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

Ashwini Deo

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 and Woisetschleger 1982). The cyclic 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. It is shown that all three parts of the trajectory can be viewed as transitions within the strategy space of a game model for communicating a given set of meanings with minimal ambiguity and cost.

1. Introduction

(1) a. Jane is sorting the mail.
    b. Jane sorts the mail.

I hear an utterance such as (1-a) and understand that the referent is engaged in a particular event that is in progress as I process the utterance. I hear (1-b) 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 (1-a) has little in common with the meaning conveyed by the simple present tense verb in (1-b).[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. Specifically, 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 emergence of progressive markers in a linguistic system and their gradual evolution into markers of imperfectivity more generally, i.e. the progressive $\gg$

[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.

imperfective diachronic shift, is only one type of crosslinguistically attested systematic semantic change. Typological and grammaticalization research on the meaning of semantic categories like tense/aspect, modality, and possession has uncovered many such systematic diachronic patterns in the linkings between form and meaning. These patterns take the form of unidirectional diachronic trajectories – recurring cross-linguistic regularities in the ways that grammatical morphemes undergo semantic change. A few such examples are given in (2).

(2) a. Progressive markers generalize to markers of imperfective aspect. (Bybee et al 1994; Comrie 1976).
b. Resultative markers generalize to markers of perfect aspect and past tense. (Bybee et al 1994; Dahl 1985, 2000).
c. Expressions encoding location evolve into expressions encoding alienable/inalienable possession. (Clark 1978; Aristar 1996; Heine 1997; Stassen 2009).
d. Expressions restricted to expressing deontic modality diachronically acquire epistemic uses, but not vice versa. (Traugott 1989; Traugott & Dasher 2002).

The broad goal of this paper is to attempt at answering 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 static, 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 Weinrich, Herzog, and Labov (1968) we can view this as the *constraints* problem for semantics – i.e. the problem of discovering the precise content of grammatically relevant elements of meaning and the constraints on their interaction. The dynamic component, on the other hand, draws from theories of language use and evolution in order to provide a plausible account for the morphosyntactic emergence of a new functional category in a language 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 Weinrich, Herzog, and Labov (1968). The account proposed in this paper for the progressive $\gg$ imperfective path illustrates the workings of these basic components in an eventual theory of semantic change.

Here then is the plan for this paper. In §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. §3 contains the structural part of the explanation, providing a semantic analysis that treats the contrast between the two categories as involving a privative opposition and derives the division of labor between them as a pragmatic outcome. §4 introduces the idea that the realization of both imperfective and progressive markers in a language allow for formal disambiguation between phenomenal and structural inquiries. It considers the dynamic aspects of the grammaticalization path – the recruitment of new progressive forms, their categoricalization, and their generalization to imperfective meaning – as being determined by considerations of economy, expressivity, and learnability. The ideas from §4 are formalized and further developed in §5. This section offers an evolutionary game-theoretic analysis of the progressive ≫ imperfective shift. The **Imperfective Game** described in this section shows how a set of concrete assumptions about imperfective/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 the conclusion (§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: (a) the progressive or *event-in-progress* reading; (b) the habitual or generic *characterizing* reading; and (c) 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.[2]

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

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

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

[2] These generalizations are based on information in Cardona (1965) and my own fieldwork in South Gujarat in 2004. Gujarati has distinct exponents of the imperfective aspect in the past and the present tenses and these combine periphrastically with past and present tense auxiliaries in past and present imperfective sentences.

as a subcategory of the more general imperfective, with a narrower, more specific meaning that can be subsumed under the broader, more general meaning of the imperfective.

(4) imperfective [progressive]

Two pieces of evidence support this analysis – blocking facts (§2.1) and grammaticalizing changes (§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 (5-a) is incompatible with the event-in-progress reading, which is uniquely expressed by the progressive (5-b).

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

b. niśā roṭi **banā rah-i** **hai**
N.NOM bread.NOM 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 permit the use of imperfective marking, alternating between (5-a) and (5-b) 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.[3] A similar situation is observed in English diachrony. Visser (1970, p. 746) observes that in Middle English, before the emergence of the *be + V-ing* construction as a grammaticalized progressive, the English Simple Present exhibited event-in-progress, stative, as well as the habitual, generic readings. The use

[3] In contrast to English and Hindi, in Italian (and Romance more generally), the presence of periphrastic progressive forms does not block the Imperfetto (and the Imparfait) from exhibiting the event-in-progress reading. Marchese (1986) reports a similar pattern of free variation for Godié, a language of the Kru family. Deo (2009, p. 516-517) offers an account of the free alternation pattern.

of the Simple Present in alternation with the Progressive continues in the Early Modern English of Shakespeare as seen in (6). In Modern English, however, the event-in-progress reading is typically unavailable to the Simple Present.

(6) a. What do you read, my lord? (Hamlet II.2.191)
b. O, I die, Horatio. (Hamlet V.2.345)

2.2. Grammaticalizing changes

The second piece of evidence is the generalizing grammaticalization path, reported extensively in the typological and grammaticalization literature, where marking restricted to descriptions of events in progress is employed in a wider range of expressive functions at a diachronically later stage.

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 reportive 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 1997:339-340). This clear-cut distribution is illustrated in (7). The examples are from Göksel and Kerslake (2005:331). In (7-a), the verb form with -*(I)yor* describes an ongoing working eventuality, while in (7-b), the -*ir* inflected verb describes a characteristic pattern of working – a characterizing reading.

(7) 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.

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 (8-a)), and is also interchangeably used with the Aorist form in characterizing contexts (8-b). The examples are from Göksel and Kerslake (2005:333).

(8) 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 M.NOM lot cigarrette smoke-IMPF-PST.3SG
At that time, Mehmet *used to smoke* (lit: was smoking) a lot.

The Aorist, on the other hand, never exhibits an event-in-progress reading. These data have been interpreted as indicating that the Turkish Progressive is being extended to the domain of the imperfective Aorist, thus instantiating the progressive $\gg$ imperfective shift. According to Comrie (1976), dialect variation in Yoruba (Niger-Congo) presents a similar case. The periphrastic locative Progressive construction is used by some speakers to describe events in progress, while other speakers have extended it to describe habits as well, suggesting an ongoing change.

2.2.2. Tigre: Two imperfective markers

Bybee et al (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.[4] In each of these cases, they find that one marker is a diachronically older form while the other is a relatively younger form, evolved from a progressive marker. Consider, as an illustration, the facts from Tigre (Semitic). All the examples in (9)-(10) are from Raz's (1983) grammar of the Tigre language (pp. 70-72). The imperfective form (labeled Imperfect by Raz) exhibits the characterizing and continuous readings.

(9) a. ...'azedi sanni **na'amrakka**
now indeed well we know-IMPF.1PL you
Now indeed, we *know* you well. (Raz 1983: 70)

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 imperfective form in periphrasis with present (*halla*) or past (*'ala*) tense auxiliaries. This use is said to resemble the use of the English Progressive to describe events in progress.

(10) a. ḥəna həd$\bar{\text{a}}$y **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

[4] Bybee et al (1994:144) 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.

A dog *was running* on the road.

However, this periphrastic construction also exhibits characterizing readings as shown in (11-a-b) with no semantic distinction from the bare Imperfect.

(11) 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
be-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 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

In addition to these cross-linguistic snapshots of progressive ≫ imperfective shifts in progress, there is also direct evidence of completed shifts in the history of some Indo-Aryan languages, one of the few language families for which we have continuous and extensive diachronic documentation. Progressive markers, innovated at earlier stages of some Indo-Aryan languages (e.g. Gujarati, Hindi, Marathi) have fully replaced the original imperfective marking to become the default markers of imperfective aspect (Deo 2006).[5] Such crosslinguistic facts about the synchronic organization and 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. Based on cases of changes in progress, forms with overlapping meaning, and replacements of an original imperfective form by a progressive form, Bybee et al make the following observation:

> 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

[5] I do not describe these changes in any detail here since even a brief exposition would require going into the details of the particular morphological paradigms and different stages that is not as central to the more general questions I address in this paper.

> 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 et al 1994: 148–149).

It is clear that the crucial questions here have to do with determining on the one hand, WHAT the content of these "small and subtle distinctions" among aspectual categories 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. In other words, 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). §3 addresses the first question, §4 addresses the second question, while §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 proposals that undertake such a comparison of the contribution of the two aspects — Ferreira (2005) and Deo (2009).[6] 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.[7]

3.1. Deo 2009

Deo (2009), building up on prior ideas (Bonomi 1997; Delfitto & Bertinetto 1995; Lenci & Bertinetto 2000; Cipria & Roberts 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:

[6] De Swart (1998) also presents a brief sketch of the typological variation according to which the progressive is said to not combine with stative predicates while the more general imperfective carries no such restriction. De Swart's analysis treats the English Progressive as denoting an aspectual modifier that stativizes dynamic predicates while a form such as the French Imparfait is an aspectually sensitive tense that presupposes that the argument it applies to is stative.

[7] For a detailed comparison of the relative empirical coverage of the proposals in Ferreira (2005) and Deo (2009), see Deo (to appear).

- 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). The repetition, though undesirable, is necessary in order to give the reader the necessary background for understanding this paper. The changes made in this version of the analysis are minimal and allow for a more transparent carryover to the model of discourse structure assumed in the dynamic component of the change in §4.

3.1.1. Semantics

The ontology includes a non-null set of intervals $\mathscr{I}$ (with points as a special case) partially ordered by the relation of temporal precedence $\prec$ and by the subinterval relation $\subseteq$. $\mathscr{W}$ is a non-empty set of worlds. $i, j, k...$ are variables over $\mathscr{I}$. 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 the final subinterval of i does not diverge from w. The function *Inr* assigns to each $i \in \mathscr{I}$ 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: 152).[8] $Hist_{i_{inr}}(w)$ is the set of **inertial alternatives** of w at i.

(12) **Inertial alternatives**
$Inr =_{def} f : I \rightarrow \wp(\mathscr{W})$
$i \mapsto Hist_{i_{inr}}(w) \subset Hist_i(w)$

$\mathscr{E}$ is a domain of eventualities, sorted into a set of events $\mathscr{E}^E$ and a set of states $\mathscr{E}^S$. The temporal trace function τ from $\mathscr{E}$ to $\mathscr{I}$ gives the run time of an eventuality. The eventuality argument of basic eventive predicates is of the sort E

[8]Dowty (1977, 1979) introduces the notions of inertia worlds and inertia futures as a means to access the set of worlds/histories that are indistinguishable from each other up until the reference interval and beyond. Much literature on the Imperfective Paradox has focused on refining Dowty's notion of inertia, particularly relativizing it to the predicate and event under question (Landman 1992; Portner 1998). It is not within the scope of this paper to contribute to these refinements to the concept of inertial futures. Both the progressive and the characterizing uses of IMPF depend on the future behaving in ways predictable from the past and the present. *Inr* is intended to be a placeholder function that allows us restrict our attention to worlds that meet this predictability requirement.

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). Aspectual modifiers such as negation, frequency and Q-adverbs, and quantified PPs 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/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 (13). A predicate of events $P^{\mathscr{E}}$ 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. A predicate of intervals $P^{\mathscr{I}}$ or of states $P^{\mathscr{S}}$ stands in the coincidence relation with i and w iff the predicate holds throughout i in w.

(13) $$\textsc{coin}(P,i,w) = \begin{cases} \forall w' \in Hist_{i_{inr}}(w) : \exists e\,[P(e)(w') \wedge \tau(e) \circ i] & \text{if } P \subseteq \mathscr{E}^{\mathscr{E}} \\ P(i)(w) & \text{if } P \subseteq \mathscr{I} \text{ or } \mathscr{E}^{\mathscr{S}} \end{cases}$$

The final notion needed in specifying the meaning of the progressive and imperfective operators is that of a *regular partition*, defined in (14). For any interval i, a partition of i is the set of the non-empty, mutually exclusive, and collectively exhaustive subsets of i.

(14) **Regular partition**

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

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

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

c. $\forall j,k \in \mathscr{R}_i \rightarrow \mu(j) = \mu(k)$ (where $\mu(x)$ stands for the Lebesgue measure of x).[9]

For any $\mathscr{R}_i$, 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, a set 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

[9]The Lebesgue measure is the standard way of assigning a length, area, or volume to subsets of Euclidean space. Intervals are a proper subset of the Lebesgue measurable subsets of the real number line.

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, $\mathscr{R}^c_j$, 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 i and properties of the event description.[10]

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

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 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, (17) 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 (16).[11]

(16) $[\![\text{-e}]\!] = \lambda P \lambda i \lambda w. \exists j[i \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \text{COIN}(P,k,w)]]$

The logical form for (17-a) is in (17-d). That is, (17-a) 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 event is fully realized in the inertial alternatives of w at k.

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

b. PRES $(\text{IMPF}\ (\lambda e[\text{Niśā-make-bread}(e)]))$

c. PRES $(\lambda P \lambda i \lambda w. \exists j[i \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \text{COIN}(P,k,w)]]\ (\lambda e[\text{Niśā-make-bread}(e)]))$
$= \text{PRES}\ (\lambda i \lambda w. \exists j[i \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \text{COIN}(\lambda e[\text{Niśā-make-bread}(e)],k,w)]])$
$= \text{PRES}\ (\lambda i \lambda w. \exists j[i \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \forall w'[w' \in Hist_{k_{inr}}(w) \rightarrow \exists e[\text{Niśā-make-bread}(e)(w') \wedge \tau(e) \circ k]]]])$

[10] 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.

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

d. $\lambda w.\exists j[\mathbf{now} \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \forall w'[w' \in Hist_{k_{inr}}(w) \rightarrow \exists e[$Niś$\bar{\text{a}}$-make-bread$(e)(w') \wedge \tau(e) \circ k]]]])$

Detailed arguments supporting this proposal have been presented in Deo (2009) which need not be repeated here. We turn now to the meaning of PROG, which, according to Deo, 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.

(18) PROG: $\lambda P \lambda i \lambda w.\forall j[j \in \mathscr{R}^c_i \rightarrow \text{COIN}\ (P, j, w)]$

By letting PROG and IMPF vary along only this parameter, Deo is able to further account for two effects: (a) the fact that the presence of the progressive blocks the imperfective from having an event-in-progress reading (described in §2.1); and (b) the inference of temporal contingence associated with progressive marking.

As seen in the examples in (19-b/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 & Woisetschlager 1982 among others.

(19) a. Mary *was biking* to work...when she got hit by a bus. *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 contingence that sometimes leads to the infelicity of the progressive with stative predicates, as in (20).

(20) 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: 174)
d. ?That argument *is resting* on an invalid assumption. (Dowty 1979: 174)

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 immoveable 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) and the inference of temporal contingence and the related effect of (in)felicity.[12]

3.1.2. The progressive–imperfective contrast

For any predicate P and interval i, $\textsc{impf}(P)(i)$ denotes the set of worlds where i is an initial subinterval of some interval j such that every cell of a small-enough partition $\mathscr{R}^c_j$ COINcides with P. For any $w \in \textsc{impf}(P)(i)$, it can be in set (21-a) or (21-b), and possibly in both. In contrast, going by the meaning in (19), $\textsc{prog}(P)(i)$ denotes the set in (21-b).

(21) a. $\{w \mid \exists j[i \subset_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \textsc{coin}(P,k,w)]]\}$
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. $\{w \mid \exists j[i = j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \textsc{coin}(P,k,w)]]\}$
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, $\textsc{prog}(P)(i)$ and $\textsc{impf}(P)(i)$ denote exactly the same set – the union of (21-a) and (21-b).[13] However, for any *eventive* predicate P and interval i, $\textsc{prog}(P)(i)$ is semantically stronger than $\textsc{impf}(P)(i)$. It is easy to see how PROG is a "semantically narrower" version of IMPF on this construal of their contribution, since $\textsc{prog}(P)(i)$ asymmetrically entails $\textsc{impf}(P)(i)$, 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 sentences containing one of

[12]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 superceded by the reasoning in §3.1.2.

[13]This is because $\textsc{coin}(P,i,w)$ 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.

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.[14]

In what follows, I will assume that the two operators are in a privative opposition, i.e. an opposition in which one element is unspecified for some semantic feature that the other element is specified for. 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 $\langle$PROG, IMPF$\rangle$ 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 strengthening 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 (21-a), 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 runtime, $\tau(e)$ (which is a continuous interval), overlaps with every cell of a small-enough partition of j. If there 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 inferences 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, i.e. they entail each other. If two expressions are semantically equivalent, there is a preference for simpler, less complex utterances – a Manner consideration. $\text{IMPF}(P)(i)$ is morphosyntactically less complex than $\text{PROG}(P)(i)$.[15] 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 sim-

[14]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.

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

pler alternative. One motivation for not choosing the simpler alternative would be that the speaker does not have evidence that P holds *beyond* the reference interval i – a salient interpretation of $\text{IMPF}(P)(i)$. In such a case, $\text{PROG}(P)(i)$ receives an upper-bounded interpretation, i.e. its use leads to the exclusion of the worlds in (21-a). The strengthened interpretation for $\text{PROG}(P)(i)$, where P is stative, is given in (22). The COIN relation has been resolved as defined for stative predicates.

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

This upper-bounded interpretation of $\text{PROG}(P)(i)$ is what the temporal contingence inference amounts to. This, 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 (23-a) or the complex Present Progressive (23-b). 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 (23-b-i) to the set in (23-b-ii), in which N does not hold at proper superintervals of the reference interval **now**.

(23) a. New Orleans *lies* at the mouth of the Mississippi River. IMPF
(i) $\lambda w.\exists j[\mathbf{now} \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}_j^c \rightarrow N(k)(w)]]$
$= \lambda w.\exists j[\mathbf{now} \subseteq_{ini} j \wedge N(j)(w)]$

b. ?New Orleans *is lying* at the mouth of the Mississippi River. PROG
(i) $\lambda w.\forall k[k \in \mathscr{R}_{\mathbf{now}}^c \rightarrow N(k)(w)]$
(ii) $\lambda w.\forall k[k \in \mathscr{R}_{\mathbf{now}}^c \rightarrow (N,k,w)] \wedge \neg\exists j[i \subset_{ini} j \wedge \forall k[k \in \mathscr{R}_j^c \rightarrow (N,k,w)]]$
$= \lambda w.N(\mathbf{now})(w) \wedge \neg\exists j[\mathbf{now} \subset_{ini} j \wedge N(j)(w)]$

This upper-bounded reading associated with the choice of (23-b) 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, (23-b) 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.[16]

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

[16] In recent years, the Mississippi has indeed 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.

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

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. On the other hand, sentences containing the morphosyntactically complex member PROG with stative predicates receive an upper-bounded reading, giving rise to the temporal contingence inference.

3.2. Summary

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. In doing so, we began with the *constraints* problem – the problem of discovering the grammatically relevant elements of meaning and the constraints on their interaction. The first step was to give an explicit semantic account of the two categories and explicate how the progressive can be seen as a semantically narrower version of the imperfective (the WHAT question). These meanings can be now straightforwardly taken as the input to and the output of the progressive to imperfective grammaticalization process. This semantic characterization underlies the grammaticalization path – giving us, for one particular path, what I am calling the static or structural component to an explanation of semantic change.

The dynamic workings of the progressive to imperfective grammaticalization path has three aspects that need to be understood: (a) the emergence of PROG as a new grammaticalized aspectual category in a language that lacks it; (b) its categoricalization in certain semantic contexts; and (c) the generalization of PROG's exponent to IMPF meaning. That is, we need to know (a) why new functional expressions emerge in an existing linguistic system, (b) become entrenched as obligatory grammatical elements, and (c) broaden in meaning over time. In addressing these questions, we are looking to understand Weinrich, Herzog, and Labov (1968)'s *actuation* and *transition* problems for semantic change – or the WHY and the HOW questions. These are addressed in §4 and §5.

4. The dynamic component

We can discern four distinct states of a linguistic system in the grammaticalization path associated with the PROG and IMPF operators. There is the initial **zero-**PROG state (25-a), in which the language possesses only a single grammaticalized device across the imperfective domain – the exponent of IMPF. The second, **emergent-**PROG state (25-b), is one in which morpho-syntactic resources of the language have been recruited in introducing a new grammaticalized exponent for PROG. In this state, the language has a progressive marker and distinguishes between progressive and non-progressive meaning but its frequency is relatively low. The third state can be called the **categorical-**PROG state (25-c). In this state, the exponents of PROG and IMPF have relatively circumscribed (though overlapping) domains of use, with at least some categorical sub-domains for each. State four, which we might call **generalized-**PROG (25-d), is the state in which the exponent of PROG loses its semantic restriction and generalizes to IMPF.[17]

(25)	a.	X_{impf}	zero-PROG
	b.	$(Y_{prog})\ X_{impf}$	emergent-PROG
	c.	Y_{prog}, X_{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:

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

b. **Categoricalization** of the new expression to obligatory use in certain contexts within the domain – the change from (25-b) to (25-c)

c. **Generalization** of this expression to the broader semantic domain – the change from (25-c) to (25-d).

The analysis of these dynamic processes proposed here builds on results from game-theoretic pragmatics and game-theory more generally that seek to model the interaction between semantic content, strategic considerations in rational communication, and acquisitional asymmetries. My broad goal will be to show that if we understand the empirical domain in the right way and make certain reasonable assumptions, the innovation of a new functional expression via recruitment, its categoricalization,

[17] As far as speakers are concerned, (25-d) 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 reinstitution of salient semantic contrasts. Some relevant discussion is in §6.

and its generalization can be viewed as transitions within the strategy space of a game model for communicating a given set of meanings with minimal ambiguity and cost. The progressive to imperfective grammaticalization path, on such a view, is reconstructible as a pattern in which alternative communication strategies rise and fall in dominance within a given population over time due to contingent as well as structural factors that effect changes in their frequency and their average utility. The analysis relies on the differential profiles of such communication strategies with respect to three factors: their communicative success, their formal complexity, and, relatedly, the probability of mis-transmission associated with each.

The full analysis is spread over two sections. This section, §4, focuses on the systematic difference between the ways in which language speakers may use progressive and non-progressive marking and the effects of these. §4.1 proposes that speakers/hearers distinguish between two types of inquiries – structural and phenomenal ones – and seek to disambiguate between these in discourse. 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. Thus, languages with and without PROG exponents have distinct communicative profiles in terms of how they solve the disambiguation problem. In §4.2, I informally outline the nature of the three transitions in a grammaticalization path: recruitment, categoricalization, and generalization. In §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 our attention to descriptions of ongoing (rather than completed) situations relative to some reference time, let us distinguish between two broad kinds of inquiries into the state of the world, any 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 (27-a) and(27-b) illustrate the difference between inquiries that I have in mind. Note that the questions in (27-a) are expressed in English using the Simple Present tense while those in (27-b) use the Progressive.

(27) a. *What characterizes the world generally?*
- (i) What problems do developing nations face?
- (ii) Why do dogs wag their tails?
- (iii) How do whales give birth?
- (iv) Does John walk to school?

(v) What does the earth revolve around?

b. *What characterizes the world at some time i?*
(i) What problems are developing nations facing in 2012?
(ii) Why is Fido wagging his tail right now?
(iii) How is that whale giving birth?
(iv) Is John walking to school?
(v) What was the earth revolving around on Tuesday evening?

Following intuitions underlying Goldsmith and Woisetschlaeger's (G&W) analysis of the English Progressive (G&W 1982), 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 underlies 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 given in §3.1 and the phenomenal/structural contrast that G&W rely on in their analysis of the progressive. Here is the connection: For any predicate *P*, interval *i*, and world *w*, $\text{PROG}(P)(i)(w)$ entails that *P* COINcides with *i* at *w*, i.e. a *P* eventuality is observable at *i*. There is no such entailment with $\text{IMPF}(P)(i)(w)$.[18] Thus, a progressive sentence entails the existence of one or more *P*-eventualities whose runtime 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 runtime of one or more *P*-eventualities. Based on this difference, I want to 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 any 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:

(28) a. Any language that realizes both PROG and IMPF can conventionally restrict the use of IMPF-marking for structural inquiries – thus max-

[18]Since every cell of $\mathscr{R}_i^c$ must COINcide with *P* in *w*, it follows that *i* itself must COINcide with *P* in *w*.

imizing the disambiguation between phenomenal and structural inquiries.

b. Any language that realizes only IMPF necessarily relies on contextual information in order to disambiguate between phenomenal and structural inquiries.[19]

The basic assumption we need to make is that there is some functional pressure to distinguish between phenomenal and structural inquiries. A linguistic system may either rely on formal disambiguation of phenomenal and structural meanings (by using PROG-marking) or on contextual disambiguation of these. These options can be regarded as distinct strategies for communication within the imperfective domain and each have their distinct profiles: one strategy uses multiple forms with disambiguated meanings, thus ensuring communication through explication; the other strategy uses a single form, but relies on the hearer's understanding of the context for successful communication.

These differences should have implications both for how communicatively successful a particular strategy is and how difficult it is to acquire. Ideally, the grammaticalization path we have set out to understand should fall out naturally as a consequence of the differential profiles of the strategies we have outlined informally. §5 offers an initial attempt at formally tackling this problem using an evolutionary game-theoretic model.

Before proceeding to the game-theoretic analysis, I will provide in §4.2 an intuitive description of the diachronic path as one in which new exponents emerge as variants, get recruited in the grammaticalized system of the language, and change in distribution and interpretation over time, as a result of the interaction between communicative goals and acquisitional biases. §4.3 briefly touches upon the puzzle of why zero-PROG languages do not seem to innovate markers that allow them to disambiguate structural inquiries, which then generalize to cover IMPF meaning.

4.2. The diachronic steps in a grammaticalization path

4.2.1. Recruitment: the emergence of variants

Consider a zero-PROG linguistic system, one that only contains marking corresponding to the broader IMPF and has no grammaticalized device to convey the narrower meaning corresponding to PROG. Speakers and listeners of such a system regularly participate in speech events which require disambiguation of phenomenal inquiries in the local context. Common knowledge of the context of interlocutors is not always enough to resolve such inquiries and in underspecified contexts, participants may undertake local efforts for distinguishing between phenomenal and

[19] This information may be pragmatically retrieved just like information about temporal location may be pragmatically retrieved in tenseless languages (See Bohnemeyer 2002, 2009; Bittner 2005; Tonhauser 2011). And optional adverbials (e.g. *right now*, *yesterday evening*, *last week*), which are not part of the grammaticalized tense–aspect system of the language are always available to formally facilitate disambiguation.

structural inquiries. 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.[20] Both adverbial and constructional linguistic devices overtly introduce the reference intervals throughout which the eventuality predicates are asserted to hold. The inquiry is determined to be phenomenal in a given context by the use of such devices because they restrict the temporal interpretation of the question or the assertion narrowly to the reference interval. Note that this is the main distinction between the semantic contribution of grammaticalized progressive marking and imperfective marking. In some zero-PROG languages, such morphosyntactically explicit local efforts may systematically accumulate and acquire high statistical frequency. The grammaticalization of a new progressive marker, on this view, is essentially the conventionalized use of one privileged exponent for conveying that the inquiry is restricted to the reference interval. Such a form is used with reliable frequency for marking phenomenal inquiries.[21] Once such a privileged exponent is chosen, the language enters the emergent-PROG state, which is characterized by the availability of a distinct morphosyntactic exponent for the PROG operator that alternates with the original under-specified (imperfective or aspectually neutral) morphology in marking phenomenal inquiries. The emergent-PROG system uses a communication strategy that is formally less economic than the zero-PROG system because it involves the use of multiple forms.[22] However, the presence of an exponent for the PROG operator increases the expressivity of such a system (and thereby chances of successful communication) relative to the zero-PROG system.

4.2.2. Categoricalization: the conventionalization of variants

Speakers of an emergent-PROG language have progressive marking to mark phenomenal inquiries but no comparable grammaticalized disambiguator for structural inquiries. Although lexicalized, progressive marking is not categorically (nor even very frequently) used in phenomenal contexts. It alternates with the older imperfective marking, which is under-specified and may mark both phenomenal and structural inquiries. This emergent-PROG state diachronically gives way to the categorical-PROG state where progressive marking is employed almost universally in phenomenal inquiries while the use of imperfective marking is conventionally associated with structural inquiries.

[20]Bybee et al (1994:127–130) 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.

[21]There is the question of whether the conventionalization happens as a result of the frequency of the exponent in the child's input or whether it may happen in an adult's grammar once sufficient frequency of exposure has been reached. The change is an addition to the functional vocabulary of the language and so it seems that both scenarios are possible.

[22]I am assuming here that economy is not only characterizable in terms of the syntactic structure of a message but can also be tied to the number of distinct messages associated with a given conceptual domain. A strategy s for communicating meanings within a meaning domain is more complex than s' iff s employs more forms than s'. This notion of economy will be used to compute complexity of strategies in §5.

Bybee (1994) labels this kind of introduction of systemic opposition the grammaticization of 'zero' morphology. Dahl (Dahl, 2000, pp. 10–11) introduces the notion of a "Doughnut Gram" to describe formal marking which fails to appear in certain semantic contexts that its meaning is compatible with, because its use is blocked by the presence of more specific formal marking. On the basis of several phenomena Horn (1984) proposed a *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. 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.[23] One of the phenomena Horn discusses involves the evolution of privative oppositions (p. 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: 37-38).[24]

As a long-term diachronic effect, it seems natural to construe both the recruitment of a new progressive marker for disambiguating phenomenal inquiries and its categoricalization as an optimizing effect of Zipfean forces – constraints on the expressivity and the economy of messages. In an emergent-PROG system, the presence of a grammaticalized progressive form allows the speaker to explicitly distinguish phenomenal inquiries from structural inquiries, i.e. be more expressive. But this specification comes at a cost – an increase in the number of messages – violating economy. However, the language does not have a grammaticalized counterpart to unambiguously mark structural inquiries. The optimally economical and optimally expressive solution would be one that allows a complete resolution of both phenomenal and structural inquiries with minimal structural complexity. This may be facilitated by the innovation of new material dedicated for structural inquiries or through the reorganization of the functions of existing material. Innovation would satisfy expressivity but would also introduce additional structure (number of messages) into the grammar, failing economy. Reorganization, in which imperfective marking is pragmatically restricted to structural inquiries (via blocking) and progressive marking is used categorically in phenomenal inquiries, satisfies both expressivity and economy. The long-term economical solution for speakers would

[23]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).

[24]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).

therefore be to reorganize rather than to innovate, that is, to restrict the semantically broader form to structural inquiries and use the innovated form with PROG meaning obligatorily in phenomenal inquiries. This is the categorical-PROG system.

4.2.3. Generalization: the failure to acquire conventionalized variants

Speakers of a categorical-PROG language have an optimally economical and optimally expressive system for disambiguating between phenomenal and structural inquiries. However, this state of the language diachronically gives way to the generalized-PROG state in which the progressive form generalizes in meaning and realizes IMPF rather than PROG. We are looking at a cyclic pattern here: 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 to distinguish between phenomenal and structural inquiries. What could possibly motivate such a counter-intuitive transition, in which the expressivity afforded by recruitment and categoricalization at prior states of the language is obliterated at a later state? It is clear that we must characterize generalization as the cumulative failure of speakers over several generations to acquire the categorical-PROG system. But what could it be that causes such systematic failure, which ultimately leads to the loss of a salient semantic distinction in the language?

There may be several ways to pursue this question, but I will propose one possible way that I find most promising. Suppose that the population of acquirers for a language is biased towards simpler single-form contextually reliant strategies for disambiguation (i.e. the zero-PROG system). This preference is defeasible and the learned grammar is ultimately determined by the structure of the input that the child receives.[25] However, as the categorical-PROG system becomes more and more established in the speech community, the frequency of overt PROG-marking in the child input would correspondingly increase. Because acquirers are biased towards a simpler single-form strategy, the increased frequency of overt PROG-marking 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 crucially tied to the frequency of the system in the population. This is only a sketch of a story and its concrete implementation in §5 is not much more than that. But it suggests that in addition to the communicative success and formal complexity associated with a given strategy for communication, transitions observed along a grammaticalization path may depend on acquisitional and learning biases. The success of an optimizing and communicatively successful strategy (measured in terms of the frequency of the PROG-exponent) may itself be

[25] As far as the Modern English goes, evidence from the first language acquisition of tense-aspect literature suggests that the use of progressive marking is much more frequent in child-directed speech than simple present marking. 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.

the basis for its demise in a population over time (through generalization of the PROG-exponent).

4.3. The non-attestation of generic to imperfective paths

In §4.2, I sketched an account of why speakers of a zero-PROG system might innovate a PROG-exponent that facilitates the disambiguation of phenomenal inquiries from structural inquiries. A logical counterpart to this grammaticalization path is one in which speakers innovate a dedicated exponent for marking 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. The *used to V* construction in Modern English is an example of such a marker. In fact, right around the time that Middle English develops the Progressive, it also develops a dedicated construction *uses to V* (which parallels the past tense construction) to express characterizing meanings in the present (Tagliamonte and Lawrence 2000).[26]

So it is not that languages do not innovate such markers. The puzzle, rather, is that there seems to be no attestation of such habitual/generic markers changing in meaning to encompass the function of the progressive. That is, there is no "generic $\gg$ imperfective path" that would be the mirror image of the progressive $\gg$ imperfective path, where the semantically more specific generic/habitual markers diachronically get extended to describe, for instance, events in progress (a salient use of progressive markers). This asymmetry in diachronic patterns deserves some comment. While I do not have a full solution to this puzzle, the basic semantics proposed in §3 may offer some direction in understanding why this might be the case.

Let us make some concrete assumptions about the contribution of GEN – the operator whose exponent would be a habitual/generic marker in a language. Within the general spirit of the aspectual semantics in §3, one could take the GEN operator to make a contribution such as in (29).

[26]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 that the *uses to 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.

(29) $\quad$ GEN: $\lambda P \lambda i \lambda w.\ \exists j[i \subset_{ini} j \wedge \forall k[k \in \mathscr{R}_j^c \rightarrow \text{COIN}(P,k,w)]]$

Compare this to the meaning of the underspecified IMPF:

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

For any stative or eventive predicate P and interval i, the denotation of IMPF$(P)(i)$ differs from GEN$(P)(i)$ in that it also includes worlds in which only the limiting case holds – i.e. where P distributes over parts of i but not over parts of proper superintervals of i.

While both PROG and GEN operators are special cases of the general IMPF, only PROG constitutes the limiting case for IMPF. The progressive $\gg$ imperfective path involves generalization from the limiting case to a wider class of cases – and is in line with the standard construal of generalization as involving the relaxation of certain parameters of definition. Specifically, in this case, the j-interval parameter is relaxed to include any superinterval of the reference interval. The generic $\gg$ imperfective path, on the other hand, would have to be effected by extension to the limiting case in IMPF (where $j = i$), which is a strengthening of the definitional parameters.

The two paths are thus qualitatively different and it is very likely that the non-attestation of the generic $\gg$ imperfective path has to do with this difference. A full treatment of the asymmetry in instantiation between the two kinds of logically possible changes must await future research. For now, I add that such a treatment should additionally take into consideration the nature of the input to the child, specifically the relative prevalence of PROG forms vs. GEN forms in caregiver speech.

5. A game-theoretic interpretation of grammaticalization paths

The dynamic interactions between expressiveness, economy, and learnability and their role in the transitions of a grammaticalization path were sketched out informally in §4.2. This section makes these interactions more precise by 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. 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 David Lewis (1969). Lewis modeled linguistic convention in terms of repeated plays of signaling games with possibly arbitrary signals. Such signals come to acquire particular meanings because of the ways in which they are used by rational agents. Later work adopted an evolutionary (Maynard Smith and Price 1973) approach to the problem of strategy selection and its propagation across a population, thus attributing a non-central role to rationalistic reasoning in how linguistic systems might come to exhibit certain tendencies (e.g. van Rooij 2004a, 2004b, Jäger 2007). This makes EGT a useful tool for modeling the kind of semantic change we see in grammaticalization paths, in which new semantic variants or new uses for older variants might come about initially as a result of pragmatic reasoning, but their conventionalization and propagation over time must rely on less rationalistic processes such as learning by imitation and adaptation over multiple generations. An excellent introduction to (evolutionary) game theory and its application to linguistics can be found in Benz, Jäger, and Van Rooij (2006b) and Jäger (2004).

The model proposed here draws on ideas in van Rooij (2004a, 2004b) and Jäger (2007), among others. The former attempts to account for the emergence of conventions corresponding to Horn's division of pragmatic labor in a language. The latter is typological in orientation and attempts to derive cross-linguistically robust case marking patterns (such as accusative systems with differential object marking and ergative systems with differential subject marking) as reflecting evolutionarily stable states. In modeling the progressive $\gg$ imperfective grammaticalization path, we are interested, on the one hand, in the typological aspect – that is, how the properties of each state in the grammaticalization path (the zero-PROG state, the emergent-PROG state, and the categorical-PROG state) might be construed as optimal patterns of functional adaptation in the game-theoretical sense.[27] On the other hand, we are interested in the cross-temporal aspect – that is, the evolutionary dynamics that can model why recruitment, categoricalization, and generalization appears to occur in cyclic fashion in the imperfective domain.

5.1. The Imperfective Game

The **Imperfective Game** developed here 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).

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. For us, these notions are relevant to understanding how a linguistic system in which imperfective marking signals both phenomenal and structural inquiries (with contextual disambiguation) may be driven out by an alternative system in which progressive marking signals phenomenal inquiries and imperfective marking signals structural inquiries (and vice versa).

[27] This is of inherent typological interest since the states identified as steps in the grammaticalization path in (25) basically correspond to a cross-linguistic typology of imperfective marking. The zero-PROG state is exemplified by languages like Russian and Modern Standard Arabic, the emergent-PROG state is found in the Modern Romance languages, while the categorical-PROG state is represented by languages like Modern English, Hindi, and Turkish (prior to the currently ongoing changes).

In order to do so, we will represent the available linguistic options in terms of speaker–hearer strategy pairs, which can be seen as linguistic conventions that are 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 (i) when it leads to successful communication, and (ii) it does so with small cost (van Rooij 2004b: 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.[28] The meaning of a form is the state that the speaker intends to communicate with the form. States and meanings are thus identified here. 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**} $\rightarrow$ {*prog*, *impf*}] while a hearer's strategy is a function from forms to meanings/states (in this case, an element of [{*prog*, *impf*} $\rightarrow$ {**phen**, **struc**}]. The utility function for the speaker and the hearer is defined with respect to such strategies. Thus, given a speaker strategy S, and 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).

(31) $$\delta_t(S,H) = 1 \text{ if } H(S(t)) = t \quad = 0 \text{ otherwise}$$

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

Furthermore, 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.

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

[28] 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.).

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

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.

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.[29] However, given that such a strategy would also be interpretively deficient – i.e. it would not allow for a subset of meanings to be distinguished – it is not comparable with hearer strategies that are able to express 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.

(33) $U_h(S,H) = \delta_t(S,H)$

Nature deals out the states **phen** and **struc** according to some probability distribution, which determines the likelihood of each state to be expressed.[30] We will make the simplifying assumption that the probability that the speaker and hearer are in the states **phen** and **struc** is the same (0.5). The speaker, however, has knowledge of the state she is in while the hearer lacks this knowledge. The average utility of a speaker or hearer strategy are then calculable as follows.[31]

(34) a. $U_s(S,H) = \sum_t P(t) \times (\delta_t(S,H) - k \times n(S))$

[29] I thank Michael Franke for pointing this out as a possible way of construing complexity over strategies.

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

[31] The average utility for a speaker or hearer strategy is computed by summing over the utility of the strategy in each state weighted by the probability of the state.

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

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_1 in which P(**phen**) = 0.9 and P(**struc**) = 0.1 and C_2 in which P(**struc**) = 0.9. Both contexts are equally likely 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.

Table 1: Speaker Strategies

	C_1		C_2	
	phen	**struc**	**phen**	**struc**
S_1	*impf*	*impf*	*impf*	*impf*
S_2	*impf*	*impf*	*prog*	*impf*
S_3	*prog*	*impf*	*prog*	*impf*
S_4	*prog*	*prog*	*prog*	*prog*

S_1 can be called a contextually disambiguating strategy in which the speaker uses the same form *impf* to convey both **phen** and **struc** relying on shared context for disambiguation. S_2 is a partially context-dependent strategy; the speaker uses the unambiguously **phen** form *prog* only in C_2, the context in which **struc** is the more probable state. S_3 is a context-independent explicit marking strategy, one in which the speaker uses the forms *prog* and *impf* across contexts to convey **phen** and **struc** respectively. Finally, S_4 is exactly like S_1, a contextually disambiguating strategy using *prog* instead of *impf*.

H_1 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_2 is still context-dependent but invariantly assigns **phen** to *prog*. In H_3, a form-dependent strategy, *prog* and *impf* are invariantly assigned **phen** and **struc** meanings respectively.[32]

[32]The strategies considered in this game model do not exhaust the logical space of strategies

Table 2: Hearer Strategies

	C_1		C_2	
	prog	*impf*	*prog*	*impf*
H_1	**phen**	**phen**	**struc**	**struc**
H_2	**phen**	**phen**	**phen**	**struc**
H_3	**phen**	**struc**	**phen**	**struc**

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.

(35) 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 (35). The results are in Table 3. The utility for hearers is exactly the same without the cost k factored in.

Table 3: Average utilities

Strategies	H_1	H_2	H_3
S_1	0.9	0.9	0.5
S_2	0.9 - k	0.95 - k	0.55 - k
S_3	0.9 - k	0.95 - k	1 - k
S_4	0.9	0.5	0.5

The game as introduced so far is an *asymmetric* game since every individual is either a speaker or a hearer and the speakers and hearers make use of disjoint sets of strategies. In an evolutionary, population-dynamic setting, this amounts to there being two populations – one corresponding to the speakers and the other to the hearers. A *symmetrized* version of this asymmetric game allows each player to switch between the role of the speaker and the hearer with equal probability. A strategy in such a symmetrized game is a pair of some speaker and some hearer strategy

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 $\gg$ imperfective cycling path. As discussed in §4.3, there is no crosslinguistically known corresponding trajectory that takes as its starting point the innovation of a generic/habitual form for marking structural inquiries and extends its use to phenomenal inquiries.

$\langle S,H\rangle$. The expected payoff of an individual's strategy $\langle S,H\rangle$ can be calculated on the basis of the average utility for her speaker strategy and her hearer strategy.

(36) **Expected payoff:** $\frac{1}{2}\times U_s(S,H)+\frac{1}{2}\times U_h(S,H)$

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. The payoffs for an individual strategy as it plays against another individual strategy are calculated on the basis of expected payoffs in Table 3. The question for us is, which of the twelve possible speaker-hearer strategy pairs can be considered to be optimal for communication?

Note that in an evolutionary setting, we are not concerned with the outcomes of a single game but with large populations of players. 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. New strategies may enter a population through unfaithful replication (**mutation**). If the mutant strategy yields a higher payoff against the incumbent population, the mutant will spread among the population and may possibly drive out the incumbent strategy(ies). Mutant strategies that yield lower than average payoffs will be wiped out over time through interaction with the incumbent strategy(ies).

Which strategy pairs from the set generated by the strategies above might successfully be adopted by a significant proportion of a population? For identifying such optimal strategy pairs, we will use the notion of an **evolutionarily stable strategy (ESS)** (Maynard Smith & Price 1973). A strategy α is said to be ESS if for all strategies β in a set of strategies A, either the condition in (37-a) or (37-b) holds. That is, either (a) the expected payoff that α gets when playing against an α-opponent is strictly greater than the expected payoff gotten by any strategy β playing against an α-opponent, or (b) the expected payoff that α gets playing against an α-opponent is equal to the payoff that β gets in playing against α, but the payoff that α gets when playing against β is strictly greater than the payoff that β gets when playing against a β-opponent.

(37) a. $(\alpha,\alpha)>(\beta,\alpha)$, **or**
b. $(\alpha,\alpha)=(\beta,\alpha)$ and $(\alpha,\beta)>(\beta,\beta)$

A weak ESS differs from an ESS in the following way: the expected payoff for α when it plays against itself is either greater than or equal to the payoff for any strategy β when it plays against α, and the payoff that α gets when playing against β is at least as great as the expected payoff received by β when it plays against itself. That is, α is a weak ESS iff for all strategies β, α does no worse against β than β against itself.

(38) a. $(\alpha,\alpha) > (\beta,\alpha)$, **or**
b. $(\alpha,\alpha) = (\beta,\alpha)$ and $(\alpha,\beta) \geq (\beta,\beta)$

Natural language grammars are stable conventions that are followed by most members of a linguistic community. The expectation therefore is that such grammars reflect ESSs that are adopted successfully by a significant proportion of the population. Diachronic changes such as those involved in the cyclic path of interest here must result at least partially from an interaction between the relative payoffs of individual strategies that allow them to change in dominance over time. In order to understand the dynamics of such changes, we must first refine the set of strategies that may be considered.

We calculate the expected payoff for each strategy as it plays against another strategy based on (36). The table of expected payoffs where the row strategy plays against the column strategy is given in the table below. The cost k, here taken to be minimal at 0.01, is only incurred for multi-form speaker strategies. From the table we see that:

- $\langle S_1, H_1\rangle$, $\langle S_2, H_2\rangle$, and $\langle S_4, H_1\rangle$ are weak ESSs.
- $\langle S_3, H_3\rangle$ is an ESS.
- $\langle S_2, H_2\rangle$ and $\langle S_3, H_3\rangle$ remain (weak) ESSs as long as k does not exceed some threshold – 0.1 in the case of $\langle S_2, H_2\rangle$ and 0.2 in the case of $\langle S_3, H_3\rangle$.

Table 4: Expected payoffs for each individual strategy

Strategies	$\langle S_1,H_1\rangle$	$\langle S_1,H_2\rangle$	$\langle S_1,H_3\rangle$	$\langle S_2,H_1\rangle$	$\langle S_2,H_2\rangle$	$\langle S_2,H_3\rangle$	$\langle S_3,H_1\rangle$	$\langle S_3,H_2\rangle$	$\langle S_3,H_3\rangle$	$\langle S_4,H_1\rangle$	$\langle S_4,H_2\rangle$	$\langle S_4,H_3\rangle$
$\langle S_1,H_1\rangle$	$\boxed{\frac{\mathbf{(0.9)+0.9}}{\mathbf{2}}}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$
$\langle S_1,H_2\rangle$	$\frac{0.9+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.5+0.95}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.5+0.95}{2}$	$\frac{0.9+0.5}{2}$	$\frac{0.9+0.5}{2}$	$\frac{0.5+0.5}{2}$
$\langle S_1,H_3\rangle$	$\frac{0.9+0.5}{2}$	$\frac{0.9+0.5}{2}$	$\frac{0.5+0.5}{2}$	$\frac{0.9+0.55)}{2}$	$\frac{0.9+0.55}{2}$	$\frac{0.5+0.55}{2}$	$\frac{0.9+1}{2}$	$\frac{0.9+1}{2}$	$\frac{0.5+1}{2}$	$\frac{0.9+0.5}{2}$	$\frac{0.9+0.5}{2}$	$\frac{0.5+0.5}{2}$
$\langle S_2,H_1\rangle$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9)}{2}$	$\frac{(0.55-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(0.55-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(0.55-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(0.55-0.01)+0.9}{2}$
$\langle S_2,H_2\rangle$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(0.55-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.95}{2}$	$\boxed{\frac{\mathbf{(0.95 - 0.01)+0.95}}{\mathbf{2}}}$	$\frac{(0.55-0.01)+0.95}{2}$	$\frac{(0.9-0.01)+0.95}{2}$	$\frac{(0.95-0.01)+0.95}{2}$	$\frac{(0.55-0.01)+0.95}{2}$	$\frac{(0.9-0.01)+0.5}{2}$	$\frac{(0.95-0.01)+0.5}{2}$	$\frac{(0.55-0.01)+0.5}{2}$
$\langle S_2,H_3\rangle$	$\frac{(0.9-0.01)+0.5}{2}$	$\frac{(0.95-0.01)+0.5}{2}$	$\frac{(0.55-0.01)+0.5}{2}$	$\frac{(0.9-0.01)+0.55}{2}$	$\frac{(0.95-0.01)+0.55}{2}$	$\frac{(0.55-0.01)+0.55}{2}$	$\frac{(0.9-0.01)+1}{2}$	$\frac{(0.95-0.01)+1}{2}$	$\frac{(0.55-0.01)+1}{2}$	$\frac{(0.9-0.01)+0.5}{2}$	$\frac{(0.95-0.01)+0.5}{2}$	$\frac{(0.55-0.01)+0.5}{2}$
$\langle S_3,H_1\rangle$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(1-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(1-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(1-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(1-0.01)+0.9}{2}$
$\langle S_3,H_2\rangle$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(1-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.95}{2}$	$\frac{(0.95-0.01)+0.95}{2}$	$\frac{(1-0.01)+0.95}{2}$	$\frac{(0.9-0.01)+0.95}{2}$	$\frac{(0.95-0.01)+0.95}{2}$	$\frac{(1-0.01)+0.95}{2}$	$\frac{(0.9-0.01)+0.5}{2}$	$\frac{(0.95-0.01)+0.5}{2}$	$\frac{(1-0.01)+0.5}{2}$
$\langle S_3,H_3\rangle$	$\frac{(0.9-0.01)+0.5}{2}$	$\frac{(0.95-0.01)+0.5}{2}$	$\frac{(1-0.01)+0.5}{2}$	$\frac{(0.9-0.01)+0.55}{2}$	$\frac{(0.95-0.01)+0.55}{2}$	$\frac{(1-0.01)+0.55}{2}$	$\frac{(0.9-0.01)+1}{2}$	$\frac{(0.95-0.01)+1}{2}$	$\boxed{\frac{\mathbf{(1 - 0.01)+1}}{\mathbf{2}}}$	$\frac{(0.9-0.01)+0.5}{2}$	$\frac{(0.95-0.01)+0.5}{2}$	$\frac{(1-0.01)+0.5}{2}$
$\langle S_4,H_1\rangle$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9+1}{2}$	$\frac{0.5+0.9+1}{2}$	$\boxed{\frac{\mathbf{0.9+0.9}}{\mathbf{2}}}$	$\frac{0.5+0.9}{2}$	$\frac{0.5+0.9}{2}$
$\langle S_4,H_2\rangle$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.5+0.95}{2}$	$\frac{0.5+0.95}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.5+0.95}{2}$	$\frac{0.5+0.95}{2}$	$\frac{0.9+0.5}{2}$	$\frac{0.5+0.5}{2}$	$\frac{0.5+0.5}{2}$
$\langle S_4,H_3\rangle$	$\frac{0.9+0.5}{2}$	$\frac{0.5+0.5}{2}$	$\frac{0.5+0.5}{2}$	$\frac{0.9+0.55}{2}$	$\frac{0.5+0.55}{2}$	$\frac{0.5+0.55}{2}$	$\frac{0.9+1}{2}$	$\frac{0.5+1}{2}$	$\frac{0.5+1}{2}$	$\frac{0.9+0.5}{2}$	$\frac{0.5+0.5}{2}$	$\frac{0.5+0.5}{2}$

In terms of the discussion in §4, each diachronic state identified in (25) can be interpreted as reflecting the relative prevalence in a population of one or more of the four evolutionarily stable strategies of the Imperfective Game. The zero-PROG state is one in which the individual strategy $\langle S_1, H_1 \rangle$ is the prevailing strategy.[33] The emergent-PROG state is the one in which $\langle S_2, H_2 \rangle$ is the prevalent strategy, while $\langle S_3, H_3 \rangle$ is the prevalent strategy in the categorical-PROG state. The generalized-PROG state reflects the prevalence of $\langle S_4, H_1 \rangle$, which is structurally identical to $\langle S_1, H_1 \rangle$ but differs in both choice of form and average payoffs when playing against $\langle S_2, H_2 \rangle$ and $\langle S_3, H_3 \rangle$. The next section presents 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.

5.2. The evolutionary dynamics

Evolutionary game dynamics has 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). Strategies that are more successful on average (have a higher average payoff) spread in the population while the less successful strategies diminish. There are two interpretations of evolutionary game dynamics. In the biological setting, particular strategies are encoded by genomes of individuals. Successful types spread in the population due to their higher reproductive rate. In the cultural setting, which is relevant to the problem here, behavioral strategies are reproduced by other individuals through imitation and learning. Successful strategies propagate through imitation while less successful strategies diminish.

Grammars, as complex strategies for communication, get replicated both in the process of acquisition, as well as through interaction between members of a population that influences linguistic behavior. In explaining any particular diachronic linguistic phenomenon, we would like to be able to understand how and under what conditions particular strategies from a given strategy set might come to become dominant in a population and give way to other strategies over time.

For explaining the cyclic structure of the diachronic trajectory, in addition to the average payoffs associated with particular strategies, it is necessary to take into consideration the relative learnability of individual strategies from the structure of the input that is available to the learner. The grammatical phenomena of 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 to the modeling of change in which we are dealing with interactions between competing forces. It has been applied to the problem of the evo-

[33] In this state, while there might be disambiguation of phenomenal inquiries in some contexts, there is no conventionalized form that has this function. In §5.3.2, the zero-PROG state is taken to be one in which $\langle S_2, H_2 \rangle$ occurs in the population at a frequency below some threshold value, which determines whether a disambiguating form is conventionalized or not.

lution of grammar (Nowak et al 2001) and can be naturally extended to the problem of how populations might evolve new rules that confer a fitness advantage (Nowak 2006: 279–282). There are two parts to the replicator-mutator dynamic. The replication rate of a strategy is determined by its overall communicative success in a population (an economy consideration). The "replicator" part of the dynamic models positive or negative changes in the frequency of a strategy over time based on how communicatively successful it is against other strategies. 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

The dynamic account here considers only the ESSs: $\langle S_1,H_1\rangle$, $\langle S_2,H_2\rangle$, $\langle S_3,H_3\rangle$, and $\langle S_4,H_1\rangle$. The expected payoffs for each interaction in this reduced game are given in (39) for quick reference.

(39)

Strategies	$\langle S_1,H_1\rangle$	$\langle S_2,H_2\rangle$	$\langle S_3,H_3\rangle$	$\langle S_4,H_1\rangle$
$\langle S_1,H_1\rangle$	0.9	0.9	0.7	0.9
$\langle S_2,H_2\rangle$	0.9 - $\frac{1}{2}k$	0.95 - $\frac{1}{2}k$	0.75 - $\frac{1}{2}k$	0.7 - $\frac{1}{2}k$
$\langle S_3,H_3\rangle$	0.7 - $\frac{1}{2}k$	0.75 - $\frac{1}{2}k$	1 - $\frac{1}{2}k$	0.7 - $\frac{1}{2}k$
$\langle S_4,H_1\rangle$	0.9	0.7	0.7	0.9

To say that $\langle S_1,H_1\rangle$, $\langle S_2,H_2\rangle$, $\langle S_3,H_3\rangle$, and $\langle S_4,H_1\rangle$ are ESSs means that each strategy is stable against invasion by small amounts of mutants. If a population consists entirely of players of one ESS, then a small number of mutants from another ESS (or any other strategy) will fail to take over this population. But the observed pattern is that a population in which $\langle S_1,H_1\rangle$is prevalent changes to one in which $\langle S_2,H_2\rangle$ is prevalent, which in turn, gives way to the prevalence of $\langle S_3,H_3\rangle$. Finally, the $\langle S_3,H_3\rangle$ population shifts to one in which $\langle S_4,H_1\rangle$ is prevalent, thus completing the cyclic trajectory.

Before we consider the more complex interactions that underlie these transitions, we will study the working parts of the replicator-mutator equation by observing how the interactions between two strategies in the abstract work. This will be made more concrete by studying the interaction between $\langle S_1,H_1\rangle$ and $\langle S_2,H_2\rangle$ as a 2×2 game in §5.3.1. Finally, in §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 contains only one type A. Now a mutant of type B enters the population. We want to know the fate of the mutant strategy: will it succeed in taking over the population or will it be opposed by the incumbent strategy? 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. In our 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]

(40) 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 vector $\overrightarrow{x} = (x_A, x_B)$ defines the composition of the population (i.e. it gives us for each strategy its frequency at a given time). The selection dynamics, i.e. the rate of change in the frequency of the A and B populations over time, can be written as follows.[35]

(41) a. $\dot{x}_A = x_A[f_A(\overrightarrow{x}) - \phi]$
b. $\dot{x}_B = x_B[f_B(\overrightarrow{x}) - \phi]$

The average fitness of the population is given by $\phi = x_A f_A(\overrightarrow{x}) + x_B f_B(\overrightarrow{x})$.

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.[36] For any strategies, i, j...n, let the probability that j mutates into i be denoted by Q_{ji}. Given these assumptions, the population dynamics are given by the "replicator-mutator" equation:

[34] These equations assume that the players meet randomly. Thus, for each player, the probability of interacting with an A player is x and the probability of interacting with a B player is $1-x$.

[35] The rate of change in the frequency of a strategy A is calculated by multiplying the frequency of the A population (x_A) with the difference between the fitness of A at a given population composition ($f_A(\overrightarrow{x})$) and the average fitness of the population (ϕ). Same for B.

[36] 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.

(42) $\quad \dot{x}_i = \sum_{j=1}^{n} x_j f_j(\overrightarrow{x}) Q_{ji} - \phi x_i$

The replicator-mutator equation gives the rate of change in frequency for any strategy i as a function of its fitness and the probability that imperfect learning leads into i. In words, the rate of change for the i-population is given by the difference between the average payoff for i times the probability that any j mutates to i and the average fitness of the population times the frequency of the i-population.

The replicator-mutator equation in (42) is in continuous-time form which assumes continuous 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 given in (43) assumes that selection and mutation occur at discrete time-steps and measures changes in frequencies of strategies at each time-step. x'_i gives the frequency of 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.[37]

(43) $\quad x'_i = \sum_{j=1}^{n} Q_{ji} \frac{x_j f_j}{\phi}$

We will use the discrete form of the replicator-mutator equation to simulate the patterns of recruitment, categoricalization, and generalization in remainder of this section.

5.3.1. *Recruitment*

In this model, recruitment, or the emergence of grammaticalized progressive marking amounts to the adoption of the $\langle S_2, H_2 \rangle$ strategy (with a particular conventionalized form) by a large proportion of the population. Suppose that there is a large population of $\langle S_1, H_1 \rangle$ players into which is introduced a small population of $\langle S_2, H_2 \rangle$ players. Let k be 0.01. Then, under what conditions will the mutant population be able to take over the incumbent population?

Let us first consider the situation in which learning is perfect – that is, the mutation matrix Q has the following form:

(44) $\quad Q =$

	$\langle S_1, H_1 \rangle$	$\langle S_2, H_2 \rangle$
$\langle S_1, H_1 \rangle$	1	0
$\langle S_2, H_2 \rangle$	0	1

Note that the replicator-mutator equation contains the replicator mutation as a lim-

$Q =$

	A	B
A	q_a	$1-q_a$
B	$1-q_b$	q_b

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

iting case (with perfect learning) (Nowak 2006: 272-73). The statement in (46) 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 $\langle S_1, H_1 \rangle$ and $\langle S_2, H_2 \rangle$. The two strategies are in a bistable relation: each is the best response to itself.

(45)

	$\langle S_1, H_1 \rangle$	$\langle S_2, H_2 \rangle$
$\langle S_1, H_1 \rangle$	0.9	0.9
	$\cup$	$\cap$
$\langle S_2, H_2 \rangle$	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 initial values for x_A (the proportion of the population using A) and x_B (the proportion of the population using B). There is an unstable equilibrium in the interior of the interval [0,1] given by:

(46) $x_A^* = \frac{d-b}{a-b-c+d}$ (Nowak 2006: 51)

If the initial condition, $x_A(0)$ is less than this value, then the system will converge to an all-B population. If the initial condition, $x_A(0)$ is greater than this value, then the system will converge to an all-A population. In the case of $\langle S_1, H_1 \rangle$ and $\langle S_2, H_2 \rangle$, plugging in the values for the payoffs, we have an unstable equilibrium at $x_A(0) = 0.9$, calculating as below.

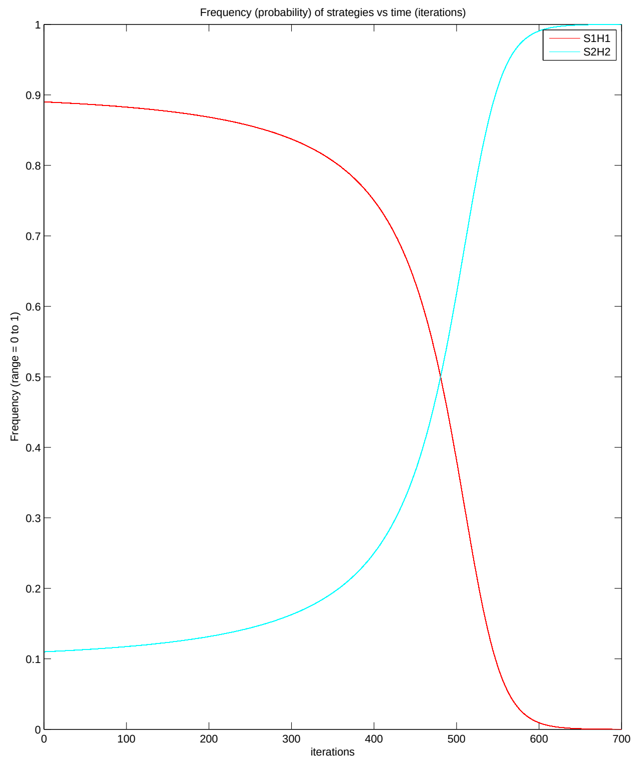
(47) $\frac{0.945-0.9}{0.9-0.9-0.895+0.945} = 0.9.$

This means that if the frequency of the mutant $\langle S_2, H_2 \rangle$ population exceeds 0.1, then the population will evolve to an $\langle S_2, H_2 \rangle$-dominant population. Recruitment or the grammaticalization of the *prog* form will be effected if the initial frequency at which $\langle S_2, H_2 \rangle$ mutants are introduced exceeds 10% of the population. Figure 1 gives the change over time when initial conditions are set to $x_{\langle S_1, H_1 \rangle}$ = 0.89 and $x_{\langle S_2, H_2 \rangle}$ = 0.11.[38]

However, it is unrealistic to assume that the mutant strategy is introduced at such a high frequency. This would mean that more than 10% of the population simultaneously mutates to the context-sensitive $\langle S_2, H_2 \rangle$, while also independently innovating the same form to realize the PROG operator. It seems more realistic that the mutant strategy is initially introduced at a low frequency by spontaneous mutation. The transition to $\langle S_2, H_2 \rangle$ can then be attributed to the likelihood that $\langle S_2, H_2 \rangle$ is a target of learning for the offspring of the incumbent $\langle S_1, H_1 \rangle$ speakers. That is, it is possible that the offspring of $\langle S_1, H_1 \rangle$ speakers interpret local disambiguation efforts undertaken by $\langle S_1, H_1 \rangle$ speakers as coventionalized, thus innovating $\langle S_2, H_2 \rangle$ grammars. Conversely, it is possible that the offspring of $\langle S_2, H_2 \rangle$ speakers introduced by such mutation fail to interpret conventionalized strategies for disam-

[38] As a reminder, all figures shown here implement the discrete-time (rather than continuous-time) version of the replicator-mutator dynamics. The x axis gives time-steps while the y axis gives the frequency for each strategy.

Figure 1: $\langle S_1, H_1 \rangle$ to $\langle S_2, H_2 \rangle$: High rate for initial mutant value



biguation as such and revert to $\langle S_1, H_1 \rangle$. This can be concretely represented by the stochastic matrix Q' below.

(48) $Q' =$

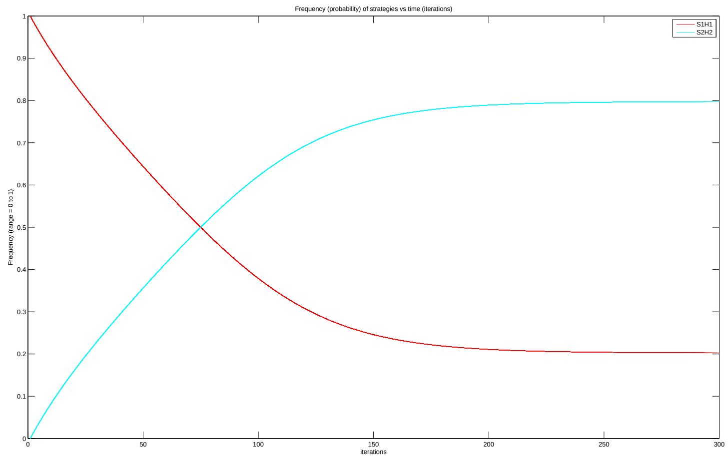
	$\langle S_1, H_1 \rangle$	$\langle S_2, H_2 \rangle$
$\langle S_1, H_1 \rangle$	0.99	0.01
$\langle S_2, H_2 \rangle$	0.01	0.99

Thus, we assume (only for expository purposes) that an $\langle S_1, H_1 \rangle$ parent is as likely to generate an $\langle S_2, H_2 \rangle$ offspring as an $\langle S_2, H_2 \rangle$ parent is to generate an $\langle S_1, H_1 \rangle$ offspring. With initial conditions set to $x_{\langle S_1, H_1 \rangle}(0) = 1$, Figure 2 gives the change in the proportion of $\langle S_1, H_1 \rangle$ and $\langle S_2, H_2 \rangle$ over time, where $\langle S_2, H_2 \rangle$ comes to dominate the population, stabilizing at $\sim$0.8, reducing the $\langle S_1, H_1 \rangle$ population to a $\sim$0.2 share.

5.3.2. Categoricalization

Categoricalization, on this model, amounts to the adoption of the context-independent explicit marking strategy $\langle S_3, H_3 \rangle$ by a large proportion of the population. In order

Figure 2: $\langle S_1, H_1 \rangle$ to $\langle S_2, H_2 \rangle$ with equal mutation rates



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 $\langle S_2, H_2 \rangle$ parent may generate an $\langle S_3, H_3 \rangle$ offspring, there will be a presence of $\langle S_3, H_3 \rangle$ players in any population mix that contains $\langle S_2, H_2 \rangle$ speakers. Moreover, the dynamics requires transitions from $\langle S_3, H_3 \rangle$ to $\langle S_4, H_1 \rangle$, which means that any population mix that contains $\langle S_3, H_3 \rangle$ players will also contain $\langle S_4, H_1 \rangle$ players in some proportion. Thus, the full game needs a modified matrix based on different assumptions that have to do with both the acquisitional properties of distinct strategies and their communicative efficiency. We will take Q'' to be the full row stochastic matrix, where transitional probabilities for $\langle S_2, H_2 \rangle$ differ from those in (48), which was only used to show how the dynamics work.

(49) $Q'' =$

	$\langle S_1, H_1 \rangle$	$\langle S_2, H_2 \rangle$	$\langle S_3, H_3 \rangle$	$\langle S_4, H_1 \rangle$
$\langle S_1, H_1 \rangle$	0.94	0.06	0	0
$\langle S_2, H_2 \rangle$	0.02	0.91	0.07	0
$\langle S_3, H_3 \rangle$	0	0	0.97	0.03
$\langle S_4, H_1 \rangle$	0	0	0	1

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

- While $\langle S_1, H_1 \rangle$ is a relatively simple grammar for a child to acquire, it leads to miscommunication in some proportion (0.1) of interactions. Thus, offspring of $\langle S_1, H_1 \rangle$ parents may infer $\langle S_2, H_2 \rangle$, a more communicatively successful grammar, from the structure of the input (which contains local disambiguating efforts) but there is no evidence to infer $\langle S_3, H_3 \rangle$ grammars in the same input.

- While $\langle S_2, H_2 \rangle$ is communicatively successful, it 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 ($\langle S_1, H_1 \rangle$) or context-invariant ($\langle S_3, H_3 \rangle$) system. This is reflected in higher mutation rates leading out of $\langle S_2, H_2 \rangle$. The mutating offspring of $\langle S_2, H_2 \rangle$ parents may go either way; they may misinterpret the input as generated by an $\langle S_1, H_1 \rangle$ grammar (given the low frequency of *prog*) or as generated by an $\langle S_3, H_3 \rangle$ grammar. But they are much more likely to do the latter than the former – given that the input provides evidence for a grammaticalized *prog* form.

- $\langle S_3, H_3 \rangle$ is both communicatively successful and easier to learn than $\langle S_2, H_2 \rangle$, requiring no context-sensitivity of either the speaker or the hearer. However, a few learners may interpret the input from $\langle S_3, H_3 \rangle$, which contains *prog* forms with higher frequency as evidence for an $\langle S_4, H_1 \rangle$ grammar – i.e. a grammar in which the *prog* form is used for both phenomenal and structural inquiries with contextual disambiguation.

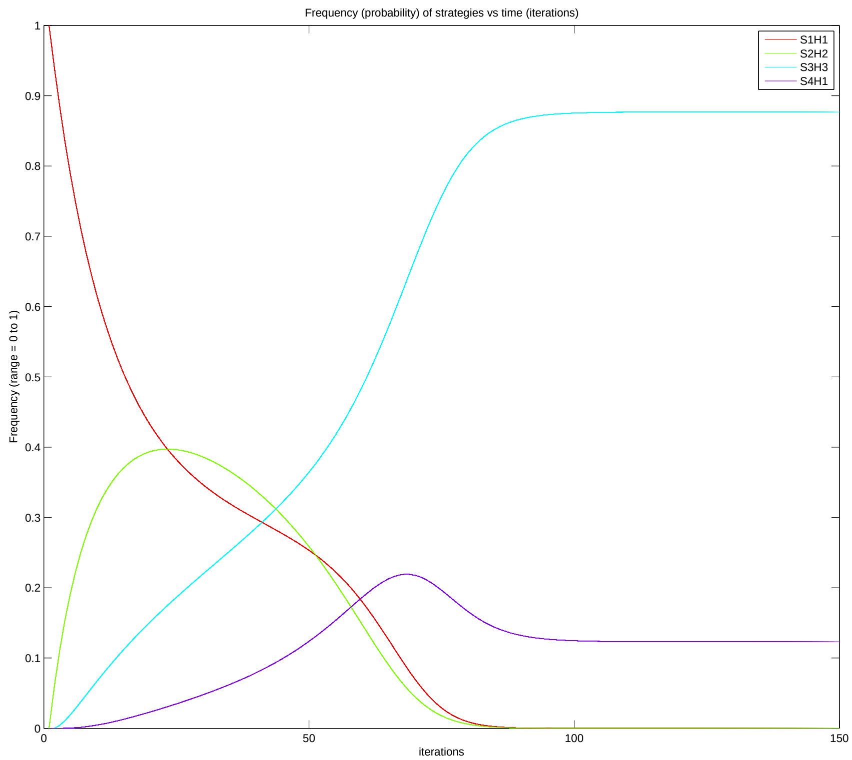
- Finally, there is no mutation out of $\langle S_4, H_1 \rangle$ into the other three strategies since an $\langle S_4, H_1 \rangle$ grammar does not provide any evidence for contrast with the older *impf* form. This will be further addressed in §5.3.3.

Q'' represents a preliminary proposal for quantizing the effect of the functional pressures of economy and expressiveness 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 $\langle S_1, H_1 \rangle = 1$. $\langle S_2, H_2 \rangle$, $\langle S_3, H_3 \rangle$, and $\langle S_4, H_1 \rangle$ are introduced by mutation. We see that the $\langle S_2, H_2 \rangle$ population advances over the $\langle S_1, H_1 \rangle$ population but is gradually taken over by a growing $\langle S_3, H_3 \rangle$ population, which stabilizes at $\sim$0.88, with elimination of $\langle S_2, H_2 \rangle$ and $\langle S_1, H_1 \rangle$ and the presence of $\langle S_4, H_1 \rangle$ at a low frequency. This can be seen as 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 to be dominant, will simply be a state in which other strategies are at a "low-enough" frequency. Before we move 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 $\langle S_1, H_1 \rangle$ system is always supplemented by some degree of $\langle S_2, H_2 \rangle$-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 $\langle S_2, H_2 \rangle$ presence in any $\langle S_1, H_1 \rangle$ popu-

Figure 3: Dynamic behavior of $\langle S_1, H_1\rangle$, $\langle S_2, H_2\rangle$, and $\langle S_3, H_3\rangle$ assuming Q''



lation.[39] This means that what has been called a zero-PROG state is really a state in which the proportion of $\langle S_2, H_2\rangle$ 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 $\langle S_1, H_1\rangle$+$\langle S_2, H_2\rangle$ speakers in any given state. We will (somewhat crudely) interpret the three apparent states in the following way:

(50) a. If $x_{\langle S_1,H_1\rangle} + x_{\langle S_2,H_2\rangle} > x_{\langle S_3,H_3\rangle}$
and $x_{\langle S_2,H_2\rangle} < \varepsilon$ zero-PROG

b. If $x_{\langle S_1,H_1\rangle} + x_{\langle S_2,H_2\rangle} > x_{\langle S_3,H_3\rangle}$
and $x_{\langle S_2,H_2\rangle} > \varepsilon$ emergent-PROG

[39] A further extension of the approach taken here could take the relevant strategies to be mixed strategies which differ only with respect to the probability at which a particular strategy is chosen. So while some players might choose $\langle S_1, H_1\rangle$ with 0.8 probability and $\langle S_2, H_2\rangle$ with 0.2 probability, other players might choose $\langle S_1, H_1\rangle$ with 0.6 probability and $\langle S_2, H_2\rangle$ with 0.4 probability. This probably approximates reality better in which markers for PROG gradually rise in frequency over time.

c. If $x_{\langle S_3,H_3\rangle} > x_{\langle S_1,H_1\rangle} + x_{\langle S_2,H_2\rangle} + x_{\langle S_4,H_1\rangle}$ categorical-PROG

For the purposes of this model, I will take ε to be 0.3. That is, if the proportion of $\langle S_2,H_2\rangle$ 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 $\langle S_3,H_3\rangle$ is the prevalent strategy. Mutations out of $\langle S_3,H_3\rangle$ into $\langle S_4,H_1\rangle$ 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 $\gg$ imperfective shift)?

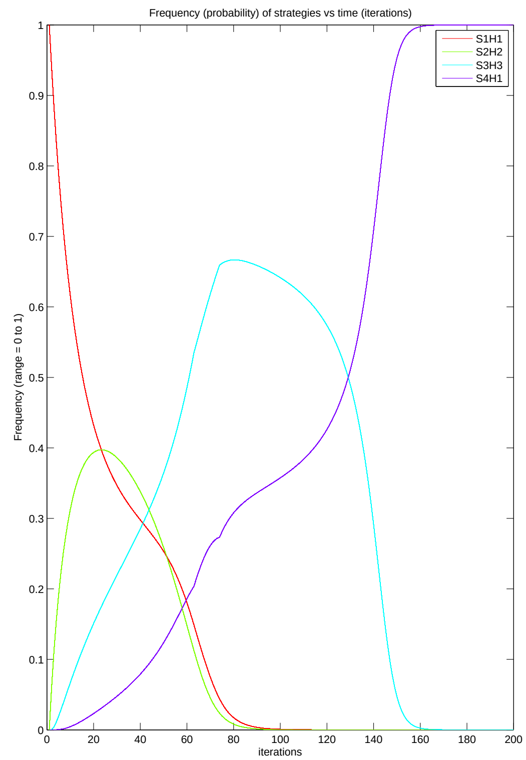
The game dynamics so far has assumed that mutations are constant (given by the mutation matrix Q'') and disconnected with the frequencies of particular strategies in a population. However, it is reasonable to assume that at least in some cases, the increase in the frequency of a strategy may be tied to an increase in the probability for mis-learning out of that strategy. In the case of $\langle S_3,H_3\rangle$, 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 $\langle S_3,H_3\rangle$ grammar and slight increase in the evidence for an $\langle S_4,H_1\rangle$ grammar – with *prog* as the new exponent of IMPF. Further, if language learners are biased towards single-form contextually reliant strategies (as suggested in §4.2.3), this should lead to increased mutations out of $\langle S_3,H_3\rangle$ into $\langle S_4,H_1\rangle$.

I will adopt this line of reasoning and assume that mutations out of $\langle S_3,H_3\rangle$ increase slightly when it reaches 0.5 frequency and slightly further when it reaches 0.65 frequency.[40] Concretely, when $x_{\langle S_3,H_3\rangle}$ exceeds 0.5, the mutation rate from $\langle S_3,H_3\rangle$ into $\langle S_4,H_1\rangle$ increases to 0.04 from 0.03. When $x_{\langle S_3,H_3\rangle}$ reaches 0.65, the mutation rate from $\langle S_3,H_3\rangle$ into $\langle S_4,H_1\rangle$ increases to 0.05. These slight increases in mutations lead to the gradual takeover of $\langle S_3,H_3\rangle$ by $\langle S_4,H_1\rangle$. This is generalization.

We can imagine that once $\langle S_3,H_3\rangle$ has declined and $\langle S_4,H_1\rangle$ is the prevalent strategy, it is fully equivalent to $\langle S_1,H_1\rangle$, since there is no alternative grammaticalized form that speakers can contrast *prog* with. A new form *prog′* may now to be innovated to realize PROG, setting into motion yet another cycle of the progressive $\gg$ imperfective path.

[40] 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: The takeover by $\langle S_4, H_1 \rangle$ with increased mutations



5.4. *Summary*

Typlogically, 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.[41] The proposed game has four evolutionarily

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

stable strategies, more than one of which may be 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 linkings 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 evolutionarily stable 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 information is less likely. Such a transition, given asymmetric payoffs and mutation probabilities, is followed by categoricalization, where the strategy in which there is no reliance on contextual recovery at all but 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 and 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 an oscillation 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:

(51)
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 (but crosslinguistically stable) 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–to–perfective/past path and the location–to–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 Weinrich, Herzog, and Labov (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.

Benz, Anton, Gerhard Jäger, and Robert Van Rooij, eds. 2006a. *Game theory and Pragmatics*. Palgrave Macmillan.

Benz, Anton, Gerhard Jäger, and Robert Van Rooij. 2006b. An introduction to game theory for linguists. In A. Benz, G. Jäger, and R. Van Rooij, eds., *Game theory and*

Pragmatics. Palgrave Macmillan.

Bittner, Maria. 2005. Future discourse in a tenseless language. *Journal of Semantics* 22:339–387.

Bohnemeyer, Jürgen. 2002. *The Grammar of Time Reference in Yukatek Maya*. Munich: Lincom Europa.

Bohnemeyer, Jürgen. 2009. Temporal anaphora in a tenseless language. In W. Klein and P. Li, eds., *Expression of Time in Language*, pages 83–128. Berlin: Mouton de Gruyter.

Bonomi, Andrea. 1997. Aspect, quantification and when-clauses in Italian. *Linguistics and Philosophy* 20(5):469–514.

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

Bybee, Joan, Revere Perkins, and William Pagliuca. 1994. *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World*. Chicago: The University of Chicago.

Cardona, George. 1965. *A Gujarati Reference Grammar*. Philadelphia: The University of Pennsylvania Press.

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

Cipria, Alicia and Craige Roberts. 2000. Spanish Imperfecto and Pretérito: Truth conditions and aktionsart effects in a situation semantics. *Natural Language Semantics* 8(4):297–347.

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

Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press.

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 J. Bybee and P. Hopper, eds., *Frequency and the emergence of linguistic structure)*, vol. 2, pages 471–80. Amsterdam: John Benjamins.

Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Basil Blackwell.

de Swart, Henriette. 1998. Aspect Shift and Coercion. *Natural Language and Linguistic Theory* 16:347–385.

Delfitto, Denis and Pier Marco Bertinetto. 1995. A case study in the interaction of aspect and actionality: The Imperfect in Italian. *Temporal reference, aspect and actionality* 1:125–142.

Deo, Ashwini. 2006. *Tense and Aspect in Indo-Aryan Languages: Variation and Diachrony*. Ph.D. thesis, Stanford University.

Deo, Ashwini. 2009. Unifying the imperfective and the progressive: partitions as quanticational domains. *Linguistics & Philosophy* 32(5):475–521.

Deo, Ashwini. 2014. Formal semantics/pragmatics and language change. In C. Bowern and B. Evans, eds., *The Routledge Handbook of Historical Linguistics*, pages 393–409. Oxford: Routledge.

Deo, Ashwini. 2015. Grammaticalization paths and diachronic semantics. *Annual Review*

of Linguistics 1:000–000.
Deo, Ashwini. To appear. The Imperfective. In L. Matthewson, C. Meier, H. Rullman, and T. E. Zimmerman, eds., *Companion to Semantics*. Oxford: Wiley-Blackwell.
Dowty, David. 1977. Toward a semantic analysis of verb aspect and the English 'imperfective' progressive. *Linguistics and Philosophy* 1(1):45–77.
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: Oxford University Press.
Ferreira, Marcelo. 2005. *Event quantification and plurality*. Ph.D. thesis, MIT.
Göksel, Asli and Celia Kerslake. 2005. *Turkish: A Comprehensive Grammar*. Routledge.
Goldsmith, John and Erich Woisetschleger. 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 and Karl Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.
Hofbauer, Joseph and Karl Sigmund. 2003. Evolutionary game dynamics. *Bulletin of the American Mathematical Society* 40(4):479–519.
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*, pages 11–42. Washington: Georgetown University Press.
Jäger, Gerhard. 2004. Evolutionary game theory for linguists. a primer. Ms. University of Tübingen.
Jäger, Gerhard. 2007. Evolutionary game theory and typology: A case study. *Language* 83(1):74–109.
Kellogg, S. H. 1893. *Grammar of the Hindi Language*. Delhi: Motilal Banarsidass. Second edition, reprinted 1990.
Kiparsky, Paul and Cleo Condoravdi. 2006. Tracking Jespersen's cycle. In *Proceedings of the 2nd international conference of Modern Greek dialects and linguistic theory*, pages 172–197.
Kornfilt, Jaklin. 1997. *Turkish*. London and New York: Routledge.
Kuryłowicz, Jerzy. 1964. *The inflectional categories of Indo-European*. Carl Winter, Universitatsverlag.
Landman, Fred. 1992. The progressive. *Natural Language Semantics* 1(1):1–32.
Leech, Geoffrey. 1970. *Towards a semantic description of English*. Indiana: Indiana University Press.
Lenci, Alessandro and Pier Marco Bertinetto. 2000. Aspect, adverbs, and events: habituality vs. perfectivity. In J. Higginbotham, F. Pianesi, and A. Varzi, eds., *Speaking of Events*. Oxford University Press, USA.
Lewis, David. 1969. *Convention*. Cambridge, MA: Harvard University press.
Li, Ping, Sarah Maher, Erica Newmark, and 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: Clarendon Press.

Maynard Smith, J. and G. R. Price. 1973. The Logic of Animal Conflict. *Nature* 146(4):15–18.

Nowak, Martin A. 2006. *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge MA: Harvard University Press.

Nowak, Martin A, Natalia L Komarova, and Partha Niyogi. 2001. Evolution of universal grammar. *Science* 291(5501):114–118.

Page, Karen and Martin Nowak. 2002. Unifying evolutionary dynamics. *Journal of Theoretical Biology* 219:93–98.

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

Portner, Paul. 1998. The progressive in modal semantics. *Language* 74(4):760–787.

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

Rubinstein, Ariel. 2000. *Economics and Language: Five Essays*. New York: Cambridge University Press.

Shirai, Yasuhiro. 1994. On the overgeneralization of progressive marking on stative verbs: bioprogram or input? *First Language* 14(40):67–82.

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

Tagliamonte, Sali and 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.

Taylor, Peter and Leo B. Jonker. 1978. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences* 40:145–156.

Tonhauser, Judith. 2011. Temporal reference in Paraguayan Guaraní, a tenseless language. *Linguistics and Philosophy* 34.3:1–47.

Traugott, Elizabeth. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65.1:31–55.

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

van Rooij, Robert. 2004a. Evolution of conventional meaning and conversational principles. *Synthese* 139(2):331–366.

van Rooij, Robert. 2004b. Signalling games select Horn strategies. *Linguistics and Philosophy* 27(4):493–527.

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

Weinreich, Uriel, William Labov, and M.I. Herzog. 1968. *Empirical foundations for a theory of language change*. University of Texas Press.

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