3

Language Variation and the Minimalist Program

3.0. Introduction

Traditionally, generative hypotheses have mainly been built on standard language data, on the basis of the grammaticality judgments of a few speakers. This approach has proven very powerful in producing a remarkable number of generalizations, which were formulated by excluding all variability complications due to performance and by exclusively focusing on native speakers’ syntactic intuitions (Barbiers 2009: 1608). On the other hand, such a methodology has often been criticized by sociolinguists, who base their observations on much larger corpora of naturalistic production data, and have developed several techniques to study the ‘real vernacular,’ or the language spoken by people when paying no metalinguistic attention to their speech (Labov 1972). In the last decade, works on microvariation attempted to combine these previously contrasting approaches to compare a speaker’s intuitions with real production data, with the goal of developing more empirically-testable generalizations (Cornips & Poletto 2005). Recent works within the minimalist framework entail a derivational approach that is inconsistent with parameter-based accounts. Thus, an extension of the parameter/microparameter idea to individual variation seems to not be a straightforward matter (Adger & Smith 2005).
This chapter provides an overview of the major approaches proposed in the literature to deal with inter- and intra-speaker variability. In doing so, it highlights the importance of the present work in combining the formal and the sociolinguistic methods to obtain a more fine-grained account of the syntactic phenomena characterizing the Afro-Bolivian Spanish Determiner Phrase.

3.1. Accounting for Variation

Inter- and intra-speaker variability has long been a hot topic of linguistic debate. Typically, formal linguists have considered such a variability as an instance of E-language, thus not concerning the core syntactic competence of the speaker and therefore, for the most part, ignorable. On the other hand, a different response, radically contrasting with the former, grew within the variationist/sociolinguistic paradigm. It consisted of positing variable rules (e.g., Cedergren & Sankoff 1974; Labov 1972), where probabilities would be built into the notion of grammar. Other models stipulated multiple grammars, where speakers have several parametric configurations (Henry 2005; Kroch 1989, 1994). Finally, a more recent account (Adger 2006; Adger & Smith 2005; Parrott 2007), developed within the architecture provided by the Minimalist Program (Chomsky 1995, 2001, 2006) and the study of microparametric variation (Barbiers & Cornips 2001; Kayne 1996, 2000; etc.), postulates that variation is the overt result of covert lexical selections. Such a model claims that there is a reduced available number of syntactic operations (Merge, Move, Agree), which are universal and constant, while the elements entering the syntactic numeration can vary in feature specification. The following sections will provide a brief overview of these competing accounts.

3.2. Approaches to the Study of Language Variation

3.2.1. THE FORMAL APPROACH

Mainstream syntactic theories (Chomsky 1957, 1965, 1986) have given little room to the study of inter- and intra-speakers’ speech variability. Within the formal approach of Principles and Parameters (P&P) (Chomsky & Lasnik 1993), languages can essentially be seen as combinations of a finite set
of innate principles, which are shared by all varieties, and a set of binary parameters that are responsible for the syntactic variability observable across human languages. Within this framework, principles are “language-invariant statements” (Chomsky 1995: 25), whereas parameters must be set for certain values. Principles are part of a genetically innate Universal Grammar (UG), which all humans possess. As such, they do not need to be learned through exposure to language. Rather, exposure to language merely triggers the parameters to adopt the correct setting.

Chomsky (2000) compares the language faculty to a switch box. This box consists of two components: a fixed network, which is the innate principles of language, and several switches, which are options determined by experience, binary parameters that can be set on or off. Different parametric combinations lead to different grammars. In Chomsky’s (2000: 8) words,

When the switches are set one way, we have Swahili; when they are set another way, we have Japanese. Each possible human language is identified as a particular setting of the switches—a setting of parameters, in technical terminology.

According to the Principles and Parameters model (Chomsky & Lasnik 1993), children are assumed to learn the language of their parents or their social environment. During this process, acquisition can be imperfect, thus involving parameter resetting and therefore cross-generational language change. Chomsky & Lasnik’s (1993) model does not allow parameter resetting during the lifespan of the speaker; in their view, once a parameter has been set, it is for good. For this reason, within the field of generative syntax, cases of language internal variation have often been disregarded as instances of E-language, not interesting from the perspective of scholars aimed to unveil the secrets of the I-language. As a result, formal syntactic theories have traditionally been built on partly-idealized standard languages, on the basis of grammaticality judgments of a reduced number of informants. As stated in the first section, parameter resetting or change is viewed in diachronic terms but never as a synchronic process.

As van Gelderen (2005: 180–181) pointed out, vast linguistic corpora have never been popular research tools among generative syntacticians. She reports Wasow (2002), who highlights Chomsky’s negative attitudes on the employment of quantitative data, which in his view would not provide any useful insight into the knowledge of the I-language. Here is a quote from Chomsky (1962: 128):
It seems that probabilistic considerations have nothing to do with grammar, e.g. surely it is not a matter of concern for the grammar of English that ‘New York’ is more probable than ‘Nevada’ in the context ‘I come from—’.

Needless to say, as we will see, sociolinguistic models of language variation have a radically different opinion on the issue.

### 3.2.2. THE SOCIOLINGUISTIC APPROACH

The Chomskyan revolution and its formal methodology produced a great deal of linguistic generalizations; the reason for its success can be found in its ability to put aside the variability derived from natural speech data, and its exclusive focus on the abstract intuitions of a restricted number of native speakers. Yet, for these same reasons, such a model has often been criticized by sociolinguists, who instead based their observations on bigger corpora of naturalistic production data, and developed several techniques to study the ‘real vernacular,’ the real language spoken by people when paying no metalinguistic attention to their speech (Labov 1972).

In the sixties, Labov (1969) extended the concept of optional rule, borrowed from the earlier transformational model (Chomsky 1957), to the idea that rules can be variable and dependent on several internal (linguistic) and external (extralinguistic) factors (Cornips 2006). In fact, Labov’s ‘variable rules’ are based on the concept of ‘orderly heterogeneity’ (Weinreich, Labov, & Herzog 1968: 100), “the idea that variation in language is not random or free, but systematic and rule-governed” (Tagliamonte 2006: 129). This notion of variable rules was first developed from the observation that people make systematic choices when they speak (Labov 1969). Because of such systematicity, statistical models can be developed; this implies postulating that language has a probabilistic component.

The sociolinguistic approach is founded on the principle of accountability (Labov 1966: 49). This principle states that all variants belonging to the same syntactic variable must be accounted for in the variable environment. In order to identify what should be considered as a variant of a specific variable, the synonymy principle (principle of sameness) must be followed. In other words, variants of the same variable are only those tokens that consist of “[different] ways of saying the same thing” (Labov 1972: 323).
Sociolinguists developed a variety of research tools that allow them to cope with the demands of their data (Chambers 2003). One of the most famous statistical instruments employed to analyze sociolinguistic variation is VARBRUL (by David Sankoff), which evolved in several more powerful software programs during the last decades (e.g., GoldVarb X by Sankoff, Tagliamonte, & Smith 2005).

Within the Labovian/sociolinguistic approach, variation has been at the center of attention for at least forty years; on the other hand, within the traditional Chomskyian/generative framework, it has generally been excluded from the research agenda, because it was considered to be a case of performance rather than competence. Nevertheless, in more recent years, scholars working within the generative framework have attempted to account for diachronic variation (Kroch 1989) and also to capture dialectal and interspeaker variation, exploring the notion of parametric variation and, in the last decade, fine-graining it to include so-called microparameters (Benincà 1989; Kayne 2000; among others).

3.2.3. THE DIACHRONIC APPROACH

Dating from the 1980s, the study of diachronic syntax has seen the rise of a particularly rich tradition of combining quantitative and formal perspectives (see Kroch 1989, 1994, for an overview). In Kroch’s view, the historical evolution of competing variants in syntactic change parallels the evolution of morphological doublets. In both cases, the competition of two forms is diachronically unstable: “One form tends to drive the other out of use and thus out of the language” (Kroch 1994: 7). This parallelism would be explained by the existence of the “Blocking Effect” (Aronoff 1976). The blocking effect, in Kroch’s words (1994:9), does not prevent doublets from arising in a language through social processes (e.g., language contact). Rather, it acts as an economy constraint on their storage.

Kroch supports his analysis of syntactic change by suggesting the general validity of the “Constant Rate Hypothesis” (Kroch 1989). The “Constant Rate Hypothesis” claims that usage frequencies change rates mirror the gradual substitution of one grammatical form by another. The formulation of the “Constant Rate Hypothesis” represents a key step toward the analysis of E-data. In fact, Kroch clearly states that statistical tendencies and patterns encountered in vast corpora can unveil the principle and parameters belong-
ing to speakers’ competence. Kroch envisions a theory of language variation that admits competing grammars. He proposes that diachronic change is triggered by variation between forms belonging to the competence realm. See also van Gelderen (2005) for an account of how the use of corpora has been integrated in syntactic theory.

3.2.4. THE MICROPARAMETRIC APPROACH

Besides these approaches to diachronic variation and change, there is also a recent and growing body of research that over the past decade or so has attempted to combine insight of generative linguistics with variationist analysis in the study of synchronic variation, especially across dialects of the same language (see Auger 1998; Barbiers 2005; Cornips & Corrigan 2005; Heap 2001; Kayne 1996, 2000; King & Nadasdi 1997; etc.).

Kayne (1996) was the first one to propose the notion of ‘microparameter.’ Microparameters represent the ‘switches’ that distinguish closely related languages. In Kayne’s (1996: xii) words,

> Comparative work on the syntax of a large number of closely related languages can be thought of as a new research tool, one that is capable of providing results of an unusually fine-grained and particularly solid character. If it were possible to experiment in languages, a syntactician would construct an experiment of the following type: take a language, alter a single one of its observable syntactic properties, examine the result to see what, if any, other property has changed as a consequence of the original manipulation. If one has, interpret that result as indicating that it and the original property that was altered are linked to one another by some abstract parameter.

Obviously, such an experiment is impossible to carry out; however, by analyzing closely related varieties, one can achieve similar results. Recent syntactic dialect atlas projects, such as the ASIS and the SAND have undertaken such a challenge. These research projects have various objectives. They not only explore the geographic distribution of syntactic variables; they also correlate them to the investigation of language change and to the broader study of universal properties of the human language, to understand the loci and limits of syntactic variation within linguistic systems (Barbiers 2009). The great improvement of this new approach on previous ones is that past
formal studies were primarily concerned with standard varieties, whereas microvariation research includes non-standard languages.

As Barbiers & Cornips (2001: 2) state,

This [approach] does not only enhance the empirical basis of syntactic theory, but it also reduces the influence of prescriptive rules and makes it possible to test potential correlations between syntactic variables while keeping other, possibly interfering factors constant.

Weiβ (2001), in fact, argues that standard languages, studied at school through formal instruction, may present quite unnatural prescriptive properties, often imposed by processes of standardization. Researchers working on this track acknowledge that dialects are heterogeneous systems, on the grounds that a dialect is constantly in contact with one or more standard varieties and sometimes also with other dialects (Barbiers & Cornips 2001: 2). Each individual may speak in a slightly different way, so that it is possible to postulate the existence of slightly different grammars and thus the existence of individual microparameters.

3.2.5. VARIABILITY IN THE MINIMALIST PROGRAM

Finally, a model that combines aspects of the microparametric framework, the sociolinguistic approach to data collection, and recent formal proposals concerning the architecture of language (the Minimalist Program) is the one presented by Adger & Smith (2005). This model, in contrast with previous formal accounts, seems particularly well suited for the study of intra-speaker variation. In fact, even in the most recent approaches to the study of dialectal variation, work on variability among the idiolects of individuals has been largely lacking. On the other hand, in Adger & Smith (2005) and related studies (e.g. Adger 2006; Parrott 2007), intra-speaker variability becomes the core of linguistic research, bringing previously disregarded phenomena, considered as belonging to E-language, to the forefront (Adger & Trousdale 2007).

The minimalist framework admits several phonological outputs for a given semantic interpretation. Adger & Smith (2005) argue for characterizing syntactic variation in terms of uninterpretable features. Certain uninterpretable features may be present in one category but absent in another. Since they are uninterpretable, they would have no semantic repercussion, thus
being equally legitimate for a convergent derivation. Therefore, variation is reduced to the specification of the uninterpretable features in a derivation (Adger & Smith 2005: 161). As expected, syntax per se remains invariable or “perfect” (Brody 2003), given that variation is located only in the lexical component. Variation will occur when one item or another enters the numeration and takes part in a syntactic derivation. Several extralinguistic factors may affect the probability of accessing one lexical item or another: social environment, social class, gender, age, education, etc. (Adger & Smith 2005: 164).

Notice that this approach is different from Kroch’s or Labov’s alternatives in that variation is located in the selection of lexical entries. On the other hand, Kroch’s model suggests that there is more than one system of grammatical knowledge available to the speaker, and variation depends on the speaker’s selection of one or the other system. The Labovian variable rules approach contrasts with this in that it states that the variation is part of the rule, tying it in more directly to grammatical competence.

Within the framework proposed by Adger and Smith, there is no notion of a probability directly linked to a particular rule; the only possible rules (Merge, Move and Agree) are invariant and apply categorically in particular cases across languages. Adger and Smith’s perspective differs from the previous approaches in that it assumes only one invariant grammatical system, containing universal mechanisms, rather than a range of systems. Overt variation is driven by the choice of different lexical items, which may be influenced by various internal and external factors. This may be considered as a “very minimal theory” (Adger & Smith 2005: 165), since the idea that speakers have to choose lexical items is one that cannot be avoided.

A concrete example of this proposal is the analysis of alternation between was and were in Buckie, a small fishing town near Aberdeen (UK). After providing accurate information on the methodology used to collect natural data and presenting quantitative analyses on internal and external factors affecting the alternation, Adger & Smith (2005) proposed a formal model to account for the results. Consider the following examples (Adger & Smith 2005: 154):

(7) Buckie English
a. He says ‘I thocht you were a diver or somethin.’
   ‘He said ‘I thought you were a diver or something.”

b. Aye, I thocht you was a scuba diver.
   ‘Yes, I thought you were a scuba diver’
The verb form *was* instead of standard *were* appears 54 percent of the time; it can be found with all persons except *they*. In terms of syntactic derivation, pronouns bear certain interpretable features for person and number. In Adger and Smith’s view, first and second person bear [+person], while third person is [–person], assuming that third person is lack of a positive specification for person (see, for example, Harley & Ritter 2002 and references therein). T bears unvalued features for number and person: T [tense: past, ucase: nom, unum:, upers:]. After merging with T, the pronoun *you* provides T with plural number and second person values, and the result is *you were*. If T is valued as third person, singular, the result is *he/she was*. Variation occurs once there is an alternative lexical item that could appear in the same syntactic context, with the same interpretable features, but with different uninterpretable ones. The authors propose the following alternative: T2 [tense: past, ucase: nom, upers:] (Adger & Smith 2005: 166), which lacks [unum:]. If T2 is selected, the derivation will remain the same except for the absence of specification for number. Since there is no specification for number in T, the verb form will be *was* by default even when the pronoun is second person plural. Therefore, this model provides a reasonable account for the variable grammatical phenomena encountered in Buckie by combining quantitative approaches to data collection and formal generative theories.

The postulation that overt variability is due to different lexical specifications, while syntax *per se* remains invariable, will be one of the assumptions of the present work. This will allow us to account for important cross-dialectal, intra-dialectal, and intra-personal instances of variation, while testing syntactic hypotheses and potential correlations between linguistic features and syntactic phenomena. In chapter 8, the study of gender/number agreement variability will provide a concrete example of how inter- and intra-speaker variation can be accounted for by Adger & Smith’s approach. In particular, I will postulate the presence or absence of certain uninterpretable gender and number features across the ABS DP to capture the agreement variability encountered in the speech of my informants.

### 3.3. Variability across Closely Related Dialects: The Afro-Hispanic Case

In recent years, syntactic studies have devoted more and more attention to dialects and substandard varieties of different languages—e.g., Italian
(Benincà 1989, 1994; Poletto 2000), English (Adger & Smith 2005), Dutch (Barbiers & Cornips 2001), etc. So far, little attention has been paid to the study of microvariation across Spanish dialects; to my knowledge, a book has yet to be published on microparametric syntax concerning Afro-Hispanic contact varieties.

The majority of the dialects that developed in Latin America from the contact of African languages and Spanish in colonial times consist of vernacular varieties. The grammars of these languages are relatively close to the one of standard Spanish. Nevertheless, they clearly present the traces of second language acquisition strategies and contain a variety of constructions that would be considered ungrammatical in standard Spanish. It is exactly this close proximity to the standard language that makes these varieties so suitable to be analyzed with a microparametric approach.

### 3.4. Data Collection and Methodology

The fieldwork was conducted during July–August 2008, December 2009, and July-August 2010. More than fifty speakers were interviewed. They were Afro-Bolivians living in the communities of Tocaña, Mururata, and Chijchipa, three villages in the municipality of Coroico, North Yungas. Data collection was carried out through sociolinguistic interviews. In particular, speakers were asked to talk freely about a variety of topics, while the interviewer would take part in the conversation by asking follow-up questions (see Labov 1984). The goal was therefore to reduce the Observer’s Paradox (Labov 1972) as much as possible. Only later, usually after one or two days from the time of the interview, the same informant would be asked for grammaticality judgments. This specific technique (sociolinguistic interviews followed by grammaticality judgments) was adopted in order not to influence the results of the interview by altering the informant in advance on the linguistic phenomena that were being studied, thus trying to reduce his/her metalinguistic awareness.

A comparison of grammatical intuitions and sociolinguistic interviews can be very helpful in studying socially stigmatized dialects like Afro-

---

1. According to the Principle of Tangential Shift, interviews may be arranged into a network of topics that do not need to be followed according to a prescribed sequence. The conversation between the interviewer and the informant should start with the least personal questions and progress, step by step, towards more intimate topics. The shift between topics should be as smooth as possible. It should be based on follow-up questions to what has just been said by the informant.
Bolivian Spanish; nevertheless, the same methodology is not equally suitable for all linguistic phenomena. In fact, while the syntactic variability of high-frequency structures can be addressed efficiently by recurring to sociolinguistic interviews, less frequent constructions are better studied by asking informants for grammaticality judgments. Therefore, while a comparison between natural data and inducted elicitations has been a prerogative in the study of the Afro-Bolivian Spanish DP, this book also considers necessary methodological adjustments imposed by the nature of the different phenomena under study.

3.5. Conclusion

To summarize, this chapter has provided an overview of the main approaches proposed to account for language variability. I focused on recent developments within the field of microparametric syntax (Barbiers & Cornips 2001). In particular, I presented the framework suggested by Adger & Smith (2005), which combines quantitative analysis and formal theory. The present book will adopt such a model to study several syntactic aspects of the Afro-Bolivian Spanish Determiner Phrase.

Microparametric syntax is a growing research field, which can be employed to test formal theories, often designed on standard language data. While several microparametric generalizations have been provided for Italian, English, and Dutch dialects (see, for example, Adger & Smith 2005; Barbiers & Cornips 2001; Benincà 1989; Poletto 2000), such research has yet to be conducted on Afro-Hispanic contact varieties. This opens up a new field of investigation, which if addressed in a methodological and systematic way, could lead to interesting discoveries. One of the main goals of this study is to place the first stone in the foundation of such a research program.