

(3) a. niśā   (atyāre) roṭli   **banāv-e**   **ch-e**
           N.NOM.SG now   bread.NOM.SG make-IMPF.3SG PRES-3SG
           Niśā *is making* bread (right now).   *event-in-progress*

     b. niśā   (roj)   roṭli   **banāv-e**   **ch-e**
           N.NOM.SG everyday bread.NOM.SG make-IMPF.3SG PRES-3SG
           Niśā *makes* bread (everyday).   *characterizing*

     c. niśā   navsāri-mā **rah-e**   **ch-e**
           N.NOM.SG Navsari-LOC live-IMPF.3SG PRES-3SG
           Niśā *lives* in Navsari.   *continuous*

Progressive marking (e.g., the English Progressive), on the other hand, saliently exhibits only the event-in-progress reading. At least since Comrie's (1976) classic text on aspect (also see Kuryłowicz 1964) the progressive has been treated as a subcategory of the more general imperfective, with a narrower, more specific meaning than the imperfective. Two pieces of evidence support this analysis — blocking facts (Section 2.1) and grammaticalizing changes (Section 2.2).

2.1 Semantic blocking

In languages where both progressive and imperfective aspects are realized with distinct morphology, the event-in-progress reading is often blocked for the imperfective form. For instance, Modern Hindi, which realizes both the imperfective and the progressive, exhibits some restriction on the distribution of imperfective marking in both the present and the past tenses (Deo 2009). For most speakers of Hindi, the imperfective form (4a) is incompatible with the event-in-progress reading, which is uniquely expressed by the progressive (4b).

(4) a. niśā   roṭi   **banā-ti**   **hai**
         N.NOM.SG bread.NOM.SG make-IMPF.F.SG PRES.3SG
         Niśā *makes/*is making* bread   *characterizing*

     b. niśā   roṭi   **banā rah-i**   **hai**
         N.NOM.SG bread.NOM.SG make PROG-F.SG PRES.3SG
         Niśā *is making/*makes* bread   *event-in-progress*

The blocking facts described above are beginning to be observed only in contemporary Hindi. In the 19th and early 20th century, there is evidence from both texts (e.g., *Premsagur*, *Rani Ketaki ki Kahani*) and grammars (e.g., Kellogg 1893) showing that the imperfective form regularly alternated with the newly evolving progressive form in describing events in progress. In fact, older speakers of contemporary Hindi even today accept the use of imperfective marking, alternating between (4a) and (4b) for describing events-in-progress. The blocking relation between the specific progressive and the general imperfective is thus not fully categorical and develops gradually along a diachronic dimension. Moreover, it is preceded by a period of free variation between both exponents, as evidenced in modern French and Italian, for instance, where the two exponents may alternate. For Middle English, Visser (1970, p. 746) observes that before the emergence of the *be+V-ing* construction as a grammaticalized progressive, the English Simple Present exhibited event-in-progress and stative readings as well as habitual, generic readings. The use of the Simple Present in alternation with the Progressive continues in the Early Modern English of Shakespeare, as seen in (5). In Modern English, however, the event-in-progress reading is typically unavailable for the Simple Present.

(5) a. What do you read, my lord?   (Hamlet II.2.191)

   b. O, I die, Horatio.   (Hamlet V.2.345)

2.2 Grammaticalizing changes

2.2.1 Turkish: Change in progress

Comrie (1976) and Dahl (1985) report that the distribution of the progressive suffix *-(I)yor* in Turkish illustrates an ongoing progressive-to-imperfective change. Based on their report and data from Turkish grammars, the situation appears to be as follows: The Turkish morpheme *-Ir* (labeled Aorist), until recently, used to exhibit characterizing and continuous readings and was also used in performative and reportative contexts. The Turkish Progressive *-(I)yor*, on the other hand, was restricted to descriptions of events in progress, as is described even in some recent grammars (e.g., Kornfilt 1997, pp. 339-340). This clear-cut distribution is illustrated in (6). In (6a), the verb form with *-(I)yor* describes an ongoing working eventuality, while in (6b), the *-ir* inflected verb describes a characteristic pattern of working — a characterizing reading.

(6) a. saat ikide   çaliş-**iyor-du**-m
       At two o' clock work-PROG-PST-1SG

       At two o' clock, I *was working.*

   b. genellikle iki saat   çaliş-**ir-di**-m
      Usually   for two hours work-IMPF-PST-1SG

      I *would* usually *work* for two hours.
      (Göksel & Kerslake 2005, p. 355)

However, recently, the Progressive *-(I)yor* has begun to appear with a wider range of readings, especially in the colloquial language. It systematically appears with lexical stative predicates (e.g., the stative *tan* 'know' in (7a)), and is also interchangeably used with the Aorist form in characterizing contexts (7b).

(7) a. sen Ömer'i benden daha iyi   tan-**iyor-du**-n
         you Omer me   better than know-PROG-PST-2SG

     You *knew* (lit. were knowing) Ömer better than me.

     b. O zamanlarda Mehmet çok sigara   iç-**iyor-du**
         At that time   Mehmet lot cigarrette smoke-IMPF-PST.3SG

     At that time, Mehmet *used to smoke* (lit. was smoking) a lot.

<div align="right">(Göksel & Kerslake 2005, p. 330)</div>

The Aorist, on the other hand, never exhibits an event-in-progress reading. These facts are taken to indicate that the Turkish Progressive is being extended to the domain of the imperfective Aorist, thus instantiating the progressive≫imperfective shift. Comrie (1976) reports that dialect variation in Yoruba (Niger-Congo) presents a similar case of ongoing change.

2.2.2 Tigre: Two imperfective markers

Bybee, Perkins & Pagliuca (1994) report on a number of languages (Tigre (Semitic), Yagaria (Papuan), Alyawarra (Pama-Nyungan), and Margi (Chadic)) which are characterized by two markers for the imperfective aspect with no apparent distinction between the two.[2] In each of these cases, they find that one marker is a diachronically older form, while the other is a younger form evolved from a progressive marker. I illustrate this with facts from Tigre (Semitic). All examples in (8–9) are from Raz's (1983) grammar (pp. 70-72). The imperfective form (labeled "Imperfect" by Raz) exhibits the characterizing and continuous readings.

(8) a. …'azedi   sanni **na'amrakka**
         now indeed well   we know.IMPF.1PL you

     Now indeed, we *know* you well.

     b. 'ana 'əb dəggalabye   **'əkkatəb**
         I   with my left hand write.IMPF.1SG

     I *write* with my left hand.

Raz further describes a compound tense, based on the Imperfect in periphrasis with present (*halla*) or past (*'ala*) tense auxiliaries. This use is

---

2 Bybee, Perkins & Pagliuca (1994, p. 144) describe these as 'present grams' rather than imperfective grams, and the data they provide is restricted to sentences with imperfective morphology and present tense marking.

said to resemble the use of the English Progressive to describe events in progress.

(9) a. ḥəna hədāy **nətfarrar** **hallena**
         we   wedding go out.IMPF PRES.1PL
         We *are going out* to the wedding.

     b. kaləb ’əb gabay **lə‘e** **‘ala**
         dog   on   road   run.IMPF PST.3SG
         A dog *was running* on the road.

However, this periphrastic construction also exhibits characterizing readings as shown in (10a–b) with no semantic distinction from the bare Imperfect.

(10) a. wa’əb lagəd’o   ’asək yom   **təmayət** **hallət**
          And of the (disease) gəd’o until today die.IMPF PRES.3SG
          And until today, they (lit. she, i.e., ‘the camels’) *die* of gəd’o disease

      b. ’ana nə’uš ’ət   ’ana kəldol   ’ət bet məhro **’əgayas**
          I   small while being I   every time to   school   go.IMPF
          ‘alko
          PST.1SG
          When I was young, I *used to go* to school every day.

While the Tigre Progressive exhibits both event-in-progress and characterizing readings, the Tigre Imperfect, which realizes imperfective aspect is not compatible with the event-in-progress reading. Bybee *et al.* conclude that the partial overlap in the distribution of the two markers is a result of the diachronic extension of the periphrastic progressive form to a wider range of contexts.

2.3   Implications for aspectual meaning

Such crosslinguistic facts about the synchronic organization and diachronic evolution of progressive and imperfective marking naturally lend themselves to an interpretation where progressive marking realizes a more specific meaning than the imperfective and gradually generalizes over time. Bybee, Perkins & Pagliuca (1994) make the following observation in connection to such semantic changes:

> The considerable overlap we find in constructions developing in the same semantic domain means that at any particular synchronic stage, the contrasts found will not necessarily represent opposite poles on an abstract semantic dimension that represents some basic dichotomy in the speaker's world view. Rather, it seems to us that there are certain major contrasts of universal validity such as the basic distinction between the perfective domain…and the imperfective…, but that within these domains, there are successive waves of grammaticizations which may follow upon one another at such a rate so as to produce only very small and subtle semantic distinctions.
>
> (Bybee, Perkins & Pagliuca 1994, pp. 148–149)

The crucial questions here have to do with determining, on the one hand, WHAT the content of these "small and subtle distinctions" is, and, on the other, WHY and HOW such "waves of grammaticizations" seem to follow upon one another, leading to morphological differentiation within a semantic domain. The explanation for Bybee *et al.*'s observation needs to be grounded in semantic theory (the WHAT question) and a theory of language usage (the WHY question) and language evolution (the HOW question). Section 3 addresses the first question, Section 4 addresses the second question, while Section 5 addresses the third question.

3   Small and subtle distinctions

What is the semantic core shared by the general imperfective attested in languages like Turkish, Tigre, Gujarati, or Romance and the more specific progressive realized in languages like English as well as in Turkish, Tigre, or Romance? And what is the content that distinguishes the two categories? A unified analysis of the two aspects must satisfy the desiderata of a single meaning that gives rise to distinct imperfective readings, and a clear source of typological variation in the manifestation of imperfectivity. To my knowledge, there exist two clear proposals that undertake this task — Ferreira 2005 and Deo 2009. In view of the broader focus of the paper, I will only describe the proposal in Deo 2009, which has the right structural properties for building an account of the observed diachronic phenomenon.

3.1 Deo 2009

Deo (2009), building up on prior ideas (Bonomi 1997, Delfitto & Bertinetto 1995, Cipria & Roberts 2000, Lenci & Bertinetto 2000) offers an account that is designed to characterize the similarities and the differences between the imperfective and the progressive. There are three main components to the account:

- The imperfective and progressive contain a universal quantifier whose domain is a regular partition of an interval. A regular partition is a set of collectively exhaustive, non-overlapping, equimeasured subsets of some set.

- The partition-measure (the length of each member of the regular partition) is a free variable with a contextually determined value. The range of readings associated with imperfective and progressive marking derive from this variability.

- The contrast between the imperfective and the progressive has to do with whether the quantifier domain is a regular partition of the reference interval (in the case of the progressive) or a superinterval of the reference interval (in the case of the imperfective).

The description here is taken almost verbatim from Deo 2009, with a few very small changes to the analysis. The repetition, though undesirable, is necessary in order to give the reader the necessary background for understanding this paper.

3.1.1 Semantics

The ontology includes a non-null set of intervals $\mathcal{I}$ (with points as a special case) partially ordered by the relation of temporal precedence $\prec$ and by the subinterval relation $\subseteq$; the variables $i, j, k \ldots$ range over $\mathcal{I}$.

$\mathcal{W}$ is a non-empty set of worlds. The historical alternatives of a world w at an interval i ($Hist_i(w)$) are those worlds w' in which the course of history up to a final subinterval of i does not diverge from w. The function Inr assigns to each $i \in \mathcal{I}$ a proper subset of $Hist_i(w)$ — the set of those worlds that continue beyond i in ways that are compatible with the normal course of events until i (Dowty 1979, p. 152). $Hist_{i_{inr}}(w)$ is the set of *inertial alternatives* of w at i.

(11)   *Inertial alternatives*

$Inr =_{def} f : I \rightarrow \wp(\mathcal{W})$

$i \mapsto Hist_{i_{inr}}(w) \subset Hist_i(w)$

$\mathcal{E}$ is a domain of eventualities, sorted into a set of events $\mathcal{E}^{\mathcal{E}}$ and a set of states $\mathcal{E}^S$. The temporal trace function τ from $\mathcal{E}$ to $\mathcal{I}$ gives the run time of an eventuality. The eventuality argument of basic eventive predicates is of the sort E while the eventuality argument of a basic stative predicate is of the sort S. *Sentence radicals* are predicates of eventualities (eventive or stative) built from such basic predicates with their individual (non-eventuality) arguments saturated (somewhat corresponding to the VP level, assuming VP-internal subjects). Quantified PPs, and aspectual modifiers such as negation, frequency and Q-adverbs, apply to such predicates of eventualities to yield predicates of intervals. Aspectual operators like the progressive or the imperfective may either apply to predicates of eventualities denoted by sentence radicals or to the predicates of intervals returned by aspectual modifiers. They map properties of eventualities or intervals to sets of intervals relative to which these predicates are instantiated via existential quantification over the Davidsonian event variable. Tense operators are functions that map predicates of eventualities or intervals to propositions, instantiating these properties at some reference time.

The instantiation of predicates at a time and a world is specified here in terms of the COINcidence relation defined as in (12). A predicate of events $P^{\mathcal{E}}$ stands in the coincidence relation with an interval i and world w iff P is instantiated in every inertial alternative of w within i or at some superinterval of i ($\circ$ stands for the temporal overlap relation). A predicate of intervals $P^{\mathcal{I}}$ or of states P^S stands in the coincidence relation with i and w iff the predicate holds throughout i in w.

(12) $\mathrm{COIN}(P, i, w) =_{def} \begin{cases} \forall w' \in Hist_{i_{inr}}(w): \\ \quad\quad \exists e[P(e)(w') \wedge \tau(e) \circ i] & \text{if } P \subseteq \mathcal{E}^{\mathcal{E}} \\ P(i)(w) & \text{if } P \subseteq \mathcal{I} \text{ or } \mathcal{E}^S \end{cases}$

The final notion needed in specifying the meaning of the progressive and imperfective operators is that of a *regular partition*, defined in (13). For any interval i, a partition of i is the set of the non-empty, mutually exclusive, and collectively exhaustive subsets of i.

(13) *Regular partition*

$\mathcal{R}_i$ is a regular partition of i if $\mathcal{R}_i$ is a set of intervals $\{j, k \dots n\}$ such that

a. $\bigcup \{j, k, \dots n\} = i$,

b. $\forall j, k \in \mathcal{R}_i \rightarrow j \cap k = \emptyset$ if $j \neq k$,

c. $\forall j, k \in \mathcal{R}_i \rightarrow \mu(j) = \mu(k)$
(where $\mu(x)$ stands for the Lebesgue measure of x).

For any $\mathcal{R}_i$, each of its subsets will have the same measure and this measure will be referred to by the term *partition measure.* Intuitively, a regular partition of i is a set of non-overlapping chunks of time of equal length partitioning i, against which predicate instantiation may be evaluated with respect to regular distribution in time.

With these notions in hand, it is possible to give appropriate meanings for the imperfective (IMPF) and the progressive (PROG) operators, which are both analyzed as universal quantifiers over times. IMPF combines with a predicate (of eventualities or intervals) P and an interval i and returns the proposition that there is some interval j which continues i such that every cell k of a "small-enough" regular partition of j, $\mathcal{R}_j^c$, COINcides with P. A "small-enough" regular partition over any interval i is a regular partition where the value of the partition measure does not exceed some contextual threshold as determined by the measure of the partitioned interval and properties of the event description.[3]

(14) IMPF: $\lambda P \lambda i \lambda w . \exists j[i \subseteq_{ini} j \wedge \forall k[k \in \mathcal{R}_j^c \rightarrow \text{COIN } (P, k, w)]]$

The characterizing and the event-in-progress readings of the imperfective depend on the context in which an imperfective sentence is uttered. The relative length of the interval j introduced by the imperfective determines what is a "small-enough" cell. If the interval under consideration is rather long relative to the typical duration of the event being described, then we obtain

---

3 The proposal in Deo 2009 takes the partition measure to be anaphoric on the context rather than vaguely determined by the measure of i and the event description. However, conversations with Lucas Champollion and the framework for measurement presented in Champollion 2010, have led me to think that the context-dependence of the partition measure is more appropriately modeled in terms of vagueness rather than the anaphoric retrieval of information.

the characterizing reading. If it is rather short relative to the typical duration of the event being described, we obtain the event-in-progress reading.

For clarity, (16) provides a step-by-step derivation demonstrating how the proposed meaning for IMPF combines with other semantic components in order to build up the meaning of IMPF-marked sentences. Let us assume that the Gujarati Imperfective (examples in (3)) realizes IMPF as given in (15).[4]

(15) $[\![-e]\!] = \lambda P \lambda i \lambda w . \exists j [i \subseteq_{ini} j \wedge \forall k [k \in \mathcal{R}_j^c \to \text{COIN}(P, k, w)]]$

The logical form for (16a) is in (16d). The sentence in (16a) denotes a proposition that holds of a world w iff there is some interval j containing *now* as its initial interval, whose every disjoint part k overlaps with an event of Niśā making bread, which is fully realized in the inertial alternatives of w at k.

(16) a. niśā   (roj)   roṭli   **banāv-e**   **ch-e**
          N.NOM   everyday bread.NOM make-IMPF.3SG   PRES-3SG

          Niśā *makes* bread (everyday).

      b. PRES(IMPF(λe[Niśā-make-bread(e)]))

      c. PRES$\Big(\lambda P \lambda i \lambda w . \exists j \Big[i \subseteq_{ini} j \wedge$
              $\forall k [k \in \mathcal{R}_j^c \to \text{COIN}(P, k, w)] \Big] (\lambda e[\text{Niśā-make-bread}(e)]) \Big)$

          $= \text{PRES}\Big(\lambda i \lambda w . \exists j \Big[i \subseteq_{ini} j \wedge$
              $\forall k [k \in \mathcal{R}_j^c \to \text{COIN}(\lambda e[\text{Niśā-make-bread}(e)], k, w)] \Big] \Big)$

          $= \text{PRES}\Big(\lambda i \lambda w . \exists j \Big[i \subseteq_{ini} j \wedge$
              $\forall k \Big[k \in \mathcal{R}_j^c \to \forall w' [w' \in Hist_{k_{inr}}(w) \to$
                  $\exists e[\text{Niśā-make-bread}(e)(w') \wedge \tau(e) \circ k]] \Big] \Big] \Big)$

      d. $\lambda w . \exists j \Big[now \subseteq_{ini} j \wedge$
              $\forall k \Big[k \in \mathcal{R}_j^c \to \forall w' [w' \in Hist_{k_{inr}}(w) \to$
                  $\exists e[\text{Niśā-make-bread}(e)(w') \wedge \tau(e) \circ k]] \Big] \Big]$

---

4 The Gujarati Imperfective paradigm is represented here by *-e*, which is the third person singular imperfective affix.

We turn now to the meaning of PROG, which differs from IMPF only in one respect. It restricts the domain of quantification to a regular partition of the reference interval, rather than a superinterval thereof.

(17)   PROG: $\lambda P \lambda i \lambda w. \forall j[j \in \mathcal{R}_i^c \rightarrow \text{COIN}(P, j, w)]$

As seen in the examples in (18b, d), progressive marking is compatible with a characterizing reading, and in such cases, licenses an inference that the situation described is temporally contingent and subject to change. This observation comes from Leech 1970, Comrie 1976, Dowty 1979, Goldsmith & Woisetschlaeger 1982 among others.

(18)   a.   Mary *was biking* to work…when she got hit.   *Event-in-progress*

       b.   Mary *was biking* to work…until she bought a car.   *Characterizing*

       c.   Mary *was baking* cookies yesterday.   *Event-in-progress*

       d.   Mary *was baking* cookies to make ends meet.   *Characterizing*

It is precisely this inference of temporal contingency that sometimes leads to the infelicity of the progressive with stative predicates, as in (19).

(19)   a. ?John *is owning* three houses.

       b. ?Mary *was knowing* the answer.

       c. ?New Orleans *is lying* at the mouth of the Mississippi River.

                                                      (Dowty 1979, p. 174)

       d. ?That argument *is resting* on an invalid assumption.

                                                      (Dowty 1979, p. 174)

The observation is that the English Progressive is acceptable with a stative predicate only when the situation denoted by the predicate is a contingent one, subject to change. More-or-less permanent situations, expressed by individual-level statives or by stage-level statives with immovable subjects, cannot be appropriately described using the Progressive.

The next section describes how the difference between the meanings and the forms of the progressive and imperfective give rise to the blocking effect (no event-in-progress reading for the imperfective), the inference of temporal contingency, and the related effect of (in)felicity.[5]

---

5 The argumentation for this part of the analysis that was given in Deo 2009 does not work exactly as was presented and is also inexplicit. It is superseded by the reasoning in Section (3.1.2).

3.1.2 The progressive–imperfective contrast

For any predicate P and interval i, $\text{IMPF}(P)(i)$ denotes the set of worlds where i is an initial subinterval of some interval j such that every cell of a small-enough partition $\mathcal{R}_j^c$ COINcides with P. For any $w \in \text{IMPF}(P)(i)$, it can be in set (20a) or (20b), and possibly in both. In contrast, going by the meaning in (17), $\text{PROG}(P)(i)$ denotes the set in (20b).

(20)   a. $\left\{ w \,\middle|\, \exists j[i \subset_{ini} j \wedge \forall k[k \in \mathcal{R}_j^c \rightarrow \text{COIN}(P, k, w)]] \right\}$

   The set of worlds w such that there is a proper superinterval j that continues i and for every k in a small-enough regular partition of j, P coincides with k in w.

   b. $\left\{ w \,\middle|\, \exists j[i = j \wedge \forall k[k \in \mathcal{R}_j^c \rightarrow (P, k, w)]] \right\}$

   The set of worlds w such that for every k in a small-enough regular partition of i, P coincides with k in w.

Whenever P is a *stative or temporal* predicate, for any interval i, $\text{PROG}(P)(i)$ and $\text{IMPF}(P)(i)$ denote exactly the same set — the union of (20a) and (20b). This is because $\text{COIN}(P, i, w)$ is defined as $P(i)(w)$ for stative and temporal predicates. Any world in which P holds throughout a (possibly non-proper) superinterval j of i is a world in which P holds at i and vice versa.

However, for any *eventive* predicate P and interval i, $\text{PROG}(P)(i)$ is semantically stronger than $\text{IMPF}(P)(i)$. It is easy to see how PROG is a "semantically narrower" version of IMPF on this construal of their contribution, since $\text{PROG}(P)(i)$ asymmetrically entails $\text{IMPF}(P)(i)$ whenever P is eventive. The reasoning is as follows: if every cell of a small-enough partition of i coincides with P in w, it follows that there is a superinterval j of i such that every cell in a small-enough partition of j coincides with an event of type P in w. The opposite does not hold, since a small-enough partition of some superinterval j of i need not correspond to a small-enough regular partition of i. This would be the case if the size of j is much larger than the size of i. As an example, consider a sentence like *John eats dairy* (which contains IMPF in its logical form) uttered in response to a question about whether we need to make a vegan dish for the dinner party. This sentence does not entail *John is eating dairy* during the reference interval — the time at which the question is asked.

These entailment relations for IMPF and PROG in composition with different sorts of eventuality descriptions lead to certain inferential patterns familiar from the domain of scalar implicatures. In particular, they allow us

to explain blocking effects and temporal contingency inferences in languages that realize both aspectual operators via distinct devices. In such a language, speakers must determine which device to use in a given context. Hearers, in interpreting a sentence containing one of the two devices, must determine its optimal interpretation given that the alternative device was not used. The reasoning involved underlies the scalar inferences that arise in this domain.[6]

In what follows, I will assume that the two operators are in a privative opposition, an opposition in which one element is unspecified for some semantic feature that the other element is specified for.[7] PROG is specified for a particular feature: it distributes the event description over the reference interval. IMPF is underspecified with respect to that feature. This allows us to construe the exponents of IMPF and PROG as forming a Horn scale ⟨PROG, IMPF⟩ with the IMPF exponent being the weaker alternative. While PROG has a more specific semantics than IMPF, we have already seen that $\text{PROG}(P)(i)$ is not always semantically stronger than $\text{IMPF}(P)(i)$ — specifically, in the case of stative predicates, where the two are equivalent. The inferences that arise come from the enrichment of literal meanings in utterance contexts.

First, an account of the blocking effect. Take an eventive predicate P and interval i. In this case, $\text{PROG}(P)(i)$ asymmetrically entails $\text{IMPF}(P)(i)$ and is therefore the stronger alternative. The use of the weaker form, $\text{IMPF}(P)(i)$, implicates that the speaker is not in a position to convey $\text{PROG}(P)(i)$, since otherwise she would have used the stronger form. $\text{IMPF}(P)(i)$ therefore conveys the strengthened proposition in (20a), which is that P is distributed over a partition of some proper superinterval j of i. This, in turn, blocks $\text{IMPF}(P)(i)$ from having an event-in-progress interpretation. Here is how the reasoning works: For the event-in-progress interpretation to obtain, there must be a single event $e \in P$, whose run-time, $\tau(e)$ (which is a continuous interval), overlaps with every cell of a small-enough partition of j. If there

---

6 Over time, these scalar inferences might get encoded as part of the usage conventions of PROG and IMPF marking. At least in the case of the English tense-aspect system, the conventionalized division of labor between the Simple Present and the Progressive seems to suggest that this is the case.

7 The notion of a privative opposition in the context of scalar meanings comes from Horn & Abbott (2012). According to them, a Horn scale is a privative dyad of the form ⟨S,W⟩, when the source of the informativeness asymmetry is that one element is more specific than the other with respect to some feature. They argue that the definite and indefinite article in English have the structure of a privative dyad. Other examples they provide include pairs of lexical expressions like ⟨thumb, finger⟩, ⟨square, rectangle⟩, and functional pairs like ⟨himself, him⟩, or ⟨this, that⟩.

is such an event e, then it follows that $\tau(e)$ also overlaps with every cell of a small-enough partition of i. But if that is the case, then the speaker *was*, in fact, in a position to convey $\text{PROG}(P)(i)$. However, the speaker chose to use the weaker form $\text{IMPF}(P)(i)$. Therefore it must be the case that the intended interpretation was not the event-in-progress interpretation. Thus, the presence of PROG as a distinct grammatical device typically blocks the availability of an event-in-progress interpretation for IMPF with eventive predicates.

Next, an account of the temporal contingency inference with stative/temporal predicates. For any such predicate P and interval i, $\text{PROG}(P)(i)$ and $\text{IMPF}(P)(i)$ denote the same set of worlds; that is, they entail each other. If two expressions are semantically equivalent, there is a preference for less complex utterances — a Manner consideration. $\text{IMPF}(P)(i)$ is morphosyntactically less complex than $\text{PROG}(P)(i)$.[8] The use of the more complex form, $\text{PROG}(P)(i)$, gives rise to the manner-based implicature that the speaker has a reason for not choosing the simpler alternative. Having made this inference, the hearer must identify the reason for the speaker's choice. The empirical observation is that the hearer, in most contexts, takes this reason to be that the speaker does not have evidence that P holds *beyond* the reference interval i. This is the temporal contingency inference. It is indeed logically possible that the hearer reasons the other way around and interprets the reason for the use of $\text{PROG}(P)(i)$ as being to convey that P *does* hold beyond the reference interval. On this reasoning, $\text{IMPF}(P)(i)$ would be reserved for conveying that P holds only at the reference interval and not beyond. But this association of forms with interpretations does not obtain. I suggest that the reason that the enrichment of the semantics takes the form of the temporal contingency inference is because it is the only one by which $\text{PROG}(P)(i)$ asymmetrically entails $\text{IMPF}(P)(i)$. This mirrors the asymmetric entailment pattern that arises obligatorily with eventive predicates.

By the above reasoning, $\text{PROG}(P)(i)$ (where P is stative) receives an upper-bounded interpretation. Its use leads to the exclusion of the worlds in (20a) as in (21). The COIN relation has been resolved as defined for stative predicates.

---

8 This is not only a fact about English; it has been observed that crosslinguistically progressive forms tend to be periphrastic (Dahl 1985, Bybee, Perkins & Pagliuca 1994) and therefore structurally more complex than imperfective forms, which are more likely to be synthetic.

(21) $\lambda w.\forall k[k \in \mathcal{R}_i^c \rightarrow \text{COIN}(P, k, w)] \wedge$

$\neg \exists j[i \subset_{ini} j \wedge \forall k[k \in \mathcal{R}_j^c \rightarrow \text{COIN}(P, k, w)]]$

$= \lambda w.P(i)(w) \wedge \neg \exists j[i \subset_{ini} j \wedge P(j)(w)]$

This upper-bounded interpretation, in turn, may give rise to infelicity. To make this concrete, consider the stative predicate *New Orleans lie at the mouth of the Mississippi river*, which we abbreviate as *N*. The semantically equivalent alternatives available for expressing the corresponding tensed proposition are either the morphosyntactically simpler Simple Present (22a) or the complex Present Progressive (22b). The speaker's choice to use the Present Progressive leads to the inference that the speaker has a reason for not using the Simple Present, likely because she does not have evidence that *N* holds beyond the reference interval **now**. This reduces the set of PROG worlds in (22b-i) to the set in (22b-ii), in which *N* does not hold at proper superintervals of the reference interval **now**.

(22) a. New Orleans *lies* at the mouth of the Mississippi River. IMPF

$\lambda w.\exists j[\textbf{now} \subseteq_{ini} j \wedge \forall k[k \in \mathcal{R}_j^c \rightarrow N(k)(w)]]$

$= \lambda w.\exists j[\textbf{now} \subseteq_{ini} j \wedge N(j)(w)]$

b. ?New Orleans *is lying* at the mouth of the Mississippi River. PROG

   i. $\lambda w.\forall k[k \in \mathcal{R}_{\textbf{now}}^c \rightarrow N(k)(w)]$

   ii. $\lambda w.\forall k[k \in \mathcal{R}_{\textbf{now}}^c \rightarrow (N, k, w)] \wedge$

$\neg \exists j[i \subset_{ini} j \wedge \forall k[k \in \mathcal{R}_j^c \rightarrow (N, k, w)]]$

$= \lambda w.N(\textbf{now})(w) \wedge \neg \exists j[\textbf{now} \subset_{ini} j \wedge N(j)(w)]$

This upper-bounded reading associated with the choice of (22b) is at odds with the expectation in most contexts about the location of cities with respect to geological bodies. Such relations are expected to be more permanent, and continue indefinitely into the future, making the use of the complex PROG alternative infelicitous in most contexts. For instance, (22b) is an infelicitous response to a question like *Where is New Orleans located?* But it would be a felicitous response in a situation where there has been a drastic change in the course of the lower Mississippi.[9]

---

9 In recent years, the Mississippi has shown a steady shift towards the Atchafalaya River channel — a course change that would prove disastrous to cities such as New Orleans and Baton Rouge.

(23) a. A: And what about the cities on the lower Mississippi? How have they fared?

  b. B: Well, New Orleans is (still) lying at the mouth of the Mississippi river, but we don't know for how much longer.

3.2 Summary

To summarize, the semantics proposed for PROG and IMPF here (and in Deo 2009) make it possible to treat cross-linguistic exponents of these meanings as members of a scale of alternatives. Combining these operators with eventive and stative predicates yields distinct entailment patterns, which result in strengthened interpretations for both IMPF-marked and PROG-marked sentences. Sentences containing the semantically under-specified member, IMPF, are blocked from receiving the event-in-progress interpretation with eventive predicates. Sentences containing the morphosyntactically complex member PROG with stative predicates receive an upper-bounded reading, giving rise to the temporal contingency inference.

To remind the reader, the goal of this section was to pin down the precise content of the "small and subtle semantic distinctions" between the general IMPF and the semantically narrower PROG. Determining the precise content of the imperfective and the progressive is in service of the larger goal of this paper: understanding how progressive markers emerge and diachronically generalize into markers of imperfective aspect. The meanings provided in this section can be now straightforwardly taken as the input to and the output of the progressive≫imperfective path. I address the dynamic workings of this path in the remainder of this paper.

4 The cyclic diachronic pattern

There are four discernible states of the linguistic system in the progressive≫ imperfective path. There is the initial *zero*-PROG state (24a), in which the language possesses only a single grammaticalized device across the imperfective domain — the exponent of IMPF. The second, *emergent*-PROG state (24b), is one in which morpho-syntactic resources of the language have been recruited in introducing an exponent for PROG. In this state, progressive marking is grammaticalized but optional. The third state is the *categorical*-PROG state (24c). In this state, the exponents of PROG and IMPF have relatively circum-

scribed (though overlapping) domains of use, with at least some categorical sub-domains for each. State four, which we might call *generalized*-PROG (24d), is the state in which the exponent of PROG loses its semantic restriction and generalizes to IMPF.[10]

(24) a. X_{IMPF} zero-PROG

   b. $(Y_{\text{PROG}}), X_{\text{IMPF}}$ emergent-PROG

   c. $Y_{\text{PROG}}, X_{\text{IMPF}}$ categorical-PROG

   d. Y_{IMPF} generalized-PROG

This is a highly schematic description and each state probably consists of several sub-states which would differ from each other in subtle, and possibly idiosyncratic ways involving aspectual properties of predicates that may/may not combine with IMPF or PROG. What is of interest here are the processes involved in the three main observable transitions in the grammaticalization path:

(25) a. *Recruitment* of existing morpho-syntactic resources to innovate a new functional category/expression within a semantic domain — the change from (24a) to (24b).

   b. *Categoricalization* of the new expression to obligatory use in certain contexts within the domain — the change from (24b) to (24c)

   c. *Generalization* of this expression to the broader semantic domain — the change from (24c) to (24d).

In this section, I give an intuitive outline of these transitions, rooting their existence in a conceptual contrast between two kinds of inquiries — phenomenal and structural inquiries. Section 4.1 describes this contrast and its connection to IMPF and PROG. It is suggested that the semantics of PROG makes progressive morphology (its overt exponent) better suited for marking phenomenal inquiries, thus allowing interlocutors to formally distinguish between the two. In contrast, languages without a PROG exponent must rely on the common contextual knowledge of interlocutors or optional disambiguators such as temporal adverbs. In Section 4.2, I informally describe the three

---

10 As far as speakers are concerned, (1d) can be interpreted as a zero-PROG stage and the language may innovate a new expressive device corresponding to PROG, resulting in yet another grammaticalization cycle. It is likely that the correct way to think about grammaticalization paths in the domain of aspectual morphology is as Jespersonian cycles with repeated processes of weakening and morphosyntactic reinstatement of salient semantic contrasts. Some relevant discussion is in Section 6.

transitions: recruitment, categoricalization, and generalization. In Section 4.3, I briefly address the question of why languages do not seem to exhibit changes in which generic/habitual markers (dedicated devices for structural inquiries) generalize to become imperfective markers.

4.1 The phenomenal/structural distinction

What sorts of inquiries are imperfective assertions useful for? Restricting attention to descriptions of ongoing (rather than completed) situations relative to some reference time, I will distinguish between two broad kinds of inquiries into the state of the world, either one of which might be the goal of a given discourse. The first is an inquiry into the stable facts and generalizations that characterize (in a relatively timeless way) the actual world, while the second concerns itself with facts of more local import, facts that pertain to specific times and the events that occupy such times. The questions in (26a) and (26b) illustrate the difference between inquiries that I have in mind. Note that the questions in (26a) are expressed in English using the Simple Present tense while those in (26b) use the Progressive.

(26) a. *What characterizes the world generally?*

What problems do developing nations face?
Why do dogs wag their tails?
How do whales give birth?
Does John walk to school?
What does the earth revolve around?

   b. *What characterizes the world at some time i?*

What problems are developing nations facing in 2012?
Why is Fido wagging his tail right now?
How is that whale giving birth?
Is John walking to school?
What was the earth revolving around on Tuesday evening?

Following intuitions underlying Goldsmith & Woisetschlaeger's (1982) analysis of the English Progressive (hereafter G&W), I will call the first kind of inquiry *structural* and the second kind of inquiry *phenomenal*. G&W proposed that the contrast between the structural and phenomenal views of the world indicates a fundamental classification of the types of knowledge we possess. Their basic idea is that language (and the conceptual structure that under-

lies it) distinguishes between properties that are seen to contingently hold of the world — attributable to the "capriciousness of nature" (p. 88) — and properties that hold non-contingently or essentially of the world. According to them, the Progressive is used to describe (and inquire about) *phenomena* or observable occurrences, particular things that happen (or are happening) in the world at particular times, while the Simple Present tends to be used to describe what we take to be facts about the *structure* of the world. It is this metaphysical distinction that they claim underlies the distribution of the English Progressive and the Simple Present.

There is a connection between the aspectual semantics of PROG and IMPF in Section 3.1 and the phenomenal/structural contrast that G&W rely on in their analysis of the progressive. For any predicate P, interval i, and world w, $\text{PROG}(P)(i)(w)$ entails that P COINcides with i at w — a P eventuality is observable at i.[11] There is no such entailment with $\text{IMPF}(P)(i)(w)$. Thus, a progressive sentence entails the existence of one or more P-eventualities whose run-time overlaps with the reference interval i, while an imperfective sentence carries no such entailment. It has the weaker entailment that some superinterval of i overlaps with the run-time of one or more P-eventualities. Based on this difference, I suggest that PROG is better suited to phenomenal inquiries than structural inquiries. This is because phenomenal inquiries are about occurrences observable at a given reference time while structural inquiries are about facts that obtain at a given reference time but for which there might not be any observable evidence (in the form of actual occurrences) within that time. If this is on the right track, then a language which realizes PROG can use it to mark phenomenal inquiries and thereby distinguish these from structural inquiries. There are two corollaries to this thesis:

(27) a. Any language that realizes both PROG and IMPF can conventionally restrict the use of IMPF-marking for structural inquiries — maximizing the disambiguation between phenomenal and structural inquiries.

b. Any language that realizes only IMPF must rely on context to disambiguate between phenomenal and structural inquiries.[12]

---

11 Since every cell of $\mathcal{R}_i^c$ must COINcide with P in w, it follows that i itself must COINcide with P in w.

12 This information may be contextually retrieved just like information about temporal location may be contextually retrieved in tenseless languages (See Bohnemeyer 2002, 2009, Bittner 2005, Tonhauser 2011. And optional adverbials (e.g., *right now*, *yesterday evening*, *last week*),

The basic assumption we need to make is that there is some functional pressure to distinguish between phenomenal and structural inquiries. Linguistic systems may either disambiguate phenomenal and structural meanings formally (using PROG-marking) or distinguish them using contextual cues. The three transitions in (25a-25c) are essentially changes from one kind of system to the other.

4.2   The three transitions

4.2.1   Recruitment: the emergence of variants

Speakers and listeners in a zero-PROG language regularly participate in speech events which require disambiguation of phenomenal inquiries in the local context. In underspecified contexts, participants may make explicit efforts towards such disambiguation. This may be effected either non-linguistically or by optional linguistic devices such as frame adverbials (*right now, last night, at that time*) and periphrastic constructions.[13] Such devices overtly introduce the reference intervals throughout which the eventuality predicates are asserted to hold. The inquiry is determined to be phenomenal by the use of such devices because they restrict the temporal interpretation of the question or the assertion narrowly to the reference interval. In some zero-PROG languages, such morphosyntactically explicit efforts may accumulate and acquire high statistical frequency. Grammaticalization of PROG amounts to the conventionalized and reliably frequent (but optional) use of one privileged exponent for conveying that the inquiry is restricted to the reference interval. Once such an exponent is chosen, the language enters the emergent-PROG state.

4.2.2   Categoricalization: The conventionalization of variants

Speakers of an emergent-PROG language have progressive marking to (optionally) mark phenomenal inquiries but no comparable device for structural inquiries. This gives way to the categorical-PROG state where progressive marking is employed almost universally in phenomenal inquiries while the

which are not part of the grammaticalized tense–aspect system of the language are always available to formally facilitate disambiguation.

13 Bybee, Perkins & Pagliuca (1994, pp. 127–130) observe that locative expressions like prepositions (e.g., *be at V-ing, be on V-ing*) and posture verbs (e.g., *stand V-ing, sit V-ing*) are frequently harnessed in the creation of new progressive marking crosslinguistically.

use of imperfective marking is conventionally associated with structural inquiries. This evolutionary pattern can be construed as an instance of Horn's (1984) *division of pragmatic labor* in which the distribution of competing expressions is determined by the dynamic between hearer and speaker oriented pragmatic principles — Horn's Q and R principles.[14] One of the phenomena Horn discusses involves the evolution of privative oppositions (pp. 33–35) such as the one characterizing the PROG–IMPF relation. In such cases, the existence of the more informative, marked form, together with the speaker's choice of the unmarked, semantically broader form in a given context allows the addressee to construct a Q-based implicature that the semantic content associated with the marked form was *not* intended by the utterance of the unmarked form (extrapolating from Horn 1984, pp. 37–38).[15] The conventionalization of this implicature pattern is the categorical-PROG system.

4.2.3 Generalization: The failure to acquire conventionalized variants

Speakers of a categorical-PROG language can unambiguously mark both phenomenal and structural inquiries. However, this state of the language diachronically gives way to the generalized-PROG state in which the progressive form changes in meaning to realize IMPF rather than PROG. This is the cyclicity of the pattern: speakers of a generalized-PROG language have exactly the same system as speakers of a zero-PROG language; they must rely on contextual information and optional disambiguators. What leads to this counter-intuitive transition, in which the expressivity afforded by recruitment and categoricalization at prior states of the language is obliterated at a later state? We must characterize generalization as the cumulative failure of speakers over

---

14 Horn's theory is inspired by Zipf (1949) who distinguishes between a speaker's economy, which is oriented towards simpler messages and auditor's economy, which is an anti-ambiguity principle, oriented towards more explicit messages. The idea that language change emerges from the interaction between two factors in communication: the speaker's need to convey a message and the principle of least effort is also found in Martinet (1962)) and goes back to Paul (1888).

15 Horn's examples involve cases of morphological blocking in the lexicon, where existence of specific forms, perhaps simplex, blocks the application of general morphological rules (e.g., *thief* blocks *stealer*, except in special cases). He also discusses language change phenomena which give examples of Q-based narrowing in the lexicon. The decision to use the term *division of pragmatic labor* for the scalar implicature arising from competition between specific and general forms is thus licensed by its original usage (and its original user, who agreed that this was not an incorrect construal of his intent).

several generations to acquire the categorical-PROG system. But what causes such failure and the resulting loss of a salient semantic distinction?

There may be several ways to pursue this question, but I will propose one way that I find most promising. Suppose that the population of acquirers for a language is biased towards formally simpler grammars – such as zero-PROG/generalized-PROG. This preference is defeasible and the learned grammar is ultimately determined by the structure of the input that the child receives. However, as the categorical-PROG system becomes established in the speech community, the frequency of overt PROG-marking in the child input would correspondingly increase. Because acquirers are biased towards simpler grammars, this would lead to increased probability of mis-learning a generalized-PROG grammar from the categorical-PROG input. Thus increased failure to learn the categorical-PROG system accurately would be tied to the success (in terms of frequency) of that system in the population.

4.3 The non-attestation of generic≫imperfective paths

A logical counterpart to the progressive≫imperfective path described above is one in which speakers start out by innovating a dedicated device for structural inquiries. We could say that such a marker realizes the operator GEN (or HAB). In the emergent-GEN state, speakers would use the innovated marker in underspecified contexts for disambiguating structural inquiries. Over time, the system would evolve towards the categorical-GEN state, in which IMPF is conventionally restricted to marking phenomenal inquiries and GEN is used categorically for structural inquiries. In the generalized-GEN state, the language would have a single exponent for IMPF realized by the GEN-marking and use contextual information for disambiguating between phenomenal and structural inquiries.

There is nothing that should prevent novel material from being recruited for marking characterizing meanings or structural inquiries. In fact, many languages do exhibit habitual/generic markers that are distinct from IMPF marking (e.g., the *used to V* construction in Modern English). Right around the time that Middle English developed the Progressive, it also developed a dedicated construction *uses to V* (which parallels the past tense construction) to express characterizing meanings in the present (Tagliamonte & Lawrence 2000).[16] So it is not that languages do not innovate such markers. The puzzle,

---

16 Thus, at this stage, the language, in addition to the simple tenses, disambiguates both phenomenal and structural inquiries with new material. It is interesting that the *uses to*

rather, is that there seems to be no attestation of such habitual/generic markers changing in meaning to encompass the function of the progressive (e.g., being used to describe events in progress). That is, there is no "habitual/generic≫imperfective path" that would be the mirror image of the progressive≫imperfective path. This asymmetry in diachronic patterns deserves some comment.

While I do not have a full solution to this puzzle, I believe that this asymmetry likely stems from the nature of the input to the child, specifically the relative prevalence of PROG forms vs. GEN forms in caregiver speech. My suggestion is that caregiver speech involves inquiries concerning the "here-and-now" more frequently than inquiries concerning stable regularities. In a categorical-PROG population, this asymmetry in the frequency of phenomenal vs. structural inquiries in child-directed speech would lead to learners generalizing the PROG form rather than any specialized GEN form, since exposure to the latter is likely to be less frequent. This of course needs to be corroborated through a study of child-directed speech that explicitly examines contexts of utterance. But if it is correct, then it predicts the unidirectionality of the path of change, rooting it in the functional properties of the progressive that effect asymmetries in frequency in care-giver input. Indirect evidence for this comes from the literature on first-language acquisition of tense and aspect, which suggests that the use of progressive marking is much more frequent than simple present marking in child-directed speech. For instance, in Li et al. 2001, which used several corpora from CHILDES to study parent input, the frequency of the progressive ($n = 2203$) in parental speech is seen to be much greater than that of the simple past ($n = 745$) and the simple present ($n = 557$) put together. Further, Shirai 1994 argues that parental input is a crucial factor in the overgeneralization of progressive marking observed in first language acquisition.

5 A game-theoretic interpretation of grammaticalization paths

The four states in (24) from Section 4 can be intuitively regarded as distinct strategies for communicating phenomenal and structural sub-meanings within the imperfective aspectual domain. The emergent-PROG and categorical-

---

V construction falls out of the language and the Simple Present tense gets conventionally restricted for expressing characterizing meanings. Describing and understanding this phenomenon would take us too far away from the goal of this paper, which is to understand how *progressive* marking grammaticalizes and generalizes in languages.

PROG states represent linguistic strategies that use multiple forms with disambiguated meanings, thus ensuring communication through formal explication. The zero-PROG and generalized-PROG states represent linguistic strategies that use a single form while relying on the hearer's understanding of contextual cues for successful communication. These differences in strategy profile should ideally be connected to the observed transitions between them. This section is an initial attempt at making this connection explicit using the resources of Evolutionary Game Theory (EGT) as they have been applied to the problem of linguistic communication and pragmatic reasoning in game-theoretic pragmatics.[17] EGT studies the general problem of strategy selection and its propagation across a population, attributing a non-central role to rationalistic reasoning in this process (Maynard Smith & Price 1973). For any population in which individuals exhibit different forms of behavior (which may or may not be the result of conscious choices), EGT allows us to determine which behaviors are able to persist, and which tend to be driven out.

We use these tools to model two aspects of the grammaticalization question. The *typological* aspect has to do with what the properties of linguistic strategies corresponding to the four states — zero-PROG, emergent-PROG, categorical-PROG), and generalized-PROG are. The *cross-temporal* aspect has to do with the evolutionary dynamics that can model why recruitment, categoricalization, and generalization appear to occur in cyclic fashion in the imperfective domain.

The analysis has two parts: Section 5.1 models the four strategies that correspond to the observed states. These strategies are shown to have differential profiles with respect to their communicative success and their formal complexity. Section 5.2 describes the behavior of these strategies over time using the replicator–mutator dynamics, which also takes into consideration the probability of imperfect transmission (or mutation) associated with each. The grammaticalization path is reconstructed as a cyclic pattern in which alternative strategies rise and fall in dominance within a given population over time due to contingent and structural factors that effect changes in their frequency.

---

17 Game-theoretic models of communication as a coordination game between the sender and the receiver of a signal can be traced to the work of Lewis 1969. Excellent introductions to game theory as applied to linguistics can be found in Benz, Jäger & van Rooij 2006 and Jäger 2004.

5.1 The Imperfective Game

The Imperfective Game builds on the basic model for communication used in the context of linguistics: the utterance situation is modeled as a game between the speaker and the hearer in which the speaker aims to convey some private knowledge to the hearer through her utterance. The game model specifies possible choices of linguistic signals for the speaker and possible interpretations of these signals for the hearer. Solution concepts for a language game can be understood as formal rules that predict how the game will be played out based on speaker and hearer preferences (signal economy, successful communication).

We represent the available linguistic options in terms of speaker–hearer strategy pairs, which can be seen as possible behavioral patterns associated with each individual in a population. If successful, a given speaker–hearer strategy pair can spread within a population through imitation or some other kind of adaptive behavior. A strategy pair is successful when (i) it leads to successful communication and (ii) it does so with small cost (van Rooij 2004b, p. 516).

We start with a system with speaker and hearer, where the speaker might be in one among two disjoint states {**phen**, **struc**}. The speaker may use one of the forms {*prog*, *impf*} to communicate the state she is in to the hearer.[18] A state here is identical to the nature of the inquiry that the speaker intends to address with the use of a given form. So the speaker is in a **phen** (or **struc**) state iff her utterance is intended to address a phenomenal (or structural) inquiry. The hearer, upon receiving the form, must choose an interpretation for it. If the hearer chooses the interpretation intended by the speaker, the communication is successful, otherwise not. In sending and receiving particular messages, the speaker and hearer must choose strategies, which determine the form chosen by the speaker in each state that is to be communicated and the interpretation given to each form by the hearer. A speaker's strategy is some function from states/meanings to forms (in this case, an element of [{**phen**, **struc**} → {*prog*, *impf*}]) while a hearer's strategy is a function from forms to meanings/states (in this case, an element of [{*prog*, *impf*} → {**phen**, **struc**}]). The utility function for the speaker and the

---

18 Throughout this paper, I have reserved small caps (PROG, IMPF) for denoting semantic operators. Italics (*prog*, *impf*) are now used to denote the forms that may be used to realize these operators. For particular linguistic forms, I have used the standard convention of capitalizing the first letter of the category (e.g., English Progressive, Hindi Imperfective etc.).

hearer is defined with respect to such strategies. Given a speaker strategy S, a hearer strategy H, and a state t, the success of communication in any given utterance situation can be measured as follows by the δ-function (Jäger 2007).

(28) $\quad \delta_t(S,H) = \begin{cases} 1 & \text{if } H(S(t)) = t \\ 0 & \text{otherwise} \end{cases}$

The δ-value is 1 if the hearer's interpretation for the form chosen by the speaker $S(t)$, $H(S(t))$, matches the meaning t intended by the speaker and 0 otherwise.

The cost associated with a strategy has implications for its use — speakers value formal economy as well as successful communication. Let us assume that the use of multiple forms within a single conceptual domain is costly, reducing the utility of a speaker strategy which employs multiple forms (Rubinstein 2000, 31ff). This can be expressed by the following speaker utility function, where n is a function that returns the number of expressions over one (the minimum necessary) employed in S for communicating the full range of meanings. Thus, for a speaker strategy that employs only a single form to communicate both **phen** and **struc**, $n(S)$ will be 0, while for a strategy that employs two forms to communicate the two meanings, $n(S)$ will be 1.

(29) $\quad U_s(t,S,H) = \delta_t(S,H) - k \times n(S)$

Following Jäger (2007), k is taken to be some parameter that modulates the expected utility for a strategy across systems. Jäger interprets this parameter in terms of the speaker's priorities – that is, in terms of how highly the speaker values linguistic clarity (disambiguation) over signal cost. In a system in which k is set to a low value, communicative success is valued more highly than signal cost. The lower/higher the value for k, the lower/higher the reduction in the speaker's expected utility for a strategy employing a costly form.[19] In the model assumed here, the value of k similarly determines the expected utility of a strategy. However, rather than tying this to the morphosyntactic complexity of a given form, it is tied to the complexity of entire strategies. Thus, a high (low) value for k corresponds to reduced (increased) utility for speaker strategies using multiple forms. For single-form strategies, the value of k makes no difference to the utility.

---

19 In the particular typological model for case marking patterns that Jäger builds, k is taken to vary across languages, and is concretely correlated with properties of linguistic systems such as degree of freedom of word order.

One might say that a hearer strategy should be taken to be more or less complex, depending on how many distinctions between meanings it makes. The idea would be that a hearer strategy that assigns to every message the same meaning is simpler than one that variably assigns distinct meanings to the message(s) that are its input. However, given that such a strategy would also be interpretively deficient — it would not allow for a subset of meanings to be distinguished — it is not comparable with hearer strategies that are able to access the full range of meanings. I will assume therefore, that any strategy that assigns the full range of meanings to messages in the input does not differ in complexity. Moreover, the hearer has no choice between or means of identifying speaker strategies, but must simply determine the speaker's intended meaning on the basis of the presented form. Hearer utility is taken to be identical to the δ-function.

(30) $\quad U_h(t, S, H) = \delta_t(S, H)$

Nature deals out the states **phen** and **struc** according to some probability distribution, which determines the likelihood of each state to be expressed.[20] The speaker has knowledge of the state she is in while the hearer lacks this knowledge. The average utility of a speaker or hearer strategy can be calculated by summing over the utility of the strategy in each state weighted by the probability of the state.

(31) $\quad U_s(S, H) = \sum_t P(t) \times (\delta_t(S, H) - k \times n(S)) \; U_h(S, H) = \sum_t P(t) \times \delta_t(S, H)$

The game model must further factor in the role of context in the disambiguation of meanings. van Rooij (2004a) proposes an enrichment of signaling games that facilitates the modeling of underspecified meanings by taking contexts into consideration. The general motivation for introducing contexts is to be able to capture the fact that the same form can be ambiguous between different interpretations, either of which might be the salient one in a given communication context. In our specific case, the form is *impf*, which is underspecified with respect to **phen** and **struc** interpretations that get resolved in context. Following van Rooij, we will assume that a context is a probability distribution over the state space {**phen**, **struc**}. We distinguish between two kinds of contexts: C_{phen} in which $P(\textbf{phen}) = 0.9$ and $P(\textbf{struc}) = 0.1$ and C_{struc} in which $P(\textbf{struc}) = 0.9$ and $P(\textbf{phen}) = 0.1$. Both contexts are equally likely

---

20 As Jäger (2007, p. 82) notes, this distribution is not a variable language-peculiar fact, but rather represents universal cognitive and communicative tendencies.

and knowledge of the context is common ground among the interlocutors. But only the speaker knows for each context the state she is in. A speaker strategy is now a function from states and contexts to forms, while a hearer strategy is a function from forms and contexts to states. The speaker may use one of the forms {*prog*, *impf*} to convey their state in each context. The hearer, correspondingly, interprets the form that she receives as conveying one of **phen** and **struc**. The speaker and hearer strategies to be considered are in Table 1 and Table 2.

	C_{phen}		C_{struc}	
	phen	**struc**	**phen**	**struc**
S_{cd}	*impf*	*impf*	*impf*	*impf*
S_{pcd}	*impf*	*impf*	*prog*	*impf*
S_{em}	*prog*	*impf*	*prog*	*impf*
$S_{cd'}$	*prog*	*prog*	*prog*	*prog*

Table 1   Speaker strategies

S_{cd} is a "context dependent" (*cd*) strategy in which the speaker uses the same form *impf* to convey both **phen** and **struc** relying on shared context for disambiguation. S_{pcd} is a "partially context dependent" (*pcd*) strategy; the speaker uses the unambiguously **phen** form *prog* only in C_{struc}, the context in which **struc** is the more probable state. S_{em} is a context independent "explicit marking" (*em*) strategy, one in which the speaker uses the forms *prog* and *impf* across contexts to convey **phen** and **struc** respectively. Finally, $S_{cd'}$ is exactly like S_{cd}, a context dependent strategy using *prog* instead of *impf*.

	C_{phen}		C_{struc}	
	prog	*impf*	*prog*	*impf*
H_{cd}	**phen**	**phen**	**struc**	**struc**
H_{pcd}	**phen**	**phen**	**phen**	**struc**
H_{em}	**phen**	**struc**	**phen**	**struc**

Table 2   Hearer strategies

H_{cd} is a context dependent hearer strategy in which the hearer is insensitive to form and relies only on context to recover the intended meaning. H_{pcd}

is partially context dependent because it invariantly assigns **phen** to *prog* but relies on context for interpreting *impf*. In H_{em}, a form dependent hearer strategy, *prog* and *impf* are invariantly assigned **phen** and **struc** meanings respectively.[21] Factoring in the role of contexts, the average speaker and hearer utilities can now be calculated in the following way. $P_c(t)$ stands for the probability that a state t is being communicated in context c.

(32) a. $U_s(S, H) = \sum_c P(c) \times \sum_t P_c(t) \times (\delta_t(S, H) - k \times n(S))$

b. $U_h(S, H) = \sum_c P(c) \times \sum_t P_c(t) \times \delta_t(S, H)$

Given the parameters above we can compute the average utility for speakers programmed for a particular strategy based on (32). The results are in Table 3. The utility for hearers is exactly the same without the cost k factored in.

Strategies	H_{cd}	H_{pcd}	H_{em}
S_{cd}	0.9	0.9	0.5
S_{pcd}	$0.9 - k$	$0.95 - k$	$0.55 - k$
S_{em}	$0.9 - k$	$0.95 - k$	$1 - k$
$S_{cd'}$	0.9	0.5	0.5

Table 3   Average utilities

The game as introduced so far is an *asymmetric* game — that is, a game in which the roles of the two players are distinguishable. This is the case here since every individual is either a speaker or a hearer and the speakers and the hearers make use of disjoint sets of strategies. In an evolutionary, population-dynamic setting, this amounts to there being two independent populations — one corresponding to the speakers and the other to the hearers. This, however, is not faithful to the actual setting we are trying to model,

---

21 The strategies considered in this game model do not exhaust the logical space of strategies for the imperfective game. For instance, we do not consider strategies in which the state **struc** is disambiguated (whether in less probable or in all contexts) using a distinct form, say *gen* either in conjunction with *prog* alone, *impf* alone, or both. A more complete game-theoretic account of changes in the imperfective domain must consider these strategic options. I do not consider these here because of the focus on the progressive≫imperfective cycling path and the non-attestation of the reverse path (Section 4.3).

where every individual in a speech community may sometimes take the role of hearer and sometimes that of a speaker. To capture this more naturalistic distribution of roles within an individual, we introduce the notion of a *symmetrized* game. A symmetrized game based on any asymmetric two-player game is one in which each player is assumed to be in one of the two roles of the asymmetric game with equal probability. Thus, in the symmetrized version of the imperfective game, each player will be in the role of the speaker half the time and in the role of the hearer for the other half. A strategy in such a symmetrized game is a pair of some speaker and some hearer strategy $\langle S, H \rangle$. The expected payoff or utility of an individual's strategy $\langle S, H \rangle$ relative to some individual strategy $\langle S', H' \rangle$ is calculated on the basis of the average utility for her speaker strategy and her hearer strategy.

(33) *Expected Utility (EU)*

$$\text{EU}(\langle S, H \rangle, \langle S', H' \rangle) = \frac{1}{2} \times (U_s(S, H') + U_h(S', H))$$

In a symmetrized game, all members of the population make use of strategies from the same set of paired speaker–hearer strategies; in the imperfective game, these will be pairs drawn from the speaker and hearer strategies in Table 1 and Table 2. There are a total of twelve such individual strategies. Of these, the strategies of interest to us are the ones that correspond to the states in (24), since they are the ones that are exhibited by populations at distinct diachronic stages. These are: $\langle S_{cd}, H_{cd} \rangle$, $\langle S_{pcd}, H_{pcd} \rangle$, $\langle S_{em}, H_{em} \rangle$, and $\langle S_{cd'}, H_{cd} \rangle$.[22] $\langle S_{cd}, H_{cd} \rangle$ is prevalent in the zero-PROG state. $\langle S_{pcd}, H_{pcd} \rangle$ is prevalent in the emergent-PROG state. $\langle S_{em}, H_{em} \rangle$ is the prevalent strategy in the categorical-PROG state. The generalized-PROG state reflects the prevalence of $\langle S_{cd'}, H_{cd} \rangle$, which is structurally identical to $\langle S_{cd}, H_{cd} \rangle$ but differs in both choice of form and average payoffs when playing against $\langle S_{pcd}, H_{pcd} \rangle$ and $\langle S_{em}, H_{em} \rangle$. The expected payoffs for each interaction between these four paired strategies are given in Table 4.

The next section presents, after some background, the evolutionary dynamics that accounts for how linguistic systems (i.e., the populations that embody them) move from the prevalence of one strategy to the other in this cyclic trajectory.

---

22 These are also exactly the strategies that come out as (weakly) evolutionarily stable strategies or ESS (Maynard Smith & Price 1973) when we consider the interactions between all twelve paired strategies. Demonstrating this is not needed for the purposes of this paper.

Strategies	$\langle S_{cd}, H_{cd}\rangle$	$\langle S_{pcd}, H_{pcd}\rangle$	$\langle S_{em}, H_{em}\rangle$	$\langle S_{cd'}, H_{cd}\rangle$
$\langle S_{cd}, H_{cd}\rangle$	0.9	0.9	0.7	0.9
$\langle S_{pcd}, H_{pcd}\rangle$	$0.9 - \frac{1}{2}k$	$0.95 - \frac{1}{2}k$	$0.75 - \frac{1}{2}k$	$0.7 - \frac{1}{2}k$
$\langle S_{em}, H_{em}\rangle$	$0.7 - \frac{1}{2}k$	$0.75 - \frac{1}{2}k$	$1 - \frac{1}{2}k$	$0.7 - \frac{1}{2}k$
$\langle S_{cd'}, H_{cd}\rangle$	0.9	0.7	0.7	0.9

Table 4   Expected payoffs for paired strategies based on (33)

5.2   The evolutionary dynamics

Evolutionary game dynamics have been used to describe and understand the behavior of large populations over time as an evolving game, and in particular, changes in the frequencies of different strategies in a population over time (Taylor & Jonker 1978, Hofbauer & Sigmund 1998). Each player plays a particular strategy and is paired at random with other players in the population. The payoff obtained from each encounter is accumulated and yields the fitness of a strategy. The average payoff of a given strategy is determined by averaging its payoffs across all encounters with all other strategies, weighted by their proportion in the population. This in turn determines the rate at which players of that strategy may replicate in the population, which may change the population composition over time. If a certain strategy yields an average payoff that is higher than the population average, we say that the fitness of this strategy is higher than the fitness of the population. This strategy will replicate at a higher rate than the population average. Thus strategies with above-average payoff (determined by the population composition) will increase their proportion in a population, while strategies with below-average payoff will decline.

There are two general interpretations of evolutionary game dynamics, and in particular, the notion of fitness. In the biological setting, particular strategies (which are heritable genetic traits) are encoded by genomes of individuals. Average payoffs, equated with fitness, give the expected number of offspring of an individual programmed for a particle strategy. Successful strategy types spread in the population due to higher reproductive rate. For the linguistic application that we are interested in, the relevant interpretation of game dynamics is the cultural one. On this interpretation, grammars are behavioral strategies that may be replicated by other individuals through imitation and learning rather than through inheritance. That is, individuals

may give up their own strategy and adopt another strategy based on its communicative success. The probability for a given strategy to be adopted is determined by the gain in payoffs caused by the strategy change. Thus, strategies that yield a higher payoff than the population average (i.e. have higher fitness than the population average) increase their share in the population while less successful strategies diminish.

In explaining the cyclic path, in addition to the average payoffs associated with particular strategies, we must consider the relative learnability of individual strategies from the structure of the input that is available to the learner. Recruitment, categoricalization, and generalization crucially presuppose mutations from one strategy to another that depend on how and under what conditions the input is (mis)-interpreted during the acquisition process. The replicator–mutator dynamic is especially well-suited for modeling change involving interactions between these two competing forces. It has been applied to the problem of the evolution of grammar (Nowak, Komarova & Niyogi 2001) and can be naturally extended to the problem of how new grammatical norms may arise and propagate in a speech community. There are two parts to the replicator–mutator dynamic. The replication rate of a strategy is the rate at which it might be adopted by players of other strategies. Thus, the "replicator" part of the dynamic models positive or negative changes in the frequency of a strategy over time based on its communicative success in a given population composition. The mutation rates associated with a strategy, on the other hand, model the barriers to the learnability of a strategy: cognitive or acquisition-related biases that might prevent its faithful transmission from parents to offspring. This is the "mutator" part of the dynamic.

5.3 The replicator–mutator dynamics

In order to understand the relevant transitions, we begin by studying the working parts of the replicator–mutator equation and observing how interactions work between any two strategies in the abstract. This will be made more concrete by studying the interaction between $\langle S_{cd}, H_{cd} \rangle$ and $\langle S_{pcd}, H_{pcd} \rangle$ as a 2×2 game in Section 5.3.1. Then in Section 5.3.2, we will examine how all four strategies evolve over time.

Consider two strategy types A and B and suppose that the population is predominantly of type A with a very small proportion of type B. We want to know the changes in the state of the population over time caused by the

differential fitnesses of the two strategies. Such 2×2 games are described by a payoff matrix of the following form:

	A	B
A	a	b
B	c	d

This matrix specifies the interactions between strategy players of any type A and B. If A interacts with A, the payoff is a and if A interacts with B, the payoff is b. B gets payoff c in interactions with A and d in interactions with B.

We will assume that payoff is equated with fitness, which means that the average payoff associated with the strategy employed by an individual equals its expected number of offspring or, on the cultural interpretation, the probability that the individual's strategy is adopted by another player by imitation. In a 2×2 game, only two strategies are present in a population at any given time. Their frequencies can be denoted by x_A and x_B respectively. The average payoff (= fitness) for A and B at a given population composition can be calculated as follows.

(34) a. $f_A = ax_A + bx_B$
      b. $f_B = cx_A + dx_B$

Frequency-dependent selection means that the rate of replication of the strategies A and B will be determined by their fitness (average payoff) relative to the fitness of the population. The selection dynamics, i.e., the change in the frequency for the A and B populations over time, can be written as in (35).

(35) a. $x'_A = x_A \dfrac{f_A}{\phi}$

      b. $x'_B = x_B \dfrac{f_B}{\phi}$

This is the discrete-time version of the replicator equation.[23] (35) says that the change in the frequency of a strategy A after one time step (x'_A) is calculated

---

[23] The continuous-time version assumes change over infinitesimal units of time, and consequently, mixing between generations of populations (parents and language-learning offspring, in this case). The discrete-time version of the equation assumes that selection and mutation occur at discrete time-steps and measures changes in frequencies of strategies at each time-step, and hence is more realistic for the linguistic scenario.

by multiplying the frequency of the A population before that time step (x_A) with the result of dividing the fitness of A before the time step (f_A) by the average fitness of the population before the time step (ϕ); and the same is true for B. The average fitness of the population for any time step is given by $\phi = x_A f_A + x_B f_B$.

In addition to the rate of replication, the frequency of a given strategy in a population also depends on the probability that the offspring of players using one strategy mutate to a different strategy. The learning process can be subject to mistakes, in which a child learning from a parent using strategy A will acquire strategy B instead. This mutation probability is represented as a row-stochastic matrix Q, which gives the transition probabilities for mutation to happen from one strategy type to another.[24] For any strategies, $i, j \ldots n$, let the probability that j mutates into i be denoted by Q_{ji}. Given these assumptions, the population dynamics are given by the discrete-time version of the "replicator–mutator" equation:

(36) $\quad x_i' = \sum_{j=1}^{n} Q_{ji} \dfrac{x_j f_j}{\phi}$

x_i' gives the frequency of strategy i after a time-step based on the average payoff for i and the average population fitness ϕ before that time-step. The mutation probabilities Q_{ji} remain unchanged across time-steps.[25] The simulation of evolution over time in the remainder of this section will be based on (36).

5.3.1 Recruitment

Recruitment amounts to the adoption of the partially context dependent $\langle S_{pcd}, H_{pcd} \rangle$ strategy (with a particular conventionalized form) by a large pro-

---

24 So, if we have two strategies, A and B, the transition probabilities can be represented as the matrix below. Here, q_a is the probability that an A parent has an A offspring, $1 - q_a$ is the probability that an A parent has a B offspring, $1 - q_b$ is the probability that a B parent has an A offspring, and q_b is the probability that a B parent has a B offspring. Saying that Q is a row-stochastic matrix means that all the rows add to 1.

$Q =$		A	B
	A	q_a	$1 - q_a$
	B	$1 - q_b$	q_b

25 This version of the equation is based on Page & Nowak (2002) and lecture notes kindly provided by Michael Franke.

portion of the population. Suppose that there is a population predominantly composed of $\langle S_{cd}, H_{cd} \rangle$ players with a small proportion of $\langle S_{pcd}, H_{pcd} \rangle$ players. Let k be 0.01. Then, how will the differential fitness of the two strategies affect the population composition over time? We are considering here the situation in which learning is perfect — that is, the mutation matrix Q has the following form:

(37) $Q = $

	$\langle S_{cd}, H_{cd} \rangle$	$\langle S_{pcd}, H_{pcd} \rangle$
$\langle S_{cd}, H_{cd} \rangle$	1	0
$\langle S_{pcd}, H_{pcd} \rangle$	0	1

Note that the replicator–mutator equation contains the replicator equation as a limiting case with perfect learning (Nowak 2006, pp. 272–273). The statement in (39) below holds only for replicator dynamics and will not apply to replicator–mutator dynamics if imperfect learning is assumed. Consider now the relative payoffs of $\langle S_{cd}, H_{cd} \rangle$ and $\langle S_{pcd}, H_{pcd} \rangle$ from Table 4. The two strategies are in a bistable relation: each is the best response to itself.

(38)

	$\langle S_{cd}, H_{cd} \rangle$	$\langle S_{pcd}, H_{pcd} \rangle$
$\langle S_{cd}, H_{cd} \rangle$	0.9	0.9
	$\cup$	$\cap$
$\langle S_{pcd}, H_{pcd} \rangle$	$0.9 - 0.005$	$0.95 - 0.005$

For any two strategies, A and B, if they are bistable, the outcome of the selection dynamics (only when there is no mutation) depends on the values for x_A (the proportion of the population using A) and x_B (the proportion of the population using B) at a given point in time. There is an unstable equilibrium in the interior of the interval $[0, 1]$ given by:

(39) $\quad x_A^* = \dfrac{d - b}{a - b - c + d}$ $\hspace{3cm}$ (Nowak 2006, p. 51)

If the proportion of strategy A players $x_A(0)$ is less than this value, then the system will converge to an all-B population. If the proportion of strategy A players $x_A(0)$, is greater than this value, then the system will converge to an all-A population. In the case of $\langle S_{cd}, H_{cd} \rangle$ and $\langle S_{pcd}, H_{pcd} \rangle$, plugging in the values for the payoffs, we have an unstable equilibrium at $x_A(0) = 0.9$, calculating as below.

(40) $\quad \dfrac{0.945 - 0.9}{0.9 - 0.9 - 0.895 + 0.945} = 0.9.$

This means that if the $\langle S_{pcd}, H_{pcd} \rangle$ share in the population is greater than 0.1, then the evolution will be to an $\langle S_{pcd}, H_{pcd} \rangle$-dominant population. Figure 1 gives the change over time when initial values for $x_{\langle S_{cd}, H_{cd} \rangle} = 0.89$ and $x_{\langle S_{pcd}, H_{pcd} \rangle} = 0.11$.
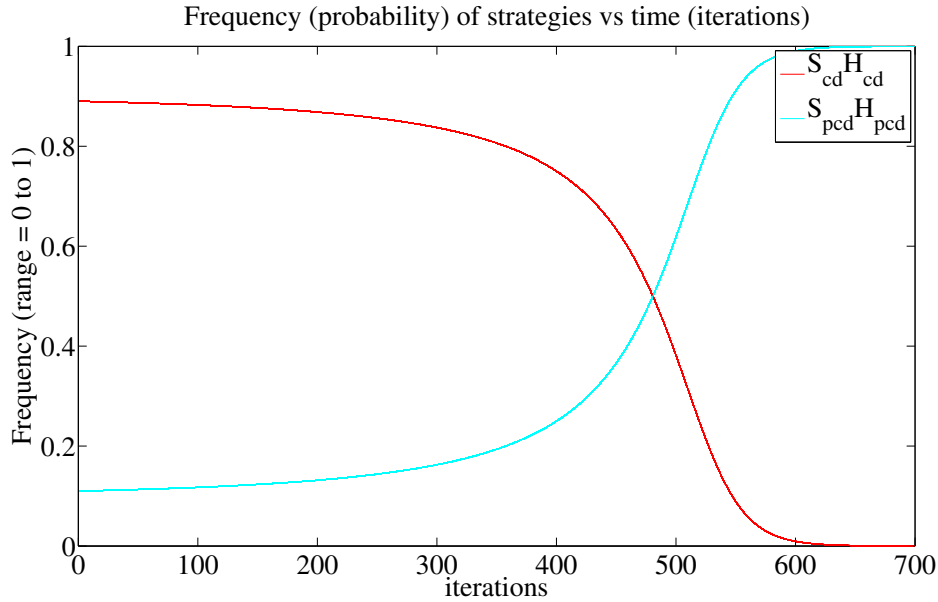

Figure 1 $\quad \langle S_{cd}, H_{cd} \rangle$ to $\langle S_{pcd}, H_{pcd} \rangle$: High initial value rate for $x_{\langle S_{pcd}, H_{pcd} \rangle}$

We consider now the possibility that $\langle S_{pcd}, H_{pcd} \rangle$ is present in the population at a frequency < 0.1 but introduce mutation into the dynamics. On this scenario, we check to see if the transition to $\langle S_{pcd}, H_{pcd} \rangle$-dominance occurs if $\langle S_{pcd}, H_{pcd} \rangle$ is a target of learning for at least some offspring of the majority $\langle S_{cd}, H_{cd} \rangle$ speakers. Linguistically, this would mean that some acquirers interpret local disambiguation efforts undertaken by $\langle S_{cd}, H_{cd} \rangle$ speakers as conventionalized, thus innovating $\langle S_{pcd}, H_{pcd} \rangle$ grammars. Conversely, it is possible that the offspring of $\langle S_{pcd}, H_{pcd} \rangle$ speakers introduced by such mutation fail to interpret conventionalized strategies for disambiguation as such and revert to $\langle S_{cd}, H_{cd} \rangle$ grammars. This can be concretely represented by the stochastic matrix Q' below.

(41) $Q' =$

	$\langle S_{cd}, H_{cd} \rangle$	$\langle S_{pcd}, H_{pcd} \rangle$
$\langle S_{cd}, H_{cd} \rangle$	0.99	0.01
$\langle S_{pcd}, H_{pcd} \rangle$	0.01	0.99

Thus, we assume (only for expository purposes) that an $\langle S_{cd}, H_{cd} \rangle$ parent is as likely to generate an $\langle S_{pcd}, H_{pcd} \rangle$ offspring as an $\langle S_{pcd}, H_{pcd} \rangle$ parent is to generate an $\langle S_{cd}, H_{cd} \rangle$ offspring. With initial values for $x_{\langle S_{cd}, H_{cd} \rangle}(0) = 1$, Figure 2 gives the change in the proportion of $\langle S_{cd}, H_{cd} \rangle$ and $\langle S_{pcd}, H_{pcd} \rangle$ over time, where $\langle S_{pcd}, H_{pcd} \rangle$ comes to dominate the population, stabilizing at ~ 0.8, reducing the $\langle S_{cd}, H_{cd} \rangle$ population to a ~ 0.2 share. Thus, assuming small, uniform mutation rates for the two strategies is already enough to drive the population towards $\langle S_{pcd}, H_{pcd} \rangle$.
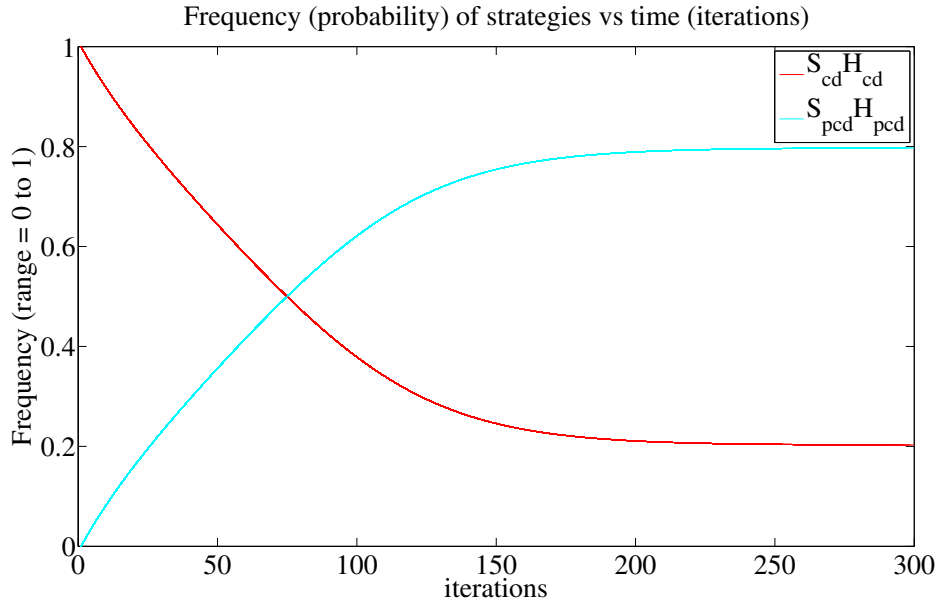

Figure 2 $\langle S_{cd}, H_{cd} \rangle$ to $\langle S_{pcd}, H_{pcd} \rangle$ with equal mutation rates

5.3.2 Categoricalization

Categoricalization amounts to the adoption of the context-independent explicit marking strategy $\langle S_{em}, H_{em} \rangle$ by a large proportion of the population. In order to model this state in the grammaticalization path we need to define a mutation matrix that gives the transition probabilities between all the

relevant strategies. In fact, given that there is non-zero probability that an $\langle S_{pcd}, H_{pcd} \rangle$ parent may generate an $\langle S_{em}, H_{em} \rangle$ offspring, $\langle S_{em}, H_{em} \rangle$ players will be present in any population mix that contains $\langle S_{pcd}, H_{pcd} \rangle$ speakers. Moreover, the dynamics requires transitions from $\langle S_{em}, H_{em} \rangle$ to $\langle S_{cd'}, H_{cd} \rangle$, which means that any population mix that contains $\langle S_{em}, H_{em} \rangle$ players will also contain $\langle S_{cd'}, H_{cd} \rangle$ players in some proportion. Thus, the full game needs a modified matrix based on different assumptions that have to do with the acquisitional properties of distinct strategies. These linguistic considerations are concretized in Q'', which we will take to be the full (and revised) row stochastic matrix.

(42) $Q'' =$

	$\langle S_{cd}, H_{cd} \rangle$	$\langle S_{pcd}, H_{pcd} \rangle$	$\langle S_{em}, H_{em} \rangle$	$\langle S_{cd'}, H_{cd} \rangle$
$\langle S_{cd}, H_{cd} \rangle$	0.94	0.06	0	0
$\langle S_{pcd}, H_{pcd} \rangle$	0.02	0.91	0.07	0
$\langle S_{em}, H_{em} \rangle$	0	0	0.97	0.03
$\langle S_{cd'}, H_{cd} \rangle$	0	0	0	1

The stochastic matrix in general is intended to reflect hypotheses about (mis)learning and optimization of successful communication. The linguistic reasoning behind Q'' is as follows:

- While $\langle S_{cd}, H_{cd} \rangle$ is a formally simple grammar for a child to acquire, it leads to miscommunication in some proportion (0.1) of interactions. Speakers may undertake local disambiguating efforts to avoid miscommunication. Offspring of $\langle S_{cd}, H_{cd} \rangle$ parents may mis-infer $\langle S_{pcd}, H_{pcd} \rangle$ from the structure of such input, which contains explicit disambiguating material. There is hardly any evidence to infer $\langle S_{em}, H_{em} \rangle$ grammars in the same input.

- $\langle S_{pcd}, H_{pcd} \rangle$ is a difficult grammar to learn because it requires the speaker to be highly attuned to the context — the speaker's choice of form depends on the speaker's assessment of whether the context is **phen**-oriented or **struc**-oriented. While child learners may acquire this system correctly, they are less likely to do so than a form-invariant ($\langle S_{cd}, H_{cd} \rangle$) or context-invariant ($\langle S_{em}, H_{em} \rangle$) system. This is reflected in higher mutation rates leading out of $\langle S_{pcd}, H_{pcd} \rangle$. The mutating offspring of $\langle S_{pcd}, H_{pcd} \rangle$ parents may go either way; they may misinterpret the input as generated by an $\langle S_{cd}, H_{cd} \rangle$ grammar (given the low frequency of *prog*) or as generated by an $\langle S_{em}, H_{em} \rangle$ grammar. But they

are much more likely to do the latter than the former because the input provides evidence for a grammaticalized *prog* form.

- $\langle S_{em}, H_{em} \rangle$ is easier to learn than $\langle S_{pcd}, H_{pcd} \rangle$, requiring no context-sensitivity of either the speaker or the hearer. However, a few learners may interpret the input from $\langle S_{em}, H_{em} \rangle$, which contains *prog* forms with higher frequency, as evidence for an $\langle S_{cd'}, H_{cd} \rangle$ grammar — a formally economic grammar in which the *prog* form is used for both phenomenal and structural inquiries with contextual disambiguation.

- Finally, there is no mutation out of $\langle S_{cd'}, H_{cd} \rangle$ into the other three strategies since an $\langle S_{cd'}, H_{cd} \rangle$ grammar does not provide any evidence for contrast with the older *impf* form. This will be further addressed in Section 5.3.3.

Q'' represents a preliminary proposal for quantifying the effect of the functional pressures of economy and learnability in the acquisition process for the imperfective domain. Change over time should reflect the interaction between asymmetric mutation rates and asymmetric payoffs. Figure 3 gives the dynamics for the four strategies over time with the stochastic matrix Q'' and initial conditions set to $\langle S_{cd}, H_{cd} \rangle = 1$. $\langle S_{pcd}, H_{pcd} \rangle$, $\langle S_{em}, H_{em} \rangle$, and $\langle S_{cd'}, H_{cd} \rangle$ are introduced by mutation. We see that the $\langle S_{pcd}, H_{pcd} \rangle$ population advances over the $\langle S_{cd}, H_{cd} \rangle$ population but is gradually taken over by a growing $\langle S_{em}, H_{em} \rangle$ population, which stabilizes at ~ 0.88, with elimination of $\langle S_{pcd}, H_{pcd} \rangle$ and $\langle S_{cd}, H_{cd} \rangle$ and the presence of $\langle S_{cd'}, H_{cd} \rangle$ at a low frequency. This amounts to categoricalization.

The structure of the mutation matrix Q'', ensures that there will never be a population consisting entirely of players of one strategy (i.e., no universal dominance). Any state in which a particular strategy appears dominant, will simply be a state in which other strategies are at a "low-enough" frequency. Before moving to accounting for the generalization of *prog*, we will interpret the notions of the zero-PROG, emergent-PROG, and categorical-PROG states in terms of strategy proportions.

Realistically speaking, any $\langle S_{cd}, H_{cd} \rangle$ system is always supplemented by some degree of $\langle S_{pcd}, H_{pcd} \rangle$-like usage. These are the local efforts at disambiguation effected by optional adverbials or periphrastic constructions. In this case, one would speak of a population using a mixed strategy rather than there being a mixed population, but this is equivalent to there being some $\langle S_{pcd}, H_{pcd} \rangle$ presence in any $\langle S_{cd}, H_{cd} \rangle$ population. This means that what has been called a zero-PROG state is really a state in which the proportion
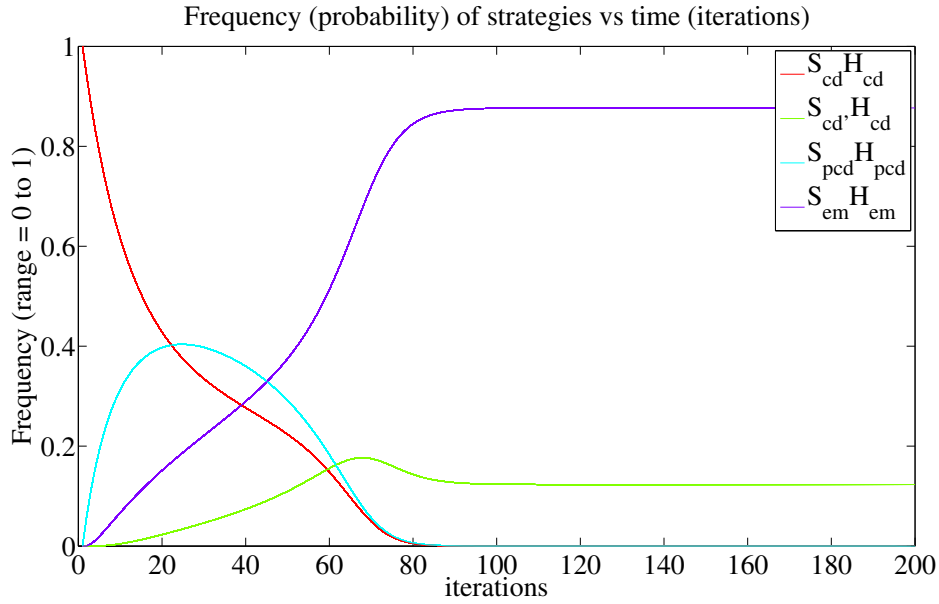
Frequency (probability) of strategies vs time (iterations)

Figure 3 Dynamic behavior of $\langle S_{cd}, H_{cd}\rangle$, $\langle S_{pcd}, H_{pcd}\rangle$, $\langle S_{em}, H_{em}\rangle$, and $\langle S_{cd'}, H_{cd}\rangle$ assuming Q''

of $\langle S_{pcd}, H_{pcd}\rangle$ remains "low enough" or below some threshold ϵ. Once its proportion exceeds this threshold and involves the use of a privileged disambiguating form, the *prog* form it employs might be said to have become grammaticalized. This means that in order to determine whether a state should be called a zero-PROG or emergent-PROG state, we need to look at the proportion of $\langle S_{cd}, H_{cd}\rangle + \langle S_{pcd}, H_{pcd}\rangle$ speakers in any given state. We will (somewhat crudely) interpret the three apparent states in the following way:

(43) a. If $x_{\langle S_{cd}, H_{cd}\rangle} + x_{\langle S_{pcd}, H_{pcd}\rangle} > x_{\langle S_{em}, H_{em}\rangle}$ and $x_{\langle S_{pcd}, H_{pcd}\rangle} < \epsilon$ zero-PROG

    b. If $x_{\langle S_{cd}, H_{cd}\rangle} + x_{\langle S_{pcd}, H_{pcd}\rangle} > x_{\langle S_{em}, H_{em}\rangle}$ and $x_{\langle S_{pcd}, H_{pcd}\rangle} > \epsilon$ emergent-PROG

    c. If $x_{\langle S_{em}, H_{em}\rangle} > x_{\langle S_{cd}, H_{cd}\rangle} + x_{\langle S_{pcd}, H_{pcd}\rangle} + x_{\langle S_{cd'}, H_{cd}\rangle}$ categorical-PROG

For the purposes of this model, I will take ϵ to be 0.3. That is, if the proportion of $\langle S_{pcd}, H_{pcd}\rangle$ in the population exceeds 0.3, then the population will be taken to have entered the emergent-PROG state from the zero-PROG state.

5.3.3 Generalization

Categoricalization results in a state in which $\langle S_{em}, H_{em} \rangle$ is the prevalent strategy. Mutations out of $\langle S_{em}, H_{em} \rangle$ into $\langle S_{cd'}, H_{cd} \rangle$ do not result in a cycling pattern from the categorical-PROG state to the generalized-PROG state (which is structurally identical to the zero-PROG state). How then can we account for the fact that the categorical-PROG state tends to give way to the generalized-PROG state, in which the *prog* form is reanalyzed as realizing IMPF (the progressive≫imperfective shift)?

The game dynamics so far has assumed that mutations are constant (given by the mutation matrix Q'') and disconnected from the frequencies of particular strategies in a population. However, it is possible that at least in some cases, the increase in the frequency of a strategy may be tied to an increase in the probability for mis-learning out of that strategy. In the case of $\langle S_{em}, H_{em} \rangle$, as more and more players adopt this strategy, the total frequency of *prog* forms in the child input would increase. This would lead to a slight decrease in the evidence for the $\langle S_{em}, H_{em} \rangle$ grammar and slight increase in the evidence for an $\langle S_{cd'}, H_{cd} \rangle$ grammar — with *prog* as the new exponent of IMPF. Further, if language learners are biased towards single-form contextually reliant strategies (as was suggested in Section 4.2.3), this should lead to increased mutations out of $\langle S_{em}, H_{em} \rangle$ into $\langle S_{cd'}, H_{cd} \rangle$.

I will adopt this line of reasoning and assume that mutations out of $\langle S_{em}, H_{em} \rangle$ increase slightly when it reaches 0.5 frequency and slightly further when it reaches 0.65 frequency.[26] Concretely, when $x_{\langle S_{em}, H_{em} \rangle}$ exceeds 0.5, the mutation rate from $\langle S_{em}, H_{em} \rangle$ into $\langle S_{cd'}, H_{cd} \rangle$ increases to 0.04 from 0.03. When $x_{\langle S_{em}, H_{em} \rangle}$ reaches 0.65, the mutation rate from $\langle S_{em}, H_{em} \rangle$ into $\langle S_{cd'}, H_{cd} \rangle$ increases to 0.05. These slight increases in mutations lead to the gradual takeover of $\langle S_{em}, H_{em} \rangle$ by $\langle S_{cd'}, H_{cd} \rangle$. This dynamic behavior, which amounts to generalization, is shown in Figure 4.

We can imagine that once $\langle S_{em}, H_{em} \rangle$ has declined and $\langle S_{cd'}, H_{cd} \rangle$ is the prevalent strategy, it is fully equivalent to $\langle S_{cd}, H_{cd} \rangle$, since there is no alternative grammaticalized form that speakers can contrast *prog* with. A new form *prog'* may now be innovated to realize PROG, setting into motion yet another cycle of the progressive≫imperfective path.

---

26 One may also model mutations as an increasing function of the frequency of a strategy, but I will leave this aspect for later exploration.
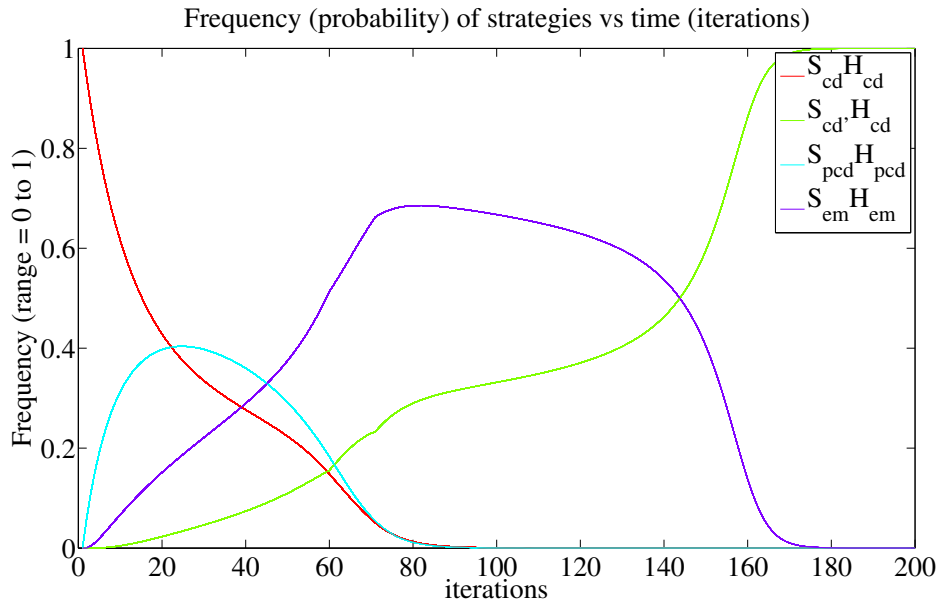
Figure 4   Takeover of $\langle S_{em}, H_{em} \rangle$ by $\langle S_{cd'}, H_{cd} \rangle$ with increased mutations

5.4   Summary

Typologically, we can divide grammars into those in which the nature of the inquiry — phenomenal vs. structural — is determined (a) contextually, and those in which it is marked linguistically, either (b) optionally or (c) categorically. The former type of grammar contains a single underspecified form that realizes the IMPF operator (e.g., Russian, Arabic, Sanskrit, Middle English) while the latter two types of grammar distinguish between exponents for PROG and IMPF by using *prog* either optionally (e.g., Romance) or categorically (e.g., Modern English, Hindi, and Turkish). Diachronically, we observe that languages move from context-dependent grammars (zero-PROG) to optional marking grammars (emergent-PROG) to categorical marking grammars (categorical-PROG) back to context-dependent (generalized-PROG) grammars.

The game-theoretic model built in this section shows that the typological patterns as well as the diachronic behavior can be shown to correspond to distinct states of a single dynamical system.[27] We examine the behavior

---

27 One other factor with respect to which grammars might vary typologically is the k factor that determines the cost of multi-form strategies. Higher values for k give rise to different equilibria and may account for the fact that some languages never seem to participate in

of four strategies in the evolutionary dynamics, where several of them are present in some proportion within a population at any given time. The payoffs of these strategies and the mutation probabilities leading from and into these strategies (rooted in acquisitional asymmetries and strategy frequencies) together lead to the cycling behavior observed in the states of the system.

6   Concluding remarks

The broad goal of this paper was to begin to understand systematic diachronic patterns in the links between the form and the meaning of functional expressions. Grammaticalization paths, as these patterns are called, are complex clusters of phenomena involving recruitment of lexical items for expressing functional meanings, the categoricalization of their functions relative to an existing grammatical system, and changes in such functions (e.g., semantic bleaching or generalization) over time. Recruitment, categoricalization, and generalization are not explanations but rather observations to be explained by theories of linguistic meaning and linguistic usage. The examination of these phenomena in the domain of imperfectivity reveals that the grammaticalization path reported for this domain is an emergent effect of the interaction between the structural and the dynamic properties of language. The relevant structural properties come from the universally shared semantic core of functional expressions, and specifically, the privative nature of the contrast between the progressive and the imperfective aspects. The privative opposition between the progressive and the imperfective mirrors the conceptual contrast between the phenomenal and the structural (or non-phenomenal). This contrast may be accessed via contextual knowledge and optional disambiguators, or via grammaticalized progressive markers that may be used optionally or categorically.

There are four game-theoretic strategies that correspond to these three ways of communicating the relevant meanings. A population contains one or more of these strategies in some proportion at any given time and mutations are continuously occurring between the four strategies. Recruitment occurs when a population in which contextual recovery of meaning (supplemented with low proportions of linguistic disambiguators) is the prevailing strategy, starts using a conventionalized form for expressing progressive meaning in greater proportions in those contexts where contextual recovery of intended

the cyclic behavior observed here, but rather maintain context-dependent systems over long periods of time. This remains an issue for further exploration.

information is less likely. Such a transition, given asymmetric payoffs and mutation probabilities, is followed by categoricalization. Here the strategy which relies on obligatory explicit marking of the two meanings, increases in proportion. This increase in the proportion of the explicit marking strategy leads to an increased frequency of *prog* marking in the child input. This, in turn, leads to increased mutations out of that strategy into a new contextually reliant strategy. This paves the path for generalization in which the specific *prog* form is generalized as the exponent of the underspecified IMPF operator. This brings the system back to its original zero-PROG stage where the cycle may begin anew.

The model proposed here can be naturally extended to any functional domain characterized by a privative semantic contrast. The immediate connection is to Jesperson's cycle in the domain of negation where material recruited for marking emphatic negation weakens to mark plain negation and new material is introduced to express emphatic negation. Kiparsky & Condoravdi (2006) analyze this process as being rooted in the privative contrast between emphatic and plain negation. The dynamic process is argued to be a semantically driven chain shift where the pragmatically motivated overuse of emphatic negation leads to increasing frequency, which in turn, leads to its weakening to plain negation. This cyclic process can be modeled as a change between context dependent and explicit marking strategies in the domain of negation, where one of the factors that would push frequency of emphatic negation markers upwards would be the inflationary use of the form chosen to express emphatic negation (Dahl 2001). Generalizing further, we might make a strong hypothesis:

(44)   a. A semantic grammaticalization path in the functional domain must be structurally underpinned by some privative contrast between a specific and a general meaning.

   b. Changes in functional domains characterized by a privative semantic contrast are cyclic in nature because increasing frequencies of (some) strategies in the population lead to increased probability of mis-learning out of that strategy.

   c. The actual occurrence of such paths would depend on contingent factors such as the cost of multi-form strategies, threshold values for grammaticalization of novel material, and possible variability in the effect of strategy frequency on mutation rates.

The perfect≫perfective/past path and the location≫possession path mentioned in (2) are instances in which the content of the privative contrast (the structural component) and the cyclicity of the observed changes appear to be quite straightforward. Further research can determine whether changes in other semantic domains can also be subsumed under this general framework for modeling semantic change. For now, we have offered a way for addressing the *constraints*, *actuation*, and *transition* problems of Weinreich, Labov & Herzog 1968 in working towards a theory of semantic change.

References

Aristar, Anthony. 1996. The relationship between dative and locative: Kuryłowicz's argument from a typological perspective. *Diachronica* 13(2). 207–224. http://dx.doi.org/10.1075/dia.13.2.02ari.

Benz, Anton, Gerhard Jäger & Robert van Rooij (eds.). 2006. *Game theory and pragmatics*. Houndsmill, NJ: Palgrave Macmillan.

Benz, Anton, Gerhard Jäger & Robert Van Rooij. 2006. An introduction to game theory for linguists. In Anton Benz, Gerhard Jäger & Robert van Rooij (eds.), *Game theory and pragmatics*. Houndsmill, NJ: Palgrave Macmillan.

Bittner, Maria. 2005. Future discourse in a tenseless language. *Journal of Semantics* 22. 339–387. http://dx.doi.org/10.1093/jos/ffh029.

Bohnemeyer, Jürgen. 2002. *The grammar of time reference in Yukatek Maya*. Munich: Lincom Europa.

Bohnemeyer, Jürgen. 2009. Temporal anaphora in a tenseless language. In Wolfgang Klein & Ping Li (eds.), *Expression of time in language*, 83–128. Berlin: Mouton de Gruyter.

Bonomi, Andrea. 1997. Aspect, quantification and when-clauses in Italian. *Linguistics and Philosophy* 20(5). 469–514. http://dx.doi.org/10.1023/A:1005388230492.

Bybee, Joan. 1994. The grammaticization of zero: Asymmetries in tense and aspect systems. In William Pagliuca (ed.), *Perspectives on grammaticalization*, 235–254. Amsterdam: John Benjamins.

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago, IL: The University of Chicago.

Champollion, Lucas. 2010. *Parts of a whole: Distributivity as a bridge between aspect and measurement*. University of Pennsylvania PhD thesis.

Cipria, Alicia & Craige Roberts. 2000. Spanish Imperfecto and Pretérito: Truth conditions and aktionsart effects in a situation semantics. *Natural Language Semantics* 8(4). 297–347. http://dx.doi.org/10.1023/A:1011202000582.

Clark, Eve. 1978. Locationals: A study of "existential," "locative," and "possessive" sentences. In Joseph Greenberg (ed.), *Universals of human language, vol. 4: Syntax*, 85–126. Stanford, CA: Stanford University Press.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems.* Cambridge, UK: Cambridge University Press.

Dahl, Östen. 1985. *Tense and aspect systems.* Oxford, UK: Basil Blackwell.

Dahl, Östen (ed.). 2000. *Tense and aspect in the languages of Europe.* Berlin: Mouton de Gruyter.

Dahl, Östen. 2001. Inflationary effects in language and elsewhere. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, vol. 2, 471–80. Amsterdam: John Benjamins.

Delfitto, Denis & Pier Marco Bertinetto. 1995. A case study in the interaction of aspect and actionality: The Imperfect in Italian. In P.M. Bertinetto, V. Bianchi, Higginbotham J & M. Squartini (eds.), *Temporal reference, aspect and actionality volume 1: Semantics and syntactic perspectives*, 125–142. Torino: Rosenberg & Sellier.

Deo, Ashwini. 2009. Unifying the imperfective and the progressive: Partitions as quantificational domains. *Linguistics & Philosophy* 32(5). 475–521. http://dx.doi.org/10.1007/s10988-010-9068-z.

Deo, Ashwini. 2014. Formal semantics/pragmatics and language change. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 393–409. Oxford, UK: Routledge.

Deo, Ashwini. 2015. Diachronic semantics. *Annual Review of Linguistics* 1. 179–197. http://dx.doi.org/10.1146/annurev-linguist-030514-125100.

Dowty, David. 1977. Toward a semantic analysis of verb aspect and the English "imperfective" progressive. *Linguistics and Philosophy* 1(1). 45–77. http://dx.doi.org/10.1007/BF00351936.

Dowty, David. 1979. *Word meaning and Montague grammar.* Dordrecht: Kluver Academic Publishers.

Eckardt, Regine. 2006. *Meaning change in grammaticalization: An enquiry into semantic reanalysis.* Oxford, UK: Oxford University Press.

Ferreira, Marcelo. 2005. *Event quantification and plurality.* Cambridge, MA: Massachussetts Institute of Technology PhD thesis.

Göksel, Asli & Celia Kerslake. 2005. *Turkish: A comprehensive grammar.* London, UK, & New York, NY: Routledge.

Goldsmith, John & Erich Woisetschlaeger. 1982. The logic of the English Progressive. *Linguistic Inquiry* 13(1). 79–89.

Heine, Bernd. 1997. *Possession: Cognitive sources, forces, and grammaticalization.* Cambridge, UK: Cambridge University Press.

Hofbauer, Josef & Karl Sigmund. 1998. *Evolutionary games and population dynamics.* Cambridge, UK: Cambridge University Press.

Hofbauer, Josef & Karl Sigmund. 2003. Evolutionary game dynamics. *Bulletin of the American Mathematical Society* 40(4). 479–519. http://dx.doi.org/10.1090/S0273-0979-03-00988-1.

Horn, Laurence. 1984. Towards a new taxonomy of pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (ed.), *Meaning, form, and use in context: Linguistic applications*, 11–42. Washington, DC: Georgetown University Press.

Horn, Laurence & Barbara Abbott. 2012. <the, a>: (In)definiteness and implicature. In William Kabasenche, Michael O'Rourke & Matthew Slater (eds.), *Reference and referring*, 325–356. Cambridge, MA: MIT Press.

Jäger, Gerhard. 2004. Evolutionary game theory for linguists: A primer. Ms. University of Tübingen. http://www.sfs.uni-tuebingen.de/~gjaeger/publications/egtPrimer.pdf.

Jäger, Gerhard. 2007. Evolutionary game theory and typology: A case study. *Language* 83(1). 74–109. http://www.jstor.org/stable/4490338.

Kellogg, S. H. 1893. *Grammar of the Hindi Language.* Second edition, reprinted 1990. Delhi: Motilal Banarsidass.

Kiparsky, Paul & Cleo Condoravdi. 2006. Tracking Jespersen's cycle. In M. Janse, B.D. Joseph & A. Ralli (eds.), *The 2nd international conference of Modern Greek dialects and linguistic theory*, 172–197. Mytilene: Doukas. http://web.stanford.edu/~kiparsky/Papers/lesvosnegation.pdf.

Kornfilt, Jaklin. 1997. *Turkish.* London, UK, & New York, NY: Routledge.

Kuryłowicz, Jerzy. 1964. *The inflectional categories of Indo-European.* Heidelberg: Carl Winter.

Landman, Fred. 1992. The progressive. *Natural Language Semantics* 1(1). 1–32. http://dx.doi.org/10.1007/BF02342615.

Leech, Geoffrey. 1970. *Towards a semantic description of English.* Bloomington, IN: Indiana University Press.

Lenci, Alessandro & Pier Marco Bertinetto. 2000. Aspect, adverbs, and events: Habituality vs. perfectivity. In James Higginbotham, Fabio Pianesi & Achilles

Varzi (eds.), *Speaking of events*, 245-287. New York, NY: Oxford University Press.

Lewis, David. 1969. *Convention*. Cambridge, MA: Harvard University press.

Li, Ping, Sarah Maher, Erica Newmark & Jennifer Hurley. 2001. The role of parental input in the acquisition of tense-aspect morphology. *Journal of Cognitive Science* 2. 119–143.

Marchese, Lynell. 1986. *Tense/aspect and the development of auxiliaries in Kru languages*. Summer Institute of Linguistics; University of Texas at Arlington.

Martinet, André. 1962. *A Functional View of Language*. Oxford, UK: Clarendon Press.

Maynard Smith, J. & G. R. Price. 1973. The logic of animal conflict. *Nature* 146(4). 15-18. http://dx.doi.org//10.1038/246015a0.

Nowak, Martin A. 2006. *Evolutionary dynamics: Exploring the equations of life*. Cambridge, MA: Harvard University Press.

Nowak, Martin A., Natalia L Komarova & Partha Niyogi. 2001. Evolution of universal grammar. *Science* 291(5501). 114-118. http://dx.doi.org/10.1126/science.291.5501.114.

Page, Karen & Martin A. Nowak. 2002. Unifying evolutionary dynamics. *Journal of Theoretical Biology* 219(1). 93–98. http://dx.doi.org/doi:10.1006/jtbi.2002.3112.

Paul, Hermann. 1888. *Principles of the history of language (Translated by H.A. Strong)*. London, UK: Swan Sonnenschein, Lowrey & Co.

Portner, Paul. 1998. The progressive in modal semantics. *Language* 74(4). 760-787. http://www.jstor.org/stable/417002.

Raz, Shiomo. 1983. *Tigre grammar and texts*. Malibu, CA: Undena Publications.

van Rooij, Robert. 2004a. Evolution of conventional meaning and conversational principles. *Synthese* 139(2). 331–366. http://dx.doi.org/10.1023/B:SYNT.0000024904.37199.6c.

van Rooij, Robert. 2004b. Signalling games select Horn strategies. *Linguistics and Philosophy* 27(4). 493–527. http://dx.doi.org/10.1023/B:LING.0000024403.88733.3f.

Rubinstein, Ariel. 2000. *Economics and language: Five essays*. New York, NY: Cambridge University Press.

Shirai, Yasuhiro. 1994. On the overgeneralization of progressive marking on stative verbs: Bioprogram or input? *First Language* 14(40). 67-82. http://dx.doi.org/10.1177/014272379401404005.

Stassen, Leon. 2009. *Predicative Possession*. Oxford, UK: Oxford University Press.

Tagliamonte, Sali & Helen Lawrence. 2000. I used to dance, but I don't dance now: The habitual past in English. *Journal of English Linguistics* 28(4). 324–353. http://dx.doi.org/10.1177/007542420002800402.

Taylor, Peter & Leo B. Jonker. 1978. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences* 40(1–2). 145–156. http://dx.doi.org/10.1016/0025-5564(78)90077-9.

Tonhauser, Judith. 2011. Temporal reference in Paraguayan Guaraní, a tenseless language. *Linguistics and Philosophy* 34(3). 1–47. http://dx.doi.org/10.1007/s10988-011-9097-2.

Traugott, Elizabeth C. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65(1). 31–55. http://www.jstor.org/stable/414841.

Traugott, Elizabeth C. & Richard B. Dasher. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Visser, Fredericus Theodorus. 1970. *An historical syntax of the English language*. Leiden: Brill Archive.

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkeil (eds.), *Directions for Historical Linguistics: A Symposium*, 95–188. Austin, TX: University of Texas Press.

Zipf, George K. 1949. *Human Behaviour and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press.

Ashwini Deo
Department of Linguistics
Yale University
Dow Hall 305
New Haven, CT 06520
ashwini.deo@yale.edu