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

Abstract: This paper offers an analysis of a robustly attested semantic change in which progressive markers "spontaneously" emerge in languages, get entrenched in the grammatical system, and diachronically grammaticalize into imperfective markers. The facts can be schematically described as follows: At Stage 0, a linguistic system *L* possesses a single imperfective or neutral aspectual marker *X* that is used to express two contextually disambiguable meanings α and β. At Stage 1, a progressive marker *Y* arises spontaneously in *L* in order to express α in some contexts. At Stage 2, *Y* becomes entrenched as an obligatory grammatical element for expressing α while *X* is restricted in use to expressing β. At Stage 3, *Y* generalizes and is used to express both α and β. *X* is gradually driven out of *L*. Stage 3 (structurally identical to Stage 0) is often followed by another instantiation of Stage 1, with the innovation of a new progressive marker *Z*. The trajectory to be explained is thus cyclic.

The analysis has a semantic component that characterizes the logical relation between the progressive and imperfective operators in terms of asymmetric entailment. 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 innovation and entrenchment of progressive marking in languages is shown to be underpinned by optimal ways of resolving both kinds of inquiries in discourse given considerations of successful and economic communication. Generalization is analyzed as the result of imperfect learning. 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.

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

[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.

be used to describe both events in progress and principled generalizations, 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 $\gg$ 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 three general questions that emerge from the kinds of empirical observations in (2).

(3) a. Why do the meanings of functional expressions change over time?
b. How do the meanings of functional expressions change over time?
c. Why do we see cross-linguistic similarities in patterns of semantic change?

These questions constitute the semantic counterpart to the question of sound change in phonology, where diachronic and synchronic patterns have traditionally been viewed together as two facets of the same explanandum. Given robustly attested sound changes such as the palatalization of consonants before front vowels or the nasalization of vowels before nasal consonants, we can ask:

(4) a. Why does sound change happen?
b. How does sound change happen?
c. Why do we see cross-linguistic similarities in patterns of sound change?

These questions have been considered integral to the study of sound systems at least since Paul (1888). Practically every theory locates the seed of systematic sound change in the variation in the acoustic realization of phonological segments introduced by coarticulation (Ohala 1981, 1993; Blevins 2004; Yu 2010, Beddor 2009,

Bermúdez-Otero 2012). The basic account is as follows: Coarticulation introduces context-dependent phonetic variants which vary systematically in a particular direction relative to the target. Although listeners typically tend to compensate for coarticulatory effects, they may sometimes attribute such effects to the grammatical component, leading to changed perceptual and production patterns. Thus, what starts off originally as a non-grammatical phonetic effect arising from the mechanics of articulation, becomes "grammaticalized" as part of the listeners' grammar in the form of gradient phonetic implementation rules (phonologization), and later as categorical phonological rules. Providing explicit, experimentally-grounded theories about the articulatory, acoustic, and cognitive underpinnings of sound change is part of the general enterprise of phonology, which concerns itself with *both* synchronic and diachronic sound patterns.

My view is that it is rather fruitful for semantics to also take a similar perspective on its domain of explanation. In general, if synchronic grammatical systems and changes to these systems are constrained by the same principles, then any theory about synchronic patterns carries implications for diachrony. The *constraints* problem (Weinrich, Herzog, and Labov 1968) for semantics is the problem of discovering the precise content of grammatically relevant elements of meaning and the constraints on their interaction. The *actuation* and the *transition* problems for this domain have to do with understanding the cognitive and communicative conditions that give rise to new grammatical elements or changes in existing ones and the propagation of these changes. Grammaticalization paths that involve systematic shifts between functional meanings or from lexical to functional meanings, raise at least the following questions that are relevant to semantic theory:

(5)
a. What leads to the emergence of new functional expressions at the left edge of a path and their generalization towards the right edge?
b. What is the semantic content of the functional expressions that constitute the input to or the output of a grammaticalization path?
c. What logical relation in the meanings of these expressions allows for a coherent account of the observed changes?
d. What sort of theories can account not only for the distribution of functional expressions at a single synchronic stage, but also for systematic changes in their distribution across stages?

Answering these questions involves integrating the empirical insights and generalizations coming from grammaticalization studies with the results of formal semantics and pragmatics. However, it is not fully transparent how such an integration might be implemented. Grammaticalization paths are often construed as being the result of *semantic bleaching* or *generalization*. The idea is that lexical material gradually loses its meaning as it gets recruited in creating a new functional category, and once it has become functional, it further bleaches and acquires a broader functional meaning. Such a characterization seems intuitive at some level of description. But it is not trivial to provide the right semantic content for such a construal of the diachronic facts, while maintaining at the same time an analysis with

adequate coverage of the synchronic facts. Moreover, neither bleaching nor generalization are construable as the *mechanisms* that effect language change. These terms can be understood as static descriptions of the relation between the meanings of an expression before and after the change. But they do not offer insight into the question of how and why the change occurs in the first place.

Any adequate explanation for grammaticalization paths involving functional categories must therefore contain two components – a static, structural one and a dynamic one. Going back to the parallel from phonology, an explanation of a synchronic pattern such as nasalization of vowels in pre-nasal context is based in the mechanics of articulation and its acoustic correlates on the one hand (the structural component) and the mechanism(s) by which these acoustic effects come to be interpreted differently and propagated through a speech community (the dynamic component) on the other. In accounting for semantic change, we will draw on semantic theory on the one hand to explicate the structural component – giving a precise characterization of the logical relation (i.e. the similarities and differences) between the meanings of related functional categories. The dynamic component, on the other hand, will draw from theories of language use and language 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. 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 findings that constitute the evidence for the progressive–imperfective connection, and in particular, the diachronic path 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 a privative opposition or a relation of asymmetric entailment. §4 introduces the idea that imperfective and progressive marking allows for a distinction between phenomenal and structural inquiries and considers the dynamic aspects of the grammaticalization path. §4.2 describes the functional pressures that lead to the emergence of progressive marking as a variant in an aspectual system that lacks it while §4.3 and §4.4 focus on why the innovation of such marking would lead, at a later stage, to a conventional system of contrasts in which the progressive is obligatorily used for phenomenal inquiries while the imperfective is reserved for structural inquiries. The ideas from §4 are formalized and further developed in §5. This section offers an evolutionary game-theoretic analysis of the progressive $\gg$ imperfective shift. The **Imperfective Game** described in this section shows how a set of assumptions about imperfective meanings, acquisition biases, and cognitive differences 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 wide range of semantic domains in which trajectorial changes have been empirically observed.

2. The progressive–imperfective connection

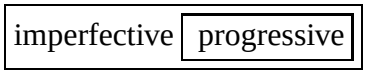
Crosslinguistically, imperfective marking is associated with at least three distinct readings: (a) the progressive or *event-in-progress* reading; (b) the habitual or generic *characterizing* reading; and (c) the *continuous* reading with lexically stative predicates. The three readings are illustrated in (6) with examples from Gujarati, an Indo-Aryan language with imperfective marking.[2]

(6) a. niś ā (aty ā re) rasoḍ ā-m ā roṭli **banāv-e**
N.NOM.SG now kitchen-LOC bread.NOM.SG make-IMPF.3.SG
ch-e
PRES-3SG
Niś ā *is making* bread in the kitchen (right now). *event-in-progress*

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

c. niś ā navs ā ri-m ā **rah-e** **ch-e**
N.NOM.SG Navsari-LOC live-IMPF.3SG PRES-3.SG
Niś ā *lives* in Navsari. *continuous*

Progressive marking (e.g. the English Progressive), on the other hand, saliently exhibits only the event-in-progress reading. At least since Comrie's (1976) classic text on aspect (also see Kuryłowicz 1964), the progressive has been treated as a subcategory of the more general imperfective, with a narrower, more specific meaning that can be subsumed under the broader, more general meaning of the imperfective.

(7) 

Two kinds of morphological evidence support this conception of a semantic organization of aspectual categories in which the progressive is a specific version of the general imperfective category – blocking facts and grammaticalizing changes.

2.1. Semantic blocking

In languages in which both the progressive and imperfective aspects are realized with distinct morphology, the event-in-progress reading is often blocked for the imperfective form. For instance, Hindi, which morphologically realizes both the imperfective and the progressive, exhibits a restriction on the distribution of imper-

[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.

fective marking in both the present and the past tenses (Deo, 2009). The imperfective form (9-a) is incompatible with the event-in-progress reading which is uniquely expressed by the progressive (9-b).

(8) a. niś ā roṭi **banā-ti** **hai**
N.NOM bread.NOM make-IMPF.F.SG PRES-3SG
Niś ā *makes/*is making* bread *characterizing*

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

These blocking facts hold for contemporary Hindi. In the 19th 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 alternated with the newly evolving progressive form in describing events in progress. The blocking relation between the specific progressive and the general imperfective thus develops along a diachronic dimension and 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 of the Simple Present in alternation with the Progressive continues in the Early Modern English of Shakespeare as seen in (9). In Modern English, however, the event-in-progress reading is typically unavailable to the Simple Present.

(9) 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 kind of morphological 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.

[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.

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 (10). The examples are from Göksel and Kerslake (2005:331). In (10-a), the verb form with *-(I)yor* describes an ongoing working eventuality, while in (10-b), the *-ir* inflected verb describes a characteristic pattern of working – a characterizing reading.

(10) 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 (11-a)), and is also interchangeably used with the Aorist form in characterizing contexts (11-b). The examples are from Göksel and Kerslake (2005:333).

(11) 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 be-

tween 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 (12)-(13) 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.

(12) 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.

(13) a. ḥəna həd ̄ay **nətfarrar hallena**
we wedding go out-IMPF PRES.1PL
We *are going out* to the wedding.

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

However, this periphrastic construction also exhibits characterizing readings as shown in (14-a-b) with no semantic distinction from the bare Imperfect.

(14) 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.

[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.

2.3. Implications for aspectual meaning

In addition to these cross-linguistic snapshots of the 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 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 functional 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 change (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

[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.

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:

- 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 question-based model of discourse structure which forms the basis of the explanation of 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

[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).

relation $\subseteq$. $\mathcal{W}$ is a non-empty set of worlds. $i, j, k...$ are variables over $\mathcal{I}$. 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 $i \in \mathcal{I}$ a proper subset of $Hist_i(w)$ — the set of those worlds that continue beyond i in ways that are compatible with the normal course of events until i. (Dowty, 1979: 152).[8] $Hist_{i_{inr}}(w)$ is the set of **inertial alternatives** of w at i.

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

$\mathcal{E}$ is a domain of eventualities, sorted into a set of events $\mathcal{E}^E$ and a set of states $\mathcal{E}^S$. The temporal trace function τ from $\mathcal{E}$ to $\mathcal{I}$ gives the run time of an eventuality. The eventuality argument of basic eventive predicates is of the sort E while the eventuality argument of a basic stative predicate is of the sort S. *Sentence radicals* are predicates of eventualities (eventive or stative) built from such basic predicates with their individual (non-eventuality) arguments saturated (somewhat corresponding to the VP level assuming VP-internal subjects). 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 (16). A predicate of eventualities P stands in the coincidence relation with i and w iff P is instantiated in every inertial alternative of w within i or at some superinterval of i. A predicate of intervals P stands in the coincidence relation with i and w iff P holds at i in w.

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

The final notion needed in specifying the meaning of the progressive and imperfective operators is that of a *regular partition*, defined in (17). For any interval

[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.

i, a partition of *i* is the set of the non-empty, mutually exclusive, and collectively exhaustive subsets of *i*.

(17) **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 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]

(18) 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 the sake of clarity, (20) 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

[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.

[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.

Gujarati Imperfective (examples in (6)) realizes IMPF as given in (19).[11]

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

The logical form for (20-a) is in (20-d). That is, (20-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 Niś ā making bread, which event is fully realized in the inertial alternatives of *w* at *k*.

(20) a. niś ā (roj) roṭli **banāv-e** **ch-e**
N.NOM everyday bread.NOM make-IMPF.3.SG PRES-3.SG
Niś ā *makes* bread (everyday).

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

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

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

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.

(21) 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 also account for the inference of temporal contingence associated with progressive marking as seen in the examples in (22).

Specifically, progressive marking is compatible with a characterizing reading as in (22-b/d), 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.

(22) 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*

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

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 (23).

(23) 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 contrast between the the meaning of the progressive and imperfective assumed here gives rise to the inference of temporal contingence and the related effect of (in)felicity.

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 is either in the set (24-a) or (24-b).[12]

(24) 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 i, P coincides with k in w.

In contrast, going by the meaning in (23), $\textsc{prog}(P)(i)$ denotes the set in (24-b). Thus, for any world w, an assertion of the form $\textsc{prog}(P)(i)(w)$ is semantically stronger than $\textsc{impf}(P)(i)(w)$. It is easy to see that PROG is a "semantically narrower" version of IMPF on this construal of their contribution, since $\textsc{prog}(P)(i)$ asymmetrically entails $\textsc{impf}(P)(i)$. It is this relation of asymmetric entailment that is at the heart of the felicitous uses of progressive and imperfective marking. The

[12] That is, $[\![\textsc{impf}(P)(i)]\!] = \{w \mid \exists j[i \subset_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \textsc{coin}(P,k,w)]]\} \cup \{w \mid \exists j[i = j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \textsc{coin}(P,k,w)]]\}$

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 *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*.

Treating PROG and IMPF as elements of a Horn scale is useful in making sense of the pragmatic effects arising from this asymmetric entailment relation. The particular kind of Horn scale instantiated by PROG and IMPF is what Horn and Abbott (2012) call a **privative dyad** of the form ⟨S,W⟩ in which S (the stronger scalar value) is marked for a feature with respect to which W (the weaker value) is unmarked, making S more informative. The use of W, the weaker value, pragmatically implicates that the speaker is not in a position to convey S.[13] Further, the use of S, the stronger value, in contexts in which the weaker value W is the expected one, gives rise to an exhaustive interpretation.

Assume that IMPF and PROG are members of a privative dyad ⟨PROG, IMPF⟩.[14] Then, in a language that has fully grammaticalized expressive devices for IMPF and PROG, the interpretation of $\text{IMPF}(P)(i)$, the weaker member, will be restricted to (24-a) in most contexts, while the use of $\text{PROG}(P)(i)$, in some contexts, will give rise to the exhaustive implicature in (25), that serves to exclude the worlds in (24-a).

(25) a. $\lambda w\, \neg\exists j[i \subset j \wedge \forall k[k \in \mathscr{R}_j^c \rightarrow \text{COIN}(P,k,w)]]$

b. The set of worlds *w* such that there is no *j* that is a proper superinterval of *i* such that for every *k* in a small-enough regular partition of *j*, *P* coincides with *k* in *w*.

This can be illustrated using English if we make the assumption that the English present tense sentences contain IMPF in their logical form, intervening between the sentence radical and tense. Consider the stative predicate *New Orleans lie at the mouth of the Mississippi river*, which we abbreviate as *N*. The alternatives available for expressing the corresponding tensed proposition are either the simple present (26-a) or the present progressive (26-b). The choice of the simple present, the weaker value, yields the strengthened proposition in (26-a-ii) by scalar implicature, conveying that the proposition is a temporally stable fact and not restricted only to the reference interval. The choice of the present progressive, on the other

[13] Dowty (1980) posits a Gricean blocking principle that conveys almost exactly the same idea: "If a language has two (equally simple) types of syntactic structures A and B, such that A is ambiguous between meanings X and Y while B has only meaning X, speakers of the language should reserve structure A for communicating meaning Y (since B would have been available for communicating X unambiguously and would have been chosen if X is what was intended." (Dowty 1980: pp 32).

[14] An assumption often made about the ordered scales of alternatives that underlie quantity implicatures is that they are equally lexicalized or syntactically complex. This condition is clearly not met in the⟨PROG, IMPF⟩ case. It is a crosslinguistically attested fact that progressive marking tends to be syntactically more complex than imperfective marking (Bybee et al 1994, Dahl 1985). The syntactic complexity of the stronger member of the dyad serves to further strengthen the exhaustive interpretation associated with PROG.

hand, leads to an exhaustive interpretation, which reduces the set of PROG worlds in (26-b-i) to the set in (26-b-ii), in which the extension of N is minimal and does not extend beyond the reference interval **now**.

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

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

The exhaustivity implicature associated with the choice of (26-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 PROG alternative infelicitous in most contexts. For instance, (26-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.[15]

(27) a. A: And what about the cities on the lower Mississippi? How have they fared?
b. B: Well, New Orleans is (still) lying at the mouth of the Mississippi river, but we don't know for how much longer.

Thus, 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, specifically, a privative dyad. Sentences containing the weaker member, IMPF, are pragmatically restricted to a non-PROG interpretation, while sentences containing PROG may be pragmatically exhaustified in certain contexts to eliminate all stronger interpretations, 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. I presented the analysis from Deo (2009) that analyzes both operators as universal quantifiers over partitions and reduces the difference between the two to a scalar contrast that shows the familiar behavioral properties of such contrasts – blocking and exhaustification. Treating PROG and IMPF as scalar alternatives accounts for the difference in available readings between imperfective

[15] 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.

and progressive marking in languages that distinguish between the two aspectual categories morphosyntactically.

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 PROG≫IMPF grammaticalization process. This semantic characterization is the underlying structural basis of 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 PROG≫IMPF 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 in simple words, the WHY and the HOW questions. These are considered in §4 and §5 respectively.

4. The dynamic component

We can discern four distinct states in the grammaticalization path associated with the PROG and IMPF operators. There is the initial **zero**-PROG state (28-a), in which the language possesses only a single grammaticalized device across the imperfective domain – the exponent of IMPF. The second, **emergent**-PROG state (28-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 comparatively low. The third state can be called the **categorical**-PROG state (28-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 (28-d), is the state in which the exponent of PROG loses its semantic restriction and generalizes to IMPF.[16]

[16] As far as speakers are concerned, (28-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

(28) a. X_{impf} zero-PROG
b. $(Y_{prog})\ X_{impf}$ emergent-PROG
c. $Y_{prog} \ll X_{impf}$ (blocking) 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:

(29) a. The **recruitment** of existing morpho-syntactic resources to innovate a new functional category/expression within a semantic domain – the change from (28-a) to (28-b)
b. The **categoricalization** of the new expression to obligatory use in certain contexts within the domain – the change from (28-b) to (28-c)
c. The **generalization** of this expression to the broader semantic domain – the change from (28-c) to (28-d)

The analysis of these dynamic processes proposed here builds on results from formal pragmatics and game-theoretic pragmatics that seek to model the interaction between semantic content and strategic considerations in rational communication. 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, 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 $\gg$ imperfective grammaticalization path can, on this view, be reconstructed 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 full analysis is spread over two sections. This section, §4, focuses on the systematic difference between the ways in which language speakers use progressive and non-progressive morphology. §4.1.1 proposes that speakers must distinguish between two types of inquiries – structural and phenomenal ones – and seek to disambiguate between these in discourse. In a system which lacks progressive marking, disambiguation between phenomenal and structural inquiries must rely on the common knowledge of interlocutors in the context of utterance, which sometimes leads to imperfect communication (§4.1.2). In contrast, in a language which has grammaticalized progressive marking, the disambiguation is formally explicit and allows for perfect communication. **Recruitment**, or the transition from the zero-PROG to the emergent-PROG state, is understood as the introduction of a

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.

conventional strategy for disambiguating phenomenal inquiries (§4.2). In **Categoricalization**, or the transition from the emergent-PROG to the categorical-PROG state, the semantically specific progressive marking becomes obligatory for marking phenomenal inquiries while the diachronically older and semantically more general imperfective marking becomes conventionally restricted to non-phenomenal or structural inquiries (§4.3). **Generalization** which is the transition from categorical-PROG back into the zero-PROG state, since it is not driven by considerations of effective communication, but rather by imperfect learning of the categorical-PROG system, is not discussed here, but in the next section, §5, which models all three transitions from the evolutionary game-theoretic perspective.

4.1. Resolving questions in context

We draw here from very general ideas about the structure of discourse as a sequence of questions and answers (Groenendijk and Stokhof 1984; Ginzburg 1995; Roberts 1998 a.o.).[17] We begin with the notion that discourse is a communal inquiry into the state of our world – a pursuit to discover the way things are (Stalnaker 1978, 1998). This over-arching goal is pursued by interlocutors only indirectly; via specific goal-oriented sub-inquiries in smaller domains of interest – subserving what Roberts (1998) calls our real world domain goals. These inquiries are guided by a partially ordered set of questions under discussion (QUD), which serve to determine what is relevant at any point in the discourse. Assertions are taken to be answers that are congruent to some such (often implicit) question. The effect of successful assertions is an update in the mutually held beliefs of discourse participants regarding the state of the world.

Dynamic models of discourse formalize these ideas within a possible-worlds framework in which the evolution of discourse leads to a progressive refinement of the set of worlds that the actual world is believed to lie in. This is modeled using the notions of the Common Ground and the Context Set. The common ground is the set of propositions taken for granted by discourse participants at any given point in time. The context set is the intersection of the propositions in the common ground, i.e. the set of those worlds that are compatible with what is taken for granted. A question (explicit or implicit) is a partition on the context set; i.e. a set of mutually exclusive alternative possibilities, one of which is realized by the actual world. Any assertion *p*, which constitutes an answer to the question under discussion, is a choice among such sets of alternatives presented by the question. *p* contributes a proposition, which, if accepted, is added to the common ground. This eliminates worlds incompatible with *p* from the context set, leaving the set of worlds compatible with *p* as the refined, post-assertion context set. An assertion that is a complete answer contextually entails an evaluation for every alternative in the QUD. That is, it selects a single cell of the QUD discarding all others, allowing for a complete resolution of the QUD. A partial answer is incompatible with one or more alternatives

[17] This proposal is also inspired by ideas about the rise of NPIs schematically presented in Condoravdi (2009).

in the QUD, but is compatible with more than one logically independent alternative (Beaver and Clark 2008).

4.1.1. Imperfective assertions and their QUDs

What sorts of questions can imperfective assertive sentences be used to answer? Alternatively, what sort of knowledge about the world do we obtain from imperfective assertions? 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 (30-a) and(30-b) illustrate the difference between inquiries that I have in mind.

(30) 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 t?*
   (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?

It is not an accident that the questions in (30-a) are expressed using the Simple Present tense while those in (30-b) use the Progressive. This is a function of the fact that English morphosyntactically realizes PROG thus allowing for a formal distinction to be made between the two kinds of inquiries within the domain of ongoing situations. Following similar intuitions as in Goldsmith and Woisetschlaeger (1982), I will call the first kind of inquiry **structural** and the second kind of inquiry **phenomenal**. Goldsmith and Woisetschlaeger 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. In their words: "one may describe the world in either of two ways: by describing what things happen in the world, or by describing how the world is made that such things may happen in it. (p. 80). It is this metaphysical distinction that they claim

underlies the distribution of the English Progressive and the Simple Present.[18]

While it seems to me that the structural/phenomenal contrast is cognitively basic, unlike them, I do not believe that languages directly lexicalize this cognitive contrast through aspectual morphology. Rather, in my view, this cognitive contrast shapes the emergence and evolution of the privative linguistic contrast between progressive and non-progressive expressions.[19] Specifically, progressive marking arises and grammaticalizes in languages because of the functional pressure to unambiguously resolve phenomenal inquiries.[20] Here is an explication of the idea.

Consider a zero-PROG language, one that only contains marking corresponding to the broader IMPF and has no grammaticalized device to convey the narrower meaning corresponding to PROG.[21] In such a language, imperfective sentences are

[18]Although the structural/phenomenal contrast has not been as well-discussed in the literature as it should have been, there is a case to be made for something similar to this contrast to underlie the stage-level/individual-level distinction and its various grammatical effects in languages. At the very least, the two categories show significant overlap. For the purposes of this paper, however, I will simply rely on this classification to distinguish between inquiries in discourse.

[19]In fact, it makes more sense to distinguish between phenomenal and non-phenomenal inquiries, where the non-phenomenal subsumes, but is not co-extensive with the structural. The generalizations expressed by non-progressive marking range from purely inductive observations (*Barns are red*, *John drives recklessly*) to those rooted within causal structures (*Americans consume vast amounts of energy*, *Hindus avoid beef*) to those that come about as a result of explicitly stipulative rule-making (*Bishops move diagonally*) or from the laws of nature (*Salt dissolves in water*). Although these cannot all be called structural generalizations in the way described by Goldsmith & Woisetschlaeger (1982), they are non-phenomenal in that they do not pertain to facts about specific times and events. Carlson (1995), based on the range of interpretations in generic sentences, draws a distinction between an inductivist approach and a rules-and-regulations based approach to generics. Cohen (2001) proposes instead that both approaches are necessary in analyzing generics and the distinction is rather between inductivist and rule-and-regulations based *readings* of characterizing sentences. While a large part of non-phenomenal knowledge can be said to be structural in nature, there is no linguistic device that is solely used to convey structural information. It therefore seems more appropriate to decouple the cognitive contrast between the phenomenal and the structural from the linguistic contrast between progressive and non-progressive aspect. The linguistic contrast is privative in nature and is employed to distinguish between phenomenal and non-phenomenal inquiries. Given that no harm is done if we keep the privative nature of the opposition in mind, I will keep referring to non-phenomenal inquiries as structural.

[20]A reviewer asks whether there is no similar pressure to innovate a designated marker for expressing characterizing meanings in a zero-PROG language. There is nothing that would prevent novel material from being recruited to express characterizing meanings. In fact, habitual/generic marking, such as the *used to V* construction in Modern English would be an example of such a dedicated device. 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). Thus, at this stage, the language, in addition to the simple tenses, disambiguates both phenomenal and structural inquiries with new material. The fact is that the *uses to V* construction falls out of the language and the Simple Present tense gets recruited for 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. Crucially, this trajectory is unidirectional: there has been no attestation of habitual/generic markers changing in meaning to encompass the function of the progressive.

[21]The zero-PROG description subsumes languages like Gujarati which have morphologically overt imperfective marking and languages like Middle English in which tense marking is aspectually

used in raising and resolving both structural and phenomenal questions about the state of the world – the nature of inquiry associated with a given utterance can only be disambiguated in context. For instance, a question like (31-a) could be interpreted, in the absence of contextual information, as either a question about Hamlet's reading habits (a structural inquiry) or a question about what Hamlet happens to be reading at that point (a phenomenal inquiry).[22] A response such as (31-b) uttered by Hamlet could be similarly interpreted as being about what Hamlet reads regularly (a structural claim), or as a claim about what he is currently engaged in doing (a phenomenal claim). On either interpretation, the semantic content of the assertion remains the same; the proposition in (32-a). Taking the existentially quantified j to be **now**, one obtains the phenomenal reading; taking it to be a proper superinterval of **now**, one obtains the structural reading. The meaning of the explicit question in(31-a), given in (32-b) as a set of propositions, remains similarly ambiguous.

(31) a. *What do you read, my Lord?* Polonius' question to Hamlet (Hamlet II, ii)
b. *I read words.*[23]

(32) a. $\lambda w.\, \exists j[\mathbf{now} \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \text{COIN}(\lambda e.\text{Hamlet-read-words}(e), k, w)]]$

b. $\{\lambda w.\, \exists j[\mathbf{now} \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \text{COIN}(\lambda e.\text{Hamlet-read-X}(e), k, w)]] \mid X \in \lambda x.\text{readable-objects}(x)\}$

4.1.2. Polonius' problem

Now consider a scenario somewhere in the Late Middle English times with Ophelia and Polonius as discourse participants.[24] In the context, it is common ground that Hamlet is reading a book at reference time. What is not commonly known is whether he is reading *Beowulf* or *The Canterbury Tales*. Polonius wishes to find out whether Hamlet is engaged in reading *Beowulf* at the reference interval **now** – that is, the (implicit) QUD is part of a phenomenal inquiry about what is happening

neutral and simple tensed sentences exhibit imperfective readings.

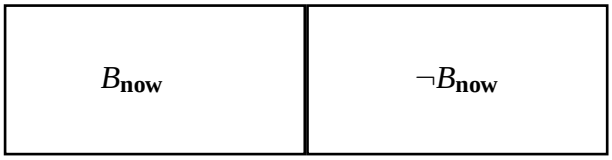
[22] In the context where it is used, it has the phenomenal interpretation. This context-dependent (rather than morphosyntactic) resolution of structural and phenomenal interpretations is a fact about Early Modern English arising from the under-specified semantics of the simple tenses. It is much later that the Progressive starts to increase in frequency and becomes an obligatory grammatical device in phenomenal inquiry contexts (Strang 1982; Denison 1993).

[23] Hamlet's actual response is the elliptical *Words, words, words.*

[24] I am supposing, for the purposes of this exposition, that both Ophelia and Polonius are fictitious speakers of Middle English living in Pre-Shakespearian times, probably a good couple of centuries before Shakespeare. That is, I want to describe the properties of and effects arising in a grammatical system which contains no distinguished exponent for PROG. Although the Progressive is relatively infrequent in the Early Modern English of Shakespeare, it is still available as a grammaticalized device. The point of this Ophelia-Polonius example is simply to illustrate the systematic discrepancy between intended questions/assertions and the existing devices for their expression that leads to the innovation and spread of a new morphosyntactic marker in the grammar of the language.

now. Let us name the proposition that Hamlet is reading Beowulf $B_{\textbf{now}}$. Then the implicit QUD divides the context-set into the partition in (33).

(33)



The content of $B_{\textbf{now}}$ is equivalent to what can be expressed explicitly by the contemporary English Progressive sentence *Hamlet is reading Beowulf*, the proposition in (34).

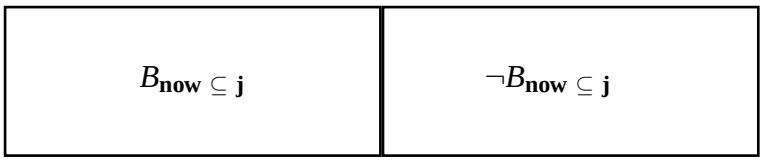
(34) $\lambda w.\ \forall k[k \in \mathscr{R}^c_{now} \rightarrow \text{COIN}(\lambda e.\text{Hamlet-read-Beowulf}(e), k, w)]$

The problem for Polonius, of course, is that he has the question but no grammaticalized device for expressing it explicitly. The strongest explicit question that he could ask given the grammaticalized linguistic resources at his disposal is (35-a), which is a made-up example intended to represent Middle English. Given the semantics assumed for simple present tense sentences, the meaning of this question is the set of propositions in (35-b). That is, the explicitly uttered question (35-a) sets up the partition in (36), where $B_{\textbf{now} \subseteq \textbf{j}}$ and $\neg B_{\textbf{now} \subseteq \textbf{j}}$ are shorthand for the two propositions in (35-b).

(35) a. Does Hamlet read Beowulf?

b. $\{\lambda w.\ \exists j[\textbf{now} \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \text{COIN}(\lambda e.\text{Hamlet-read-Beowulf}(e), k, w)]],$
$\lambda w.\ \neg\exists j[\textbf{now} \subseteq_{ini} j \wedge \forall k[k \in \mathscr{R}^c_j \rightarrow \text{COIN}(\lambda e.\text{Hamlet-read-Beowulf}(e), k, w)]]\}$

(36)



An affirmative linguistic response in this context will resolve the explicit question completely, eliminating $\neg B_{\textbf{now} \subseteq \textbf{j}}$ worlds and reducing the updated context set to the set of $B_{\textbf{now} \subseteq \textbf{j}}$ worlds. What is crucial for our purposes is that the content of the assertion only partially resolves the implicit but grammatically ineffable (in Middle English) QUD in (33) which is whether $B_{\textbf{now}}$ holds in the actual world. Being given the information that $B_{\textbf{now} \subseteq \textbf{j}}$ holds in our world is not enough to eliminate all $\neg B_{\textbf{now}}$ worlds since the updated context set is compatible with such worlds.

This imaginary problem faced by Polonius is not an isolated one, of course. It also raises its head in the literarily more authentic case in (31-a). In fact, it appears systematically and is endemic to any language that instantiates a zero-PROG system. In such a language, the available grammatical strategies of the language fail to fully disambiguate between (and therefore fail to completely resolve) phenomenal and structural inquiries. This is not to say, by any means, that the addressee fails to resolve the fine-grainedness of the partition set up by the implicit QUD or that the response remains under-specified and is entered as such in the Common Ground. It simply means that the linguistic discourse between interlocuters must be heavily supported by contextual knowledge about the intent of the question (whether it is part of a phenomenal or structural inquiry) and the relevance of the answer to the intended question. Such information may be pragmatically retrieved in a zero-PROG language much 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 facilitate disambiguation.

4.2. Recruitment: the emergence of variants

Polonius' problem reflects a systematic expressiveness deficit in the grammaticalized tense-aspect system of a zero-PROG language. The actuation problem in this domain is the problem of how and why variants that might emerge in discourse in a zero-PROG language get grammaticalized. There are a number of nuanced issues that need to be addressed in understanding actuation, but here is how the basic process must work.

Speakers and listeners in a zero-PROG language regularly participate in local speech events which require disambiguation of phenomenal inquiries. Common knowledge of the context of interlocutors is not always enough to resolve such inquiries and in ambiguous contexts, participants must undertake local efforts for distinguishing between phenomenal and structural inquiries.[25] 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. 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.[26] Both adverbial and constructional linguistic devices overtly introduce the reference intervals throughout which the event predicates are asserted to hold. The inquiry is determined to be phenomenal in a given context by the use of such devices precisely because they restrict the temporal interpretation of the question or the assertion to the reference

[25]The term "ambiguous context" is shorthand for a context that lacks information that can distinguish between the structural and the phenomenal reading of imperfective utterances.

[26]The periphrastic constructions listed here are English calques of grammaticalized progressive marking in Scots-Gaelic, Spanish, and Dutch respectively. Dutch uses two constructions – one with a locative preposition and the other with the postural verb *zitten* in expressing PROG meaning.

interval – the main semantic contribution of grammaticalized progressive marking. Such morphosyntactically explicit local efforts will systematically accumulate in a zero-PROG language. The grammaticalization of a new progressive marker, on this view, is essentially the conventionalization of the "temporal restriction" function on one privileged exponent.[27] 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. That is, although the language now has a distinguished exponent for the PROG operator, the use of this exponent is not yet obligatory in contexts of phenomenal inquiry. This is the state instantiated, for example, by Late Middle English, where scholars have noted the optionality and comparatively low occurrence of the English Progressive construction and its steady increase until a great spurt in its frequency in the 19th century (e.g. Strang 1982; Denison 1993; Arnaud 1998; Smitterberg 2005). One might also assign the same status to the Spanish Progressive, which still exhibits optionality of use in phenomenal contexts while systematically increasing in frequency (see Torres Cacoullos 2012 for an account of the diachronic development of the Spanish Progressive).

There is a strong body of quantitative scholarship on the development of the English Progressive which is no doubt relevant to understanding the propagation of the change in different linguistic and sociological contexts. Among other things, the lexical aspect of predicates, the animacy and agentivity of subject denotations, voice, and future-oriented reference have been shown to be significant linguistic variables (Denison, 1998; Strang, 1982; Hundt 2004; Nesselhauf 2007), while gender, degree of intimacy, emotional involvement, and subjective attitudes have been studied as non-linguistic variables (Arnaud 1998; Smith 2002; Smitterberg 2005) with respect to the choice of the Progressive form. If the account proposed here for the emergence of progressive marking is on the right track, then we expect to see these observed correlations in English to be epiphenomenal effects (rather than causal factors) of the nature of the inquiry in discourse, ultimately underpinned by the phenomenal/structural contrast. A substantive examination of this hypothesis for languages that come with accessible diachronic records is naturally what should be done next. Here I have focused on outlining only the essential structure of and impetus for the diachronic process.

4.3. Categoricalization: the conventionalization of variants

Once a grammaticalized functional category is part of the grammar, it is always available in principle to speakers for use in the appropriate speech contexts. In this

[27] 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. In §5, I will assume that the change is a result of misinterpretation of the input by child acquirers.

case, it might be logical to expect that post-grammaticalization, progressive marking is used categorically in every context in which interlocutors undertake a phenomenal inquiry. The facts, however, clearly indicate otherwise. Historical studies of English have shown that the near-categorical use of the Progressive in describing events in progress only goes back to the 19th century (a review of the relevant literature can be found in Smitterberg 2005), while the grammaticalized Spanish Progressive is still only optionally used to describe events in progress, alternating with the Simple Present and the Imperfect in this function.[28] Grammars and texts of 19th century Hindi similarly indicate that the Hindi Progressive, which is categorical in phenomenal contexts in contemporary Hindi, was an optional device in the 19th century (Kellogg 1893; Beames 1966).

Generalizing across languages, we systematically observe a temporal lag between the grammaticalization of a progressive marker (the onset of the emergent-PROG state) and its near-categorical distribution in phenomenal contexts (the categorical-PROG state). Such lags are the norm in language change and historical linguists refer to long periods of "layering" in which innovated expressions occur side-by-side with older expressions in a given domain (Bybee et al 1994: 21; Hopper 1991).[29] The layering principle is one way of describing structured variation in language – the coexistence of functionally similar but subtly distinct variants in a linguistic subsystem – that often, but not always, signal changes in progress.

The details of the path of development from optional use of progressive marking to its categoricalization via increasing frequency and extension to new linguistic contexts (e.g. passives, subordinate clauses, inanimate subjects) are not the focus here. Rather, the question is, why does progressive marking become obligatorily used in certain [i.e. phenomenal] contexts within the domain? The answer is that the change is an instance of Horn's pragmatic division of labor. A categorical-PROG system, in which progressive marking is categorically used in phenomenal inquiries and imperfective marking is pragmatically restricted to structural inquiries, emerges because it offers the optimal solution to the communicative problem of obtaining complete resolution to *both* phenomenal and structural inquiries.

4.4. Zipfean forces in disambiguation

Speakers in 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

[28] Obviously, none of the research makes reference to contexts of phenomenal inquiry but speaker intuitions about contemporary English and Spanish make it clear that the class of contexts in which Progressive marking is expected to show categoricality is a subset of the phenomenal inquiry contexts.

[29] "Within a broad functional domain, new layers are continually emerging. As this happens, the older layers are not necessarily discarded, but may remain to coexist with and interact with the newer layers" (Hopper 1991: 22).

structural inquiries.[30] 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.

Bybee (1994) labels this kind of introduction of systemic opposition the grammaticization of 'zero' morphology.[31] 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.[32] 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). The discussion in §3.1.2 is based on exactly this Neo-Gricean story.[33]

[30] Just as a reminder, the older form may be aspectually imperfective as in the case of Gujarati or Spanish or neutral as in the case of English simple tenses.

[31] In her words: " The fact that the [English] Simple Present is now incompatible with temporal phrases such as "right now" attests to the fact that it is explicitly expressing habitual meaning. The lack of a tense or aspect marker in the verb phrase conveys this particular meaning." (Bybee 1994: 239). Bybee's example involves a dynamic predicate (**I drink a cup of coffee right now*). But it is not clear that the facts straightforwardly extend to all predicates, since it is indeed possible to express a non-habitual proposition like *My head hurts right now* or *I love John right now* with the Simple Present. Although the English Progressive must be categorically used with dynamic predicates in describing events-in-progress, the Simple Present is not obligatorily paired with habitual/generic meaning. Larry Horn (p.c.) reminds me of the structural uses of the Simple Present in descriptions of recipes and experiments, where it cannot be interpreted as habitual. The relevant contrast is between "And now I fold in the eggs and stir the mixture gently" (uttered by a TV chef while demonstrating a recipe) and #"And now I step out of the kitchen and I go outside and smoke a cigarette." (as a description of the chef's activities while demonstrating the recipe).

[32] 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).

[33] 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).

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 – a more complex message – 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 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 therefore be to reorganize rather than to innovate, that is, to restrict the structurally simpler and semantically broader form to structural inquiries and use the structurally complex form obligatorily in phenomenal inquiries. This is the categorical-PROG system.

The Neo-Gricean story proposed here, however, poses a new question. What is the level at which these Zipfean considerations of expressiveness and economy in language change can be construed as applying? Is it the local *conversational* choices of speakers and hearers, based on considerations of rational communication, that lead to the diachronic emergence of a conventionalized categorical-PROG system? Or do expressiveness and economy constraints operate upon languages and shape their evolution without requiring inferential reasoning to obtain in each individual speaker–hearer interaction? This problem is a specific instance of the general problem of how semantic variants emerge, get selected, propagated, and conventionalized in communities of speakers, and the level of consciousness at which these processes take place.

I propose that the changes we observe in the grammaticalization domain involve a cycling between alternative solutions to a communication problem that have the right properties to reproduce and propagate among a population of speakers. In §5, I will demonstrate this by constructing a game with three evolutionarily stable strategies and proceed to describe the evolutionary dynamics that can explain the three observed transitions in this domain – **recruitment**, **categoricalization**, and **generalization**.

5. A game-theoretic interpretation of grammaticalization paths

The dynamic interaction between expressiveness and economy and some diachronic changes it may engender were sketched out informally in the last section. This interaction can be made more precise using the resources of evolutionary game theory as they have been applied to the problem of linguistic communication and pragmatic reasoning in game-theoretic pragmatics. This scholarship is centrally concerned with language use, in particular, with modeling how interlocutors establish conventional ways of signaling meaning given their information states and available alternative options for conveying the same information. 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 (e.g. van Rooij 2004a, 2004b, Jäger 2007) investigates the question of why particular strategies of conventionalization or marking are reflected robustly in natural language systems. This research adopts an evolutionary 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. It is this property that makes the evolutionary approach 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.

The model proposed here draws on ideas in van Rooy (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 in terms of Evolutionary Game Theory (Maynard Smith and Price 1973). The latter is typological in orientation and attempts, also within an evolutionary paradigm, to derive typologically 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.[34] On the other hand, we are interested in the cross-temporal aspect – that is,

[34]This is of inherent typological interest since the states identified as steps in the grammaticalization path in (28) 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).

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).

Evolutionary Game Theory shows that the basic concepts of game theory can be applied even to situations in which no individual is reasoning rationally or making explicit decisions. The idea is to determine for a population in which individuals exhibit different forms of behavior (which may or may not be the result of conscious choices), which forms of behavior are able to persist, and which forms of behavior 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 support) may be driven out by an alternative system in which progressive marking signals phenomenal inquiries and imperfective marking signals structural inquiries (and vice versa).

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 Rooy 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.[35] 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*}

[35]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.).

$\rightarrow$ {**phen**, **struc**}]. The utility function for the speaker and the hearer is defined with respect to such strategies.

Thus, given a state t, a speaker strategy S, and a hearer strategy H, the success of communication in any given utterance situation can be measured as follows by the δ-function:

(37) $\delta(t,S,H) = 1$ if $H(S(t)) = t$
$= 0$ 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. 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 and $P(t)$ is the probability of the state t. 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.

(38) $U_s(t,S,H) = \delta(t,S,H) - k \times P(t) \times n(S)$

Following Jäger (2007), k is taken to be some parameter that modulates the expected utility for a strategy across systems. Jäger interprets this parameter in terms of the speaker's priorities – 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, I take the linguistic interpretation of k to be similar but tied to the complexity of entire strategies rather than particular forms. 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.

Cost is not taken to be reflected in any way in the hearer's utility function, since the hearer has no choice between or means of identifying more/less complex strategies, but must simply determine the speaker's intended meaning on the basis of the presented form. Hearer utility is thus identical to the δ-function.

(39) $U_h(t,S,H) = \delta(t,S,H)$

Nature deals out the states **phen** and **struc** according to some probability distribu-

tion, which determines the likelihood of each state to be expressed.[36] The speaker has knowledge of the state she is in while the hearer lacks this knowledge. The average utility of a speaker or hearer strategy are then calculable as:

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

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

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). Further, in an evolutionary setting with a single population, every individual can be assumed to be in the role of a speaker half the time and the role of a hearer half the time. The expected payoff of an individual's strategy can be then calculated on the basis of the average utility for her speaker strategy and her hearer strategy.

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

The game model must further factor in the role of context in the disambiguation of meanings. Van Rooy (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, we deal with *impf*, which is underspecified with respect to **phen** and **struc** interpretations that get resolved in context. Following Van Rooy, 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.

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

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

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*

strategy using *prog* instead of *impf*.

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**

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.[37]

Given the parameters above, the average utility for speakers programmed for a particular strategy is as in Table 3 (based on (40-a)). The utility for hearers is exactly the same without the cost *k* factored in.

An individual's strategy is a pair $\langle S, H \rangle$ and each individual is assumed to be a speaker half the time and a hearer half the time. Given the payoffs in Table 3, 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 as fitness, which determines the replication rate of players. If a certain strategy yields an average payoff that is higher

[37]The strategies considered in this game model do not exhaust the logical space of strategies for the imperfective game. For instance, we do not consider strategies in which the state **struc** is disambiguated (whether in less probable or in all contexts) using a distinct form, say *hab* 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. There is no crosslinguistically known corresponding trajectory that takes as its starting point the innovation of a habitual/generic form for marking structural inquiries and extends its use to phenomenal inquiries.

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	0.9	0.5

than the population average, 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).

The question for us is: 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)**. A strategy α is said to be ESS if for all strategies β in a set of strategies *A*, either the condition in (42-a) or (42-b) holds. That is, either (a) the expected payoff of playing α against itself is strictly greater than the expected payoff of playing α against any other strategy β, or (b) the expected payoff of playing α against itself is equal to playing α against some β, but this payoff is strictly greater than the expected payoff of playing β against β.

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

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 (41). The table of expected payoffs where the row strategy plays against the column strategy is given in the table below.[38] The cost *k*, here taken

[38]Individual strategies based on speaker strategy S_4 are not considered in the game since S_4 is

to be minimal at 0.01, is only incurred for multi-form speaker strategies. From the table we see that:

- $\langle S_1, H_1\rangle$ and $\langle S_2, H_2\rangle$ are weak ESSs.[39]
- $\langle S_3, H_3\rangle$ 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	$\langle S_1, H_1\rangle$	$\langle S_1, H_2\rangle$	$\langle S_1, H_3\rangle$	$\langle S_2, H_1\rangle$	$\langle S_2, H_2\rangle$	$\langle S_2, H_3\rangle$	$\langle S_3, H_1\rangle$	$\langle S_3, H_2\rangle$	$\langle S_3, H_3\rangle$
$\langle S_1, H_1\rangle$	**$\frac{(0.9)+0.9}{2}$**	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$
$\langle S_1, H_2\rangle$	$\frac{0.9+0.9}{2}$	$\frac{0.9+0.9}{2}$	$\frac{0.5+0.9}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.5+0.95}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.9+0.95}{2}$	$\frac{0.5+0.95}{2}$
$\langle S_1, H_3\rangle$	$\frac{0.9+0.5}{2}$	$\frac{0.9+0.5}{2}$	$\frac{0.5+0.5}{2}$	$\frac{0.9+0.55}{2}$	$\frac{0.9+0.55}{2}$	$\frac{0.5+0.55}{2}$	$\frac{0.9+1}{2}$	$\frac{0.9+1}{2}$	$\frac{0.5+1}{2}$
$\langle S_2, H_1\rangle$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(0.55-0.01+0.9}{2}$	$\frac{(0.9-0.01+0.9}{2}$	$\frac{(0.95-0.01+0.9}{2}$	$\frac{(0.55-0.01+0.9}{2}$	$\frac{(0.9-0.01+0.9}{2}$	$\frac{(0.95-0.01+0.9}{2}$	$\frac{(0.55-0.01+0.9}{2}$
$\langle S_2, H_2\rangle$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(0.55-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.95}{2}$	**$\frac{(0.95 - 0.01)+0.95}{2}$**	$\frac{(0.55-0.01)+0.95}{2}$	$\frac{(0.9-0.01)+0.95}{2}$	$\frac{(0.95-0.01)+0.95}{2}$	$\frac{(0.55-0.01)+0.95}{2}$
$\langle S_2, H_3\rangle$	$\frac{(0.9-0.01)+0.5}{2}$	$\frac{(0.95-0.01)+0.5}{2}$	$\frac{(0.55-0.01)+0.5}{2}$	$\frac{(0.9-0.01)+0.55}{2}$	$\frac{(0.95-0.01)+0.55}{2}$	$\frac{(0.55-0.01)+0.55}{2}$	$\frac{(0.9-0.01)+1}{2}$	$\frac{(0.95-0.01)+1}{2}$	$\frac{(0.55-0.01)+1}{2}$
$\langle S_3, H_1\rangle$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(1-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(1-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(1-0.01)+0.9}{2}$
$\langle S_3, H_2\rangle$	$\frac{(0.9-0.01)+0.9}{2}$	$\frac{(0.95-0.01)+0.9}{2}$	$\frac{(1-0.01)+0.9}{2}$	$\frac{(0.9-0.01)+0.95}{2}$	$\frac{(0.95-0.01)+0.95}{2}$	$\frac{(1-0.01)+0.95}{2}$	$\frac{(0.9-0.01)+0.95}{2}$	$\frac{(0.95-0.01)+0.95}{2}$	$\frac{(1-0.01)+0.95}{2}$
$\langle S_3, H_3\rangle$	$\frac{(0.9-0.01)+0.5}{2}$	$\frac{(0.95-0.01)+0.5}{2}$	$\frac{(1-0.01)+0.5}{2}$	$\frac{(0.9-0.01)+0.55}{2}$	$\frac{(0.95-0.01)+0.55}{2}$	$\frac{(1-0.01)+0.55}{2}$	$\frac{(0.9-0.01)+1}{2}$	$\frac{(0.95-0.01)+1}{2}$	**$\frac{(1 - 0.01)+1}{2}$**

In terms of the discussion in §4, each diachronic state identified in (28) can be interpreted as reflecting the relative prevalence in a population of one or more of the three 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.[40] 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 identical to $\langle S_1, H_1\rangle$ other than with respect to the choice of form (*prog* instead of *impf*). 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.

structurally identical to S_1.

[39] A strategy is said to be a weak ESS if

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

[40] 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.

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 densities of different strategies in a population over time. 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 on a shorter time-scale 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 the input is (mis)-interpreted during the acquisition process. In particular, they also depend on *changes* in the mutation rates for strategies that exceed some threshold in the population. What follows is a preliminary sketch using replicator-mutator dynamics that could account for the observed behavior.

5.3. The replicator mutator dynamics

The dynamic account here considers only the ESSs: $\langle S_1, H_1\rangle$, $\langle S_2, H_2\rangle$, and $\langle S_3, H_3\rangle$. The expected payoffs for each interaction in this reduced game are given in (44) for quick reference.

(44)

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

To say that $\langle S_1, H_1\rangle$, $\langle S_2, H_2\rangle$, and $\langle S_3, H_3\rangle$ are ESSs means that each strategy is stable against invasion by small amounts of mutations. 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 back to one in which $\langle S_1, H_1 \rangle$ is prevalent.

These transitions in actuality involve interactions between all three strategies. But before we consider this more complex interaction, we will first study the replicator-mutator equation by breaking down the Imperfective game into sequential 2×2 games, that involve only two strategies at any given time. We will consider 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, before turning to the more realistic (and more complex) situation in which all three strategies are always present in the population in some proportion.

In each case, the population is dominated by some incumbent type A and 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 can be 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 simplified 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 for *A* and *B* can be calculated as follows.[41]

(45) 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. The selection dynamics, i.e. the rate of change in the frequency of the *A* and *B* populations over time, can be written as:

(46) 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 $\phi = x_A f_A(\overrightarrow{x}) + x_B f_B(\overrightarrow{x})$.

[41] 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 $1 - x$.

In addition to the rate of replication, the frequency of a given strategy in a population also depends on mutations from one strategy to the other. 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 stochastic row matrix *Q*, which gives the transition probabilities for mutation to happen from one strategy type to another.[42] For any strategies, *i, j...n*, let the mutation probability be denoted by Q_{ij}. Given these assumptions, the population dynamics are given by the "replicator-mutator" equation:

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j(\overrightarrow{x}) Q_{ji} - \phi x_i \tag{47}$$

The replicator-mutator equation gives the rate of change 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 *j* mutates to *i* and the average fitness of the population times the frequency of the *i*-population.

The replicator-mutator equation, together with some assumptions about the structure of the stochastic matrix *Q*, allow us to model recruitment and categoricalization. Further assumptions will be needed for modeling generalization and a more nuanced model will be introduced in §5.3.3.

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:

(48) $Q =$

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

[42] So, if we have two strategies, *A* and *B*, the transition probabilities can be represented as the matrix in (47). Here, *m* is the probability that an *A* parent has an *A* offspring, *n* is the probability that an *A* parent has a *B* offspring, *o* is the probability that a *B* parent has an *A* offspring, and *p* is the probability that a *B* parent has a *B* offspring. Saying that *Q* is a stochastic row matrix means that all the rows add to 1.

$Q =$

	A	B
A	*m*	*n*
B	*o*	*p*

Note that $\langle S_1, H_1\rangle$ and $\langle S_2, H_2\rangle$ are in a bistable relation: each is the best response to itself.

(49)

	$\langle S_1, H_1\rangle$	$\langle S_2, H_2\rangle$
$\langle S_1, H_1\rangle$	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 (without considering 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:

(50) $x_A^* = \frac{d-b}{a-b-c+d}$ (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 $x_A(0) = 0.9$.[43] 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, with population set at 10000.

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

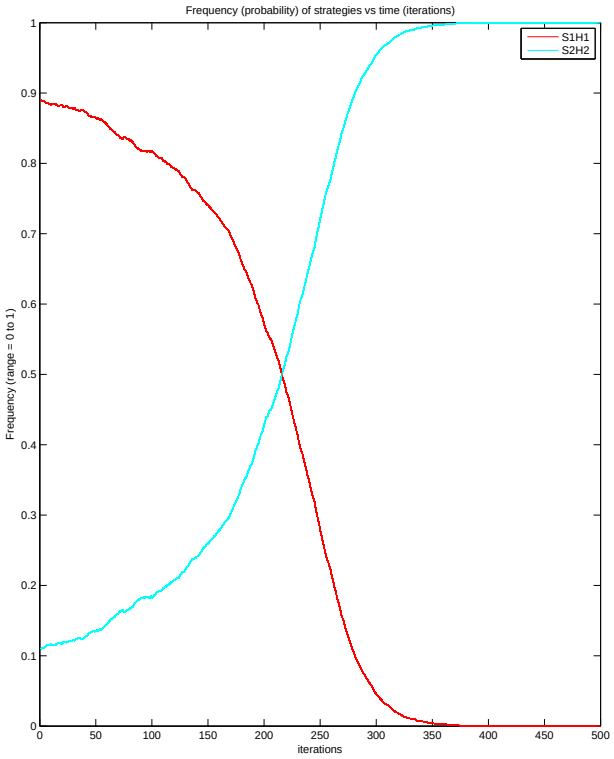
(51) $Q' =$

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

[43] $\frac{0.945-0.9}{0.9-0.9-0.894+0.945} = 0.9$.

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

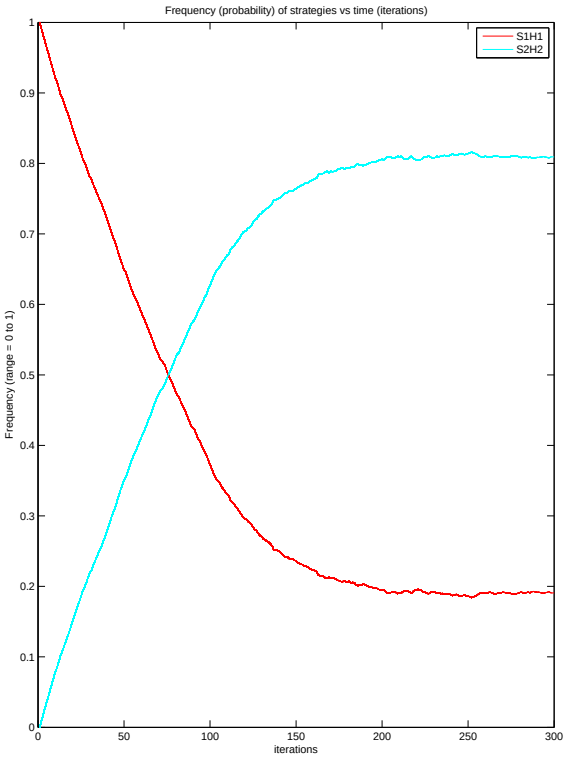


as likely to generate an $\langle S_2, H_2 \rangle$ offspring as an $\langle S_2, H_2 \rangle$ parent is to generate an $\langle S_1, H_1 \rangle$ offspring. With initial conditions set to $x_{\langle S_1, H_1 \rangle}(0) = 1$, Figure 2 gives the change in the proportion of $\langle S_1, H_1 \rangle$ 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 to model this state of the grammaticalization path, we need to consider a stochastic matrix that gives the transition probabilities between all three 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. Thus, the full game needs a modified matrix based on different assumptions that have to do with both the relative learnability of distinct strategies and their communicative efficiency. We will take Q'' to be the full stochastic matrix, where transitional probabilities for $\langle S_2, H_2 \rangle$ differ from those in (51), which was only used to show how the dynamics work.

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



(52) $Q'' =$

	$\langle S_1, H_1 \rangle$	$\langle S_2, H_2 \rangle$	$\langle S_3, H_3 \rangle$
$\langle S_1, H_1 \rangle$	0.93	0.07	0
$\langle S_2, H_2 \rangle$	0.02	0.91	0.07
$\langle S_3, H_3 \rangle$	0.03	0.02	0.95

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.

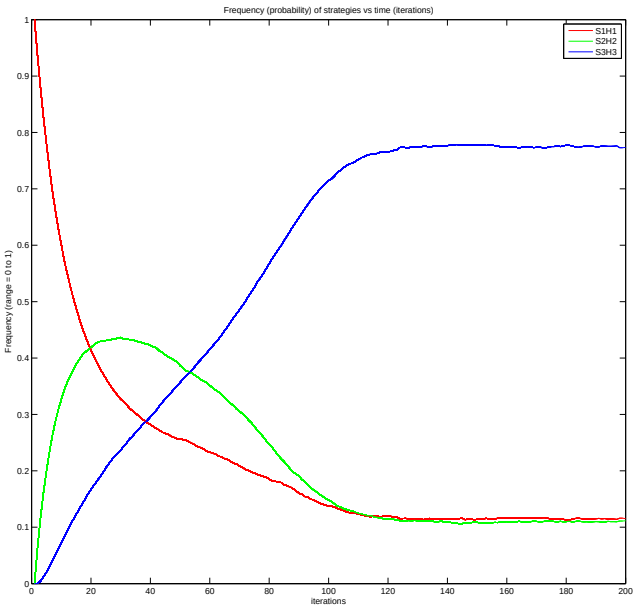
- Finally, $\langle S_3, H_3 \rangle$ is both communicatively successful and easier to learn than $\langle S_2, H_2 \rangle$, requiring no context-sensitivity of either the speaker or the hearer. This is reflected in lower mutations leading out of $\langle S_3, H_3 \rangle$ – slightly more into $\langle S_1, H_1 \rangle$ (due to its relative simplicity) than $\langle S_2, H_2 \rangle$.

The matrix Q'' thus assumes an ordering of the three grammars on a scale of relative difficulty from the perspective of acquisition:

(53) $\langle S_2, H_2 \rangle \gg \langle S_1, H_1 \rangle \gg \langle S_3, H_3 \rangle$

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 will reflect the interaction between asymmetric mutation rates and asymmetric payoffs. Figure 3 gives the dynamics for the three strategies over time with the stochastic matrix Q'' and initial conditions set to $\langle S_1, H_1 \rangle$ = 1 with population set at 10000. $\langle S_2, H_2 \rangle$ and $\langle S_3, H_3 \rangle$ are introduced by mutation. We see that the $\langle S_2, H_2 \rangle$ population quickly advances over the $\langle S_1, H_1 \rangle$ population and is gradually taken over by a growing $\langle S_3, H_3 \rangle$ population, which stabilizes at $\sim$0.77, with low proportions of $\langle S_2, H_2 \rangle$ and $\langle S_1, H_1 \rangle$. This is categoricalization.

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''



Given the structure of the stochastic matrix Q'', there can never be a population consisting entirely of $\langle S_1, H_1 \rangle$, $\langle S_2, H_2 \rangle$, or $\langle S_3, H_3 \rangle$ speakers – 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 $\langle S_1, H_1 \rangle$ population. 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 be 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:

(54) 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

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

For the purposes of this model, I will take ε to be 0.4. That is, if the proportion of $\langle S_2, H_2 \rangle$ in the population exceeds 0.4, then the population will be taken to have entered the emergent-PROG state from the zero-PROG state.

5.3.3. Generalization

There is a single stable equilibrium in the replicator-mutator dynamics which is given by Figure 3. The dynamics are guaranteed to lead to this equilibrium for any initial conditions. So if the population achieves this equilibrium, then it will stay there indefinitely and there will be no further change.

The cycling behavior from zero-PROG to emergent-PROG to categorical-PROG *back* to the zero-PROG state would appear to be quite a mystery in this game. How does the population repeatedly move away from the only stable equilibrium in this game only to come back towards it? In order to model the cycling behavior, I will start with a simple but non-traditional assumption:

(55) The cycling behavior obtains precisely because it is impossible for a population to reach the equilibrium.

Suppose that it is impossible for 77% of the population to successfully acquire the formally complex strategy which requires mastery over the use of two distinct forms. Then, mis-learning would increase drastically once the proportion of $\langle S_3, H_3\rangle$ speakers reaches some absolute threshold and speakers would increasingly shift to alternative strategies. Of the two alternative strategies, $\langle S_1, H_1\rangle$ is easier to acquire than $\langle S_2, H_2\rangle$, which would mean increased mutations towards $\langle S_1, H_1\rangle$. This would result in cycling back from the categorical-PROG state to the zero-PROG state. This is generalization.

I will make the assumption in (55) and concretely assume that $\langle S_3, H_3\rangle$ is optimally learnable for 0.5 of the population and maximally learnable for 0.75 of the population. If $x_{\langle S_3,H_3\rangle}$ exceeds 0.75, the mutation rate from $\langle S_3, H_3\rangle$ into $\langle S_1, H_1\rangle$ doubles (from 0.03 to 0.06) and this increased mutation rate persists until $x_{\langle S_3,H_3\rangle}$ reaches 0.25. This is an arbitrary lower bound that can go as high as 0.3 but not much beyond it for the cycling behavior to still obtain.[44] The reasoning is that once the system is pushed away from the equilibrium, the outward mutations will continue until some lower bound is reached that signifies the proportion of population for which $\langle S_1, H_1\rangle$ and $\langle S_2, H_2\rangle$ are unlearnable.[45] Figure 4 shows this behavior of the system for the described parameters with initial conditions set at $x_{\langle S_1,H_1\rangle} = 1$.

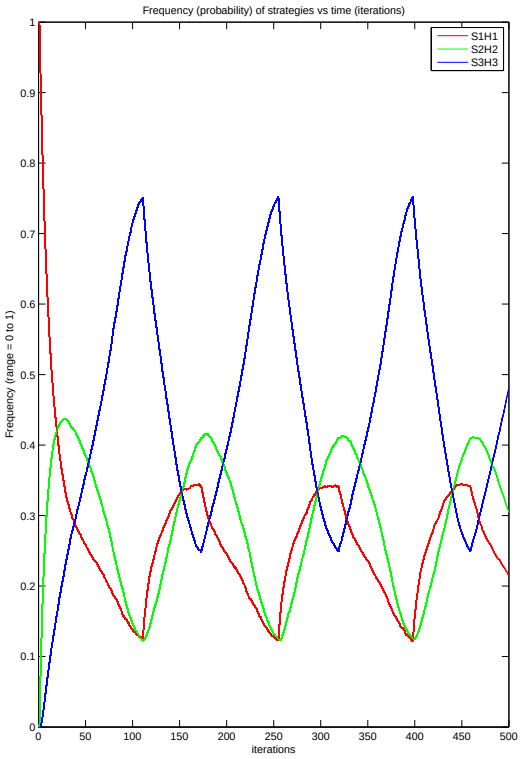
5.3.4. Choice of form in generalization

In the progressive $\gg$ imperfective grammaticalization path, the progressive form is said to generalize in meaning. For us, this means that mutant learners that acquire an $\langle S_1, H_1\rangle$ or $\langle S_2, H_2\rangle$ grammar from $\langle S_3, H_3\rangle$ and $\langle S_2, H_2\rangle$ parents tend to choose *prog* rather than *impf* as the general form. This can be accounted for if we make one more assumption about the structure of the input during acquisition: parent-child interactions during the acquisition process are more likely to occur in phenomenal contexts than in structural contexts. This assumption needs to be corroborated through a study of child-directed speech that explicitly examines contexts of utterance. But there is already indirect evidence in the first language acquisition of tense-aspect literature that 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,

[44] If no lower bound is set below which the mutation probability from $\langle S_3, H_3\rangle$ into $\langle S_1, H_1\rangle$ lowers back to its earlier frequency, $\langle S_3, H_3\rangle$ stabilizes at $\sim$0.25.

[45] These hypothetical constraints amount to a claim about differences in cognitive processing and learning styles within a population – which certainly need to be substantiated (in future research). But we do know that individual cognitive functioning exhibits differences with respect to acquisition and processing of information. Although one tends to not to take into consideration such preferences and biases in determining the relative learnability of particular grammars, there is no justification for ignoring the possible role of such structured differences in determining the outcomes of competition between grammars. Yu (2010)'s innovative study convincingly demonstrates that variability in ability to compensate for context-induced variations in speech perceptually is governed by sex and cognitive processing style. If constraints such as the one I have proposed structure at (at least some) populations, then they would guarantee that some kinds of equilibria are never achieved and lead us to the cognitive underpinnings of cycling patterns in semantic domains.

Figure 4: Cycling Behavior



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.

All this to say that child learners in (mis-)acquiring their own grammars from $\langle S_3, H_3 \rangle$ parents will receive more evidence for choosing *prog* than *impf* as the generalized form in $\langle S_1, H_1 \rangle$ and $\langle S_2, H_2 \rangle$. This, in due course, leads to the progressive $\gg$ imperfective shift.

5.4. Summary

Typologically, we can divide grammars into those in which the nature of the inquiry – phenomenal vs. structural – is determined (a) contextually and those in which it is marked linguistically – either (b) optionally or (c) categorically. The former type of grammar contains a single underspecified form that realizes the IMPF operator (e.g. Russian, Arabic, Sanskrit, Middle English) while the latter two types of grammar distinguish between exponents for PROG and IMPF by using *prog* either optionally (e.g. Romance) or categorically (e.g. Modern English, Hindi, and Turkish). Diachronically, we observe that languages move from context-dependent grammars (zero-PROG) to optional marking grammars (emergent-PROG) to categorical marking grammars (categorical-PROG) back to context-dependent 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.[46] The proposed game has three evolutionarily stable strategies which are always present in some proportion within a population. The payoffs of these strategies, mutation probabilities leading from and into these strategies (rooted in acquisitional asymmetries), and threshold frequencies for some strategies (hypothesized to be rooted in cognitive differences within a population) together lead to the cycling behavior observed in the states of the system. Crucially, the replicator-mutator dynamics proposed here has a single, stable equilibrium with $\langle S_3, H_3 \rangle$ as the most frequent strategy. It was hypothesized that the cycling behavior occurs because it is impossible for the population to reach that equilibrium. Once the population reaches close to the equilibrium, mis-learning increases and these increased mutation rates lead to the increase of $\langle S_1, H_1 \rangle$ and $\langle S_2, H_2 \rangle$ share in the population. The chosen default form in such mis-learned grammars is likely to be *prog* rather than *impf* because of its greater frequency in the acquisition input.

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 three evolutionarily stable strategies that correspond to these three ways of communicating the relevant meanings. A population contains each of these strategies in some proportion at any given time and mutations are continuously occurring between the three strategies. Recruitment occurs when a population in which contextual recovery of meaning (supplemented with low proportions of lin-

[46] 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.

guistic 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 quickly followed by categoricalization, in which the strategy in which there is no reliance on contextual recovery at all but obligatory explicit marking of the two meanings, increases in proportion. Despite this increase, which drives the system very close to the only stable equilibrium in the imperfective game, the population cannot reach it. The increased mutation rates, which follow when the strategy reaches its threshold proportion result in decreasing frequency of this strategy. This paves the path for generalization, which is the increase in the frequency of $\langle S_1, H_1\rangle$+$\langle S_2, H_2\rangle$ with $\langle S_2, H_2\rangle$ remaining below the threshold value that corresponds to recruitment. Once $\langle S_2, H_2\rangle$ reaches the threshold value, we again get recruitment, followed by categoricalization, followed by generalization and so on and so forth.

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 the 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:

(56)
a. A semantic grammaticalization path in the functional domain must be structurally underpinned by some privative semantic contrast between a specific and a general meaning.
b. Changes in functional domains characterized by a privative semantic contrast are cyclic in nature because these domains are structured as games which give rise to equilibria that can be almost, but not quite, reached by the population.
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 threshold values for the frequency of the explicit marking strategies.

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 of 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.

Arnaud, René. 1998. The development of the progressive in 19th century English: A quantitative survey. *Language Variation and Change* 10:123–152.

Beames, John. 1872-79. *A Comparative Grammar of Modern Indo-Aryan Languages of India*. Delhi: Munshiram Manoharlal. Republished 1966.

Beaver, David I. and Brady Z. Clark. 2008. *Sense and sensitivity: How focus determines meaning*. Oxford: Blackwell Publications.

Beddor, Patrice Speeter. 2009. A coarticulatory path to sound change. *Language* 85(4):785–821.

Bermúdez-Otero, Ricardo. Forthcoming. Amphichronic explanation and the life cycle of phonological processes. In P. Honeybone and J. Salmons, eds., *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press.

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

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

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.

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. 2003. Cognitive processes in grammaticalization. In M. Tomasello, ed., *The new psychology of language (Cognitive and functional approaches to language structure)*, vol. 2, pages 145–167. Laurence Erbaum.

Bybee, Joan. 2006. From uage to grammar: The mind's response to repetition. *Language* 82(4):711–733.

Bybee, Joan and Östen Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in language* 13(1):51–103.

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.

Carlson, Greg. 1995. Truth-conditions of generic sentences: two contrasting views. In M. Krifka, F. J. Pelletier, G. Carlson, A. ter Meulen, G. Chierchia, and G. Link,

eds., *The Generic book*, pages 224–237. Chicago: The University of Chicago Press.
Champollion, Lucas. 2010. *Parts of a whole: Distributivity as a bridge between aspect and measurement*. Ph.D. thesis, University of Pennsylvania.
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.
Cohen, Ariel. 2001a. On the generic use of indefinite singulars. *Journal of Semantics* 18(3):183–209.
Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press.
Condoravdi, Cleo. 2009. The Informativity of Negative Assertions and the Rise of NPIs. Handout of talk presented at 3^{rd} International Conference of Modern Greek Dialects and linguistic theory.
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.
Denison, David. 1993. *English historical syntax: verbal constructions*. London: Longman.
Denison, David. 1998. Syntax. In S. Romaine, ed., *The Cambridge history of the English language, Vol. 4: 1776–1997*, pages 92–329. Cambridge: Cambridge University Press.
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 quantificational domains. *Linguistics & Philosophy* 32(5):475–521.
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.
Dowty, David. 1980. Comments on the paper by Bach and Partee. In *Proceedings of CLS 16, Papers from the Parasession on Pronouns and Anaphora*.
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.
Ginzburg, Jonathan. 1995a. Resolving questions, i. *Linguistics and Philosophy* 18(5):567–609.
Ginzburg, Jonathan. 1995b. Resolving questions, ii. *Linguistics and Philosophy* 18(6):459–527.
Givòn, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.
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.

Groenendijk, Jeroen and Martin Stokhof. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6):1043–1068.

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

Heine, Bernd. 2003. Grammaticalization. In B. Joseph and R. Janda, eds., *The handbook of historical linguistics*, pages 575–601. Oxford: Blackwell.

Hopper, Paul J. 1991. On some principles of grammaticization. In E. C. Traugott and B. Heine, eds., *Approaches to Grammaticalization, Volume 1*, pages 17–36. Amsterdam: John Benjamins.

Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

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.

Horn, Laurence and Barbara Abbott. 2012. ⟨the, a⟩: (In)definiteness and Implicature. In W. Kabasenche, M. O'Rourke, and M. Slater, eds., *Reference and Referring*, pages 325–356. Cambridge: MIT Press.

Hundt, Marianne. 2004. Animacy, agentivity, and the spread of the progressive in modern english. *English Language and Linguistics* 8(1):47–69.

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. 2005. The Vedic injunctive: Historical and synchronic implications. *The Yearbook of South Asian Languages and Linguistics* .

Kiparsky, Paul. 2012. Grammaticalization as optimization. In D. Jonas, J. Whitman, and A. Garrett, eds., *Grammatical change: Origins, nature, and outcomes*, pages 15–51. Oxford: Oxford University Press.

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.

Koontz-Garboden, Andrew. 2004. Language contact and Spanish aspectual expression: a formal analysis. *Lingua* 114:1291–1330.

Kornfilt, Jaklin. 2005. *Turkish*. 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.

Levinson, Stephen. 2000. *Presumptive meanings*. Cambridge, MA: MIT Press.

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.

Nesselhauf, Nadja. 2007. The spread of the progressive and its 'future' use. *English Language and Linguistics* 11(1):191–207.

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

Ohala, John. 1981. The listener as a source of sound change. In *Proceedings of The Chicago Linguistics Society, 1981*.

Ohala, John. 1993. The phonetics of sound change. In C. Jones, ed., *Historical linguistics: Problems and perspectives*, pages 237–278. London: Longman.

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.

Roberts, Craige. 1998. Information structure in discourse: Towards an integrated formal theory of pragmatics. In J. H. Yoon and A. Kathol, eds., *OSUWPL, 49*, pages 91–136. The Ohio State University, Department of Linguistics.

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

Smitterberg, Eric. 2005. *The progressive in 19th-century English: a process of integration*. New York: Rodopi.

Stalnaker, Robert. 1978. Assertion. In P. Cole, ed., *Syntax and Semantics 9: Pragmatics*. New York: Academic Press.

Stalnaker, Robert. 1998. On the representation of context. *Journal of Logic, Language and Information* 7(1):3–19.

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

Strang, Barbara. 1982. Some aspects of the history of the be+*ing* construction. In J. Anderson, ed., *Language Form and Linguistic Variation*, pages 427–474. John Benjamins.

Swart, Henriette De. 1991. *Adverbs of quantification: a Generalized Quantifier Approach*. Ph.D. thesis, Rijksuniversiteit Gröningen.

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

Torres-Cacoullos, Rena. 2012. Grammaticalization through inherent variability: The development of a progressive in Spanish. *Studies in Language* 36.1:73–122.

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.

Yu, Alan. 2010. Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PloS One* 5(8).

Zeevat, Henk. 2007. Simulating recruitment in evolution. In G. Bouma, I. Kramer, and J. Zwarts, eds., *Cognitive Foundations of Interpretation*, pages 175–194. Amsterdam: KNAW.

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