CHAPTER 2

Grammaticalization paths as variable contexts in weak complementarity in Spanish

Scott A. Schwenter and Rena Torres Cacoullos
Ohio State University / Pennsylvania State University

2.1 The form-function problem

The analysis of variation begins with noticing “alternative ways of saying the same thing” within a speech community (Labov 2008:2).¹ For grammatical constructions, however, some have argued that what appear to be social variants may not be equivalent in meaning and thus should not be analyzed as linguistic variables (e.g. Lavandera 1978). For example, García (1985:220) censures variationist studies for “[leaving out] […] the communicative value of the forms […] the difference it makes to what one says whether one chooses one form or another.” The problem is that most attributions of meaning differences between different forms have relied on intuitions and cherry-picked examples. In order to empirically establish linguistic differences (semantic-pragmatic, lexical, syntactic) as well as differences in social distributions between grammatical constructions, we have to show co-occurrence patterns. To do that, we have to know what to count.

In this chapter we return to the notion of “weak complementarity” (Sankoff & Thibault 1981) under the lens of grammaticalization. We propose a grammaticalization-path approach to delimiting the envelope of variation, examining two cases of grammatical variation in Spanish aspect, progressive and perfect/perfective. This approach is both form-based and function-based (see Introduction, this volume). It is form-based in that the forms show weak complementarity along

¹. The analysis of variation also includes “alternative meanings conveyed by the same form” (Labov 2008:2).
a social class dimension in the first case and a geographical dimension in the second. It is function-based in that the “same thing” expressed by the variants is a set of diachronically related functions along a hypothesized grammaticalization path: the set, or stages, of temporal-aspectual meanings along cross-linguistically recurring diachronic paths of given source-construction meanings.

2.2 Weak complementarity as a diagnostic for tense-aspect variables

In their 1981 paper “Weak complementarity”, David Sankoff and Pierrette Thibault point out that controversies about the semantic equivalence of different constructions usually cannot be settled by the data. In reviewing cases whose status as linguistic variables has been disputed, such as the Futur (je ferai) / aller + infinitive (je vais faire) alternation in Montreal French, they demonstrate that these cases exhibit weak complementarity, a complementary increase and decrease in absolute (not relative) frequency of two constructions over time or along another extralinguistic dimension.

Our main claim is that the syntactic variable is largely recognizable through its distribution properties … where one variant is used less, the other is used more. [Weak complementarity] is this quantitative inverse relationship between the rate of usage of two forms across a community of speakers.

&Sankoff & Thibault 1981:207, our emphasis

Weak complementarity in social distribution is not sufficient, though, for treatment of two constructions as variants of a grammatical variable. Some linguistic criterion is needed. Rather than syntactic relatedness or a common underlying structure according to some theory of autonomous syntax, Sankoff and Thibault (1981:208) propose that the variants must “serve one or more generally similar discourse functions”. That is, grammatical variables may or may not have different structural origins or semantic distinctions in some theory. “All that is of interest is that for some important discourse function, one form appears to be replacing the other, either in time or along some socioeconomic or demographic dimension in the speech community” (Sankoff & Thibault 1981:213).

The issue we address is the one posed by Sankoff and Thibault (1981:215) at the end of their paper: “An important problem […] is that of identifying, in a systematic way, the universe of basic discourse functions ….” How do we demarcate the similar discourse functions served by the two candidate variants?

2.3 Grammaticalization paths

The origins of tense-aspect-mood expressions are shared by many languages, genetically related or not. For example, in several Germanic and Romance languages, periphrastics such as the passé composé in modern spoken French originate in the periphrastic ‘have’ or ‘be’ plus past participle resultative construction, which evolved to a perfect and thence to a perfective (Heine & Kuteva 2002:232).

In usage-based theory, the major source for the creation of tense-mood-aspect expressions is grammaticalization (Bybee 2006:719–721). Grammaticalization is the diachronic process whereby existing constructions with particular lexical items gain frequency and become new constructions, following cross-linguistic evolutionary paths. Common imperfective and perfective sources and evolutionary paths are shown in (1) and (2), respectively, from Bybee, Perkins, and Pagliuca’s (1994) cross-linguistic survey.

(1) locative/movement > progressive > imperfective (Bybee et al. 1994:Ch. 3)
(2) ‘be’, ‘have’ + Past Participle > resultative > perfect > perfective/past (Bybee et al. 1994:Ch. 5)

The contribution of grammaticalization to the circumscription of the variable context lies precisely in the discovery of cross-linguistic evolutionary paths. Typological studies have revealed synchronic universals in aspect categories such as progressive and perfect (e.g., Dahl 1985), but stronger cross-linguistic patterns may be the diachronic processes and grammaticalization paths which create such universals (Bybee 2006a). The finding that source constructions tend to follow similar paths of development is pertinent to the problem of defining aspect variables. The two aspect variables we will discuss are defined by the grammaticalization path along which we know (at least one of) the proposed variants have evolved. In each case, both variants cover a range of aspectual meanings along a single grammaticalization path.

2.4 Spanish Progressive estar vs. andar + VERB-nndo

In Mexican varieties of Spanish, there appear to be at least two gerund (-ndo) periphrases expressing progressive aspect, one composed of estar ‘be (located)’ + VERB-nndo (3), the other andar ‘go around’ + VERB-nndo (4). Both gerund constructions, estar and andar, can be translated by the English Progressive.

2. On the cohort of gerund periphrases, see Torres Cacoullos (2000).
Table 2.1 Token frequencies of estar and andar + Verb-ndo in Mexico Spanish (per 100 pages of printed text) (from Torres Cacoullos 2001, Table 4)

<table>
<thead>
<tr>
<th></th>
<th>estar</th>
<th>andar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normalized rate</td>
<td>N</td>
</tr>
<tr>
<td>Educated speakers</td>
<td>90</td>
<td>277</td>
</tr>
<tr>
<td>Popular speakers</td>
<td>68</td>
<td>306</td>
</tr>
<tr>
<td>Essays – academic prose</td>
<td>5</td>
<td>67</td>
</tr>
</tbody>
</table>

PROGRESSIVE (ongoing at speech time)

(3) Pero estás hablando de una forma de vida, Gordo (UNAM 1971: 261)

‘But you are (estar) talking about a way of life, Gordo’

(4) Ando buscando unas tijeras, porque se me rompió una uña (UNAM 1976: 415)

‘I am (andar) looking for a scissors, because I broke my nail’

Table 2.1 shows token frequencies of estar + Verb-ndo and andar + Verb-ndo in three corpora of Spanish. The first two corpora consist of oral interview data, from Mexico City speakers of educated (Habla culta) and popular (Habla popular) varieties (Lope Blanch 1971, 1976), while the third corpus contains essays and academic prose by Mexican authors. In the Educated and Popular speech corpora (top two rows), the combined normalized rate of estar and andar + Verb-ndo is around 90 occurrences per 100 pages (94 Educated, 87 Popular), whereas in Essays-academic prose the combined rate of estar and andar + Verb-ndo is approximately 5. This large rate difference between the Essays-academic prose and the two sets of oral interview data can be attributed to genre differences. Essays concern gnomic or generic situations, propositions for which the predicate holds for all time for a class of entities. Therefore, a greater proportion of stative predicates and habitual aspect situations are expected in this genre than in interview or conversational data (Givón 1990:963).

However, there is no reason to believe that the distribution of discourse functions is different in the Educated and Popular corpora, which belong to the same genre and were assembled in the same way by the same researchers. We have no reason to think that there are differences in “topic of conversation, modes of discourse, or other pragmatic considerations” (Sankoff & Thibault 1981:209). That is, we can assume that whatever aspect is being expressed at the same rate (approximately 90 times per 100 pages of transcribed text) it is the same one in

3. See Torres Cacoullos (2001:475–476) for data sources. We normalized by pages rather than word count since electronic versions were not always available.

these two groups of speakers. Nevertheless, the rate of the andar construction in Popular speakers is nearly five times greater than in Educated. The counts in Table 2.1 thus satisfy the criterion for weak complementarity; that is, a correlation between occurrence rates (token-text frequency, not variant proportions) and an extralinguistic aspect of speakers, here level of education. This weak complementarity suggests that if there is an aspectual distinction between the estar and andar + Verb-ndo constructions, “it serves no basic discourse function” (Sankoff & Thibault 1981:210).

The extralinguistic distribution facts thus suggest a grammatical variable. Nevertheless, a variety of aspectual and non-aspectual meaning nuances have been attributed to andar + Verb-ndo (cf. Torres Cacoullos 2000:7–8, 166). For example, it has been assigned “iterative” (Squartini 1988:263) or “distributive” meaning (“occurs here and there”) (Olbertz 1998:437–438) as opposed to the estar gerund construction’s progressive meaning. How do we define the envelope of variation – is it progressive or continuous or frequentative or some other aspect-type? Examples of continuous and habitual aspect with either construction are shown in (5)–(8).

CONTINUOUS (not necessarily in progress at speech time)

(5) Y están trabajando allá ahórrito andan, pues andan […] en el ajo creo (Chih’97#2bPH)

‘And they are (estar) working there now, they are […] with garlic I think’

(6) Ahorita andan trabajando en las pizcas (Chih’97#1aCJ)

‘Now they are (andar) working in the crops’

HABITUAL (regularly repeated)

(7) ahora se están gastando sobre diez y siete millones de pesos diarios, (UNAM 1971:115)

‘Now more than seventeen million pesos are spent (literally: are (estar) being spent) daily’

(8) En los domingos … toda la gente anda dándose no más la vuelta por esa calle (Chih’97#17aDor)

‘On Sundays … everybody goes (literally: is (andar) going) up and down just that street’

Evidently neither form has a single invariant meaning. Both cover a range of similar meanings in the domain of imperfective aspect (e.g. Camus Bergareche 2004).

4. But see Myhill (1996) on cases of co-variation of form and function in the development of the strong obligation system in American English.
aspect variable (variable context) = locative > progressive > general imperfective grammaticalization path

Figure 2.1 The variable context for grammaticalizing variants encompassing stages along cross-linguistic grammaticalization paths: estar / andar + Verb-ndo

Estar and andar + Verb-ndo have origins in an Old Spanish (12th–15th century) general gerund construction, in which finite forms of spatial (locative, postural, or motion) verbs combine with another verb in gerund (-ndo) form to mean 'be/go Verb-ing' (9):

(9) Old Spanish gerund construction:

\[
\text{[Verb locative-postural-motion] + gerund (-ndo) = 'be/go Verb-ing'}
\]

Particular instances of this general gerund construction grammaticalize, yielding a set of aspectual constructions. In their diachronic trajectory, both estar and andar + Verb-ndo evolve from these origins along the path outlined in (1) (Torres Cacoullos 2000).

We submit that the basic discourse function served by these variants is the set of diachronically related aspectual meanings on this path. Thus, we define the variable broadly, "the largest environment in which this variation occurs" (Labov 2008:2), as the imperfective grammaticization path.

Having applied the heuristic method of weak complementarity and having delimited a variable context, the resulting linguistic variable is reanalyzed in terms of variant proportions (Sankoff and Thibault 1981:214–215). Multivariate analysis shows that the relative frequency of estar and andar + Verb-ndo is conditioned by the lexical type and semantic class of the verb: andar is more likely to be chosen with motion or physical activity verbs, especially when talking about outdoor or rural activities, and andar buscando in particular is a collocation – the conventional way of expressing 'be looking for'. This distribution is the residue of source-construction meanings, 'be located (stationary)' for estar and 'go around' for andar. Torres Cacoullos (2001) suggests that the social meaning of andar + Verb-ndo is related to the lexical associations observed, since rural activities in large spaces are more compatible with 'going around.' Thus, the educated/popular social stratification of estar/andar + Verb-ndo may derive from particular instances of use, developing from an indoor/outdoor and urban/rural difference.

2.5 Spanish Present Perfect vs. Preterit variation

The variation between expressions along the locative to imperfective grammaticalization path in Mexican Spanish discussed above is most likely a case of stable variation. The grammaticalization approach to the variable context is particularly apt for cases of ongoing change.

The "synchronic reflex" of linguistic change is "dialect differentiation" (Sankoff 1988:147; Silva-Corvalán 2001:16), which can reflect different degrees of grammaticalization (e.g., Schwenter 1994; Poplack & Tagliamonte 1999). One notorious locus of dialect differentiation in Spanish is the use of the Present Perfect (10), abbreviated as PP in the English translation, and the Preterit (11), abbreviated as PRET. Previous research reveals that the Present Perfect is more frequent in Peninsular (Spain) than in Mexican Spanish varieties (e.g., Lope Blanch 1961; Penny 2000:160; Company 2002).

(10) ayer he comprado un aire acondicionado y me da calor (Spain, BCON014B)

'yesterday I bought (PP) an air conditioner and I'm getting heat [from it]' (Spain, BCON014B)

(11) Estas son prácticamente igual que las que compramos ayer (Spain, CCON013C)

'These are practically the same as the ones we bought (PRET) yesterday'

Table 2.2 shows the normalized token frequency of the Present Perfect and the Preterit in Peninsular and Mexico corpora. Tokens were exhaustively extracted from an approximately 100,000-word sample of the conversational portion of the COREC Peninsular Spanish corpus (Marcos Marín 1992), and from a similar sample of the Habla culta (educated speech) and Habla popular (popular speech) Mexican Spanish corpora (Lope Blanch 1971, 1976). The rate of the PP in Peninsular is three and a half times greater than in Mexico (928/296 = 3.1) and conversely the rate of the Preterit in Mexico is double that in Peninsular (1699/803 = 2.1). An objection against a weak complementarity interpretation of these occurrence rates may posit a functional difference, such that the Present Perfect, but not the Preterit, signals some kind of current relevance (e.g. Fleischman 1983). However, although the Mexico material comes from interviews and the Peninsular corpus is more conversational, the combined normalized rate of the Present Perfect and

---

Table 2.2 Token frequencies of Present Perfect and Preterit in Spanish (per 100,000 words) (from Schwenter & Torres Cacoullos 2008)

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Present Perfect</th>
<th>Preterit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peninsular Spain</td>
<td>928</td>
<td>803</td>
</tr>
<tr>
<td>Mexico</td>
<td>296</td>
<td>1699</td>
</tr>
</tbody>
</table>

---

5. See Schwenter and Torres Cacoullos (2008:13) for data sources and exclusions.
the Preterit together is very close (1731 in Peninsular, 1995 in Mexico). The close combined rate of PP and Preterit in the two corpora strongly suggests that a basic discourse function — of past (non-imperfective) temporal reference — is being fulfilled at approximately the same rate, despite the distinct characteristics of the corpora. Thus, what differs in the two dialects is the construction speakers choose to mark this function (or as we will see, this set of functions).

Having established weak complementarity, this time along national-regional dialect lines, how do we define the envelope of variation? Perfect aspect signals a past situation that is related to (the discourse at) speech time, whereas perfective aspect conveys strictly that the situation is viewed as bounded temporally; thus, cross-linguistically it is used for narrating sequences of discrete events in the past (Bybee et al. 1994:54; Comrie 1976:5; Fleischman 1983:194; Hopper 1982). However, defining the linguistic variable narrowly in terms of one of the members of the aspect category such as perfect or perfective aspect is contradicted by the data. We find examples of both perfect (12) and perfective (13) uses of the peninsular Spanish Present Perfect, including a range of perfect uses (the meaning labels are adopted from Comrie’s (1976:56–61) distinct types of perfects and Dahl’s (1985:132) prototypical uses). Moreover, we find the same range for the Preterit, shown together with the Present Perfect in the Mexican examples in (14).

(12) a. Perfect of result
   Mira, la han puesto a Vanesa aparato
   ‘Look, they have put (PP) braces on Vanesa.’
   (CCON018C)

b. Experimental (or existential) perfect
   bueno, yo ya he comprado y por ahi cadenas de esas
   ‘well, I already have bought (PP) around there chains of that kind’
   (BCON015B)

c. Continuative perfect (or perfect of persistent situation)
   No, eso lo han comido siempre los chicos nuestros.
   ‘No, our kids have always eaten (PP) that’
   (CCON022E)

(13) Perfective
   O sea ha esperado a acabar de hablar con Nicolás, lo que había empezado,
   ha tardado su minuto y luego ya ha cogido la llamada.
   ‘I mean he waited (PP) to finish talking with Nicolás what he had started,
   he took (PP) his minute and then he finally answered (PP) the call’
   (CCON016A)

(14) Perfect of result
   a. ya levantaron un gran edificio. Ya esta toda la estructura
   ‘they already put up (PRET) a big building. The whole structure is already up’
   (MexCult, 428)

(15) a. Temporal distance: hodiernal, proximate past
   lo he escuchado esta mañana
   ‘I heard (PP) it this morning’
   (CCON028A)

b. Temporal distance: pre-hodiernal
   Liebres sí se ven algunas. Y zorras- y zorras muchas. Y jabalíes el año pasado han matado uno o dos.
   ‘Hares you see some. And foxes – lots of foxes. And wild boars last year they killed (PP) one or two’
   (CCON019A)
c. **Temporal Distance: Irrelevant**

Hay gente que se muere con noventa años y *nunca ha madurado* (BCCN014D)

“There are people who die at ninety years old and they *never have matured* (PP)”

**d. Temporal Distance: Indeterminate**

con papá no montaban más en el coche porque no, se lo *regalé* a mi hija y no he cogido el coche, ahora *le he comprado* […] y ahora *le he comprado* a mi nieto uno. Eh - un coche muy bueno ¿eh? y tal,  (CCCON004C)

‘they wouldn’t get in the car with dad anymore because they wouldn’t, I *gave* (PRET) it to my daughter and I haven’t taken the car, now I’ve bought […] and now I *have* bought (PP) one for my grandson’

Similarly, there is no replicable way to delimit the variable context to past perfective situations that are currently relevant. A defining meaning component of perfects cross-linguistically is said to be current (or present) relevance of a past situation (e.g., Bybee et al. 1994:61; Comrie 1976:52; Dahl 1985:134; Fleischman 1983:194; Li & Thompson 1982). Per Dahl and Hedin (2000:391), however, current relevance is a graded concept that can be more or less evident in a given example. Moreover, the criterion for determining relevance need not be a condition on the world as in a tangible “continuance of a result” but rather a condition on the discourse: speakers present the consequences of a past event as important to what they are saying (Dahl & Hedin 2000:392; cf. Li, Thompson & Thompson 1982). Thus, determining which tokens are aspectually perfective – but *not* currently relevant – is unverifiable. For example, in (9)–(10) we have no empirically motivated reason to consider the Present Perfect more currently relevant than the Preterit, except for the circular argument that the Present Perfect signals current relevance and the Preterit does not. An argument for current relevance is plausible in each case. In (9) (with the Present Perfect) the interlocutors are talking about the new air conditioner producing hot air, a condition on the world, or materially relevant, and in (10) (with the Preterit) they are talking about the practice of price-gouging, a condition on the discourse, or discursively relevant, in Dahl and Hedin’s (2000) terms.

In sum, the Present Perfect and the Preterit have a range of uses, just as we saw with *estar* and *andar* + *Verbs-*ndo constructions, and putative semantic invariants such as current relevance are empirically unverifiable: illustrative examples cannot resolve whether or not there is a common underlying basis or semantic equivalence. Thus, we adopt a grammaticalization-path approach to the variable context. As depicted in Figure 2.2 the variable context encompasses the stages along the cross-linguistic perfect-to-perfective grammaticalization path ((2), above) (for Romance, see Harris 1982).

Similarly, there is no replicable way to delimit the variable context to past perfective situations that are currently relevant. A defining meaning component of perfects cross-linguistically is said to be current (or present) relevance of a past situation (e.g., Bybee et al. 1994:61; Comrie 1976:52; Dahl 1985:134; Fleischman 1983:194; Li & Thompson 1982). Per Dahl and Hedin (2000:391), however, current relevance is a graded concept that can be more or less evident in a given example. Moreover, the criterion for determining relevance need not be a condition on the world as in a tangible “continuance of a result” but rather a condition on the discourse: speakers present the consequences of a past event as important to what they are saying (Dahl & Hedin 2000:392; cf. Li, Thompson & Thompson 1982). Thus, determining which tokens are aspectually perfective – but *not* currently relevant – is unverifiable. For example, in (9)–(10) we have no empirically motivated reason to consider the Present Perfect more currently relevant than the Preterit, except for the circular argument that the Present Perfect signals current relevance and the Preterit does not. An argument for current relevance is plausible in each case. In (9) (with the Present Perfect) the interlocutors are talking about the new air conditioner producing hot air, a condition on the world, or materially relevant, and in (10) (with the Preterit) they are talking about the practice of price-gouging, a condition on the discourse, or discursively relevant, in Dahl and Hedin’s (2000) terms.

In sum, the Present Perfect and the Preterit have a range of uses, just as we saw with *estar* and *andar* + *Verbs-*ndo constructions, and putative semantic invariants such as current relevance are empirically unverifiable: illustrative examples cannot resolve whether or not there is a common underlying basis or semantic equivalence. Thus, we adopt a grammaticalization-path approach to the variable context. As depicted in Figure 2.2 the variable context encompasses the stages along the cross-linguistic perfect-to-perfective grammaticalization path ((2), above) (for Romance, see Harris 1982).
paths. Nevertheless, in these data we first determined weak complementarity and also that the Preterit covers the same uses as the Present Perfect, albeit at different rates. In cases where both putative variants (virtually) do not cover the same range of uses, to include all occurrences of a form can distort our view of the patterning of variability. Aaron (2006) demonstrates this in her study of two future expressions in Spanish, one of which has developed epistemic uses: when the variable context is circumscribed to future temporal reference, co-occurring temporal adverbials favor the morphological (-re) modal-origin form; inclusion of epistemic together with future contexts gives the result that temporal adverbial co-occurrence favors the periphrastic ‘go’-based future. Thus, the grammaticalization-path approach to the variable context is function- rather than just form-based (Introduction, this volume) because both proposed variants cover, albeit to different extents, the same territory on a single grammaticalization path.

2.6 Conclusion

Sankoff and Thibault (1981:214) view weak complementarity as a proper diagnostic for incipient variables; over time, semantic distinctions between the alternating constructions may become neutralized in spontaneous discourse. The neutralization-in-discourse hypothesis states that while contexts can almost always be found in which different forms have different meanings, there are alternations in which the full accompaniment of meaning distinctions is not pertinent either for the speaker or the interlocutor (Sankoff 1988:153). This neutralization of aspectual meaning distinctions appears to have occurred in the expressions we have examined, to a greater extent for estar and andar + VerB-nDo than for the Present Perfect and Preterit in Mexican Spanish.

Sankoff and Thibault (1981:207) insist that the true origin of much grammatical variation lies in discourse: change occurs by juxtaposition of constructions which are used for similar discursive functions. Sociolinguistic variation theory is equipped to handle scientifically this type of change, “which is so important for the relationship of linguistic and interactional processes” (Sankoff & Thibault 1981:207).

In the study of grammaticalizing aspectual constructions, the contexts for the analysis of variation include not only all the forms fulfilling a given function, but also the range of functions along the grammaticalization path that can be expressed by those forms. This grammaticalization-path approach to the variable context surmounts, or rather bypasses, the question of equivalent meaning for grammaticalizing variants, without incurring in unwarranted, empirically unverifiable, semantic/aspectual classifications of the alternating expressions.

The adoption of a grammaticalization-path perspective on the envelope of variation leads to better empirical coverage of the data, as well as a principled model for situating the variables in both synchronic and diachronic perspective. The basic discourse function in weak complementarity (Sankoff & Thibault 1981) may actually be identified as a set of diachronically related functions. Since “the definition of a linguistic variable is the first and also the last step in the analysis of variation” (Labov 2008:2), it is fitting, then, to conclude that aspectual variables may be defined as grammaticalization paths.

References


CHAPTER 3

Aspectual periphrases and syntactic variation in Brazilian Portuguese

Ronald Beline Mendes
University of São Paulo

3.1 Introduction

The variation between the periphrastic constructions estar + geríndio ‘be + GERUND’ (EG) and ter + participio ‘have + PARTICIPLE’ (TP) to express durative and iterative aspect in Brazilian Portuguese (BP) raises interesting methodological issues. However, rather than arguing that the constructions themselves constitute variants of a single variable, in this chapter I argue instead that the sentences in which EG and TP occur are the variants: that is, EG may express not only progressive but also durative and iterative aspect; and TP does not express perfective or past tense (as do equivalent constructions in other Romance languages), but rather expresses iterative and sometimes durative aspect. A first methodological issue concerns locating sentences with instances of EG whose meaning could be expressed by TP (and vice versa), a question of how to define the variable context (or the envelope of variation). Although this question is methodological, it intersects with analysis to the extent that it raises the question of whether the variable context should be defined on formal or functional grounds (see Introduction, this volume).

This chapter begins by presenting synchronic and diachronic data to make explicit the methodological decisions involved in defining the variable context and selecting tokens for quantitative analysis. This discussion will shed light on the difficulties involved in interpreting form-meaning relations when the meaning is not uniquely associated with a single form, but rather consists of a number of constituents inside and outside the sentence. The next section discusses the question of how to operationalize aspectual distinctions as factor groups in a quantitative analysis, the problem of apparent neutralization of aspectual distinctions in discourse, and the relationship between sentential aspect and verb-class effects. In the final section, I present a novel method of conducting quantitative...