From lexical to grammatical to social meaning

R E N A   T O R R E S   C A C O U L L O S

*Department of Spanish and Portuguese, University of New Mexico,
Albuquerque, NM 87131-1146, rcacoull@unm.edu*

ABSTRACT

It is argued that a grammaticization perspective affords insights on the problem of functional equivalence in variables "above and beyond phonology." This is a study of variation between aspectual expressions involving Spanish *estar* 'be located' and *andar* 'go around' + gerund. An analysis of Mexican oral corpora shows both linguistic and social conditioning: Each auxiliary is favored by certain classes of main verbs, and *andar* is favored in popular varieties. Semantic bleaching along parallel paths results in layering in the domain of progressive aspect, while retention results in synchronic distribution patterns congruent with the original meaning of the source constructions. Thus, once meaningful aspectual differences become distributional routines. At the same time, the patterning of verbs denoting outdoor activities with *andar* leads to social differences and the association of this variant with rural and popular varieties. (Grammaticization, sociolinguistic variation, progressives, gerundial periphrases, Spanish, Mexico.)*

GRAMMATICIZATION AND SOCIOLINGUISTIC VARIABLES

This study draws on grammaticization to account for sociolinguistic patterns of variation in the domain of progressive aspect in Spanish. The extension of variation theory to variables "above and beyond phonology" (G. Sankoff 1986[1973]) has been a contentious issue, revolving around the problem of functional equivalence among various forms (e.g. Lavandera 1978, Labov 1978, Romaine 1984, García 1985, Cheshire 1987, Milroy 1987, D. Sankoff 1988, Winford 1996, Silva-Corvalán 1997). I will show how grammaticization provides insights on the development of socially stratified grammatical variables. On the one hand, formal and semantic reduction along parallel evolutionary paths results in the availability of different forms for the same grammatical function, or variants of a grammatical variable. On the other hand, retention of meaning features in the grammaticizing forms results in distribution differences, which in turn acquire social meaning.

The forms under study are two Spanish aspectual periphrastic expressions, *estar* 'be located' + gerund and *andar* 'go around' + gerund. I will refer to *estar* and *andar* as "auxiliaries" for convenience, without a commitment to the degree of their auxiliation.[1] Both expressions cover a range of uses in imperfective territory, from progressive to habitual (as defined by Comrie 1976:24–5). Cross-linguistically, progressives derive from constructions involving a locative element

© 2001 Cambridge University Press 0047-4045/01 $9.50   443

(Bybee, Perkins & Pagliuca 1994:127–37). As I will show, both *estar* and *andar* in construction with a gerund show synchronic variation between more lexical and more grammatical uses.

My claim will be that *estar* + gerund and *andar* + gerund are variants of a single socially stratified variable. An example of the *estar/andar* + gerund variation appears in ex. (1). Both (1a), with *estar*, and (1b), with *andar*, express present progressive meaning; that is, in both cases the actions take place simultaneously with the moment of reference. Notice that both are translated with the English Present Progressive. (Sources of data cited in this article are listed following the endnotes.)

(1) a. Present progressive with *ESTAR* + *-NDO*
         *Pero **estás hablando** de una forma de vida, Gordo*. (UNAM 1971:261)
         'But **you are** (ESTAR) **talking** about a way of life, Gordo.'
     b. Present progressive with *ANDAR* + *-NDO*
         *¡Ay! **Ando buscando** unas tijeras, porque se me rompió una uña*. (UNAM 1976:415)
         'Oh dear! **I am** (ANDAR) **looking** for some scissors, because I broke a nail.'

The title of this essay represents a diachronic order: from lexical to grammatical to social meaning. First, I review diachronic changes in *estar/andar* + gerund sequences that lead to their emergence as conventionalized units. The synchronic result of these changes is variation between these grammaticizing forms as aspectual morphemes (layering as formal diversity). I then show that *estar/andar* + gerund variation is socially stratified in that a higher frequency of *andar* is associated with popular varieties. Next, I tackle the problem of meaning differences in grammatical variation and offer the solution of variation within grammaticizing forms, showing that both *estar* and *andar* + gerund cover the same aspectual territory (layering as polysemy). Following that, I show distribution differences between *estar* and *andar* in co-occurring main verbs; these co-occurrence patterns reflect once meaningful differences that have become nonmeaningful routines (nonmeaningful aspectually). Finally, I provide evidence that the social meaning of *andar* + gerund follows from the original lexical meaning *andar* 'go around'; specifically, from an early association of *andar* + gerund with outdoor activities follows an association with rural and popular urban speech.

FORMAL AND SEMANTIC REDUCTION (DIACHRONY) →
LAYERING (SYNCHRONY)

Formal and semantic reduction

Estar and *andar* + gerund both start out as free combinations of autonomous lexical items. The meaning of these free combinations of *estar* and another verb in gerund form was 'be located in a specific place verb-*ing*'. In (ex. 2), for example, the subjects are located (*estar*) talking (*hablando*). Locative meaning is shown by the following clause, *ningun otro no y estaua* 'no one else was there'. In contrast, the meaning of *andar* + gerund combinations was 'go around verb-

ing'. In (ex. 3), the two brothers are going around (*andar*) looking for (*buscando*) someone or something. Motion meaning is confirmed by the occurrence of the motion verbs *passaron* 'passed' and *uinieron* 'came' in the two clauses following the *andar* + gerund sequence. Both examples are from Alfonso X's *Estoria de España* (EE1), a 13th-century prose chronicle:

(2) *E cuenta ell estoria q⟨ue⟩ quando gelo el dixo q⟨ue⟩* **estauan** *amos* **fablando** *solos q⟨ue⟩ ningun otro no y estaua.* (EE1 19r37)
   'And the story goes that when he told him that, they **were** both **there talking** alone, no one else was there.'

(3) *Onde estos dos sos hermanos que la* **andauan buscando** *passaron a affrica e uinieron a siria.* (EE1 22r62)
   'where those two his brothers who **were going around looking for** her crossed to Africa and came to Syria'

That both *estar* or *andar* and the cooccurring gerund (-*ndo* form) were combining as independent lexical items is shown by several indices in Old Spanish data (12th–15th centuries). First, there was positional variation in the order of the two items, as shown in ex. (4) below. Second, open class material often intervened between the two items, as in ex. (5). Third, multiple parallel gerunds combined with the same auxiliary, as in ex. (6). Fourth, there was variation in object clitic position: The clitic occurred preposed to the auxiliary, as in ex. (3) above, between the auxiliary and gerund, as in ex. (4a), or postposed to the gerund, as in ex. (6). Ex. (4) is from the *Poema de mio Cid* (PMC), a 12th-century verse epic; exx. (5) and (6) are from the *Estoria de España*.

(4) Positional variation in Old Spanish
   a. *tornaua la cabec'a &* **estaua** *los* **catando** (PMC 0002)
      'he turned his head and **was** (stood) **there looking** at them'
   b. **catando estan** *a myo c'id q⟨u⟩a'ntos ha en la cort* (PMC 3123)
      'everybody in the court [**looking**] **is** (stands) **there** looking at My Cid'

(5) Intervening open class material in Old Spanish
   estaua **una uez en el te⟨m⟩plo** *faziendo sacrificio* (EE1 73v101)
   'he was **once in the temple** offering a sacrifice'

(6) Multiple parallel gerunds in Old Spanish
   E estaua **fabla⟨n⟩do** *ante tod el pueblo y* **falagando** *los* (EE1 74v32)
   'And he was there **speaking** before all the people and **flattering** them'

Over time, lexical combinations of *estar* or *andar* with a gerund evolve into grammatical constructions. Grammaticization of function is concurrent with grammaticization of form, as predicted by the "parallel reduction hypothesis" (Bybee et al. 1994:106ff.). In present-day Spanish, indications of grammaticization of form are positional fixing to a categorical auxiliary + gerund order and categorical preposing of object clitic pronouns to finite forms of the auxiliary. In addition, relative to Old Spanish data, there has been a reduction in the occurrence and amount of intervening material and a decrease in multiple parallel gerunds. All

ESTAR & *-NDO* = be located verb-*ing*   > *ESTAR* + *-NDO* = be verb-*ing*
Estauan ... fablando 'were ... there talking' (2) > *estás hablando* 'you're talking' (1a)
ANDAR & *-NDO* = go around verb-*ing*   > *ANDAR* + *-NDO* = be verb-*ing*
andauan buscando 'were going around looking   > *ando buscando* 'I'm looking for' (1b)
   for' (3)

FIGURE 1: Free lexical combinations > conventionalized units.

these changes point to the emergence of *estar* + gerund and *andar* + gerund as fused units (Torres Cacoullos 2000:31–55).

Since Meillet 1958 (1912) proposed the term, grammatic(al)ization has generally been defined as the evolution of lexical into grammatical material (e.g. Hopper & Traugott 1993:xv). A growing body of work, however, suggests that structure is emergent from actual usage patterns and, furthermore, that there is no sharp distinction between grammar and lexicon (Hopper 1987, 1998; Bybee 1998). Bybee (in press) proposes a more precise and compelling definition: Grammaticization is "the process by which a frequently-used sequence of words or morphemes becomes automated as a single processing unit." This definition draws attention to the conventionalization of usage patterns and the fusion of formerly autonomous elements.

In the Old Spanish examples, *estar* 'be located' and *hablando* 'talking' are two independent lexical items that combine freely, as are *andar* 'go around' and *buscando* 'looking for'. *Estar hablando* here means 'be somewhere talking' and *andar buscando* means 'go around looking for'. In present-day Spanish, *estar hablando* and *andar buscando* are no longer free lexical combinations but, rather, conventionalized units, with a stronger aspectual than spatial meaning component. This change from free lexical combination to conventionalized unit is apparent when we compare the Old Spanish examples of exx. (2) and (3) with the present-day Spanish examples of exx. (1a) and (1b). Figure 1 schematizes the emergence of conventionalized units from lexical combinations.

Layering as formal diversity in a functional domain

The grammaticization of these constructions involves the loss of meaning features through a process I will refer to as SEMANTIC REDUCTION (Bybee et al. 1994:6) or BLEACHING (Givón 1975). Other familiar terms are semantic generalization (Bybee & Pagliuca 1985) and desemanticization (Lehmann 1995 [1982], Heine 1993). Figure 2 represents the parallel paths followed by the grammaticizing constructions. Locative meaning is bleached from *estar* + gerund, while nondirectional motion meaning is bleached from *andar* + gerund. In both cases, "being located" and "going around," we have the bleaching of spatial meaning. As specific features of SPATIAL meaning are lost, the constructions generalize to
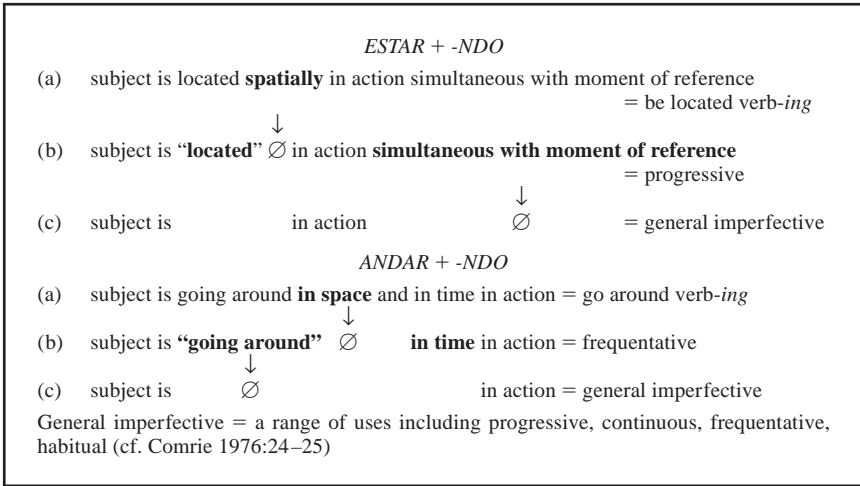
ESTAR + -NDO

(a) subject is located **spatially** in action simultaneous with moment of reference
= be located verb-*ing*

↓

(b) subject is "**located**" ∅ in action **simultaneous with moment of reference**
= progressive

↓

(c) subject is   in action   ∅   = general imperfective

ANDAR + -NDO

(a) subject is going around **in space** and in time in action = go around verb-*ing*

↓

(b) subject is **"going around"** ∅   **in time** in action = frequentative

↓

(c) subject is   ∅   in action = general imperfective

General imperfective = a range of uses including progressive, continuous, frequentative, habitual (cf. Comrie 1976:24–25)

FIGURE 2: Semantic reduction in ESTAR + -NDO *and* ANDAR + -NDO.

more contexts of use, with a more TEMPORAL (aspectual) meaning. The result of bleaching along parallel paths is that the meanings of the two constructions converge, as depicted in lines (c) of Figure 2.

Semantic reduction or bleaching as outlined in Figure 2 results in what has been called "layering" in grammaticization. Hopper (1991:22–24) defines layering as the availability of different forms to serve "similar or even identical functions," as newer layers emerge without displacing older ones within a "functional domain" (in the sense of Givón 1984:32–35). An example from English is the Past Tense, where ablaut (e.g. *snuck*) represents an older layer and affixation (*sneaked*) a more recent layer of grammaticized forms (Hopper 1991:24). Another example is the variation between expressions of future temporal reference *will*, *be going to*, Present Progressive, and futurate Present (Poplack & Tagliamonte 1999). Layering in the functional domain of progressive aspect in Spanish was illustrated in the first set of examples, *estás hablando* and *ando buscando*, 'you are talking' (ESTAR) and 'I am looking for' (ANDAR).

The principle of layering belies structuralist views of grammatical morphemes as based on maximal contrasts (cf. Bybee et al. 1994:22, 148). As Hopper (1991:24) nicely puts it, the "'cluttering' of grammar with functionally similar constructions . . . is not easy to reconcile with any picture of a language as a homogenous, architectured, and delimited object." This view of layering as formal diversity converges with the variationist approach, which has demonstrated that distinctions in grammatical function among different forms can be "neutralized" in discourse (D. Sankoff 1988:153).

TABLE 1. *Relative frequencies of* ESTAR *and* ANDAR + -NDO *in Mexico City*
"habla culta" and "habla popular." [a]

Corpus[b]	ESTAR + -NDO		ANDAR + -NDO		Total
	N	%	N	%	N
habla culta	482	95	24	**5**	506
habla popular	412	80	102	**20**	514
Total	894	88	126	12	1020

[a] Chi-square = 53.70726 (p = .0000).
[b] *habla culta* = educated México, D.F. (capital city) speech, cf. UNAM 1971; figures for 30 hours of recording from Luna (1980:13, 211); *habla popular* = popular México, D.F. speech, cf. UNAM 1976; figures for 25 hours of recording from Arjona (1991:8,116).

SOCIAL STRATIFICATION

In this article, I use the term "social meaning" as in traditional large-scale socio-linguistic surveys where the social significance of language variation is measured by correlations with social attributes such as class, sex, and age (Chambers 1995:1). More recently, sociolinguists have used ethnographic methods to explore the expressive function of variation, or "linguistic variation as social practice" (Eckert 2000). Here, however, I restrict "social meaning" to the covariation of *estar/andar* + gerund with social class.

Let us return to the first set of examples, repeated below:

(7) a. *Pero **estás hablando** de una forma de vida, Gordo.* (MexCult – UNAM 1971:261)
      'But **you are** (ESTAR) **talking** about a way of life, Gordo'
     b. ***Ando buscando** unas tijeras, porque se me rompió una uña.* (MexPop–UNAM 1976:415)
      '**I am** (ANDAR) **looking** for some scissors, because I broke a nail.'

The *estar* example (7a) is from a volume of transcribed educated Mexico City speech, *El habla de la Ciudad de México* (UNAM 1971; cf. Lope Blanch 1986). The *andar* example (7b) is from that volume's popular-speech companion, *El habla popular de la Ciudad de México* (UNAM 1976). A not immediately noticeable but nevertheless real difference between the two corpora is in the frequency of *andar* + gerund. As shown in Table 1, in educated Mexico City data, the frequency of *andar* relative to *estar* + gerund is 5%, while in popular speech data it is 20%. That is, the relative frequency of *andar* + gerund is four times greater in the popular corpus.

In his diachronic study of Spanish progressives, Spaulding (1926:259) observed that "*andar* often seems to be merely a lively, and COLLOQUIAL, substitute for *estar*" (my emphasis). Similarly, in her study of educated Mexico City data, Luna (1980:206) found that *andar* + gerund could be replaced, with loss of what she terms "expressive nuances," by *estar* + gerund, but not by other auxiliaries

448

TABLE 2. *Relative frequencies of* ESTAR *and* ANDAR + -NDO *by speaker level of education in Chihuahua data.*[a]

Speaker education level	ESTAR + -NDO		ANDAR + -NDO		Total
	N	%	N	%	N
University	158	91	15	**9**	173
Primary or secondary	198	73	73	**27**	271
Totals	356	80	88	20	444

[a] Chi-square = 20.32413 (p = .0000).

(*ir* 'go' or *venir* 'come' + gerund). Arjona (1991:125), who studied popular Mexico City data, commented: "A los hablantes cultos el uso frecuente de *andar* + gerundio les parece poco elegante, al tiempo que a los informantes de habla popular les resulta sumamente expresivo" ('the frequent use of *andar* + gerund seems not elegant to educated speakers, while it turns out to be highly expressive for popular speech informants'). Comments such as these are in line with the results in Table 1, which show that the relative frequency of *andar* and *estar* + gerund is socially stratified.

These scholars' observations lend support to the hypothesis that *andar/estar* + gerund is a sociolinguistic variable in the classical Labovian sense. In other words, *estar/andar* + gerund are alternative ways of saying the same thing that distinguish social groups, and the meaning difference is social rather than functional (aspectual).

The effect of social class as indexed by speaker occupation and level of education in the *habla culta* and *habla popular* data operates beyond México, D.F., the capital city. Table 2 shows the distribution of *estar/andar* + gerund according to level of education in a corpus of data I recorded in 1997 in the northern Mexican state of Chihuahua. Of 22 speakers, six had a university education. Four of these were teachers, two in secondary schools and another in a post-secondary institution; one was a lawyer and another a writer. All the university-educated speakers were recorded in the state capital, Chihuahua, an urban center. By contrast, the group with primary or secondary school education includes speakers recorded in the village of Ascención, Chihuahua, where almost all families are involved in agriculture and livestock raising. University-educated speakers show a relative frequency of 9% for *andar* + gerund. This compares to 27%, three times greater, for speakers with at most a secondary school education.

Stylistic differences between more casual and more careful varieties have been interpreted in terms of social class (cf. Bell 1984). To compare the oral, more or less conversational data in Table 1 with formal, written data, I draw on a corpus of essays and academic prose published between 1927 and 1995 by well-known

TABLE 3. *Relative frequencies of* ESTAR *and*
ANDAR + -NDO *in essays and academic prose.*[a]

ESTAR + -NDO		ANDAR + -NDO		Total
N	%	N	%	N
67	96	3	**4**	70

[a] Castellanos (1966:13–113), García Canclini (1990:31–93, 1993:86–196), Lope Blanch (1983:11–175), Martínez, ed. (1958:9–409), Monsiváis (1995:15–113), Montemayor (1985:9–111), Paz (1959:9–105), Reyes (1927:5–89, 133–268). Total: 1,344 pages.

Mexican authors, including Rosario Castellanos, Néstor García Canclini, Juan M. Lope Blanch, Carlos Monsiváis, Carlos Montemayor, Octavio Paz, and Alfonso Reyes.[2] Table 3 shows that the relative frequency of *andar* + gerund is 4% in this corpus. This is virtually identical to *andar* + gerund's relative frequency in the Mexico City educated speech sample, at 5% (Chi-square = .028792, p = .8653, n.s.) but five times smaller than in the popular Mexico City data, at 20% (Table 1).

Token or text frequency figures for the two constructions are equally revealing. Table 4 compares the average occurrence of *estar* and *andar* + gerund per 100 pages of printed text for the written essays and academic prose corpus and for the two Mexico City oral corpora. The essays and academic prose corpus totals 1,344 pages. MexCult, the educated variety (*habla culta*), represents 307 pages of transcribed oral data from *El habla de la Ciudad de México* (UNAM 1971:11–317), about two-thirds of the entire volume. MexPop, the popular variety (*habla popular*), represents all 448 pages of *El habla popular de la Ciudad de México* (UNAM 1976:14–461).[3] The token frequency of *estar* and *andar* + gerund combined is virtually identical in the two Mexico City oral corpora, at about 90 tokens per 100 pages, as depicted in the Total-Average column in Table 4. In sharp contrast, the frequency of these aspectual expressions combined is only about 5 tokens on average per 100 pages in the written essays and academic corpus.

We may attribute the tremendous difference between the formal, written data and the oral, conversational data in the token frequency of *estar/andar* + gerund to genre differences. The distributional norm for tense-aspect-modality is genre-dependent (Givón 1990:943ff.). That is, the kinds of things most talked (written) about in essays are gnomic or generic situations, propositions in which the predicate holds for all time for a class of entities. For English academic prose, Givón (1990:963) found that predicates are overwhelmingly stative and verb forms are overwhelmingly habitual. So it is not surprising that we find fewer progressives in the written corpus than in the two oral corpora. In short, with respect to overall

TABLE 4. *Token frequencies of* ESTAR *and* ANDAR + -NDO
(per 100 pages of printed text).

Corpus[a]	no. pages	ESTAR		ANDAR		Total	
		N	Aver	N	Aver	N	Aver
MexCult	307	277	**90**	13	**4**	290	**94**
MexPop	448	306	**68**	86	**19**	392	**88**
Essays and academic	1,344	67	**5**	3	**.2**	70	**5**

[a] MexCult = pp. 11–317 of *El habla de la Ciudad de México* (UNAM 1971), educated Mexico City speech; MexPop = pp. 14–461 of *El habla popular de la Ciudad de México* (UNAM 1976), popular Mexico City speech. For essays and academic corpus, see Table 3.

estar/andar + gerund token frequency, the two oral corpora line up together in opposition to the written corpus.

Despite the similarity between the two Mexico City oral corpora in the overall token frequency of *estar/andar* + gerund, there is a big difference if we look at *estar* and *andar* separately. As shown in the *ANDAR* – Average column in Table 4, the average occurrence of *andar* + gerund in the educated speech data is 4 tokens per 100 pages, while in the popular speech data there are 19 tokens. However, while the overall token frequency of *estar/andar* + gerund combined is much lower in essays and academic prose than in the educated speech data, the token frequency of *andar* + -ndo relative to *estar* + -ndo is virtually the same in the two corpora.[4]

This can be seen if we compare *estar/andar* ratios, depicted in Table 5. The ratio in the educated speech corpus is about 21 (277/13), and in the essays and academic prose data about 22 (67/3). These figures contrast with a ratio of less than 4 (306/86) in the popular corpus. This is parallel to what we found when we compared *andar* + gerund relative frequencies (tables 1 and 3). In short, with respect to the relative frequency of *andar* + gerund, the written and educated speech corpora line up together in opposition to the popular speech corpus.

The three corpora line up in two different ways. In overall *estar/andar* + gerund token frequency, the lineup is formal, written data in opposition to conversational, oral data (both educated and popular) (Table 4). But in the relative frequency of *andar*, educated speaker data (both written and oral) line up against the popular data (Table 5). This pair of results supports the view that *andar* + gerund frequency is socially rather than functionally stratified. That is, whether educated speakers are conversing or writing formal prose, they use *andar* + gerund less frequently than do speakers of popular varieties. At the same time, conversational data, whether from educated or popular-variety speakers, display combined frequencies of *estar/andar* + gerund are about the same, which lends

TABLE 5. ESTAR/ANDAR + -NDO *ratios*.

	ESTAR	*ANDAR*	Ratio
Corpus	N	N	
MexCult	277	13	21.3
MexPop	306	86	3.6
Essays – Academic	67	3	22.3

further support to a claim for social rather than functional differences. That is, the difference between educated and popular conversations appears to lie not in the occurrence of occasions to talk about situations as progressive, but rather in distinct preferences for certain variants.

THE PROBLEM OF FUNCTIONAL EQUIVALENCE AND ITS SOLUTION, LAYERING AS POLYSEMY

The above results are supportive but not conclusive with respect to the problem of meaning differences. The mere finding that *estar*/*andar* + gerund have the same combined token frequency but have different relative frequencies in educated and popular oral varieties does not preclude the possibility that there are aspectual differences between the two constructions. That is, the difference in these frequencies between the educated and popular conversational data might still result from differences in what is talked about, and how.

Frequentative-habitual meaning

What exactly does *andar* + -*ndo* mean? "Going around verb-*ing*" in space and time first grammaticizes into frequentative meaning (line b in Fig. 2). If someone "goes around" places doing something, it may be inferred that this "doing" is repeated. With frequentative aspect, a situation is repeated frequently, but not necessarily all on one occasion, as in iterative aspect (Bybee et al. 1994:160). Instead, frequentative – like habitual – indicates repetition on different occasions and describes a situation "which is characteristic of an extended period of time, so extended that … [it] is viewed … as a characteristic feature of a whole period" (Comrie 1976:27–28). Frequentative further specifies that the repetition is frequent during that period; that is, the habitual situation occurs often (Bybee et al. 1994:127, 165).

Exx. (8a–c) are instances of frequentative aspect. All three are taken from versions of the legend of La Llorona, a woman who is said often to be heard wailing at night. The first example is from the Chihuahua corpus, the second from New Mexico data (see note 14), and the third from the educated Mexico City data.

452

In telling the story, all three speakers use *andar* to refer to La Llorona's crying and lamenting: *anda llorando*, *anda penando*.

(8) *La Llorona "anda penando"* = goes around mourning (in space and time)

    a. *Ya casi no me acuerdo nada más me acuerdo de eso, que mm se oía por las calles porque había matado a mm a sus hijos pero no me acuerdo por qué. Había matado a su esposo y a sus hijos. Y que, como se había arrepentido, que en las noches **andaba … llorando** por lo mismo, porque pues estaba arrepentida.* (Chih97#1bPH)

    'I don't really remember, I just remember this, that she would be heard in the streets because she had killed her children but I don't remember why. She had killed her husband and her children. And that, since she had repented, that at nights **she would go around crying** because of that, well because she repented.'

    b. E: *Oh, sí, es que, ¿cómo cómo estuvo? Que esta mujer mató, no quería a su bebé. Y lo lo y lo mató. Y se lo dio al cochino se lo dio al cochino que se lo comiera el cochino. Es ansina la oí yo. Y luego por eso **anda llorando**. Todo el tiempo.*

    F: ***Anda penando***. (NM.Edw)

    E: 'Oh yes, what happened was that, how did it go? That this woman, didn't want her baby. And she killed it. And she gave it to the pig to eat. That's how I heard the story. And that's why **she goes around crying**. All the time.'

    F: **She goes around mourning**.'

    c. . . . . . . *y dicen que desde entonces, de remordimiento, que el alma de esta señora **anda penando** hasta la fecha, y que grita: "¡Aaay, mis hijooos!"* (MexCult – UNAM 1971:51)

    'And they say that since then, out of remorse, that the soul of this woman **goes around mourning** up until now even, and that she shouts "Oh, my chiiiiiildren!"'

In these examples, physical motion in space merges with an aspectual frequentative meaning. *Anda llorando* expresses that she is literally, physically going around, and at the same time it expresses that she can OFTEN be heard crying. It is not surprising that, in recounting the story, the speakers use the same expressions, *andar llorando*, *andar penando*. These are formulaic uses that have been routinized in the context of a traditional legend.

Co-occurrence patterns in Old Spanish (12th–15th centuries) are consistent with the use of *andar* and gerund combinations to express frequentative aspect. I will limit myself here to two indices of frequentative meaning, direct object number and co-occurring locative type.

First, most direct objects in Old Spanish *andar* and gerund combinations are plural. In the following Old Spanish example, the Vandals (plural) are going around (*andar*) destroying all France, demolishing churches (plural), and killing saints (plural). Here, plural objects and multiple parallel gerunds are highly congruent with frequentative-habitual aspectual meaning, as the situations are evidently repeated on several occasions. In Langacker's (1996:301) terms, "plurality reflects multiple instances of the event type."

(9) *E aquell anno andauan los Vuandalos destruyendo toda tierra de francia & desfaziendo las yglesias. & matando los sa⟨n⟩tos.* (EE1 126r14)

    'And that year the Vandals were going around destroying the entire land of France and demolishing the churches and killing the saints.'

In a study based on an Old Spanish corpus of over one-half million words, I found that the proportion of plural objects was 55% (37/67), or conversely, the propor-

tion of singular objects was 45% (Torres Cacoullos 2000:159). This is quite a striking result, since singular is generally the unmarked member of the nominal category of number. Thus, Greenberg (1966:32) reports singular proportions ranging from 74% to 85% in three different language samples of nouns.

An index of semantic bleaching in *andar* + gerund is the diachronic reversal in direct object number. In present-day oral data (MexCult, MexPop, and MexRep), singular has the higher proportion of direct objects, at 73% (45/62). The proportion of singular direct objects for *estar* + gerund is identical, also at 73% (204/280).[5] This figure is close to cross-linguistic averages (Greenberg 1966:32).

Co-occurring locatives

A second index of frequentative uses in Old Spanish is the co-occurrence of *por* 'along, around, all over' locative expressions. In the corpus I studied, this was the most frequent locative type co-occurring with *andar* and gerund: fully 20% of all *andar* and gerund tokens had a co-occurring *por* locative. *Por* locatives are consonant with the 'go around' lexical meaning of *andar*. Furthermore, more than one-third of those *por* locatives had a plural NP, indicating multiple locations (Torres Cacoullos 2000:81). Multiple locations are also highly congruent with 'going around' and with frequentative-habitual meaning. Going around (*andar*) along an extensive location or locations (*por*) while doing things (plural objects) may mean a repeated doing. In ex. (10), Pompey goes around moving people *por las tierras* 'all over the lands'.[6]

(10) *E tu ponpeyo magno q⟨ue⟩* **andas mouiendo** *las yentes* **por las ti⟨er⟩ras** *& sacando huestes pora parar azes. […] muy poco sabes de lo q⟨ue⟩ los fados te guisan*. (EE1-47r69)
  'And you great Pompey who **goes around** moving people **all across the lands** and taking out armies to stand troops. […] very little do you know of what fate holds for you.'

It is important that gerund combinations with *estar* pattern very differently from *andar* with respect to co-occurring locatives in the Old Spanish data. In contrast with *andar*'s 20%, fewer than 1% of tokens of *estar* co-occur with a *por* locative. Instead, *estar* and gerund most frequently co-occur with *en* 'in' expressions, with about 8% of all tokens. *En* is consonant with the meaning 'be located in a specific place verb-*ing*'. In a complementary fashion, *andar* and gerund combinations co-occur with an *en* locative less than 1% of the time (Torres Cacoullos 2000:75). It is evident, then, that the distribution of these two locative types was greatly skewed in Old Spanish.

Over time, locative and motion meaning became bleached, and both grammaticizing constructions now cover the same territory in the functional domain of imperfective aspect, including progressive and habitual (Fig. 2). Following a spectacular diachronic rise in token frequency, *estar* + gerund has spread from progressive/continuous meaning into frequentative-habitual territory in present-day Spanish. That is, *estar* + gerund has spilled into *andar* + gerund territory. *Andar* + gerund, which has also increased in token frequency, though not so

TABLE 6. *Co-occurring locatives in* ESTAR/ANDAR + -NDO *in popular Mexican Spanish corpora.*[a]

Locative type	ESTAR + -NDO		ANDAR + -NDO		Total	
	N	%	N	%	N	%
None	665	80	163	20	828	78
Deictics[b]	86	77	26	23	112	11
En 'in'	22	81	5	19	27	3
Con 'with'	20	87	3	13	23	2
Donde 'where'	10	71	4	29	14	1
Por 'around'	3	27	8	**73**	11	1
Other[c]	8	57	6	43	14	1
More than one of above[d]	24	92	2	8	26	3
Totals	838	79	217	21	1055	100

[a] Corpora: Chihuahua, MexPop, MexRep.
[b] Deictics include *aquí* 'here', *allí* 'there', *ai* 'around here/there'.
[c] See endnote 8.
[d] See endnote 7.

sharply, also shows semantic generalization in that present-day habitual uses may lack the more specific frequentative meaning component.

If all this is true – that is, if locative meaning is bleached from *estar* + gerund and motion meaning is bleached from *andar* + gerund, and if they cover the same aspectual territory – then we would expect not to find the same skewing in the distribution of *en* and *por* locatives in present-day data. Table 6 shows the distribution of co-occurring locatives in a combined corpus comprising the Chihuahua corpus, the popular Mexico City corpus (MexPop), and a northern states corpus, MexRep; the last is 247 pages of transcribed oral data from northern Mexican states, published in *El habla de la República Mexicana* (UNAM 1995:10–257). The reason for combining the corpora is to expand the data base, given the low token frequency of these constructions (see note 4). *Andar* + gerund has a relative frequency of around 20–23% in all three data sets, so we can assume that *estar/andar* + gerund variation is conditioned in the same way by the same factors.

Unlike the Old Spanish data, *en* 'in' locatives are now fairly evenly proportioned between *estar* and *andar* + gerund, with 81% and 19%, respectively. If we look at this result from the perspective of each construction, we see that 2.6% (22/838) of *estar* + gerund tokens co-occur with an *en* locative; similarly, 2.3% (5/217) of *andar* + gerund tokens co-occur with this same locative type (Chi-square = .071288, p = .7895, n.s.).[7] Indeed, as shown in Table 6, all locative types are fairly evenly apportioned between the two constructions (the proportions, viewed from the perspective of each construction, are not significantly different, at p > .05 in all cases).[8]

TABLE 7. *Proportion of* ESTAR *and* ANDAR + -NDO *co-occurring with frequentatives in popular Mexican Spanish corpora.*[a,b]

	Total N	Frequentative N	Frequentative %
ESTAR + -NDO	838	23	2.7%
ANDAR + -NDO	217	4	1.8%

[a] Chi-square = .5615 (p = .4537, n.s.)
[b] Corpora: Chihuahua, MexPop, MexRep.

The exception is *por* locatives, which, as in Old Spanish, favor *andar*. However, fewer than 4% (8/217) of all *andar* + gerund tokens have a *por* locative; in the Old Spanish data, a substantial 20% do. Furthermore, these *por* locatives are overwhelmingly the deictic *por aquí*, *por allí* 'around here, around there' (6/8=75%), compared to *por* + NP$_{(plural)}$ in the Old Spanish data, as in ex. (10) above. The even distribution of *en* 'in' locatives between *estar* and *andar* and the decline of *por* locatives with *andar* are congruent with semantic reduction – loss of spatial meaning features – in both constructions, along parallel paths. Over all, 80% (665/838) of *estar* and 75% (163/217) of *andar* + gerund have no co-occurring locative at all (Chi-square = 1.835329, p = .1755, n.s.).

Cooccurring temporals

In short, judging by the indices of object number and co-occurring locative types, the meaning of the two periphrases has converged. Let us now look at co-occurring temporal expressions. If *andar* + gerund is more of a frequentative-habitual than is *estar* + gerund, we might expect to find differences in co-occurring lexical expressions of frequentative aspect.

Frequentative adverbials found in the data include *cada rato* 'every so often', *constantemente* 'constantly', *diario* 'daily', *mes con mes* 'every month', and *seguido* 'often', as well as *siempre* 'always' and expressions with *todos* 'all', as in *todos los días* 'every day'. Table 7 shows the proportion of *estar* and *andar* + gerund tokens co-occurring with a frequentative in the combined Chihuahua, popular Mexico City (MexPop), and northern states (MexRep) data set.

The average occurrence of a frequentative expression with *estar* + gerund is about 3% (23/838) and about 2% (4/217) with *andar* + gerund. Contrary to what we might expect if we suppose that *estar* + gerund is basically a progressive while *andar* + gerund is basically a frequentative, *estar* shows a slightly higher proportion with co-occurring frequentatives than does *andar*, although the difference is not statistically significant (Chi-square = .5615, p = .4537, n.s.). These results also point to a convergence of meaning.

TABLE 8. *Co-occurring temporals in* ESTAR/ANDAR + -NDO *in popular Mexican Spanish corpora.*[a]

Temporal type	ESTAR + -NDO		ANDAR + -NDO		Total	
	N	%	N	%	N	%
None	620	78	180	22	800	76
Ahora 'now'	65	92	6	8	71	7
Ya 'already'	46	81	11	19	57	5
Durative	27	93	2	7	29	3
Locating	26	90	3	10	29	3
Cuando 'when'	21	72	8	28	29	3
Frequentative	23	85	4	15	27	3
Other	10	77	3	23	13	1
Totals	838	79	217	21	1055	100

[a] Corpora: Chihuahua, MexPop, MexRep.

Table 8 shows the distribution of all co-occurring temporal adverbials. The combined corpus is the same as in Tables 6 and 7. In addition to frequentative expressions, another group of expressions is "durative" adverbials, which indicate 'for x time' or answer the question 'for how long?' (Binnick 1991:300–10; Smith 1991:156). Examples are *menos de un año* 'for less than a year', *como ocho meses* 'for about eight months', and *toda la noche* 'all night'. Another group is "locating" adverbials, which answer the question 'when?' (Smith 1991:151). Examples are *el otro día* 'the other day', *entonces* 'then, at that time', *hace ratito* 'a little while ago', and co-occurring *cuando* 'when' clauses, as in *cuando nos platicó estábamos riendo de la tragedia* 'when she told us the story we were laughing at the tragedy' (Chih'97#11a.CJ). Both durative and locating temporal expressions are compatible with either progressive or habitual meaning. As shown in the rightmost column in Table 8, durative and locating temporal expressions each make up 3% of all the data.[9]

Frequentative adverbials make up only 3% of the data; that is, they are not very frequent. As can be inferred from Table 8, these adverbials are fairly evenly distributed between *estar* and *andar* + gerund, as 85% (23/27) co-occur with the former and 15% (4/27) with the latter, in tandem with the relative frequency of the two constructions (79% and 21%, respectively). As in the case of locatives, all temporal types are fairly evenly apportioned between the two constructions (the proportions of *estar* and *andar* with each adverbial type are not significantly different, at p > .05 in all cases) – with the exception of *ahora* 'now'.

As observed with the locatives, most tokens occur without a co-occurring temporal adverbial – 74% and 83% of *estar* and *andar* + gerund, respectively. Nevertheless, *estar* + gerund has a higher average of co-occurring temporals, at

26% (218/838, subtracting "None" tokens) than does *andar* + gerund, at 17% (37/217) (Chi-square = 7.556031, p = .0060, significant at p < .01). This higher average may be attributed, at least in part, to *ahora* 'now' adverbials.

Let us therefore take a look at *ahora*. Silva-Corvalán (1999:73) observes that *ahora*, both as a temporal and a discourse deictic, "occurs most frequently in contexts where a situation is explicitly compared or contrasted with a previous one." The adversative sense of *ahora* is evident here, too. The proportion of *estar* + gerund with an *ahora* adverbial is 8% (65/838), compared to just 3% (6/217) for *andar* (Chi-square = 6.841951, p =.0089). The relatively high average occurrence of *ahora* 'now' with *estar* + gerund is related, I think, to the spread of this construction into frequentative-habitual territory (Torres Cacoullos 2000:177ff.). When *ahora* co-occurs with *estar* + gerund, it may express that a habitual situation represents a new development, a change with respect to the state of affairs holding previously. In ex. (11), watching a *novela* 'television novel' is not an act in progress but a regularly repeated situation, i.e. a habitual. This situation, which characterizes the present, contrasts with the past: NOW we are watching a *novela*, but before we didn't use to watch any (see Blansitt 1975:3 on "generic progressive" uses).

(11) *Bueno, las novelas casi no me gustan. […] Mire,* **ahorita** *la única que* **estamos viendo***, y eso de casualidad, porque … pues yo creo que nos llamó la atención, es la que sale a las siete en el canal cuatro*. (MexRep – UNAM 1995:197)
'Well, I don't really like *novelas*. […] Look, **right now** the only one **we are watching**, and this just by chance, because … well I think it caught our attention, is the one that comes on at seven on channel four.'

Similarly, in ex. (12), "spending more than 17 million pesos daily" is not progressive but habitual, indeed frequentative, as indicated by *diarios* 'daily': every day 17 million pesos are spent. *Ahora* indicates that this amount represents an increase with respect to the past. The contrast is made explicit in the sentence that follows, 'before … today …':

(12) **Ahora se están gastando** *sobre diez y siete millones de pesos diarios. Lo que antes se gastaba en quince años, hoy se gasta en un día*. (MexCult – UNAM 1971:115)
'**Now** more than seventeen million pesos **are being spent** daily. What before would be spent in fifteen years, today is spent in one day.'

The relatively high proportion of *estar* + gerund tokens co-occurring with *ahora* suggests that newer uses of a grammaticizing expression rely more on linguistic context than do older uses. That is, *andar* + gerund's older frequentative-habitual uses, consistent with its original "go around verb-*ing*" meaning, do not seem to depend on co-occurring lexical elements such as frequentative adverbials or *ahora*; however, *estar* + gerund's frequentative-habitual uses, which are at some distance from its original "be located verb-*ing*" meaning, may be more dependent on lexical expressions of frequentative-habitual.

Nevertheless, the distribution of temporal expressions shown in Table 8 does not indicate that any particular temporal adverbial type – with the exception

perhaps of *ahora* – favors one auxiliary over the other. The fairly even distribution of temporal adverbials, especially of frequentative expressions, between *estar* and *andar* + gerund supports the claim that the aspectual meanings of the two constructions have converged. Similarly, the distribution of locatives in Table 6 does not suggest an effect on the choice of auxiliary, with the exception of *por* 'around' expressions. In a variable rule analysis (see next section), neither co-occurring locative nor temporal was selected as a significant factor contributing to the choice between *andar* and *estar* + gerund.[10] Of course, statistical significance for these two factor groups might be established in a larger sample. Nonetheless, given the low token frequency of these constructions and the very low frequency of co-occurring locatives and temporals, I suspect that even a larger sample would not show strong effects.

In summary, these objective measures of direct object number, locatives, and temporals support the claim of a convergence of aspectual meanings. These findings confirm observations of other scholars on the "interchangeability" of *estar* and *andar* + gerund, albeit with expressive nuance differences (Squartini 1998:278; cf. Spaulding 1926:259; Yllera 1980:76-82; Luna 1980:206; Arjona 1991:125).[11]

Layering as polysemy

Let us now look at some examples that show overlapping functions and contexts. In ex. (13), present-tense *andar trabajando* refers to an activity in progress at reference time. Speaker A and his companion are visiting outside P's house. Speaker P calls out to family member X to come over and greet A (line 1). P jokes that X *anda metido ahí* 'is hiding there' (line 2). But A replies *anda trabajando* 'he's working', i.e. that X is not being rude but is presently engaged in doing something (line 3):

(13) *ANDAR trabajando* = progressive

  1 *P*: [Calls out to X] *Ven a ven a conocer a estos muchachos. Ja ja ja.* [ ]
  2   [To A] *Anda metido ahí.*
  3 *A: No,* **anda trabajando**.
  4   [To X, who comes up to greet him] *Buenas tardes.*
  5 *X: Buenas tardes.* (Chih'97#23)

  1 P: [Calls out to X] 'Come over and meet these people. Ha ha ha.' [ ]
  2   [To A] 'He's hiding there somewhere.'
  3 A: 'No, he **is** (ANDAR) **working**.'
  4   [To X, who comes up to greet him]. 'Good afternoon.'
  5 X: 'Good afternoon.'

Here, *anda trabajando* is a true progressive. By contrast, in ex. (14) present-tense *andar trabajando* clearly refers to a habitual situation, one that is a "characteristic feature" in Comrie's (1976:27–28) terms. Speaker R says *andamos tomando pisto y andamos trabajando* 'we (literally) go around drinking and we go around working' (lines 1–2). Frequentative-habitual meaning is shown by speaker C's use of the simple Present with the same predicates, *usted toma y trabaja* 'you

drink and work', to refer to the same situation (line 6). Habitual meaning is further confirmed by her comment that this situation does not mean that R is *de baja alcurnia* 'low class', where she uses *ser*, the copula of "essence" (as opposed to *estar*'s "accident"; see Bolinger 1973).

(14) *ANDAR tomando y trabajando* = habitual

   1 R: *Nosotros **andamos tomando pisto**, y **andamos-***
   2 C: ***trabajando**.*
   3 R: *Sí.*
   4 C: *Mhm*
   5 R: *Y yo todo el tiempo le he dicho a usted-*
   6 C: *Y porque usted toma y trabaja, ¿es de baja alcurnia?* (Chih'97#10Aa)

   1 R: '**We drink** (ANDAR, literally: we go around drinking) and **we**' (ANDAR)
   2 C: '**work**.'
   3 R: 'Yes.'
   4 C: Mhm.
   5 R: 'And I've always told you–'
   6 C: 'And because you drink and work, does that mean you are low class?'

In ex. (15), the same speaker C was later commenting to a friend about R, the person who habitually drinks and works. Here she uses *estar* with the same predicates, *trabajando* 'working' and *tomando* 'drinking'. That the meaning is habitual is evident, since the speaker is telling her friend what kind of person R is. In fact, *estar* + gerund occurs in a definition: they are (*ser*) among those who work and drink. A frequentative meaning component is added by the repetition of *trabajando y* 'working and'.

(15) *ESTAR trabajando y tomando* = habitual
    C: *Son de los que **están trabajando y tomando**, trabajando y-. Entonces ya cuando fuimos por él . . .* (Chih'97#10Ab)
    'He's one of those who **work** (ESTAR, literally: they are working) **and drink** (ESTAR), works and– So by the time we went to pick him up . . .'

Thus, *estar*, as well as *andar* + gerund, can refer to a frequentative or habitual situation. But *estar* + gerund can also be a true moment-of-speech progressive, as we saw in ex. (1a), *pero estás hablando de una forma de vida, Gordo* 'but you are talking about a way of life, Gordo'. Ex. (16) is also a true present progressive. Here *usted está pisteando* 'you are drinking' is in progress at reference time[12]:

(16) *ESTÁ pisteando* = progressive
    *Pues no más usted **está pisteando**, y Luis nada, no más lo ve.* (Chih'97#10Aa)
    'It's just you that **is drinking**, and Luis nothing, he just looks at you.'

However, in the next exchange, with the same main verb, the meaning is frequentative-habitual. Again R and C are talking. That *está pisteando* this time means a frequently repeated customary situation is indicated by the temporal expression *hasta qué horas* 'until what time' in C's question (lines 1 and 3) and *todos los días . . .* (*hasta las*) *siete de la tarde* 'every day . . . (until) seven in the evening' in R's reply (lines 2 and 4). It is confirmed by C's promise to give R a

ride back at 7 P.M., *pa que no le pongan falta* 'so that you don't get marked off as absent' (line 5).

(17) *ESTÁ pisteando* = frequentative-habitual
1 C: *¿Hasta qué horas **está pisteando** ahí usted?*
2 R: *Todos los días.*
3 C: *¿Hasta qué horas?*
4 R: *Las siete de la tarde.*
5 C: *A las siete yo lo dejo ahí donde se tiene que ir. ¿Verdad? Pa que no le pongan falta.* (Chih'97#10Aa)

1 C: 'Till what time **are you drinking** there?'
2 R: 'Every day.'
3 C: 'Till what time?'
4 R: 'Seven in the evening.'
5 C: 'At seven I'll leave you there where you have to be. So you don't get an absent mark.'

Ex. (17) is especially interesting because it also shows retention of locative meaning in *hasta qué horas está pisteando*, something like 'until what time are you there drinking?'. Locative meaning is reinforced by *ahí* 'there' and the context of the exchange, in which C is telling R that she will drive him back to the location of habitual drinking. Here *estar* + gerund means both 'you are there, at the place where you drink' (locative, lexical) and 'you do so habitually' (habitual, grammatical).

The case of *estar/andar* + gerund variation seems to exemplify a view of grammar as not separable from the lexicon (cf. Hopper 1987, Bybee 1998). It is important that synchronic variation may include cases of ambiguity or merger between more lexical and more grammatical uses (Coates 1983:15–17, 1995:61; Heine 1993:52). Although much work has been done to distinguish auxiliary from lexical periphrases, or to determine degree of grammaticization (cf. Olbertz 1998), it seems that linguistic reality is more scalar. A scalar view of periphrastic constructions based on a diachronic perspective (cf. Heine 1993, Bybee et al. 1994), rather than strict classifications between lexical or auxiliary, would allow precisely for cases in which auxiliary and lexical meaning merge, as in the example above. Even *estar* + gerund, which is considered the most "highly" grammaticized of Spanish gerundial periphrases (cf. Quesada 1995; Olbertz 1998:299–301, 479), presents synchronic variation, including spatial uses very close to the lexical end of the scale. Cases of merger belie the view that lexical and auxiliary readings are exclusive.

It has been argued (e.g. Lavandera 1978) that grammatical expressions cannot be true sociolinguistic variables, like phonological variables, because variant expressions have different referential-functional-pragmatic meanings. The problem with semantic interpretations is that they are difficult, if not impossible, to operationalize and test objectively. As Poplack & Tagliamonte (1999:321) point out in their study of *going to* and other future expressions, given the analyst's lack of access to speaker intent and hearer inference, "attributions of semantic moti-

vations or interpretations of variant selection are no more valid than the alternative assumption . . . of the 'neutralization' of any functions carried by these variants in 'unreflecting discourse' (see D. Sankoff 1988)." What we can objectively do in cases of grammatical variants is examine distribution constraints (Poplack & Tagliamonte 1999:322).

In the case of grammaticizing forms in the same functional domain, a more fundamental problem with the insistence on differences in meaning between variant forms is the assumption that each form has a "basic" meaning that can be counterposed to the basic meaning of another form. In this case, the claim would have to be that *estar* + gerund is basically a progressive whereas *andar* + gerund is basically a frequentative. As the present data indicate, it is hard to pinpoint a single invariant meaning for either *estar* + gerund or *andar* + gerund. As we have seen, both cover a range of uses, from locative and motion to progressive-continuous to frequentative-habitual; that is, both expressions are polysemous. Moreover, a particular occurrence of either one can itself be ambiguous, merging a range of meanings along its grammaticization path. As we saw in ex. (17), *está pisteando* 'you are drinking' merges locative and aspectual meaning.

The term LAYERING has been used in two different ways in grammaticization studies, and both kinds of layering are crucial for variationism. In Hopper's (1991) original discussion, layering was explained as formal diversity, and this is how many scholars use the term (e.g. Bybee et al. 1994:21–22). Others have drawn more on what I am calling LAYERING AS POLYSEMY. An example is Thompson and Mulac's (1991:325) discussion of the grammaticization of epistemic parentheticals in English, where *I think* as a subject-verb phrase (*I think that* . . .) coexists with *I think* as an epistemic phrase (when it occurs without *that*, either before a clause or in some other position). In variation studies, both kinds of layering come into play. In their study of *be like*, Tagliamonte & Hudson (1999:149–50, 152–53) describe variation among a cohort of quotatives – *say*, *go*, *be like* – (= layering as formal diversity); they also discuss the multi-functionality of *(be) like* itself, as an introducer to dialogue as well as a marker of focus or new information (= layering as polysemy).

To summarize, grammaticization results in layering of two complementary kinds: formal and semantic. These are depicted in Figure 3. On the one hand, we find layering as formal diversity, or synchronic variation among different forms in the same functional domain. We saw that *estar/andar* + gerund are functionally (aspectually) equivalent, by the measures of direct object number and co-occurring locative and temporal expressions. On the other hand, we have layering as polysemy, or synchronic variation among different meanings in the same form. We have seen that *estar/andar* + gerund may both have locative-motion uses, and that both cover a range of aspectual meanings from progressive to habitual. Thus, a diachronic perspective provides a solution to the problem of meaning differences in grammatical(izing) variables.
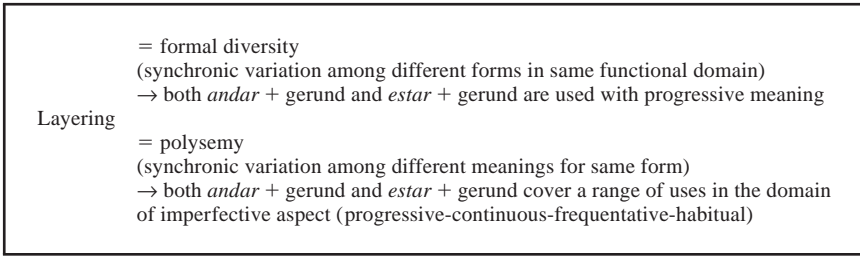
FIGURE 3: Layering as formal diversity and polysemy.

RETENTION: EARLIER MEANINGFUL DISTRIBUTION
PATTERNS → COLLOCATIONAL ROUTINES

So, if it is not aspectual meaning differences that determine the distribution of *estar/andar* + gerund, what linguistic factors might constrain their occurrence? It turns out that the most important linguistic constraint in the distribution of *estar/andar* + gerund is co-occurring main verb classes and main verb types. A second, related conditioning factor is whether the activity referred to is an outdoor or indoor one. These results point to a second major principle of grammaticization – retention. However, this is not so much a case of retention of earlier meaning differences as of the retention of patterns of distribution.

Variable rule analysis

I submitted data from the educated Mexico City (MexCult), popular Mexico City (MexPop), and northern states (MexRep) corpora to variable rule analysis using GoldVarb (Rand & Sankoff 1990), a type of multivariate analysis that considers factors (constraints) simultaneously and picks out the ones that contribute a statistically significant effect to the choice of variants. Table 9 displays the results of a variable rule analysis of the contribution of factors to the choice of *andar* + gerund vs. *estar* + gerund. Factor groups included in the run were corpus, main verb class, and location of the activity. Weights are interpreted thus: values above .500 mean that *andar* is favored, values below .500 indicate that *estar* is favored, and a value of .500 indicates that the constraint does not strongly favor one variant over the other.

The results for corpus are as expected, based on earlier comparisons of the relative frequency of *andar* + gerund in popular and educated data (see above). It is interesting, though, to note that the weight for Interviewer (*Encuestador*) is .459, compared with .188 for the educated corpus. The interviewers for these corpora were linguists or linguistics students, so we might expect that, as speakers of the educated variety, they would favor *estar* more strongly. When we look

TABLE 9. *Variable rule analysis of the contribution of factors selected as significant to the probability of* ANDAR + *gerund*[a]

	Percent *andar*	Factor Weight	Percent of data
Main verb (gerund) class			
non-directional motion verb	87%	.981	3%
other motion verb	21%	.641	2%
physical activity verb	23%	.665	20%
verb of speech	11%	.508	17%
general activity verb	13%	.495	35%
bodily activity verb	8%	.351	4%
mental, perception, stative-locative	4%	.173	19%
	Range	*808*	
Corpus[b]			
Popular	22%	.690	59%
Educated	3%	.188	31%
Interviewer	12%	.459	10%
	Range	*502*	
Location of activity			
Outdoors	36%	.624	7%
Indoors	3%	.150	10%
Indeterminate	15%	.544	82%
	Range	*474*	
Overall	15%		

N = 920, Input .085, Log likelihood = −297.800, Significance = .000
Total Chi-square = 31.5425, Chi-square/cell = .6309
[a] In an earlier run with the factor groups verb class, locatives, temporals, and location of activity, neither temporals nor locatives were selected (see note 10).
[b] Popular = MexPop (392 tokens) and MexRep (151) combined; Educated = MexCult (287); Interviewer = *Encuestador* in MexPop, MexRep, MexCult (90) (see note 3).

at each corpus separately, it turns out that the relative frequency of *andar* for the interviewer was 0 (0/22) in the educated Mexico City corpus, 15% (6/41) in the popular Mexico City corpus, and 20% (4/20) in the northern states data. Assuming that the same people, or people from the same group, were conducting the interviews, the adjustments in relative frequency suggest that the interviewers were accommodating (Giles 1980) to the different informant groups. This result supports the hypothesis of social significance for *andar* + gerund.

Let us now consider the results for main verb class, the factor group that was selected first in the GoldVarb step-up analysis and that shows the greatest range. *Andar* is favored by motion verbs such as *yendo* 'going' and most strongly (.981) by nondirectional motion verbs such as *dando la vuelta* or *dando vueltas* 'going around', as in ex. (18):

(18) *En los domingos, no se puede caminar porque es una aglomeración. Hay mucho– O sea, [ ]*
como que la gente nada más circula por esa calle y todos andamos por esa calle. Ves. Las
*otras calles, están libres, pero toda la gente **anda dándose** no más **la vuelta** por esa calle.*
(Chih'97#17a.Dor)
'On Sundays, you can't walk [along the main street in Ascención] because it's a throng.
There's a lot of– I mean, [ ] it's like people only use that street and we all go by that street.
The other streets are empty but everybody just **goes around going around** (going up and
down, back and forth) along that street.'

The selection of nondirectional motion *andar* with another verb of nondirec-
tional motion implies a diachronic process of bleaching. This echoes Poplack
& Tagliamonte's (1999:335–36) findings on the use of English future *going to*
with main verbs of motion. However, it is unlike the case of *going to*, where
use with motion verbs apparently was avoided in earlier varieties (Poplack &
Tagliamonte 1999:338). The diachronic change does not lie in the extension of
andar to use with motion verbs. *Andar* has co-occurred with motion-verb ger-
unds since its earliest attestations in Old Spanish, as shown in ex. (19). Nor is
the change a quantitative increase in the relative frequency of motion-verb ger-
unds cooccurring with *andar*. In the Old Spanish corpus I studied, 14% (16/
117) of *andar* tokens pair up with a motion-verb gerund such as *corriendo*
'running', *fuyendo* 'fleeing', or *siguiendo* 'following, chasing' (Torres Cacoul-
los 2000:162). In the present-day popular data (Chihuahua, MexPop, MexRep),
the proportion of *andar* with motion-verb gerunds is about the same as in the
Old Spanish corpus, at 11% (24/217). These combinations with another mo-
tion verb were originally harmonic motion verb expressions, where *andar* was
the main verb and the gerund was a manner adverbial (Torres Cacoullos
1999a:34–36; cf. Lyons 1977:807, Bybee et al. 1994:214). For example:

(19) *come⟨n⟩c'o a **andar corriendo** por toda la casa dando loor a dios* (EE1-118r70)
'he began to go around running all over the house praising God'

In the example above, the subject literally (physically) goes around the house
(*andar . . . por toda la casa*), and the motion verb gerund *corriendo* 'running' tells
the manner in which he was going around.

The diachronic change is that *andar* + motion verb combinations have be-
come collocational routines where *andar* is bleached of motion meaning. *Dando*
la vuelta/vueltas 'going around, strolling around, driving around' alone makes up
25% (6/24) of the *andar* + motion-verb gerunds. In ex. (18), *toda la gente anda*
dándose no más la vuelta 'everybody just goes around going around', *andar*
contributes mostly aspectual meaning: Omitting it would not reduce the physical-
motion meaning component of *andar* + *dando la vuelta*. Collocational routines
play a big role in *estar/andar* + gerund variation, as we will see below.

Table 9 indicates that also favoring *andar* are verbs of physical activities, such
as *jugando* 'playing'. Playing may be congruent with 'going around'. Further-
more, this activity is often realized outdoors, a point to which I will return. In
contrast, perceptible bodily activities such as *durmiendo* 'sleeping', *esperando*
'expecting a child (being pregnant)', and *llorando* 'crying' favor *estar*. Also fa-

voring *estar* are mental activity verbs (*pensando* 'thinking'), perception verbs (*viendo* 'seeing, watching'), stative-locatives (*viviendo* 'living'), and "abstract" verbs (*tendiendo hacia* 'tending toward').

Most numerous is what I call the "general activity" verb class. These verbs refer to an activity composed of more specific sub-activities none of which alone constitutes the general activity. A general activity need not be restricted to a particular location or a particular physical manifestation. Examples are *ayudando* 'helping', *estudiando* 'studying, going to school', *trabajando* 'working', and different *haciendo* 'doing' predicates. This group makes up 35% of the data. That the biggest verb class does not seem to favor one auxiliary over the other, as indicated by a factor weight very close to .500, supports the claim that the meaning of the auxiliaries has converged. Similarly, verbs of speech, the third most numerous class with 17% of the data, also do not seem to strongly favor one over the other.

However, there is variability within the verb classes themselves. Different verb types within both these classes show clear patterns of preference for a particular auxiliary. These differential preferences are related to retention in grammaticization.

Retention in main verb type distribution

Grammaticizing constructions retain features or nuances of meaning of the source construction. This is known as the RETENTION (Bybee & Pagliuca 1987) or PERSISTENCE (Hopper 1991) hypothesis. Retention is reflected in differences in the main verb types with which the two auxiliaries are used. The skewed distribution of gerunds is evident in Old Spanish texts. For example, in the 13th-century chronicle *Estoria de España*, every time the gerund *fablando* 'talking' co-occurrs with an auxiliary, it is with *estar*. 'Talking', 'saying', 'chatting' and other verbs of speech are activities usually circumscribed in one location, which is consistent with the 'be located' meaning of *estar*. On the other hand, *buscando* 'looking for' tends to co-occur with *andar*. This predicate is consistent with the nondirectional motion, "going around" meaning of *andar*: one may have to go around to look for something. It is interesting that the verbs of speech associated with *andar* in *Estoria de España* – *predigando* 'preaching' and *preguntando* 'asking (for), inquiring' – are also congruent with physical motion. In ex. (10), the subject literally goes *a todas partes* 'all over the place' inquiring:

(20) **ando** *tanto* **preguntando** *a todas partes' q⟨ue⟩ fallo un iudio q<ue> auie nombre iudas.* (EE1-118r26)
'he **went** (walked) so much **inquiring** all over the place that he found a Jew whose name was Judas'

The collocational patterns evident in the *Estoria de España* are vigorous in present-day data as well. Table 10 shows the main verbs that occur most frequently with *estar* and *andar* in the popular Mexico City (MexPop) and the Chihuahua corpus. Both corpora represent about 17 hours of recordings, and the total

TABLE 10. *15 most frequent gerunds in two popular Mexican Spanish oral corpora (in descending order).*[a]

	Gerund totals		ESTAR		ANDAR	
	MexPop	Chih	MexPop N = 340	Chih 356	MexPop 89	Chih 88
	N	N	%	%	%	%
hacer 'do'	23	41	70	73	30	27
trabajar 'work'	32	27	**97**	**74**	3	26
decir 'say'	15	30	**80**	**97**	20	3
platicar 'chat'	11	25	**91**	**100**	9	0
pagar 'pay'	19	7	**100**	**100**	0	0
esperar 'wait'	18	7	**100**	**100**	0	0
hablar 'talk'	6	16	**100**	**94**	0	6
tomar 'drink'	14	8	**100**	**88**	0	12
jugar 'play'	11	10	73	20	27	80
dar – various	12	8	67	75	33	25
estudiar 'study'	13	5	**92**	**100**	8	0
ver 'see'	6	12	**83**	**83**	17	17
vivir 'live'	7	10	**100**	**100**	0	0
buscar 'look for'	10	6	20	17	**80**	**83**
grabar 'record'	2	7	**100**	**100**	0	0

[a] **Bold** indicates verbs occurring overwhelmingly with one as opposed to the other auxiliary. Underline in *andar* column indicates differences between the two corpora: for *trabajar* and *jugar*, the relative frequency of *andar* is higher in Chihuahua than in MexPop (p = .0108, p = .0157, respectively). Chi-square tests for equality of proportions indicate no other significant differences in relative frequencies.

number of tokens is 340 and 356, respectively, for *estar*, and 89 and 88, respectively, for *andar*. The Mexico City corpus represents an urban variety. The Chihuahua corpus has both rural and urban speakers, though even many of those recorded in the state capital maintain regular contact with family in rural parts.

The results for both corpora with respect to the distribution of main verb types are nearly identical. Gerunds in the verbs of speech class tend to occur with *estar*: *diciendo* 'saying', *platicando* 'chatting' and *hablando* 'talking' have a relative frequency for *estar* close to 100%. On the other hand, main verb *buscando* 'looking for' tends to occur with *andar*, with a relative frequency of 80% in the Mexico City and 83% in the Chihuahua corpus. This is the same pattern found in the 13th-century *Estoria de España* text.

Retention also shows up in the patterning of what we might call "verbs of warring" with *andar*. The *Estoria de España*, being a history of heroes and great events, abounds with "destruction" and "conquering" verbs such as *astragando*

tierra 'ravishing the land', *destruyendo* 'destroying', *conquiriendo* 'conquering', *guerreando* 'warring'. For example (see also ex. [9]):

> (21) Association of *andar* with fighting and "warring" in *Estoria de España*
> *Auino assi que ell emperador diocleciano que **andaua guerreando & conquiriendo** toda tierra de egipto* (EE1-108r97)
> 'It happened that the emperor Diocleciano that he **was going around fighting and conquering** all the land of Egypt'

The "warring" class of verbs makes up about one-third of gerund tokens combining with *andar* in this text. Such verbs are highly compatible with the "going around" meaning of *andar*, since conquering involves outdoor expanses and many (plural) victims.

The association of *andar* with struggle and war continues in present-day varieties. Ex. (22) is about the activities of Pancho Villa, a central figure in the Mexican revolution of 1910. The speaker, from the village of Ascención, Chihuahua, was about ninety years old at the time of the recording.

> (22) Association of *andar* with fighting and "warring" in Chihuahua corpus
> *Y así fue. E. Cuando la revolución, cuando Villa. […] **Andaba andaba pegando** muy duro. Ves. A todos les **andaba pegando**. A los ricos. A los pobres casi no les hacía nada. No a los pobres no les hacía nada. En lo más a los puros ricos […] que decían ellos querían mandar ellos más que de nos[j]otros. Ese es lo, es lo que **andaba peleando** Villa, e, muy cierto. **Anduvo peleando** muy duro. Les echó unos fregazos aquel carajo. Sí.* (Chih'97#23A)
> 'And that's how it was. At the time of the revolution, at the time of Villa. He **was (going around) was (going around) hitting** very hard. He **was (going around) hitting** all of them. The rich. To the poor he really didn't do much. No, he didn't do anything to the poor. Just the rich, the ones who said they wanted to boss around more than us. That's what, that's what Villa **was (going around) fighting**, it's very true. He **was (going around) fighting** very hard. He really gave them some blows, that devil. Yes.'

Collocational routines

I propose that the pattern of distribution of *estar/andar* + gerund in present-day Spanish reflects retention, but not so much the retention of original meaning differences as the retention of distribution patterns. In other words, it is the retention of collocational routines that constrains variation between the grammaticizing constructions.

High-frequency collocations are highly "entrenched" in Langacker's (1987) or Bybee's (1985, 1998) terms. The verb types in Table 10 make up a big chunk of the data: 50% of all *estar* tokens in both corpora and 33% and 43% of *andar* tokens in popular Mexico City and Chihuahua, respectively. Strong evidence for collocational routines emerges from the combined Chihuahua, MexPop, MexRep, and MexCult data. On the one hand, high-frequency verbs of speech pattern with *estar*: 98% (55/56) of *hablando* tokens pair up with *estar*, as do 98% (50/51) of *platicando* and 92% (61/66) of *diciendo* tokens. These three collocations alone – *estar hablando* 'be talking', *estar platicando* 'be chatting', and *estar diciendo* 'be saying' – add up to 15% (166/1136) of all *estar* + gerund tokens, about one-sixth

of the data. On the other hand, 80% (20/25) of *buscando* tokens pair up with *andar*. *Andar buscando* 'be looking for' alone makes up 8% (19/228) of all *andar* + gerund tokens. This collocation is so frequent that Spaulding (1926:259) calls it "a set phrase," and, more recently, Squartini (1998:261) refers to it as a case of "lexical specialization."

The fact that some main verb types almost always occur with one as opposed to the other auxiliary supports the claim that it is not aspectual differences that distinguish the two. Otherwise, we would have to defend the position that *hablando* 'talking' inherently lends itself to progressive, while *buscando* 'looking for' inherently lends itself to frequentative. In the absence of cross-linguistic evidence, this seems indefensible.

Instead, *estar/andar* + gerund variation may be described as the residue of older patterns of use that were once meaningful. Thus, main verb *buscando* tends to pair with *andar*, even when there is no physical "going around" motion involved. In ex. (23), the speaker is talking about her granddaughter looking for a date in the calendar. Physical motion is precluded here, and there do not seem to be any connotations of intensity or negative attitude (see note 11). The meaning is past progressive or continuous, not frequentative. The use of *andar buscando* here is best described as a collocational routine.[13]

> (23) *Y esta mañana andaba con el almanaque, dijo "¿qué día es [ ] mamá?" [...] "Dice el 24. Pero dice aquí Bautista" dice, "no más dice Bautista." Le digo "pues esa es Juan Bautista" le dije. Aha. [Risa] Yo no sabía por qué* **andaba buscando** *el día de San Juan.* (Chih'97#17a.Sab)
> 'And this morning she was (ANDAR) with the calendar, she said "what day is [] mama?" [...] "It says the 24th. But it says here Bautista" she says, "it only says Bautista." I tell her, "Well that is Juan Bautista" I told her. Umhm. [Laughter] I didn't realize why she **was** (ANDAR) **looking** for the day of San Juan.'

The little girl was looking for the day of San Juan because she wanted her mother to get her a water pistol, since the custom that day is to throw water at people.

The results of this study with respect to the patterning of main verbs are reminiscent of findings on French Subjunctive use. Poplack (1992:255) showed that it is not so much semantic classes of matrix verbs – such as those with a volition, emotion, or doubt component – but particular lexical items that trigger the subjunctive. This scholar (Poplack 1992:246,249) found that most occurrences of the Subjunctive were cases of impersonal verb *falloir* 'have to', such as Present *il faut* and Imperfect *il fallait*, followed by a highly frequent irregular verb such as *avoir* 'have', *être* 'to be', *aller* 'to go', or *faire* 'to make, do'. In other words, mood choice does not appear to be semantically motivated (Poplack 1992:256–57). As Bybee & Thompson (2000:384) have pointed out: "While grammatical analysis should proceed with the working hypothesis that formal distinctions represent functional distinctions, we also have to bear in mind that not all contrasts and distributions are meaningful or functional. Some patterns represent a lexically arbitrary residue of formerly productive patterns."

ORIGINAL LEXICAL MEANING → SOCIAL MEANING

In summary, collocational routines largely account for the distribution of *estar/ andar* + gerund. But we also know that social factors are important, since we have seen differences between popular and educated varieties. Where does this social meaning come from? Part of the answer, at least, is provided by a grammaticization perspective. I suggest that this social difference derives from the original motion meaning of *andar*. The "going around" meaning leads to the co-occurrence of *andar* with main verbs referring to outdoor and rural activities, and thus to the association of *andar* + gerund with rural and popular varieties.

Returning to Table 10, we see that there are two points in which the popular Mexico City and the Chihuahua corpora differ. One is with main verb *trabajando* 'working'. *Trabajando* is the single most frequent gerund in the combined Chihuahua, MexPop, MexRep, and MexCult data, with 7.5% (102/1364) of all *estar/ andar* tokens. In the popular Mexico City corpus, *andar trabajando* has a relative frequency of only 3% (1/32), but in the Chihuahua corpus, it is 26% (7/27) (Chi-square = 6.49547, p = .0108). If we look closely at what kind of 'work' *andar trabajando* refers to, we see that in most cases it is outdoor work, usually in agriculture. For example:

> (24) *Ahorita **andan trabajando** en las pizcas y allá andan* (Chih'97#1aCJ)
>   'Right now **they are working** (ANDAR) in the crops and they are there'

More than half of the *andar trabajando* tokens (4/7) in the Chihuahua corpus refer to outdoor, agricultural work. The one token of *andar trabajando* in the Mexico City corpus also refers to working in the fields (UNAM 1976:78). Thus, one way that we might explain the difference between the two corpora in the relative frequency of *andar trabajando* is by the number of rural speakers in the Chihuahua corpus as opposed to the urban speakers in the popular Mexico City corpus.

However, it is important that agricultural or rural work does not exclude *estar*. 25% (5/20) of the *estar trabajando* tokens in the Chihuahua corpus also refer to work in agriculture. Ex. (25) is from the sister of the person who gave ex. (24) and refers to the same people, their parents, who are – *estar* or *andar* – working in the fields:

> (25) *Y **están trabajando** allá ahorita andan, pues andan en qué. En el ajo creo* (Chih'97#2bPH)
>   'And **they are working** (ESTAR) there right now, they are, well they are in what? In garlic I think'

The variation between *andar* and *estar* + *trabajando* in these examples provides another nice illustration of layering and overlapping contexts of occurrence.

The second point in which the Mexico City and the Chihuahua corpus differ is with main verb *jugando* 'playing'. Again, the difference seems to lie in outdoor versus indoor playing. In the Chihuahua corpus, 75% (6/8) of *andar jugando* tokens refer to playing outdoors. In contrast, one *estar jugando* token refers to playing with dolls and the other is a figurative use of 'playing' in reference to

politicians. Exx. (26–27) show the outdoor/indoor contrast between *andar* and *estar + jugando*:

> (26) ... *de chamaco que **andábanos jugando** los caballitos* (Chih'97#18aCh)
> '... as a boy when **we would be playing** (ANDAR) at horses'

> (27) *Porque hasta bien noche **estábamos jugando** con las muñecas.* (Chih'97/#2aPH)
> 'Because until late at night **we would be playing** (ESTAR) with our dolls'

The same outdoor/indoor contrast shows up below, where *limpiando* 'cleaning' the dishes is with *estar*, but *limpiando* 'cleaning, clearing' the land is with *andar*[14]:

> (28) *una hermana de mi mamá **estaba lavan– limpiando** los trastes [ ] Y ella oía bulla. Pero no se daba cuenta ella que [] que los niños andaban jugando* (NMmon/Mar)
> 'a sister of my mother **was wash– cleaning** (ESTAR) the dishes. And she heard noise. But she didn't realize that the children were playing'

> (29) *Tenía como– como ocho hombres conmigo cuando **andábanoj limpiando** el onde iban a hajer laj casas y todo* (NMmon/NMCSS#311)
> 'I had about– about eight men with me when **we were cleaning** (clearing) (ANDAR) the– where they were going to build the houses and all that'

With both *estar* and *andar*, the meaning is past progressive or continuous, 'was/were cleaning'. Thus, the meaning difference is not aspectual. However, the first activity is washing dishes, which is usually done indoors, while the second is clearing land, which is an outdoor activity.

The outdoor/indoor distinction also appears if we look more closely at *en* 'in' locatives co-occurring with *andar* + gerund. Here, 50% (4/8) were wide, outdoor spaces: *en las pizcas*, *en el campo*, *en la calle*, *en el recreo* 'in the fields, in the countryside, in the street, in the playground'; another 25% were cities, *en Cuauhtemoc*, *en Los Angeles y dondequiera*; and the remainder were vehicles, *en el camión* 'in the bus', *en un coche* 'in a car' (combined Chihuahua, MexPop, MexRep, and MexCult data).

It is not surprising that agricultural and outdoor activities in general should favor *andar*, given its original lexical meaning of 'going around'. Indoor activities in a circumscribed space, by contrast, should favor *estar*, originally meaning 'be located'. The results of the variable rule analysis in Table 9 indicate that *andar* is selected with outdoor activities and *estar* with indoor activities, while with activities of indeterminate location one is not particularly favored over the other. In most cases the outdoor/indoor distinction was not relevant or determinable (82%), but with some predicates the activity clearly happened indoors, as in *planchando* 'ironing' and *bañandose* 'bathing', while in others it was outdoors, as in *sembrando* 'planting' and *ordeñando las vacas* 'milking the cows'.

Table 11 depicts the proportion of *andar* + gerund and *estar* + gerund occurring with an outdoor activity, in combined popular Mexico City, educated Mexico City, and northern states data. For *andar*, 18% of all tokens were outdoor, compared to just 2% indoor activities. For *estar*, the skewing is reversed, though it is not as great, with 5% outdoors and 12% indoors. Thus, the distinction between

TABLE 11. ESTAR *and* ANDAR + -NDO *with outdoor vs. indoor activities.*[a]

Location of activity:	Outdoors	Indoors	Indeterminate	Total
	%	%	%	N
ESTAR + -NDO	5	12	83	777
ANDAR + -NDO	**18**	**2**	80	143
Total	7	11	82	920

[a] MexPop, MexRep, MexCult data combined.

outdoor and indoor activities turns out to be important in accounting for present-day patterns of variation between the two.

The question now is whether social differences in the relative frequency of *andar* + gerund are a straightforward reflection of differences in topics of conversation. In other words, the difference between popular and educated varieties might be due to differences in the kinds of activities working-class as opposed to educated speakers talk about. If educated speakers spend more time talking about abstract things and less time talking about outdoor activities, then it would not be surprising that they use *andar* less. Thus, the social difference revealed earlier might be a simple epiphenomenon of functional-lexical differences (cf. Poplack 1997).

However, the data suggest that there really is a social difference. At least with respect to the outdoor/indoor activity distinction, popular and educated varieties are about the same. The first two rows in Table 12 show that, in both the educated and popular Mexico City corpus, *estar/andar* + gerund outdoor activities make up 5% of the data. Indoor activities make up 9–13% (the indoor difference is not statistically significant at p < .05).

There does seem to be an urban-rural difference, however. The third row in Table 12 shows that, in the northern states corpus, outdoor activities make up 15% of the data, more than in either the educated or popular capital city corpus (comparing MexRep and MexPop, Chi-square = 17.63768, p = .0000).

Here again we have two different line-ups. The popular Mexico City and northern states data are different in the overall proportion of outdoor activities, but they line up in the relative frequency of *andar*, as does the Chihuahua corpus. The urban educated Mexico City and popular Mexico City corpora line up together in the amount of time *estar/andar* + gerund overall are used for outdoor activities, but they differ sharply in the relative frequency of *andar*. Thus, even though there is an urban/rural difference among the popular varieties in the proportion of outdoor activities, the relative frequency of *andar* is the same. And even though the urban varieties have the same proportion of outdoor activities, the relative frequency of *andar* is higher in the popular than in the educated variety. This pair of results points to social rather than functional-lexical differences.

TABLE 12. *Outdoor vs. indoor activities in popular and educated Mexico City corpora.*

Location of activity:	Outdoors		Indoors		Indeterminate		Total
	N	%	N	%	N	%	N
MexPop	22	**5**	41	9	376	86	439
MexCult	17	**5**	41	13	252	81	310
MexRep	26	**15**	15	9	130	76	171

Functional-lexical differences can acquire social meaning. The case of *andar* + gerund is reminiscent of clitic position in Spanish, a sociolinguistic variable subject to register effects, where postposed clitics (*voy a decirte* 'I am going to tell **you**') are more frequent in formal than informal varieties. Yet there is a diachronic dimension to variation in clitic position, since preposed clitics increase as periphrastic constructions grammaticalize (Myhill 1988). What has happened is that the persistence of older patterns in written varieties has become a mark of formality in its own right (Torres Cacoullos 1999b). Similarly, a diachronic association with outdoor and rural activities has led to the development of *andar* + gerund as a popular variant.

In sum, we have seen how diachronic bleaching results in synchronic layering in the same functional domain. Both kinds of layering are important: formal diversity (different forms in the same functional domain) and polysemy (one form with different meanings or functions). Expressions grammaticizing along parallel paths thus become variants of a single variable. At the same time, retention of original meaning features shows up in synchronic distribution patterns and collocational routines. These results suggest that the social stratification of *estar/andar* + gerund may originate as an indoor/outdoor, urban/rural difference: Rural activities in large outdoor spaces are more compatible with the original meaning of *andar*. But what began as a functional lexical difference has now acquired social meaning.

NOTES

[1] The term "auxiliation" is taken from Benveniste 1968. See Hopper & Traugott (1993:25).

[2] Lope Blanch was born in Spain but works in Mexico.

[3] Note on data sources: The results in Tables 4 and 5 exclude data from the interviewer (*Enc = encuestador*); similarly, quoted material was excluded in the Essays-Academic Prose corpus (Table 3). Tables 6–8 and 10–12 include interviewer data, which is considered separately in Table 9. Table 1 figures from Luna 1980 and Arjona 1991 are based on expanded versions of MexCult and MexPop, respectively.

[4] In regard to Table 5, it is fair to state that the ratios in MexCult and Essays-Academic are "virtually" the same, given the small number of tokens, which is a true reflection of the low token frequency of these constructions. Token frequency is usually stated in terms of the average number of occurrences per a certain number of words of text, as in Mossé's frequency coefficient, "instances of x/ 100 000 words of text" (Mair & Hundt 1995:113). To give an idea of *estar/andar* + gerund token frequency compared with the English Progressive, I calculated a frequency coefficient for the popular Mexico City corpus (MexPop), based on Clegg & Rodriguez's (1993) count of 172,699 running words of text and their figures for *estar* (366) and *andar* (103). This produces a Mossé coefficient of about 270 for *estar* and *andar* combined. Figures for the English Progressive range from the 300s in journalism to the 700–800s in fiction and drama (Mair & Hundt 1995:113,118).

[5] Both constructions most frequently occur without a direct object of any kind, 50% (70/140) of the time for *andar* + gerund and 58% (449/780) for *estar* + gerund.

[6] *Andar* + gerund + *por* locative + plural object is also compatible with progressive meaning. In the following Old Spanish example, the cowherd is literally (physically) going around the mountains watching his cows, and nondirectional motion (*andar* and *por*) blends not with frequentative but rather with progressive meaning, as indicated by the temporal expression *una vez* 'once':

> (i) *Vn pastor **andaua** una uez **por unos montes guardando sus uacas**. & uio ell una dellas que traye e[l] pie corto & marauillosse mucho quie⟨n⟩ gelo cortara.* (EE1-146v18)
> 'A cowherd **was going around** (ANDAR) once **on some mountains watching his cows**. And he saw one of them, that its foot was short, and he wondered who had cut it.'

[7] Adding 11 cases of *aquí, allí* 'here, there' + *en* (counted as "more than one") to the *en* locative count for *estar* still does not make the difference in *estar* (33/838) vs. *andar* (5/217) proportions with *en* statistically significant (Chi-square = 1.325102, p = .2497).

[8] In Table 6, Other locatives are: *a* 'to'(2), *adentro* 'inside', *afuera* 'outside', *de* 'from' (2), *pallá* 'over there' (2), for *estar*; and *adentro* 'inside', *alrededor* 'around', *arriba* 'above', *dondequiera* 'all over', *pallá y pacá* 'here and there', *para el lado de* 'toward', for *andar*. The difference in "Other" locatives is significant at p < .05, though not at p < .01 (Chi-square = 4.314094, p = .0378).

[9] In Table 8, the *Cuando* row is cases where *estar/andar* + gerund occurred in a clause headed by *cuando* 'when', as in *cuando anda uno trabajando ni el sueño le da* 'when you are working you don't feel sleepy' (Chih'97#17b.Sab). Cases of *ya* plus another adverbial were counted with the other type, e.g., *ya cuando* was counted toward the *cuando* total. Co-occurring *desde* 'since' + time phrases and *desde que* … 'since' clauses, as well as tokens occurring in *desde cuando* 'since when', *¿cuánto tiempo* … *?* 'for how long?', and *hace*-plus-time *que* 'it's been x time that' clauses, were included in the durative adverbial count.

[10] Neither temporals nor locatives were selected in a GoldVarb analysis of tokens from MexPop, MexRep, and MexCult. Other factors in the run were main verb class and outdoor vs. indoor activity.

[11] Nonaspectual meaning nuances are beyond the scope of this article, though they deserve further study. An example of what Yllera (1980:77) calls "intensive activity viewed negatively by the speaker" might be:

> (ii) *"Ay! […] voy a decirle a Martita que tú **no más andas jugando** y que tú **no más andas haciendo** esto y que l'otro, y que **andas dando no más la vuelta**, y nada que trabajas"* (MexRep – UNAM 1995:80)
> "'Oh! […] I'm going to tell Martita that you **just go around playing** and that you **just go around doing** this and that, and that you **go around just going around**, and don't work at all'"

Notice in the example above that negative attitude may be attributed to the co-occurrence of *no más* 'just'. Three operational measures of intensity and negative attitude in *andar* + gerund we could consider are the co-occurrence of *siempre* 'always', tense form distribution, and the proportion of negative polarity cases. *Siempre*, described by Squartini (1998:264–66) as a hyperbolic iteration adverbial, pairs up with *estar* proportionally, at 83% (5/6) in the present data (Chihuahua, MexPop, MexRep). The occurrence of Infinitive forms, though, is not proportional: 49% (40/81) are *andar*. Present Subjunctive and Imperative forms also seem to favor *andar*, with 39% (7/18) and 57% (4/7), respectively. The negative attitude in *andar* + gerund may be related to its occurrence in what I have

called the "negation or rejection of a situation construction," which is of the form *qué* or *cómo* + *ir a* + *andar* (Infinitive) + gerund (Torres Cacoullos 2000:167); for example:

(iii) *Cómo se va a andarse uno paseando, pues no oye* (Chih'97#22b.LT)
    'how could one be strolling around, no, that's not right'

All Imperative forms in the data, both *andar* and *estar*, were negated; for example:

(iii) "*no diga eso, no me ande diciendo eso," me dice* (Chih'97#7a.Ch)
    '"don't say that, don't be saying that," she tells me'

However, whether intensive or negative connotations should be attributed to *andar* + gerund or whether they arise from the constructions or contexts it occurs in, the case remains that it is used less frequently in educated than in popular varieties of Mexican Spanish.

[12]*Pistear* is not so much the physical action of drinking as a social activity.

[13] Parallel processing or repetition effects (see ex. [23]) may also constrain *estar/andar* + gerund variation, and they may contribute to the social differences reported in this article, if *andar* has a lower token frequency overall (in whatever kind of construction) in educated than in popular varieties.

[14] Examples (28) and (29) are from recordings of older, near-monolingual speakers of traditional New Mexican Spanish; Ex. (29) is from the New Mexico-Colorado Spanish Survey (see Bills 1997).

DATA SOURCES

Chihuahua: 1997 recordings by author in Ascención and Chihuahua, Chihuahua, Mexico (17 hours).

MexPop: UNAM 1976: 14–461.

MexRep: UNAM 1995: 10–257.

MexCult: UNAM 1971: 11–317.

EE1 (*Estoria de España*): Kasten, Lloyd; Nitti, John; & Jonxis-Henkemans, Wilhelmina (1997). *The electronic texts and concordances of the prose works of Alfonso X, El Sabio*. Madison: HSMS. (CD-ROM).

PMC (*Poema de Mio Cid*): Menéndez Pidal, Ramón (1964). *Cantar de Mio Cid. Texto, gramática y vocabulario*. 4th ed., 3 vols. Madrid: Espasa-Calpe.

Essays – Academic:

Castellanos, Rosario (1966). *Juicios sumarios: Ensayos*. Xalapa: Universidad Veracruzana.

García Canclini, Néstor (1990). *Culturas híbridas: Estrategias para entra y salir de la Modernidad*. Mexico City: Grijalbo.

García Canclini, Néstor, (1993) ed. *El consumo cultural en México*. Mexico City: Consejo Nacional para la Cultura y las Artes.

Lope Blanch, Juan M. (1983). *Análisis gramatical del discurso*. Mexico City: UNAM.

Martínez, José Luis, (1958) ed. *El ensayo mexicano moderno*. Mexico City: Fondo de Cultura Económica.

Monsiváis, Carlos (1995). *Los rituales del caos*. Mexico City: Era.

Montemayor, Carlos (1985). *El oficio literario*. Xalapa: Universidad Veracruzana.

Paz, Octavio (1959). *El laberinto de la soledad*. Mexico City: Fondo de Cultura Económica.

Reyes, Alfonso. (1927). *Cuestiones gongorinas*. Madrid: Espasa-Calpe.

REFERENCES

Arjona Iglesias, Marina (1991). *Estudios sintácticos sobre el habla popular mexicana*. Mexico City: UNAM.

Bell, Alan (1984). Language style as audience design. *Language in Society* 13:145–204.

Benveniste, Emile (1968). Mutations of linguistic categories. In W. P. Lehman & Yakov Malkiel (eds.), *Directions for historical linguistics: A symposium*, 85–94. Austin: University of Texas Press.

Bills, Garland D. (1997). New Mexican Spanish: Demise of the earliest European variety in the United States. *American Speech* 72:154–71.

Binnick, Robert I. (1991). *Time and the verb*. Oxford: Oxford University Press.

Blansitt, E. L., Jr. (1975). Progressive aspect. *Stanford University Working Papers on Language Universals* 18:1–34.

Bolinger, Dwight (1973). Essence and accident: English analogs of Hispanic *ser – estar*. In B. B. Kachru et al. (eds.), *Issues in Linguistics: Papers in Honor of Henry and Renee Kahane*, 58–69. Chicago: University of Illinois Press.

Bybee, Joan L. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.

———— (1998). The emergent lexicon. *Chicago Linguistic Society* 34 (*The Panels*):421–35.

———— (in press). Mechanisms of change in grammaticization: The role of frequency. In Richard Janda & Brian Joseph (eds.), *Handbook of historical linguistics*. Oxford: Blackwell.

Bybee, Joan, & Pagliuca, William (1985). Cross-linguistic comparison and the development of grammatical meaning. In Jacek Fisiak (ed.), *Historical semantics, historical word formation*, 59–83. Berlin: Mouton de Gruyter.

————, ———— (1987). The evolution of future meaning. In A. G. Ramat et al. (eds.), *Papers from the 7th International Conference on Historical Linguistics*, 109–22. Amsterdam: John Benjamins.

————; Perkins, Revere; & Pagliuca, William (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

————, & Thompson, Sandra (2000). Three frequency effects in syntax. *Berkeley Linguistics Society* 23:378–88.

Chambers, J. K. (1995). *Sociolinguistic theory*. Oxford: Blackwell.

Cheshire, Jenny (1987). Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics* 25:257–82.

Clegg, J. Halvor, & Rodríguez, J. David (1993). Progressive constructions in the speech of U.S. Hispanics. Paper presented at XIV Spanish in the United States Meeting, San Antonio.

Coates, Jennifer (1983). *The semantics of modal auxiliaries*. London: Croom Helm.

———— (1995). The expression of root and epistemic possibility in English. In Joan Bybee & Suzanne Fleischman (eds.), *Modality in grammar and discourse*, 55–66. Amsterdam: Benjamins.

Comrie, Bernard (1976). *Aspect*. Cambridge: Cambridge University Press.

Eckert, Penelope (2000). *Linguistic variation as social practice*. Oxford: Blackwell.

García, Erica (1985). Shifting variation. *Lingua* 67:189–224.

Giles, Howard (1980). Accommodation theory: Some new directions. *York Papers in Linguistics* 9:105–36.

Givón, T. (1975). Serial verbs and syntactic change: Niger-Congo. In Charles N. Li (ed.), *Word order and word order change*, 47–112. Austin: University of Texas Press.

———— (1984, 1990). *Syntax: A functional-typological introduction*. 2 vols. Amsterdam: John Benjamins.

Greenberg, Joseph H. (1966). *Language universals*. The Hague: Mouton de Gruyter.

Heine, Bernd (1993). *Auxiliaries: Cognitive forces and grammaticalization*. Oxford: Oxford University Press.

Hopper, Paul J. (1987). Emergent grammar. *Berkeley Linguistics Society* 13:139–57.

———— (1991). On some principles of grammaticization. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 1, 17–35. Amsterdam: Benjamins.

———— (1998). Emergent grammar. (Revised version of Hopper 1987.) In Michael Tomasello (ed.), *The new psychology of language: Cognitive and functional approaches to language structure*, 155–76. Mahwah, NJ: Lawrence Erlbaum.

————, & Traugott, Elizabeth Closs (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Labov, William (1978). Where does the sociolinguistic variable stop? A response to Beatriz Lavandera. *Sociolingustic Working Papers*, 44. Austin: Southwest Educational Development Laboratory.

Langacker, Ronald (1987). *Foundations of cognitive grammar*. Vol. 1. Stanford: Stanford University Press.

———— (1996). A constraint on progressive generics. In Adele Goldberg (ed.), *Conceptual structure, discourse and language*, 289–302. Stanford: CSLI.

Lavandera, Beatriz (1978). Where does the sociolinguistic variable stop? *Language in Society* 7:171–82.

Lehmann, Christian. (1995 [1982]). *Thoughts on grammaticalization*. Munich & Newcastle: Lincom Europa.

Lope Blanch, Juan M. (1986). *El estudio del español hablado culto: Historia de un proyecto*. Mexico City: UNAM.

Luna Trail, Elizabeth (1980). *Sintaxis de los verboides en el habla culta de la ciudad de México*. Mexico City: UNAM.

Lyons, John (1977). *Semantics*. 2 vols. Cambridge: Cambridge University Press.

Mair, Christian, & Hundt, Marianne (1995). Why is the progressive becoming more frequent in English? *Zeitschrift für Anglistik und Amerikanistik* 2:112–22.

Meillet, Antoine (1958 [1912]). L'évolution des formes grammaticales. In *Linguistique historique et linguistique générale*, 130–48. Paris: Champion.

Milroy, Leslie (1987). *Observing and analyzing natural language: A critical account of sociolinguistic method*. Oxford: Blackwell.

Myhill, John (1988). The grammaticalization of auxiliaries. *Berkeley Linguistics Society* 14:352–63.

Olbertz, Hella (1998). *Verbal periphrases in a Functional Grammar of Spanish*. Berlin: Mouton de Gruyter.

Poplack, Shana (1992). The inherent variability of the French subjunctive. In Christiane Laeufer and Terrell A. Morgan (eds.), *Theoretical analyses in Romance linguistics*, 235–63. Amsterdam: Benjamins.

———— (1997). The sociolinguistic dynamics of apparent convergence. In Gregory Guy et al. (eds.), *Towards a social science of language: Papers in honor of William Labov*, *Vol 2: Social interaction and discourse structures*, 285–309. Amsterdam: Benjamins.

————, & Tagliamonte, Sali (1999). The grammaticization of *going to* in (African American) English. *Language Variation and Change* 11:315–42.

Quesada, Juan Diego (1995). *Estar + -ndo* y el aspecto progresivo en español. *Iberoromania* 42:8–29.

Rand, David, & Sankoff, David (1990). GoldVarb: A variable rule application for the Macintosh. Version 2. Montreal: Centre de recherches mathématiques, Université de Montréal.

Romaine, Suzanne (1984). On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. *Folia Linguistica* 18:409–37.

Sankoff, David (1988). Sociolinguistics and syntactic variation. In Frederick Newmeyer (ed.), *Linguistics: The Cambridge survey*, vol. 4, 140–61. Cambridge: Cambridge University Press.

Sankoff, Gillian (1986 [1973]). Above and beyond phonology in variable rules. In *The social life of language*, 81–93. Philadelphia: University of Pennsylvania Press. Originally published in C.-J. N. Bailey and Roger Shuy (eds.), (1973) *New Ways of Analyzing Variation in English*, 44–61, Washington, DC: Georgetown University Press.

Silva-Corvalán, Carmen (1997). Variación sintáctica en el discurso oral: Problemas metodológicos. In Francisco Moreno Fernández (ed.), *Trabajos de sociolingüística hispánica*, 115–35. Alcalá, Spain: Universidad de Alcalá.

———— (1999). *Ahora*: From temporal to discourse deictic. In Robert J. Blake, Diana L. Ranson, & Roger Wright (eds.), *Essays in Hispanic linguistics dedicated to Paul M. Lloyd*, 67–81. Newark, DE: Juan de la Cuesta.

Smith, Carlotta S. (1991). *The parameter of aspect*. Dordrecht: Kluwer Academic Publishers.

Spaulding, Robert K. (1926). History and syntax of the progressive constructions in Spanish. *University of California Publications in Modern Philology* 13:229–84.

Squartini, Mario (1998). *Verbal periphrases in Romance: Aspect, actionality, and grammaticalization*. Berlin: Mouton de Gruyter.

Tagliamonte, Sali, & Hudson, Rachel (1999). *Be like* et al. beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics* 3:147–72.

Thompson, Sandra A., & Mulac, Anthony (1991). A quantitative perspective on the grammaticization of epistemic parentheticals in English. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 2, 313–29. Amsterdam: Benjamins.

Torres Cacoullos, Rena (1999a). Variation and grammaticization in progressives: Spanish *–ndo* constructions. *Studies in Language* 23:25–59.

———— (1999b). Construction frequency and reductive change: Diachronic and register variation in Spanish clitic climbing. *Language Variation and Change* 11:143–70.

———— (2000). *Grammaticization, synchronic variation, and language contact*. Amsterdam: John Benjamins.

UNAM (1971). *El habla de la Ciudad de México: Materiales para su estudio*. Mexico City: Centro de Lingüística Hispánica.

———— (1976). *El habla popular de la Ciudad de México: Materiales para su estudio*. Mexico City: Centro de Lingüística Hispánica.

———— (1995). *El habla de la República Mexicana: Materiales para su estudio*. Ed. Juan M. Lope Blanch. Mexico City: UNAM/Colegio de México.

Winford, Donald (1996). The problem of syntactic variation. In Jennifer Arnold et al. (eds.), *Sociolinguistic variation: Data, theory, and analysis* (NWAV 23). Stanford: CSLI.

Yllera, Alicia (1980). *Sintaxis histórica del verbo español: Las perífrasis medievales*. Zaragoza: Universidad de Zaragoza.