

Linguistic emergence on the ground: a variationist paradigm

SHANA POPLACK
University of Ottawa

RENA TORRES CACOULOS
Pennsylvania State University

Abstract

Observing that linguists’ intuitions about the function of forms may be far removed from the motivations for variant choice operating in real interactions, we propose that the structure of variability inherent in speech can be marshaled to detect emergence on the ground. In discourse, emergence is observable in the favoring effect of contextual factors that operationalize putative functions on the selection of a given form over any of its competitors in the same (“variable”) context. We show how situating emergent forms in the social and linguistic structures in which they are embedded helps pinpoint the extent of, competing motivations for, and mechanisms involved in emergence.

Keywords

Variation, neutralization-in-discourse, linguistic variable, speech community, competing motivations, analogy, grammaticalization, priming, probabilistic constraints.

Linguistic emergence on the ground: a variationist paradigm*

SHANA POPLACK
University of Ottawa

RENA TORRES CACOULOS
Pennsylvania State University

1. Introduction

Among the core facts about language that emergentism seeks to account for are the probabilistic rather than deterministic nature of linguistic behavior, the gradual progression of language change within speakers and across speech communities, and the conventionalization of grammatical regularities, or patterns, from speaker interactions. These are all linked to the ubiquitous variation observable at all levels of linguistic organization. Much theorizing and experimentation has gone into elucidating the emergence of linguistic behavior, bringing a variety of tools, from sophisticated computational models to eye-tracking studies, to bear on the goal of providing empirically testable descriptions of the mechanisms at play (MacWhinney 2001, 449). Conspicuously absent from these efforts has been any sustained consideration of language emergence *on the ground*, i.e. what it looks like and how it works in the course of actual interactions among members of a real speech community. Indeed, although sociolinguistics is often cited approvingly as an allied approach (Beckner et al. 2009, 15; MacWhinney in press), emergentist work has proceeded largely independently of over half a

* This research has been supported by grants from the Social Sciences and Humanities Research Council of Canada and the Killam Foundation to Poplack and the National Science Foundation (BCS 1019112) to Torres Cacoullos. Poplack holds the CRC in Linguistics. Our foray into the world of emergentism has been greatly facilitated by incisive comments from Peter Auer, Esther Brown, Joan Bybee, Paul Hopper, Brian MacWhinney and William O’Grady, which we gratefully acknowledge.

century of cumulative advances in that field. Where sociolinguistics *is* acknowledged, it is largely to bolster support for emergentist critiques of generative tenets like the homogeneity of linguistic structure. The fact that the specifically variationist branch has been modeling, analyzing and, most importantly, amassing key evidence about the very questions preoccupying emergentists has not been capitalized on.

With this chapter we hope to remedy this situation by highlighting the many kindred assumptions shared by both camps, albeit sometimes masked by very different sets of data and methods. We rephrase the emergentist perspective in variationist terms, paying special attention to core notions like neutralization in discourse, the linguistic variable and orderly, or constrained, heterogeneity (§ 2). Focusing on the domain of morphosyntactic variation and change, we suggest how variationist methodology can be marshaled to further the emergentist goal of “discovering the exact shape of emergent mechanisms” (MacWhinney 2001, 449). We first argue for the primacy of community-based spontaneous speech data by displaying the sharp breach between both prescriptive and theoretical accounts on the one hand, and actual usage, here instantiated by the French subjunctive and English subject-verb agreement, on the other (§ 3). We then raise the important issue of how *much* a form has to be associated with a given function to qualify as “emergent”, exemplifying with the behavior of Nigerian Pidgin English pre-verbal *bin* (§ 4). The remainder of the paper is devoted to the question of what it means — empirically—for a linguistic form or construction to emerge. We show how emergence can be detected “on the fly” both in the speech community and in the linguistic system, making use of multivariate models of natural production data. In the speech community, social meaning has emerged in the associations of Mexican Spanish imperfective auxiliaries with social class, while new stylistic meaning has been assumed by the erstwhile negation marker *ne* and the

interrogative particle *est-ce que* in Québec French (§ 5). Probing mechanisms of emergence *internal* to the linguistic system, we demonstrate priming effects in the grammaticalization of the Spanish progressive, and language-internal analogy as the mechanism giving rise to proscribed preposition stranding in Québec French (§ 6). In concluding, we submit that the insights offered by the data and analysis of unreflecting language production are unavailable from, and an indispensable complement to, the intuited, elicited, and constructed materials on which much work in the emergentist paradigm has hitherto been based.

2. A variationist perspective on emergentism

2.1 Inherent variability

Usage-based theories postulate that grammatical knowledge is founded on speakers’ linguistic experience—the frequency and contexts of use of forms (e.g., Bybee 2010). Variationists share the assumption that the primary object of study is language as it is actually *used*, as opposed to linguistic idealizations. Language use, especially as instantiated in spontaneous speech, is characterized by the property of *heterogeneity*, which, as Weinreich, Labov and Herzog (1968, 101) put it, “is not a matter of multidialectalism or ‘mere’ performance, but is part of unilingual linguistic competence.” Yet while situationally-determined heterogeneity, ranging from choice of language to choice of words, is regularly invoked by emergentists (e.g. MacWhinney 2001; Beckner et al. 2009; Ellis 2012), the kind of heterogeneity most pertinent to the issues at hand tends to be given short shrift. This is the *inherent variability* manifest in the myriad form-function asymmetries which are the hallmark of spontaneous speech: variation among different morphosyntactic forms serving generally similar grammatical functions (Labov 1969; 1972;

Sankoff and Thibault 1981; Sankoff 1988a). This is illustrated in (1), where *will* (a) and *be going to* (b) both express future tense.

- (1) a. He *’ll* be two in August. (QEC.023.1247)ⁱ
b. Same age as Emily, he *’s gonna* be six in- August. (QEC.119.1470)

Such inherent variability has been the stock-in-trade of the variationist study of language over the last five decades. “The variationist viewpoint on language is determined first by a scientific interest in accounting for grammatical structure *in discourse* [...] and second by a preoccupation with the polyvalence and apparent instability *in discourse* of linguistic form-function relationships” (Sankoff 1988a, 141, italics in original).

2.2 *Neutralization in discourse and the linguistic variable*

The working hypothesis is that of *neutralization in discourse*: while contexts can almost always be found in which different morphosyntactic expressions have different meanings, as when *will* conveys desire, and *going to*, motion towards a goal, in other well-defined contexts, such meaning distinctions do not come into play (Sankoff 1988a, 153).

The key construct in variation theory is the *linguistic variable* (Labov 1969), a set of *variants* which “are used interchangeably to refer to the same states of affairs” (Weiner and Labov 1983, 31). Identification of a linguistic variable rests on the possibility of circumscribing the sum of contexts in which potential meaning differences embodied by the alternating forms elsewhere are not operative. These constitute the “variable context”. Here, semantic or functional distinctions between variant forms are *neutralized*, and such neutralization is the “fundamental discursive mechanism of (nonphonological) variation and change” (Sankoff 1988a, 153). In a cognitive linguistics framework, a compatible idea is Croft’s (2010, 42) proposal that language

change is possible because of “indeterminacy in verbalizing human experience”. In example (1) above, the variable context is the functional domain of future temporal reference, and *will* and *be going to* are two of its variant expressions.

2.3 *Competing motivations*

The variationist notion of the linguistic variable relates forms to the linguistic structures in which they are embedded. This entails that alongside an emerging form (or construction), other variant forms will be jockeying for the same linguistic work. The resulting competition naturally raises the question of the choice mechanisms speakers employ in deciding amongst forms.

Neutralization of semantic distinctions in the variable context does not entail that variant selection is random. On the contrary, variationists recognize that any number of factors in the linguistic and social environments may conspire or conflict in tipping the balance toward one choice over another. The ultimate shape of linguistic structure emerges from the interplay of such competing motivations (e.g., Bates and MacWhinney 1987; Haiman 2011), though further cumulative progress is still to be made in determining these on a case-by-case basis before venturing cross-linguistic generalizations.

The contour of these motivations corresponds to what variationists call the (socio)linguistic *conditioning* of variation, that is, the configuration of constraints—as expressed by their significance, magnitude and direction of effect—on the selection of one variant of a linguistic variable rather than another (Poplack and Tagliamonte 2001, 93-95). These may be online, context-dependent factors (such as the type of clause in which a variant occurs or its proximity to another element in the clause), cumulative, storage-based factors, reflecting speakers’ overall prior experience with a form (such as token frequency or frequency of

occurrence in a particular context), or operationalizations of processing considerations, among others. For example, for subject-verb agreement, factors favoring the variable selection of plural verb morphology in the Spanish *haber* ‘there is/are’ + NP construction (prescriptively, always singular) are informal speech style, an online constraint, and preponderance of the NP in subject role, a cumulative usage effect (Brown and Rivas 2012; in English existentials, significant is the presence of a plural –s, which may be interpreted as a processing constraint (Walker 2007; cf. O’Grady 2008).

The analysis of linguistic variables provides rich evidence for the competing motivations central to emergentist accounts. This is because the linguistic variable embodies a way of measurably attributing usage tendencies to (extra)linguistic constraints, by establishing their effects on variant choice through multivariate quantitative models of production data (cf. Labov 1969; Cedergren and Sankoff 1974). These constraints often act independently, in the mathematical sense. For example, in the social and stylistic stratification of *-ing* in New York City, the effect whereby the alveolar variant is more frequent in casual than in careful speech is replicated for each social class (Labov 1966). However there may also be interaction, such as the “cross-over pattern” reported for the rate of syllable-final (r)-pronunciation, which is directly related to social status in general, but which—in the most formal style—is greater in the second-highest than in the highest status social group (cf. Labov 1966; Labov 2004, 10).

Such effects cannot even be contemplated in the laboratory. Even studies of large-scale corpora, typically constituted of amalgamated texts, often fail to incorporate the social component that would allow for an empirical test of the speech-community-based motivations comprising the actual linguistic experiences invoked by emergentists as shaping linguistic structure.

2.3 *The principle of accountability*

From the notion of the linguistic variable follows the methodological principle of *accountable reporting*, which requires that values be reported for *every* case where the variable element occurs in the variable context as we have defined it. This means that an analysis must account not only for the cases in which the process of interest applied, but also for all the cases where it *could have* applied even if it did not (Labov 1982, 30; Labov 2004, 7). For example, in order to study the emergence of the *going to* future, we consider tokens of *going to* as well as instances where *will* or another variant occurred instead.

The consequences of applying the principle of accountable reporting are theoretical as well as methodological. Emergence has been construed as the association of *one* form with a new meaning or function, as in the well-documented (e.g., Bybee et al 1994) development of future meaning from purposive motion source constructions (*going to*) or expressions of desire (*will*). But such accounts often ignore the role of the other forms (e.g., the simple present or present progressive) coexisting in the same context. The cumulative findings of variationist research show that without an understanding of how existing variants accommodate to the incursion of the construction-in-becoming—and how the emerging variant is itself affected by variation with the older variant(s)—we obtain only a very partial view of this process (Poplack and Malvar 2007; Torres Cacoullos and Walker 2009; Poplack 2011; Torres Cacoullos 2011).

2.4 *Confronting emergentist and variationist accounts*

In sum, though emergentists and variationists share nearly all their core assumptions, variationists approach them quite differently. The recognition that the (complex adaptive) system

which is language consists of *multiple agents in interaction* is investigated directly (rather than by modeling) through construction of speech corpora collected from speakers specifically sampled so as to embody at least some of the competing social motivations (or “external determination” [MacWhinney 2001, 450]) for variant selection. The goal is to document as wide a variety of actual interactions as possible, but minimally including at least some instantiation of the *vernacular*, which provides the most systematic data for linguistic analysis (Labov 1972, 208). This is where unreflecting use of linguistic forms, least affected by self-monitoring or hyper-correction, may be studied. The resulting material constitutes the raw data of “experience” (which figures so prominently in emergentist accounts), on which analyses of patterning, frequencies or statistical regularities are performed.

The property of “intrinsic diversity” (Ellis 2012) or “variation” (MacWhinney 2001; Beckner et al. 2009) is dealt with head-on through the construct of the linguistic variable, constituted not only of the variant of interest but also of all the forms with which it alternates in the “variable context”. The construct of the linguistic variable offers the distinct advantage of allowing the analyst to *situate* the variant of interest within the system in which it is emerging, providing a more accurate model of actual behavior, since linguistic forms do not develop in isolation.

Speaker behavior is construed as the *product of competing factors*. Variationists view speakers’ choices as subject to constraints bound to features of the linguistic and extra-linguistic environments in which they occur. These contextual features define factors, which thus operationalize hypotheses about what motivates variant choice. Multivariate analysis of the contribution of these factors allows us to make probabilistic statements about the relative frequency of co-occurrence of linguistic forms and elements of the linguistic context (Poplack

and Tagliamonte 2001, 88-95), which in turn form the basis for predicting which of a number of competing motivations will win out in a given context.

These methodological and analytical procedures are empirical applications of the emergentist commitments to calculating distributional contingencies, identifying local co-occurrence relationships, and more generally, to extracting statistical regularities from experience (O’Grady 2008, 456; O’Grady 2010, 275). But in contrast with reservations regarding the lack of consensus on the form of knowledge resulting from emergentist *models* of the mechanisms that extract statistical regularities (O’Grady 2008, 456), decades of variationist study of the *behavior* of the speech community has yielded a well-defined and amply replicated characterization of the patterning of linguistic variation in many areas of the grammar.

A final shared assumption is that *language is not static*. Emergentists characterize both the language of the community and that of the individual as being in a state of *constant change* and reorganization (Beckner et al. 2009; Ellis 2012). Change, emergent or otherwise, can only be adequately assessed in the context of a diachronic component, as illustrated by the case studies in ensuing sections. The tools of variationist sociolinguistics offer a means to evaluate emergentist construals of change in both real and apparent time, and crucially, distinguish variability from change. Inherent variability of the type we have been describing is a necessary precursor to change, but should not be confounded with it. Indeed, the cumulative results of the variationist research program show that it is *variation* that is the pervasive characteristic of language, while (completed) *change* is a good deal rarer.

In the remainder of this chapter, we illustrate what the variationist framework and the construct of the linguistic variable can contribute to studying the emergence of linguistic

structure from usage. We turn first to the question of what constitutes appropriate data for such an endeavor.

3. Community-based spontaneous speech as the data of emergence: an apologia

In the preceding section we highlighted the many core assumptions emergentists share with variationists. We are nonetheless struck by what we see as the great divide between the emergentist insistence on language *use*, and the nature of the data typically brought to bear on its study. Despite the professed predilection for corpora (Beckner et al. 2009), with a few notable exceptions, including work in discourse syntax (e.g., Thompson and Hopper 2001), grammaticalization (e.g., Bybee 2010) and child language (e.g., Dąbrowska and Lieven 2005), the data of morphosyntactic emergence is either unsystematically cherry-picked, constructed for purposes of experimentation or computer-generated. As such, they hardly differ from the materials on which the generative approaches they dismiss are based. Recent variationist work grounded in the usage of the speech community is showing that many of the most regular and firmly entrenched patterns of variation lie so far below the level of consciousness that they cannot be intuited, let alone modeled, even by native speakers (Poplack and Dion 2009; Torres Cacoullos and Walker 2009; Poplack, Zentz and Dion 2012; Poplack, Lealess and Dion 2013). Indeed they are invisible to any but systematic quantitative analysis of spontaneous speech. One such example of the disconnect between what is and what is taken to be a linguistic fact comes from the treatment of the French subjunctive.

3.1 *The French subjunctive*

In Romance, the subjunctive is considered the grammatical mood *par excellence*, and centuries of prescriptive and descriptive linguistic effort have gone into delimiting the conditions requiring it and the readings it conveys. While the general consensus is that the overriding determinant is semantic, controversy rages over what its contribution actually is. Nominally, this is because the commitment to meaning as explanatory of variant choice is at odds with the many (prescriptively) licensed but semantically “unwarranted” uses of the subjunctive (e.g. the well-documented realis-subjunctive and irrealis-indicative mismatches). Even greater discrepancies vis-à-vis the way the subjunctive is “supposed” to be used arise from the way it is *actually* used, although these have not been formally recognized.

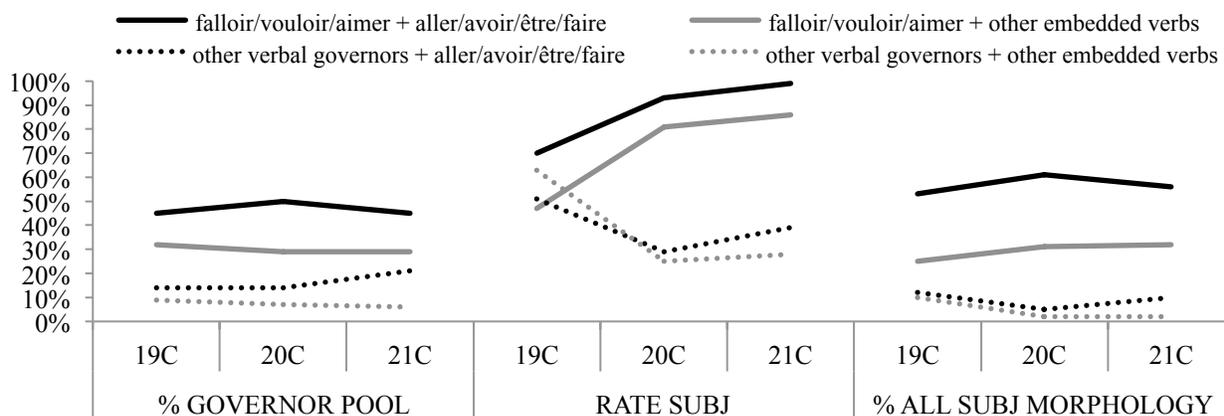
Prominent among the disparities between linguistic theorizing and linguistic fact is the finding that no “subjunctive-selecting” governor of even moderate frequency occurs categorically with that variant (Poplack 1992). The major characteristic of subjunctive usage is its *inherent variability*, as exemplified by the sentences in (2), in which subjunctive, indicative and conditional morphology all alternate under the same governor *aimer* ‘like’, with no change in reading of the embedded verb.

- (2)
- a. Tu sais, tu aimerais mieux qu’ils *soient* (SUBJ) pas là. (OH.25.608)
‘You know, you’d like it better if they weren’t there.’
 - b. J’aimerais ça que vous *verriez* (COND) mon mari. Vous allez voir comme c’était un bel homme. (OH.072.1335)
‘I’d like it if you would see my husband. You’ll see how handsome he was.’
 - c. Tu aimerais qu’il *guérisse* (IND) ta paralysie, parce je suis paralysée dans le visage. (OH.044.1554)
‘You’d like for him to cure your paralysis, because my face is paralyzed.’

Moreover, in stark contrast to the hundreds of lexical governors that have been assumed to select for this variant, almost *all* of the subjunctive usage in a vast corpus containing over 4000 contexts where it *could* have occurred is concentrated under a small handful (Poplack 1992; Poplack, Lealess and Dion 2013). Likewise, though every verb in the language is eligible (indeed required) to carry subjunctive morphology when embedded under a prescribed governor, very few distinct lexical types ever do, and still fewer do so with any regularity. Such huge deviations from prescribed usage are coupled with severely skewed distributions in actual usage of governors and embedded verbs, both in terms of proportion of the data accounted for and strength of association with the subjunctive. For example, a single lexical governor, *falloir* ‘be necessary’, accounts by itself for approximately 2/3 of all the governors in the data, and 3/4 of all subjunctive tokens!

Such patterns have never been invoked in any but the sociolinguistic literature on the topic. Yet these robust community norms for subjunctive use, all the more striking for being untaught and unremarked, have not only been firmly in place since the 19th century (at least), but are becoming more entrenched over time (Figure 1).

Figure 1: *Distribution of subjunctive selection in matrix + embedded verb collocations* (adapted from Poplack, Lealess and Dion 2013, Figure 10).



Any apparent semantic motivation for mood choice, operationalized along with other competing motivations as factors in a multivariate analysis, was shown to be an epiphenomenon of this overriding lexical effect. The only other factors affecting variant choice are structural, pertaining to the construal of the context as a stereotypical subjunctive construction: morphological irregularity of the embedded verb, presence of the complementizer *que*, and adjacency of main and embedded clauses. Only when these favorable conditions are present is selection of the subjunctive at all likely. There is little reason to believe that anything new is *emerging* here, since no change is apparent, at least over the time frame studied. Instead, what we observe is a lexical/structural *schema* involving some canonical subjunctive components.

Detailed analysis of usage on the ground thus confirms that the constraint uncritically *assumed* by prescriptivists and theoretical linguists alike to govern subjunctive use -- meaning -- is simply not operative in speech. Yet in a parallel and completely independent development, the received wisdom that subjunctive selection is semantically motivated has morphed into widely shared “intuitions”, which are now serving as the linchpin for growing numbers of elaborate and highly complex semantic (and syntactic) theories (see e.g. the papers in Quer 2009). These have little if any basis in fact and are thus of little utility in evaluating emergence in this area of the grammar.

3.2 *Agreement in English*

A major interest of emergentists lies in identifying mechanisms that move toward creating structure. But the explanatory power of such mechanisms must ultimately reside in the extent to which they capture the facts of usage. Consider, for example, O’Grady’s (2008) intuitively attractive processing mechanism for subject-verb agreement in English. He proposes an

efficiency principle that requires that dependencies be resolved at the earliest opportunity. But variability in subject-verb agreement abounds in non-standard dialects all over the English-speaking world; in many, agreement in the third singular, as in (3), is the minority variant, while non-concord –s, as in (4), is robust, yielding just the sorts of patterns that, it is claimed, are “not found in any known human language” (O’ Grady 2008, 457).

(3) She comes sometimes and *passes* the night with me. (SE.009.171)

(4) Oh I live my life. I and Emma, and Aunt B. all- we all *lives* our life. (NPR.014.3424)

With existentials, community-based studies have shown that the standard agreement that the proposed processor is predisposed to generate has now been almost entirely replaced by *singular* agreement, regardless of the number of the noun, as in (5). Rates of occurrence of examples like (5a) reach as high as 80% (e.g. Meechan and Foley 1994; Adams 2005).

(5) a. I’m here and *there’s* things that I have to do. (QEC.172.366)

b. *There’s* a broken-down apartment building that people like to call the ghetto. (QEC.031.2245)

This is not to say that processing considerations are not operative; only that their contribution must be tested, along with that of other competing motivations for selecting standard singular agreement over plural, in multivariate models of spontaneous speech production data. We know for example, that the presence of a plural –s on the noun (e.g., *persons* vs. *people*) and the presence of a phrasal extension (6a) favor plural agreement, while adverbial extensions (6b) do not (Walker 2007).

(6) a. *There were* just a few of us sent back. (QEC.002.160)

b. *There was* no Irish people around him. (QEC.077.1921)

Such examples could be multiplied. (For phonological illustrations of “the gap between the analyst’s conception of the data and the actual data themselves”, see Ernestus and Baayen 2011: 374 ff.) Accommodating real variation and competing motivations can only refine the proposed processor, and in so doing, enrich our understanding of processing mechanisms. The point is that intuitions cannot begin to capture the robust and systematic -- but implicit -- community norms that dictate usage. It is to such community-based norms that we turn in the following sections.

4. Gauging emergence in discourse through the linguistic variable

How does emergence relate to variation, stable or changing? If emergence of a grammatical construction can be equated with an association between a form and a meaning, how strong does such an association have to be to qualify as having “emerged”? Any quantitative approach to linguistic usage involving counting will reveal associations. But the variationist construct of the linguistic variable, made up of competing variants, and the consequent principle of accountable reporting, which requires counting non-occurrences of the variant of interest as well as occurrences, enables us to assess both the extent to which a given form, once selected, actually signals a given function, as well as the extent to which it is associated with a given (variable) context. These measures need not be coterminous.

4.1 Marking tense and aspect in creoles

That emergence must be gauged not only from overall frequency of use, but from the competing motivations for variant choice is exemplified with the behavior of putative creole tense-aspect

markers. Aspectual distinctions in creole varieties are widely held to be marked by preverbal particles (e.g., Bickerton 1975; Bickerton 1984; Faraclas 1987). In English-based Creoles, these typically include forms like *kɔm* (7a) and *bin* (7b) as well as a zero variant (in (7a) and (7c)) in the following examples from Nigerian Pidgin English (Poplack and Tagliamonte 1996, ex. 7b, 7a, 6). On the aspectual account, *bin* in (7b) would be construed as expressing anterior or remote meaning, while zero (7c) would indicate a simple past.

- (7) a. *i kɔm go fɔ dat ples wey i ø kɔt di tri* (NPE.004/1019)
he went to the place where he had cut the tree from’
- b. *a no wan kɔm. a bin de kɔm go* (NPE.008.147)
‘I did not want to come. I had been coming and going’
- c. *imidetly wey de ø si de demɔsretɔs, de ø ŋut* (NPE.010.270)
‘immediately they saw the demonstrators, they shot’

In their study of the grammaticalization of zero and other forms into past markers in Nigerian Pidgin English, Poplack and Tagliamonte (1996) tested these claims in spontaneous speech data. Defining the variable context broadly as the functional domain of past temporal reference and applying the principle of accountability, six distinct options were found to be available, summarized in Table 1. To assess the claimed association between *bin* and anterior or remote meaning, all tokens of all six forms were coded as occurring in one of two temporal relationship contexts, Anterior (when Event 2 was ordered before Event 1, as in (7b)) or Sequential (when Event 1 was ordered before Event 2, as in (7c)), with the residue classed as other non-anterior.

Table 1 reproduces the relevant parts of a “variable rule” analysis (Rand and Sankoff 1990; Sankoff, Tagliamonte and Smith 2005) of competing motivations for choice of pre-verbal marker. The procedure uses logistic regression to perform binomial multivariate analysis for a

choice of the “1” (application) variant vs. the “0” variant (the associated non-occurrences); this determines the factor groups that together account for the largest amount of variation, in terms of stepwise increase of log likelihood, such that the addition of any of the remaining factor groups does not significantly improve the fit to the model (Sankoff 1988b). The corrected mean indicates the overall probability that the application variant will occur when no competing factors are considered. Probabilities vary between 0 and 1; the higher the value, the more favorable the effect.

Table 1 confirms that the *relative* probability that *bin* will be selected in anterior, as opposed to other, temporal relationship contexts is in fact very high, as indicated by the factor weight of .90 in the left *bin* column in the Anterior row. Note, however, that its probability of occurrence in *non*-anterior contexts (other than Sequential) is also quite high (.65). More important, its *absolute* probability of occurring in that (or any other!) context is extremely low, as shown by the corrected mean of .004 (corresponding to a frequency relative to its competitors of 1.5% [73 tokens/4759 past temporal reference contexts]).

Table 1: *Variable rule analyses of the contribution of temporal relationship to selection of pre-verbal markers in Nigerian Pidgin English* (adapted from Poplack and Tagliamonte 1996; Shading indicates the probability of occurrence when the combined effects of corrected mean and factor weight are considered)

Total N = 4692	<i>kɔm</i>		<i>dɔn</i>		<i>Bin</i>		<i>fɪniʃ</i>		<i>de</i>		<i>Zero</i>	
<i>Corrected mean</i>	.19	.19	.07	.07	.004	.004	.012	.011	.07	.07	.57	.57
TEMPORAL RELATIONSHIP												
Anterior	.20	.06	.76	.20	.90	.04	[]	.01	.50	.07	.55	.62
Sequential	.70	.35	.28	.03	.21	.00	[]	.01	.40	.05	.47	.53
Non-anterior	.41	.14	.65	.12	.65	.01	[]	.01	.64	.12	.51	.58

Only once *bin* is situated with respect to the six other variants with which it competes, and the combined effect of frequency and probability of occurrence taken into account (the shaded columns in Table 1), can it be seen to be less likely to occur in this context than every other

variant but one (*finiš*). A narrow focus on the “emerging” construction alone would have missed the crucial fact that anteriority is overwhelmingly *not* expressed by *bin* – indeed, given its infinitesimal rate of occurrence, how could it be? Nonetheless, it already displays that aspectual function, even at such an incipient stage: on the rare occasions where it is selected, it will almost certainly be to express anteriority (independent of the fact that at this point, nearly every other variant is more likely to do so). *Bin* is *emerging* as an expression of anteriority, but has not yet achieved the status of a “marker” (where one would expect the meaning to be overwhelmingly [if not obligatorily] expressed by that form). The stark contrast between this example of emergence and the case of the subjunctive variant is particularly instructive: though the latter is far more robust, at an overall rate of 76% under subjunctive triggers, it is *never* used to signal the readings analysts attribute to it.

5. Characterizing emergence in discourse

5.1 *The emergence of social meaning*

Social context is an oft-cited source for the emergence of linguistic structure (e.g. MacWhinney 2001, 453-4). But what is the mechanism involved? In this section we consider how the linguistic variable helps gauge the emergence of social meaning in the speech community. The most striking evidence of the links between social categories and linguistic ones has come from the investigation of sound change beginning in the 1960s (e.g. Labov 1963) and more recently from sociophonetic research, (e.g. Foulkes and Docherty 2006; Foulkes and Hay this volume). Here we illustrate social values as embodied by morphosyntactic constructions.

Consider the Spanish periphrases with a gerund, whose suffix is *-ndo*. In Mexican varieties, *andar* ‘go (around)’ + VERB-*ndo* (8a) competes with *estar* ‘be (located)’ + VERB-*ndo* (8b) as an expression of progressive and other imperfective aspects.

- (8) a. *Ando busca-ndo* unas tijeras. (UNAM 1976, p. 415)
‘I am looking for a scissors.’
- b. *Estás habla-ndo* de una forma de vida. (UNAM 1971, p. 261)
‘You are talking about a way of life.’

Choice between the variant gerund constructions is not semantically conditioned, but rather is constrained by collocations that are the residue of the uses of the source constructions and by social distributions (Torres Cacoullos 2001). With respect to collocations, multivariate analysis shows that the *andar* construction is favored with classes of gerunds denoting motion and physical activities, while the *estar* variant is more probable with verbs of speech like *hablando* ‘talking’, as well as verbs denoting perceptible bodily activities (*llorando* ‘crying’) and mental activities (*pensando* ‘thinking’). Similarly to the role of *falloir* in the French subjunctive, the variation here is in part shaped by particular collocational routines, such that *andar* + *buscando* ‘looking for’ is the conventional way to express ‘be looking for something’ (81% (13/16) *buscando* tokens occur with *andar*, though this is the minority variant, representing 20% (177/873) of all gerunds). With respect to social distributions, the *andar* variant has a higher probability of occurrence among working class speakers.

These collocational and social associations are related, deriving from the lexical sources of the two gerund constructions. Both have grammaticalized from spatial expressions. In Old Spanish, *andar* + VERB-*ndo* meant literally ‘to go around VERB-ing’, tending to co-occur with locative expressions with the preposition *por* ‘along, around, all over’ (Torres Cacoullos 2000). Today, it is simply another way of expressing imperfective aspects in Mexican varieties.

Nevertheless, retention of the source-construction ‘going around’ meaning in *andar* + VERB-*ndo* is reflected in its contemporary tendency to co-occur with verbs denoting motion and physical activities, particularly outdoor activities, such as *dando la vuelta* ‘going / strolling / driving around’ and *trabajando* ‘working’ when the work is in the fields, since rural activities in large spaces are more compatible with the original movement meaning of *andar* + VERB-*ndo*. These distributions suggest that what has emerged as a socially stratified variant in Mexican Spanish—with a relative frequency higher among the working classes—originated in an indoor/outdoor, urban/rural difference, which itself reflects retention of source-construction meaning (Torres Cacoullos 2001). Neutralization-in-discourse of the original spatial meaning distinction enabled the eventual aspectual variants to develop new social associations, albeit with collocational preferences.

Exemplar models, which hold that representations of general constructions are built up from instances that include real-world information about speakers and situations, might predict the emergence of such a social meaning for a linguistic form. But we would not have been able to ascertain its specific trajectory without systematically tracking correlations with linguistic and social contextual features in historical and present-day community-based production data.

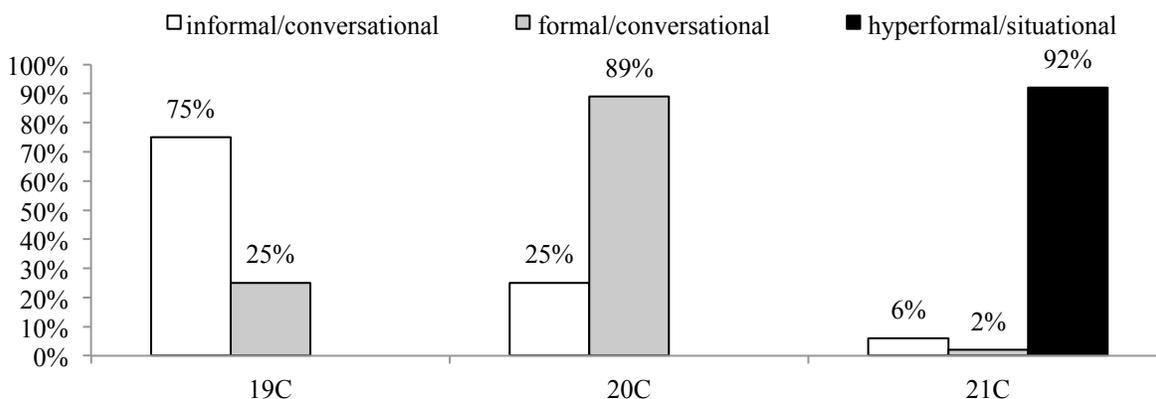
5.2 *The emergence of stylistic meaning*

5.2.1 The French negative “marker” *ne*

Another example of emergence in the speech community, this time of stylistic meaning, comes from the discontinuous French negative construction *ne ... pas*. The grammaticalization of *pas* ‘step’ from negative intensifier to the major if not sole negator and erosion of *ne* in many dialects of Modern French is a well-documented development. Most scholars think the story ended there.

But in spoken Québec French, not only did *ne* not disappear, it has been hanging on, against all odds, at an infinitesimal rate of approximately .02% (178 tokens of *ne* out of 78,399 negative contexts in which it could have occurred) since the latter half of the 19th century at least (Poplack and St-Amand 2007; Poplack to appear). Figure 2 depicts its emergence from obsolescing negation marker to stylistic marker of formal speech.

Figure 2: *Distribution of ne according to context over time*



In the 19th-century corpus the distribution of *ne* shows no particular association with topic of conversation (nor indeed, with any other explanatory factor, consistent with linguistic obsolescence). It simply appeared, if very sporadically, in garden-variety negative sentences like (9).

- (9) Il cherche une secousse, mais il *ne* trouve rien. (RFQ.018.1067)
 ‘He looks for a while, but he doesn’t find anything.’

By the 20th century, despite no change in overall miniscule rate, 89% of the 152 tokens of *ne* were concentrated in formulaic uses like (10a), or in formal conversations dealing with topics like religion, education or language itself (10b), where speakers are known (Labov 1972) to monitor their variant choices.

- (10) a. Si vous désirez plus de renseignements, n’hésitez pas à téléphoner.
 (OH.119.1941).
 ‘If you wish more information, do not hesitate to phone.’

- b. On a vieilli en faisant le choix de bien parler notre français ou de *ne* pas bien le parler. (OH.115.1217)
‘We grew up making the choice between speaking our French well, or not speaking it well.’

We can no longer reconstruct the exact events that triggered this situation, but the advent of compulsory education in the second half of the 20th century surely played a role. Once *ne* had disappeared from speech, it could only be acquired through formal instruction, to which the 19th-century speakers had virtually no access. The normative dictate, explicitly taught in school, that every postverbal negator requires a preverbal *ne*, coupled with the extreme rarity of the latter, together conspired in its recent conversion from highly sporadic negative morpheme to emergent marker of formal speech. This particular stylistic function of *ne* has become further entrenched in the 21st century, as evidenced by the behavior of contemporary high school students (Poplack to appear). At 0.2% (14/7,645), their overall rate of *ne* has remained unchanged since the 19th century; if anything Figure 2 shows a *decrease* with formal topics in everyday conversation. This is because they have promoted the formality condition to a *hyperformality* condition, reserving their selection of *ne* for in-school use under extreme circumstances, as when the quality of their French is being explicitly tested, exemplified in (11).

- (11) La planète est déjà détruite avec tout ce qui est- s’est passé dans les années antérieures. Il n’y a rien à faire. (FeC.546.547)
‘The planet is already destroyed with everything that is- has occurred in previous years. There is nothing to be done.’

5.2.2 The French interrogative particle *est-ce-que*

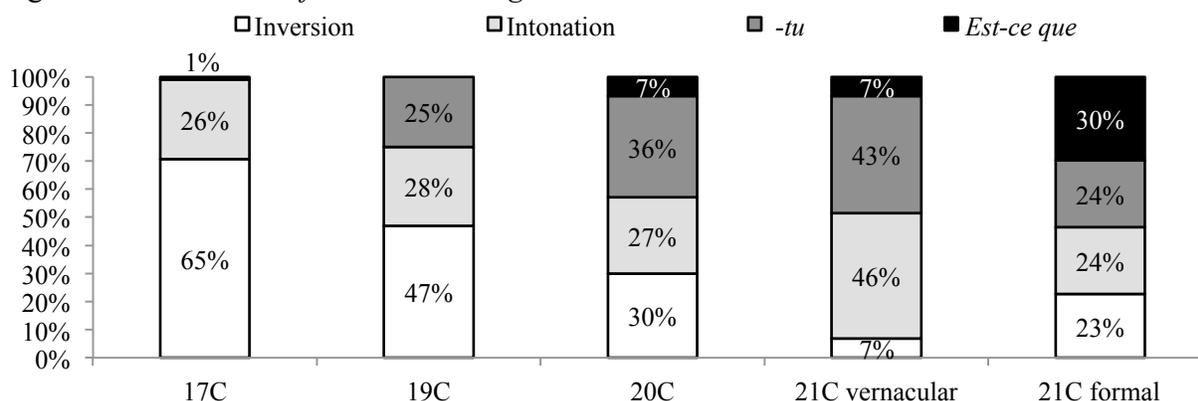
A similar situation has arisen with the preposed interrogative particle *est-ce que*, lit. ‘is it that’ (12a), one of four options for forming basic polarity questions including verb-subject inversion

(12b), postposition to the verb of the (currently non-standard) interrogative marker *-tu* (12c), and rising intonation (12d).

- (12) a. Mes bombes *est-ce que* je les largue ici? (OH.078.1502)
 ‘My bombs, do I throw them here?’
- b. *As-tu* déjà parlé avec un vrai Français de France là? (OH.105.2768)
 ‘Have you ever spoken to a real Frenchman from France?’
- c. *Tu vas-tu* être plus marié oubedonc moins marié? (OH.079.1471)
 ‘Are you gonna be more married or less married?’
- d. Ah, toi *tu restes* pas avec tes parents? (OH.112.1819)
 ‘Oh, you don’t live with your parents?’

Est-ce que emerged as an interrogative particle in the 16th century, and although it was soon promoted by the *Académie Française*, it was still extremely rare in the 17th-century language presumably transplanted to Canada. It remained that way in the vernacular 19th-century Québec French materials studied by Elsig and Poplack (2006), at an overall rate of only 1%. By the 20th century, although still the minority variant by far, *est-ce que* had enjoyed a small increase (to 7%; Figure 3).

Figure 3: *Distribution of French interrogative variants over time*



Now, however, whether spurred by the same considerations abetting the conversion of *ne* into a stylistic marker, or in imitation of the (revered) example of Metropolitan French, the factor of style exerts the one of the strongest influences on selection of interrogative marker in a

multivariate analysis of the contribution of a number of competing motivations for variant choice (Table 2).

Table 2: *Variable rule analyses of the contribution of factors to selection of interrogative variant in 20th-century Québec French* (excerpted from Tables 5 and 6, Elsig and Poplack 2006)

	Inversion	Intonation	-tu	<i>Est-ce que</i>
STYLE				
Careful	.44	.52	.46	.64
Casual	.61	.45	.58	.24
AGE				
35+	.56	[]	.42	1 (knockout)
15-34	.37	[]	.69	
SOCIOECONOMIC CLASS				
Upper middle	[]	[]	[]	.85
Working	[]	[]	[]	.16

The prestige connotations of *est-ce que* underlying its selection in formal speech styles emerge from its very strong associations with the upper middle classes (Probability .85) and older speakers (Probability 1). The succeeding generation of 21st-century students maintains the 20th-century community rate of 7% in their normal everyday vernacular but augments it to 30% in the most formal French they can muster – the variety deemed appropriate for an oral French exam (Figure 3). As with *ne*, here too a quantitatively minor variant is marshaled in the conversion of a formality marker into a hyperformality marker over time.

Summarizing, the case studies in this section illustrate how embedding forms with respect to both the variable linguistic structure and the speech community in which they are used yields a fine-grained characterization of social and stylistic emergence. Which variant will be pressed into service for which function is an empirical question, and may well differ from community to community. Our studies suggest that the mechanisms underlying the assumption of social and stylistic meanings include neutralization of erstwhile semantic distinctions, shifts in

overall relative frequency as one variant overtakes the other(s), and strengthening associations with a specific class of speakers or extra-linguistic contexts.

In the next two sections we employ the notion of the linguistic variable to probe measures and mechanisms of internal emergence from diachronic and synchronic perspectives, by examining the processes of grammaticalization and analogical extension.

6. Linguistic structure from linguistic structure: mechanisms of language-internal emergence

6.1 Diachronic priming and loss of compositionality-analyzability

In this section we apply the variationist method to determine the role of structural priming in the emergence of new units, which involve shifts in internal (constituent) structure. As seen in the pair of examples in (13), variant expressions of present progressive in Old Spanish texts were the Progressive construction, *estar* ‘be (located)’ + VERB-*ndo* (gerund) (13a), and the simple Present (13b). Latin did not have a dedicated morpheme or construction for progressive aspect, the Present serving this function, among others (Allen and Greenough 1916, 293 §465).

- (13) a. no sabemos quién está dentro; *habla-ndo están* (PROG). (Celestina, Act XIV, 15th c.)
‘we don’t know who is inside; they *are talking*’
- b. escucha, que *hablan* (PRS) quedito (Celestina, Act XII)
‘listen, they *are speaking* softly’

Defining the variable context broadly as the domain of present temporal reference and adhering to the principle of accountable reporting (Labov 1972, 72), Torres Cacoullos (2012; to appear) submitted occurrences of the Progressive, as well as *non*-occurrences, where the simple Present

materialized instead (as in (13b) and (14b)), to independent multivariate analyses for three time periods—the 13th-15th, 17th, and 19th centuries. We review three competing motivations for variant selection: co-occurrence of locative expressions, aspectual reading, and structural priming.

The locative co-occurrence factor group probes the crosslinguistic grammaticalization path according to which progressives originate as constructions meaning “be located at” (Bybee, Perkins and Pagliuca 1994, Ch. 4). The prediction that follows from the twin hypotheses of retention of lexical source construction meaning, on the one hand, and on the other, of semantic bleaching with advancing grammaticalization, is that a co-occurring locative will favor choice of the Progressive, but that this favoring effect should weaken over time.

The aspect factor group operationalizes the grammaticalization hypothesis on the development of zero morphemes (Bybee 1994). According to this, as *estar* + Verb-*ndo* is used more frequently, the inference develops that *not* using it means non-progressive meaning, with the result that the simple Present is restricted to habitual contexts (e.g., (14), that is, it becomes a zero morpheme indicating habitual aspect (as presumably is already the case in English, where Present tense *I drink decaf* = habitually).

- (14) a. quando non han con quién fablar, *están fabla-ndo* (PROG) consigo mesmas entre sí
 (Corbacho, Part II, XII, 15th c.)
 ‘when they don’t have anyone to talk to, they *talk* to themselves’
- b. Que a quien más quieren, peor *hablan* (PRS) (Celestina, Act VI, 15th c.)
 ‘To those they love best they *speak* worst’

As indicated by the Probability values in Table 3, direction of effect has generally remained stable, indicating continuity in the linguistic conditioning of the Progressive, which is favored in the presence of a locative expression and situations of limited duration (as in (13)).

The main locus of change is in the relative *magnitude* of effect, indicated here by the *Range* between the highest and lowest Probability within a factor group (shown in italics). The *Range* values for locative co-occurrence and aspect are close in the 13th-15th century analysis, but in the 17th century the *Range* for aspect is nearly twice as great ($54:29 = 1.9$) and in the 19th century, it is three times greater ($59:19 = 3.1$) than that of locatives. Clearly the aspectual effect has grown stronger over time, while the locative effect has weakened.

Table 3: *Variable rule analyses of the contribution of factors to selection of Progressive over time*

	13 th -15 th centuries	17 th century	19 th century
LOCATIVE CO-OCCURRENCE			
Present	.90	.76	.67
Absent	.47	.47	.49
<i>Range</i>	<i>43</i>	<i>29</i>	<i>18</i>
ASPECT			
Limited duration	.68	.70	.71
Extended duration	.35	.16	.12
<i>Range</i>	<i>33</i>	<i>54</i>	<i>59</i>
STRUCTURAL PRIMING			
Preceding <i>estar</i> construction	.76	.69	[.49]
Preceding “other” tenses	.54	.53	[.56]
Preceding simple Present	.46	.47	[.46]
<i>Range</i>	<i>30</i>	<i>22</i>	

What the analysis of variation highlights is that *estar* + VERB-*ndo* is increasingly *disfavored* in extended duration (habitual, indefinitely existing state) contexts, as Probability values get closer to zero (at .35, .16 and .12, respectively, in the 13th-15th, 17th and 19th centuries). This shows that *estar* + VERB-*ndo* does not come in as a full-blown progressive, merely increasing in rate in progressive contexts, thus concomitantly restricting the zero-marked simple Present to habitual. Rather, comparison of linguistic conditioning over time shows us that the originally more locative *estar* + VERB-*ndo* is increasingly used as an aspectual expression of limited—not extended—duration, emerging in this way as a progressive.

The weakening of the favoring effect of co-occurring locatives (“semantic bleaching”; e.g., Givón 1975) provides a measure of the loss of source-construction meaning in the course of grammaticalization. We see in Table 3 that the Probability values for presence of a locative get farther from 1 over time (at .90, .76 and .67, respectively). We construe this as loss of *compositionality*, which has to do with transparency, or the degree to which the meaning of the whole is predictable from the meaning of the component parts (Bybee 2010, 45). If in the early locative construction the meaning was ‘be at while VERB-ing’, the semantic contribution of *estar* (< Latin *stare* ‘stand (still), remain’) would have been more apparent. The heightened aspectual meaning of *estar* + VERB-*ndo* is less transparent than the locative meaning of the source construction.

We turn now to structural priming, a mechanical effect which is pervasive in language variation, as shown in variationist (beginning with Poplack 1980; Weiner and Labov 1983; Scherre and Naro 1991) and psycholinguistic (e.g., Bock 1986) studies. To obtain a measure of *analyzability* (i.e., the degree to which internal structure is discernable (Bybee 2010, 45)), tokens were coded as having an *estar* construction, a simple Present, or another finite verb form in the immediately preceding clause. If *estar* + VERB-*ndo* is analyzable—with discernable internal structure and component parts that are recognizable as individual words, namely *estar* and the gerund of another verb—it should be primed by other constructions composed of *estar* and another element. *Estar* constructions of the schematic form *estar* + X include locative, predicate adjective, and resultative (15) constructions.

- (15) están cocidas con sus garbanzos, cebollas y tocino, la hora de ahora *están dicie-ndo* (PROG): “¡cómeme!, cómeme!” (Quijote II, LIX, 17th c.)
‘They are cooked with their chickpeas, onions and bacon and now *are saying* “eat me, eat me!”’

Returning to Table 3, we see that the Progressive is in fact favored by a preceding non-Progressive *estar* construction through the 17th century. This is as predicted, if the Progressive has an analyzable internal structure. Since *estar* constructions have independently increased in frequency (at least to the detriment of copula *ser* ‘be’ [Silva-Corvalán 1994, 94-95]), the priming of the Progressive by other *estar* constructions revealed here may have spurred its emergence.

In the 19th century, however, which, as we saw, is when the aspectual constraint is strongest as the Progressive is most disfavored in extended duration contexts, the priming effect is no longer significant. The disappearance of the earlier priming effect is consonant with diminished analyzability of the whole. Together with attenuated compositionality, which was manifested in the weakening of the locative co-occurrence effect, this result is indicative of increasing cohesion of *estar* + VERB-*ndo* as a new unit.

In this section, then, we have marshaled evidence from changes in the *strength* of competing motivations for variant choice to illustrate the emergence of a progressive-nonprogressive opposition from the gradual aspectual divergence between variants, with accompanying loss of compositionality and analyzability of the emerging form.

6.2 *Emergence through analogical extension*

Another internal avenue for the emergence of linguistic structure is through analogical extension of an existing pattern to a novel context. The pattern examined in this section involves the variable occurrence of phrase-final (“stranded”) prepositions in Québec French, as illustrated by *avec* ‘with’ in the relative clause in (16).

- (16) J’avais pas personne à parler *avec* (STRD). (OH.013.1964)
‘I had no one to talk *with*.’

This strategy is inadmissible in Standard French, which requires instead, for relative clauses, that the preposition be preposed (“pied-piped”) to its complement, as is *avec* in (17).

- (17) Les anglaises *avec* (PP) qui je parlais, ils le croyaient pas. (OH.082.1695)
‘The Anglophones *with* whom I was talking, they didn’t believe it.

Large-scale study of patterns of preposition placement in the speech community (Poplack, Zentz and Dion 2012) shows that speakers do in fact prepose their prepositions much (37%) of the time, but even more frequently (51%) resort to an alternate non-standard strategy of omitting (“absorbing”) them altogether, as in (18).

- (18) (IVer: Pourquoi tu as changé [d’école]?) Pour faire quelque chose de différent, du nouveau monde à parler [] (ABSORBED). (OH.002.040)
(Why did you change [schools]?) ‘To do something different, new people to talk [].’

Indeed, stranding is only a minor phenomenon in Québec French, accounting for no more than 12% of all preposition placement in relative clauses. Nonetheless, it is a highly salient and stigmatized phenomenon in the community, thanks in large part to formal similarity to the ubiquitous (99%), stranded prepositions in English, illustrated in (19).

- (19) And this is the guy I’ve always had a crush *on*. (QEC.301.1372)

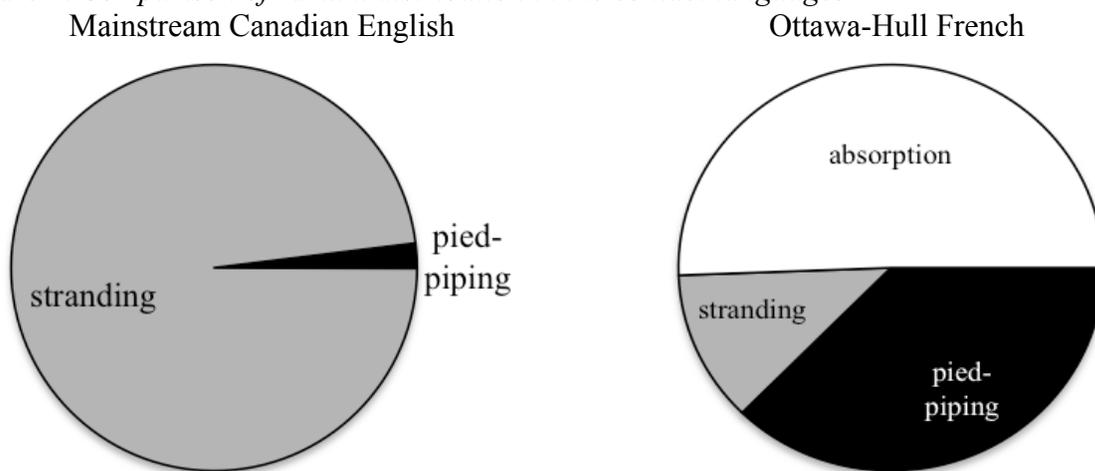
The fact that English is the majority language of Canada, and that Québec French has been in intense contact with it for over two centuries, coupled with the twin assumptions that language contact begets change in the minority language and that stranding constitutes one such change, together conspire in the widespread belief that these bare prepositions emerged in the French relative system through contact-induced change.

Poplack, Zentz and Dion (2012) applied variationist methodology to test that hypothesis, by confronting it with other possible motivations for the selection of stranding over the competing strategies of pied-piping and absorption. Some of those motivations relate to the

preposition itself (e.g. lexical identity, semantic weight), others relate to the verb (semantic contribution, lexical identity), and still others to the complement (type, humanness of the complement NP, proximity and place of preposition with respect to it). Variable rule analyses were employed to reveal which of them contributed statistically significant effects to variant choice when all competitors were considered simultaneously, thus contextualizing the behavior of the stranded variant with respect to the remainder of the linguistic system.

Results showed that the overwhelming determinant was a combined factor of semantic weight/lexical identity of the preposition: only “strong” prepositions are stranded, while “weak” prepositions virtually never are. Systematic comparison with patterns of preposition placement in a variety of mainstream Canadian English which would arguably have constituted the model for the putative contact-induced change showed a very different picture. In terms of rates of occurrence, here stranding is almost categorical, while in French it is the minority variant. The French majority variants, pied-piping and absorption, are virtually if not entirely non-existent in English (Figure 4). In terms of lexical identity of the relativizer, stranding occurs freely in English while in French it is highly restricted.

Figure 4: *Comparison of variant distribution in the contact languages*



Most important is lexical identity of the preposition, depicted in Table 4. In French, the few that are ever stranded figure among the cohort labeled *strong* (which in turn is almost entirely constituted of the four lexical prepositions *avec, pour, dedans, dessus*), while the two prepositions making up most of the *weak* class (*à* and *de*) are never stranded, but overwhelmingly absorbed.

Table 4: *Distribution of preposition placement strategies by semantic weight of preposition in the contact languages* (Poplack, Zentz and Dion 2012, Table 12)

WEIGHT	ENGLISH		FRENCH					
	Stranding		Stranding		Absorption		Pied-piping	
	%	N	%	N	%	N	%	N
Strong	99	68/69	50	39/78	19	15/78	31	24/78
Weak	97	186/254	1	2/262	60	157/262	39	103/262
Total	98	254/260	12	41/340	51	172/340	37	127/340

No such distinction operates in English, where stranding is equally probable with all prepositions, weak ones (20) included. These results indicate that despite superficial similarities across the two languages in structural *form*, the *pattern* of preposition stranding could not have come from English.

(20) So- that’s- pretty much all I could think *of* (STRD). (QEC.046.806)

What then is its source? In attempting to situate bare prepositions within the French grammatical system, we come across a superficially similar native process, exemplified in (21), known as *orphaning* (e.g. Bouchard 1982; Zribi-Hertz 1984; Roberge 1998).

(21) Oui mais, il veut pas payer *pour* (ORPHAN). (OH.013.260)
 ‘Yes but, he doesn’t want to pay *for*.’

Orphaning is an independent process, differing from stranding not only in terms of the contexts in which it appears, but also in terms of the variants with which it competes. Poplack, Zentz and Dion (2012) made use of this (perfectly standard) model for bare prepositions in *non-contact* French to determine the trajectory by which stranding emerged in relative clauses. Comparing

the grammar of native orphaning, as instantiated by the quantitative conditioning of variant choice, with that of the putatively borrowed stranding, they reasoned that if the two phenomena obeyed the same linguistic constraints, this would support the analysis that phrase-final placement of prepositions in relative clauses is an internally-motivated extension of orphaning to a new context, and not the product of contact-induced change.

Multivariate analysis of the competing motivations for orphaning revealed that greatest effect by far was contributed by semantic weight, in the same pattern observed for stranding: weak prepositions are never orphaned; only their semantically strong counterparts are. And here too, the label *semantic weight* reflects the same idiosyncratic lexical effect observed for prepositions in relative clauses. The pivotal result is that only a few of the many theoretically eligible prepositions are *actually* orphaned in orphaning contexts, and these are the *same* few that tend to be stranded in relative clause contexts.

Having operationalized the notion of analogy in terms of parallel contextual effects on variant choice, it was possible to test this competing motivation and demonstrate that the emergence of bare prepositions in the context of French relative clauses proceeded on the basis of analogy with the pre-existing conventionalized phrases with orphan prepositions in the transitive context. Again, it is the fine conditioning of variant choice in discourse—and its parallels across two independent preposition placement strategies—that enabled us to characterize the mechanism at work.

7. Conclusion: emergence via the linguistic variable

In this chapter we have attempted to evaluate the role of variability in understanding the phenomenon of linguistic emergence. We have argued that going beyond the variability in *rates* readily observable amongst speakers and situations to the *structural* variability inherent in the utterances of a single individual in a single situation opens the door to observing emergence on the ground. A key prerequisite to this endeavor is real speech, situated in social context. We have shown, here and elsewhere, that linguistic lore based on intuitions about forms and functions may be far removed from distributions and motivations for variant choice actually operative in discourse.

We first took on the fundamental, though underexplored, question of what the process of emergence looks like in actual interaction. We proposed that in discourse, morphosyntactic emergence is detectible—and measurable—in the favoring effect of contextual factors that operationalize specific meanings or functions on the selection of a given variant over any of its competitors in the same (“variable”) context. Approached empirically in this way, we could then turn to the crucial issues of the extent of, the competing motivations for, and the mechanisms involved in social and linguistic emergence.

Consideration of both relative and absolute probability of occurrence enabled us to characterize the *extent* of emergence of a form-meaning pairing that has not yet achieved the status of a grammatical marker. This is the case of Nigerian Pidgin English pre-verbal *bin*, with its strong, but crucially non-privative, association with anteriority.

Weighing *competing motivations*, we have shown that putative semantic or functional factors often conceal lexically particular constructions, which turn out to contribute the pertinent

effects, as with Québec French [*falloir* + subjunctive] and Mexican Spanish [*andar* + gerund of outdoor physical activity]. And even when the grammatical or semantic role of a general construction is obsolete (or never obtained, as with the French subjunctive), forms may acquire social-stylistic meaning, as with erstwhile preverbal negator *ne*, now emergent hyperformality marker in Québec French.

Finally, confronting the *structure* of inherent variability directly allowed us to characterize *mechanisms* of emergence. Evidence from the diachrony of the Spanish Progressive suggests that emergence of a grammatical unit-in-becoming is manifested in loss of compositionality and analyzability, as gauged by structural priming from related constructions, which we found to apply in earlier rather than later stages of grammaticalization. In the case of much maligned preposition stranding in Québec French relative clauses, comparison of the fine details of linguistic conditioning of variable preposition placement in this context with those of a similar pre-existing native pattern in another context revealed the operation of language-internal constructional analogy. Similarly, by assessing shifts in the strength of probabilistic contextual constraints, we were able to propose that emerging tense-aspect expressions, here the Spanish Progressive, develop not only via increased frequency, but also via increasing functional differentiation from an existing alternative form.

We stress that all of these patterns would have been invisible had we limited our view of linguistic structure to the emergent variant alone, and simply to its overall rates of occurrence. Labov (2001, xvi) reminds us that “it is hard to understand the world by rising above it”. Emergentism can only be enriched by embracing inherent variability and the well-honed variationist tools for tackling it, especially if the interplay between production and processing is to be firmly established. We hope to have made a case for the imperative of a quantitative view

from the ground of natural language production, based on speech situated in its social and linguistic context.

References

- Adams, Jackie. 2005. “Concord variation, convergence, and Québec English: “*There’s lots of things to consider*”.” M.A. mémoire, University of Ottawa.
- Allen, J. Henry, and James B. Greenough. 1916. *Allen and Greenough’s new Latin grammar for schools and colleges, founded on comparative grammar*. Boston [etc.]: Ginn & Company.
- Bates, Elizabeth, and Brian MacWhinney. 1987. “Competition, variation, and language learning.” In *Mechanisms of language acquisition*, edited by Brian MacWhinney, 157-194. Hillsdale, NJ: Lawrence Erlbaum.
- Bickerton, Derek. 1975. *Dynamics of a creole system*. New York: Cambridge University Press.
- Bickerton, Derek. 1984. “The language bioprogram hypothesis.” *The Behavioral and Brain Sciences*, 7(2): 173-221.
- Beckner, Clay, Rickard Blythe, Joan L. Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann. 2009. “Language is a complex adaptive system.” Position paper, *Language Learning*, 59, Supplement 1: 1-27.
- Bock, Kathryn J. 1986. “Syntactic persistence in language production.” *Cognitive Psychology*, 18: 355-387.
- Bouchard, Denis. 1982. “Les constructions relatives en français vernaculaire et en français standard: étude d’un paramètre.” In *La syntaxe comparée du français standard et populaire: approches formelle et fonctionnelle, volume 1*, edited by Claire Lefebvre, 103-133. Québec: Éditeur officiel du Québec.
- Brown, Esther, and Javier Rivas. 2012. “Grammatical relation probability. How usage patterns shape analogy.” *Language Variation and Change*, 24(3): 317-341.
- Bybee, Joan L. 1994. “The grammaticalization of zero: Asymmetries in tense and aspect systems.” In *Perspectives on Grammaticalization*, edited by William Pagliuca, 235-254. Amsterdam and Philadelphia: John Benjamins.
- Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan, Revere Dale Perkins, and William Pagliuca. 1994. *The Evolution of Grammar*. Chicago: University of Chicago Press.
- Cedergren, Henrietta, and David Sankoff. 1974. “Variable rules: performance as a statistical reflection of competence.” *Language*, 50(2): 233-255.
- Cervantes Saavedra, Miguel de. 1547-1616. *Don Quijote de la Mancha*. Biblioteca Virtual Miguel de Cervantes, <http://www.cervantesvirtual.com/>.
- Croft, William. 2010. “The origins of grammaticalization in the verbalization of experience.” *Linguistics*, 48: 1-48.
- Dąbrowska, Eva, and Elena Lieven. 2005. “Towards a lexically specific grammar of children’s question constructions.” *Cognitive Linguistics*, 16: 437-474.

- Ellis, Nick. C. 2012. “Emergentism.” In *The Encyclopedia of Applied Linguistics: Cognitive Approaches*, edited by Carol A. Chapelle (General Editor), Marianne Gullberg and John Williams (Area Editors), 1873-1882. Wiley-Blackwell.
- Elsig, Martin, and Shana Poplack. 2006. “Transplanted dialects and language change: question formation in Québec.” *Penn Working Papers in Linguistics*, 12(2) (Selected Papers from NWAV 34): 77-90.
- Ernestus, Mirjam, and R. Harald Baayen. 2011. “Corpora and exemplars in phonology.” In *The handbook of phonological theory*, 2nd edition, edited by John A. Goldsmith, Jason Riggle, and C. L. Alan, 374-400. Wiley-Blackwell.
- Faraclas, Nicholas G. 1987. “Creolization and the tense-aspect-modality system of Nigerian Pidgin.” *Journal of African Languages and Linguistics*, 9(1): 45-59.
- Foulkes, Paul, and Gerard Docherty. 2006. “The social life of phonetics and phonology.” *Journal of Phonetics*, 34: 409-438.
- Gerli, Michael, ed. 1978. *Corbacho* (1438, Alfonso Martínez de Toledo). Madrid: Cátedra.
- Givón, Talmy. 1975. “Serial verbs and syntactic change: Niger-Congo.” In *Word order and word order change*, edited by Charles N. Li, 47-112. Austin: University of Texas Press.
- Haiman, John. 2011. “Competing motivations”. In *Oxford Handbook of Typology*, edited by Jae Jung Song, 148-165. Oxford: Oxford University Press.
- Labov, William. 1963. “The social motivation of a sound change.” *Word*, 19: 273-309.
- Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, William. 1969. “Contraction, deletion and inherent variability of the English copula.” *Language*, 45(4): 715-762.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 1982. “Building on Empirical Foundations.” In *Perspectives on historical linguistics*, edited by Winfred P. Lehmann and Yakov Malkiel, 17-92. Amsterdam and Philadelphia: John Benjamins.
- Labov, William. 2001. *Principles of Linguistic change, volume III: Cognitive and cultural factors*. Oxford: Wiley Blackwell.
- Labov, William. 2004. “Quantitative Analysis of Linguistic Variation.” In *Sociolinguistics/Soziolinguistik: an international handbook of the science of language and society, 2nd edition, volume 1*, edited by Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier and Peter Trudgill, 6-22. Berlin: de Gruyter.
- MacWhinney, Brian. 2001. “Emergentist approaches to language.” In *Frequency and the emergence of linguistic structure*, edited by Joan L. Bybee and Paul Hopper, 449-470. Amsterdam: Benjamins.
- MacWhinney, Brian. In press. “Competition across time.” In *Competing motivations in grammar and usage*, edited by Brian MacWhinney, Andrej L. Malchukov and Edith A. Moravcsik. Oxford: Oxford University Press.
- Meechan, Marjory, and Michele Foley. 1994. “On resolving disagreement: Linguistic theory and variation – There’s bridges.” *Language Variation and Change*, 6: 63-85.
- O’Grady, William. 2008. “The emergentist program.” *Lingua*, 118: 447-464.
- O’Grady, William. 2010. “Emergentism.” In *The Cambridge Encyclopedia of the Language Sciences*, edited by Patrick Hogan, 274-76. Cambridge, UK: Cambridge University Press.

- Poplack, Shana. 1980. “Deletion and disambiguation in Puerto Rican Spanish.” *Language*, 56(2): 371-385.
- Poplack, Shana. 1989. “The care and handling of a mega-corpus.” In *Language change and variation*, edited by Ralph Fasold and Deborah Schiffrin, 411-451. Amsterdam: Benjamins.
- Poplack, Shana. 1992. “The inherent variability of the French subjunctive.” In *Theoretical analyses in Romance linguistics*, edited by Christiane Laeufer and Terrell A. Morgan, 235-263. Amsterdam: Benjamins.
- Poplack, Shana. 2011. “A variationist perspective on grammaticalization.” In *Handbook of grammaticalization*, edited by Bernd Heine and Heiko Narrog, 209-224. Oxford: Oxford University Press.
- Poplack, Shana. To appear. “L’École, la communauté et la relation symbiotique entre la diaphasie et la diastratie”. In *Variations diasystémiques et leurs interdépendances*, edited by Lene Schøsler and Jan Lindschouw. Cambridge Scholars Publishing.
- Poplack, Shana, and Johanne S. Bourdages. 2005. *Corpus du français en contexte: milieux scolaire et social*. University of Ottawa. Social sciences and humanities research council of Canada Research grant #410-2005-2108.
- Poplack, Shana, and Nathalie Dion. 2009. “Prescription vs. praxis: The evolution of future temporal reference in French.” *Language*, 85(3): 557-587.
- Poplack, Shana, and Elisabete Malvar. 2007. “Elucidating the transition period in linguistic change.” *Probus*, 19(1): 121-169.
- Poplack, Shana, Allison Leales, and Nathalie Dion. 2013. “The evolving grammar of the subjunctive.” *Probus*, 25(1): 135-195.
- Poplack, Shana, and Anne St-Amand, Anne. 2007. “A real-time window on 19th century vernacular French: *The Récits du français québécois d’autrefois*.” *Language in Society*, 36(5): 707-734.
- Poplack, Shana, and Sali Tagliamonte. 1996. “Nothing in context: variation, grammaticization and past time marking in Nigerian Pidgin English.” In *Changing meanings, changing functions: Papers relating to grammaticalization in contact languages*, edited by Phillip Baker and Anand Syya, 71-94. Westminster, UK: University Press.
- Poplack, Shana, and Sali Tagliamonte. 2001. *African American English in the diaspora*. Oxford: Basil Blackwell.
- Poplack, Shana, James A. Walker, and Rebecca Malcolmson. 2006. “An English “like no other”?: Language contact and change in Québec.” *Canadian Journal of Linguistics*, 51(2/3): 185-213.
- Poplack, Shana, Lauren Zentz, and Nathalie Dion. 2012. “Phrase-final prepositions in Québec French: An empirical study of contact, code-switching and resistance to convergence.” *Bilingualism: Language and Cognition*, 15(2): 203-225.
- Quer, Josep, ed. 2009. Special issue, *Lingua*, 12 (Twists of Mood: The Distribution and interpretation of indicative and subjunctive): 1779-1968.
- Rand, David, and David Sankoff. 1990. *GoldVarb: A variable rule application for the Macintosh*. Montreal, Canada: Centre des recherches mathématiques, Université de Montréal.

- Roberge, Yves. 1998. “Les prépositions orphelines dans diverses variétés de français d’Amérique du Nord.” In *Français d’Amérique: variation, créolisation, normalization*, edited by Patrice Brasseur, 49-60. Avignon: Centre d’études canadiennes.
- Sankoff, David. 1988a. “Sociolinguistics and syntactic variation.” In *Linguistics: The Cambridge survey*, edited by Frederick J. Newmeyer, 140-161. Cambridge: Cambridge University Press.
- Sankoff, David. 1988b. “Variable rules.” In *Sociolinguistics: an international handbook of the science of language and society*, edited by Ulrich Ammon, Norbert Dittmar and Klaus J. Mattheier, 984-997. Berlin: Walter de Gruyter.
- Sankoff, David, and Pierrette Thibault. 1981. “Weak complementarity: tense and aspect in Montreal French.” In *Syntactic change, vol. 25*, edited by Brenda B. Johns and David R. Strong, 205-216. Ann Arbor: Department of Linguistics, University of Michigan.
- Sankoff, David, Sali Tagliamonte, and Eric Smith. 2005. *GoldVarb X: a variable rule application for Macintosh and Windows*. Department of Mathematics, University of Ottawa, and Department of Linguistics, University of Toronto.
- Scherre, Maria Marta Pereira, Anthony J. Naro. 1991. “Marking in discourse: “Birds of a feather”.” *Language variation and change*, 3(1): 23-32.
- Severin, Dorothy S., ed. 1987. *La Celestina* (1499, Fernando de Rojas). Madrid: Cátedra.
- Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.
- Thompson, Sandra A., and Paul J. Hopper. 2001. “Transitivity, Clause Structure, and Argument Structure: Evidence from Conversation.” In *Frequency and the Emergence of Linguistic Structure* edited by Joan L. Bybee and Paul J. Hopper, 27-60. Amsterdam: John Benjamins.
- Torres Cacoullous, Rena. 2000. *Grammaticization, synchronic variation, and language contact. A study of Spanish progressive -ndo constructions*. Amsterdam/Philadelphia: John Benjamins.
- Torres Cacoullous, Rena. 2001. “From lexical to grammatical to social meaning.” *Language in Society*, 30(3): 443-478.
- Torres Cacoullous, Rena. 2011. “Variation and grammaticalization.” *The Handbook of Hispanic Sociolinguistics*, edited by Manuel Díaz-Campos, 148-167. Malden, MA/Oxford: Wiley-Blackwell.
- Torres Cacoullous, Rena. 2012. “Grammaticalization through inherent variability: The development of a progressive in Spanish.” *Studies in Language*, 36(1): 73-122.
- Torres Cacoullous, Rena. To appear. “Gradual loss of analyzability: diachronic priming effects.” In *System, Usage, and Society*, edited by Aria Adli, Göz Kaufmann, and Marco García. Berlin: De Gruyter.
- Torres Cacoullous, Rena, and James A. Walker. 2009. “The present of the English future: Grammatical variation and collocations in discourse.” *Language* 85(2): 321-354.
- UNAM. 1971. *El habla de la Ciudad de México: Materiales para su estudio*. Mexico City: Centro de Lingüística Hispánica.
- UNAM. 1976. *El habla popular de la Ciudad de México: Materiales para su estudio*. Mexico City: Centro de Lingüística Hispánica.
- Walker, James A. 2007. “‘There’s bears back there’: Plural existentials and vernacular universals in (Quebec) English.” *English World-Wide*, 28(2): 147-166.

- Weiner, Judith and William Labov. 1983. “Constraints on the agentless passive.” *Journal of Linguistics*, 19(1): 29-58.
- Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. “Empirical foundations for a theory of language change.” In *Directions for historical linguistics*, edited by Winfred P. Lehmann and Yokov Makiel, 95-188. Austin: University of Texas Press.
- Zribi-Hertz, Anne. 1984. *Orphan prepositions in French and the concept of ‘null pronoun’*. Bloomington, IN: Indiana University Linguistics Club.

Cross-references

Foulkes and Hay

Notes

ⁱ Examples are reproduced verbatim from audio recordings. Codes in parentheses following French and English examples refer to corpus, speaker number and line number. Corpora referenced: RFQ: *Récits du français québécois d’autrefois* (Poplack and St-Amand 2007); OH: *Corpus du français parlé à Ottawa-Hull* (Poplack 1989); FeC: *Français en contexte: milieux scolaire et social* (Poplack and Bourdages 2005); QEC: *Québec English Corpus* (Poplack, Walker and Malcolmson 2006); NPE: *Nigerian Pidgin English Corpus* (Poplack and Tagliamonte 1996); NPR: *Corpus of African American Nova-Scotian English* (North Preston) (Poplack and Tagliamonte 2001); SE: *Corpus of Samaná English* (Poplack and Tagliamonte 2001). For Spanish examples, see list of publications.