

Addressing the problem of intra-speaker variation for parametric theory¹

Christina Tortora, CUNY (College of Staten Island & The Graduate Center)

1. Introduction

In this paper I address the challenges that intra-speaker variability presents for establishing, within the theory of parameters, which surface syntactic phenomena cluster. Although the discussion mainly involves an exploration of the theoretical relevance of certain variable syntactic phenomena exhibited by Appalachian speakers, as we shall see, I believe that the significance of the problems raised will be made clearer through a comparison of the Appalachian facts with variable syntactic phenomena found in African American and Belfast English, as well.

Before entering into the particulars of the discussion, I would like to begin with an illustration of the phenomenon of intra-speaker variability in English, as a way of establishing the fundamental assumptions I will be making in this paper. To this end, consider some pairs of examples which are possible for speakers of different varieties of North American and British English. The first set of examples ((1)-(2)) comes from African American English, the second set from Belfast English ((3)-(6)), and the third set from Appalachian English ((7)-(13)):

African American English (exx. (1a-d) & (2) from Green, Roeper, & Terry 2007; (1e,f) are attested):

- (1) **a. She can leave?** (= ‘Can she leave?’)
 b. Can she leave?
c. Why she can have that book? (= ‘Why can she have that book?’)
 d. Why can she have that book?
e. What that gots to do with me? (= ‘What does that have to do with me?’)
f. What does that got to do with me? (= ‘What does that have to do with me?’)
- (2) **a. I wonder can she leave.** (= ‘I wonder if she can leave’)
 b. I wonder if she can leave.

Belfast English (Henry 1995):

- (3) **a. The eggs is cracked.**
 b. The eggs are cracked.
 c. *Is the eggs cracked?
 d. Are the eggs cracked?
- (4) **a. Any country couldn’t stand it.** (= ‘No country could stand it’)
 b. No country could stand it.

¹ I started presenting some of the ideas in this paper at a 2008 University of Massachusetts workshop on variability, organized by Lisa Green; I presented a more developed version of the paper in a colloquium at SUNY Stony Brook in the same year, and then again at the 2011 *Irish Network on Formal Linguistics* conference in Belfast, organized by Alison Henry. I would like to thank Lisa and Alison for the opportunity, and also the various audiences at those talks for questions and comments which forced me to refine and clarify my ideas (especially Ellen Broselow, Andrew Carnie, Daniel Finer, Heidi Harley, Robert Hoberman, William Labov, Richard Larson, Cristina Schmitt, and Charles Yang). I would also like to thank Paola Benincà, Frances Blanchette, Anthony Kroch, Larry Horn, Michael Montgomery, Jeffrey Parrott, Beatrice Santorini, Jiahong Yuan, and Raffaella Zanuttini for help, ideas, and critical suggestions, and also for listening. Naturally, all persistent shortcomings are purely my own responsibility. The research for this work was supported by National Science Foundation Grant #BCS-0617197, a PSC-CUNY Grant (#61341-00 39), and a grant from the Office of the Provost at the College of Staten Island.

- (5) **a. I have a sister ____ lives in Dublin.**
b. I have a sister who lives in Dublin.
- (6) **a. See my brother, __ never stops talking.** (= ‘See my brother, he never stops talking’)
b. See my brother, he never stops talking.

*Appalachian English.*²

- (7) **a. The eggs is cracked.**
b. The eggs are cracked.
- (8) **a. Is the eggs cracked?**
b. Are the eggs cracked?
- (9) **a. Nobody couldn’t stand it.** (= ‘Nobody could stand it’)
b. Nobody could stand it.
- (10) a. He is happy.
b. **He are happy.**
- (11) **a. They is another one down the street.** (= ‘There is another one down the street’)
b. They are another one down the street. (= ‘There is another one down the street’)
- (12) **a. They was happy.**
b. They were happy.
- (13) **a. They used to get a royalty off the timber ____ was cut for the mines.**
b. They used to get a royalty off the timber which was cut for the mines.

The main observation to be made regarding these examples is the following: each member of a pair represents the semantic/functional equivalent of the other member in the pair. That is, we can describe each pair of sentences as “variants” of one another, where each variant is possible in the same semantic contexts as the other. Take for example (1a) vs. (1b): the sentence in (1a) (without T-to-C movement) is one possible way of forming an unmarked yes-no question in African American English (Green et al. 2007);³ the sentence in (1b) is a semantic equivalent, also possible for African American English

² The Appalachian sentences here are abstractions of examples found throughout the literature (e.g., Wolfram & Christian 1976 and Montgomery & Hall 2004), and of speakers’ grammaticality judgments (given in the context of joint research with Judy Bernstein, Marcel den Dikken, and Raffaella Zanuttini). I chose to abstract away from the real examples (e.g., compare (9a) with an example taken from Wolfram & Christian: *Nobody didn’t see him*), for the purposes of highlighting the morpho-syntactic patterns, comparing the data with those from the other varieties (e.g. compare (9) with Belfast English (4)), and minimizing possible distractions induced by irrelevant differences. The example in (10b) is a variation on the sentence in (32) (section 3.1 below), from Montgomery & Hall (2004). The use of *are* and *were* with third singular subjects is comparatively rare in Appalachia (with the forms *is* / *was* being much more frequent), but is attested nevertheless.

³ As discussed in detail in Green et al. (2007) and Roeper & Green (2007), and also in section 3.3.2 below, the operant word here is *unmarked*. That is, in contrast with other varieties of English, the sentence in (1a) is possible as a “real” question (i.e., a real request for information), hence its exhibition above as a semantically equivalent variant of (1b). There are other varieties of English (e.g., mainstream), in which the string in (1a) is possible as a “special” question, i.e., one which carries with it a presupposition (as Green et al. 2007 discuss), something which is not relevant to the present discussion. To be noted in this

speakers.⁴ The possibility of variant syntactic forms can also be seen in the Belfast English pairs in (3) through (6) (from Henry 1995 and A. Henry, p.c.); thus, speakers who allow (4a) also allow (4b), with the same meaning (in terms of both truth conditions, and pragmatics).

The examples in (7) through (13) illustrate that the English spoken in the Southern Appalachians in the United States is no different from African American English or Belfast English, in terms of the pervasiveness of intra-speaker variability: thus, (12a) is possible alongside (12b), and so forth. As noted in footnote 2, the pair of examples in (11) are of particular interest, in that neither member of the pair is recognizable as a kind of “standard” counterpart to the other, something which will be addressed in greater detail in section 3.2.

Perhaps the most obvious question which arises with respect to these examples is the following: how do speakers who exhibit this kind of variability model the variant forms? One approach is to take each variant in a pair as the reflex of a different grammar — an idea which was pursued by Henry (1995) for Belfast English, and by Kroch (1989; 1994; 2001) for syntactic variation and change in the history of English in general. Under this view, which I will term the “Multiple Grammars” hypothesis (borrowing from Kroch), a speaker who allows both (13a) and (13b), for example, would be bi-dialectal (to be defined, within the theory of parameters, in a highly specific way; see section 4). In this paper, I will be assuming this approach to intra-speaker variability.⁵

In these introductory observations, then, I hope to have clearly established (a) my understanding of what intra-speaker variability is, and (b) my adoption of the “multiple grammars” approach of intra-speaker variability. These basic assumptions in place, then, what I would like to do in this paper is address the challenges we face when hypothesizing that certain syntactic phenomena cluster, in situations with robust intra-speaker variability. This challenge can be preliminarily illustrated, for example, by Henry’s (1995) data in (3) and (4): specifically, while Henry may have very good theory-internal reasons for arguing that (3a) and (4a) represent surface reflexes of the same underlying parameter (i.e., that they “cluster”), finding independent empirical support for this hypothesis is another matter altogether, and one, I will argue, which becomes more difficult to establish, given the existence of (3b) and (4b). Even if we are lucky enough to have informants who are highly linguistically aware and good at giving judgments, at the very best, speaker judgments can only ever tell us which forms are acceptable; unfortunately, the judgments can never reliably tell us which phenomena belong to which dialect (and furthermore, as should be thoroughly obvious, no matter what the reliability of the speaker’s intuitions, their judgments can never tell us which phenomena cluster). As we will see, Appalachian English presents a case where it is particularly difficult to sort out the data simply by appealing to intuitions, since in some cases we are dealing with variant forms which are both highly regional.

I organize my discussion as follows: in section 2 I review the concept of “clustering,” with an eye towards contextualizing the arguments in the remaining sections; in section 3 I illustrate in detail, with a

regard is the following: the fact that AAE allows (1a) as a “real,” unmarked question is consistent with the fact that (1c) is likewise possible; that is, they both fall under the rubric of “no T-to-C movement in real questions.” In contrast, varieties which only allow (1a) as a “special” question, do not allow (1c) at all — real, special, or otherwise. For example, (1a) is possible for me as a “special” question (a la Green et al. 2007), while (1c) is impossible for me in any context.

⁴ Also worth noting is the fact that (1b) is identical in form and meaning to the syntactic construction for yes-no questions in mainstream varieties of English. The **bolding** in the examples in (1)-(13), then, is intended to indicate the variant in each pair which could be taken to be the “non-standard” one (as we will see later, however (sections 3.2 and footnote 10), it is not always the case that for every pair (or set) of variants, one is necessarily “standard;” neither example in the pair in (1e,f), or in the pair in (11), for instance, could be taken to be standard).

⁵ Under this view, then, using a label like “Appalachian English” to identify the sentences in (7) through (13) — which suggests the identification of a single grammar — is in a sense incoherent, precisely because the underlying assumption is that the existence of semantically equivalent syntactic variants entails the existence of two different grammars.

specific example from Appalachian English, the problem that intra-speaker variation presents for parametric theory; in this section, I also support the conclusions drawn with some illustration from African American and Belfast English. In sections 4 and 5 I discuss how Kroch's (1989; 1994) diachronic studies can offer a way around the problem of synchronic intra-speaker variation. In section 6, I provide what I hope is an optimistic conclusion, which calls for an approach to research on micro-parametric variation, in situations involving intra-speaker variability, which appeals to the combined use of both speaker intuitions, as well as statistical analysis of corpora.

2. Parametric theory and “clustering”

One of the goals of the theory of parameters is to account for syntactic phenomena which are linked to one another. The idea is that, for a given grammar, there might be a single abstract property which gives rise to two or more surface phenomena which might not at first glance appear like they should be connected. A clear example of a parameter which yields a cluster of phenomena is what Pollock (1989) called the “opacity” or “transparency” of Agreement in French and English, and which we can just call the verb-movement parameter. Even though the example is well-known, I would like to review it here, to provide a context for the discussion of Appalachian, to follow in section 3.

2.1 The verb movement parameter

As is well known, Modern English exhibits appearance of lexical verbs to the right of adverbs like *never*; this can be seen in (14), where (14b) shows that the lexical verb cannot appear to the left of this adverb:

- (14) a. Mary never smokes.
 b. *Mary smokes never.

At the same time, we observe a seemingly unrelated set of facts in (15), where interrogatives formed on sentences without auxiliaries exhibit the phenomenon of *do*-support; so, in (15b) we see that appearance of the lexical verb *smokes* to the left of the subject *Mary* is disallowed in Modern English interrogatives:

- (15) a. Does Mary smoke?
 b. *Smokes Mary?

The facts in (14) and (15) would not seem, in the absence of any theory, to necessarily be connected to each another; as A. Kroch has pointed out, for example, even though Ellegård's (1953) study quantified both verb-adverb syntax in English, *and* the phenomenon of *do*-support, he never theorized that the two had anything to do with one another. Of course, we know, though, that modern generativists have had reason to hypothesize that the phenomena in (14) and (15) are both linkable to a single parameter, which has been characterized in the literature in subtly different ways, but for the present purposes, let us just think of it in terms of the strength of the Inflectional head of the clause. So if Infl is “strong,” it attracts the verb, which has to overtly move to adjoin to it. This movement would be responsible (i) for the appearance of the verb to the left of the relevant adverbs, and (ii) for the movement of the verb to Comp in interrogatives (Infl-to-C movement). If Infl is “weak,” on the other hand, it does not attract the verb, and so the verb remains to the right of the relevant adverbs. And furthermore, since the verb never resides in Infl in the overt syntax, it does not move to Comp in interrogatives, since Infl-to-C movement would not implicate the lexical verb.

Regarding the question of empirical support for the claim that the patterns in (14a) and (15a) are attributable to a single parameter, it is tempting to claim that the grammaticality judgments reflected in (14) and (15) are enough to tell us that the adverb-verb word order and *do*-support are surface reflexes of the same underlying property; but we know that the fact that speakers accept both (14a) and (15a) (and reject (14b) and (15b)) cannot alone serve as empirical confirmation that the two syntactic phenomena “cluster.” That is, while speaker judgments can provide indications for certain directions of study, they cannot serve as the sole empirical support for the hypothesis that they suggest. In cases like this one, we

need to find additional empirical support for the hypothesis; the empirical support in this case could come from comparison of English to other languages, such as French, and it could also come from analysis of the empirical facts revolving around the diachronic change in English verb/adverb syntax, and the historical rise of *do*-support (something which was in fact done by Kroch 1989;1994, as we will see in section 4).

This is not to say that the judgments are not useful, and that they should not point towards certain directions of study; in fact, it is these very judgments — together with judgments from French — which suggested the clustering hypothesis in the first place. And certainly, if speakers had the judgments exactly as in (14), but *accepted* (15b), then we would be much less inclined to hypothesize that the patterns arise from properties of Infl.

3. The problem of intra-speaker variation for finding clusters of phenomena

Within the context of the above discussion, let us now address the problem of native speaker judgments in situations of robust intra-speaker variation, where sorting out the data is not as straightforward. To illustrate, let us imagine a speaker of a variety of English who allows appearance of the lexical verb both to the right and to the left of the adverb, and who allows both *do*-support and movement of the lexical verb to C⁰ in interrogatives; in other words, imagine a speaker who allows all four possibilities seen in (16) through (19):

(16) Mary never smokes.

(17) Mary smokes never.

(18) Does Mary smoke?

(19) Smokes Mary?

There are two comments that we can make in this regard: first, this situation is not purely hypothetical; in fact, this exact example of mixed phenomena is exhibited in earlier stages of English, as Ellegård (1953) and Kroch (1989;1994) have shown. Here, we can imagine a generative syntactician alive at the time Middle English was spoken, who wanted to understand the nature of the intra-speaker variability observed; this generativist would have found him or herself in the middle of the change, and so would have been witness only to the existing variability amongst speakers; the linguist would not have had the advantage of observing the trajectory of the change, or of knowing what its conclusion would be. And although — as we said — it is true that the *Modern* English judgments in (14) and (15) alone cannot tell us if (14a) and (15a) *cluster*, these data do not *preclude* a clustering hypothesis, since we at least know that (14a) and (15a) are possible in the same system. Our 15th Century generativist would not have had this advantage, however: witnessing the fact that speakers use all four grammatical options (16) through (19), they would not have had a way of knowing whether the clustering hypothesis was even precluded, given that there would have been no way to know whether (16) and (18) were even possible in the same system; it might have been just as reasonable to hypothesize that (16) and (19) belong to the same system, or that (17) and (18) belong to the same system, or that all four belong to four different systems, for that matter.

The point is, while we do not have to worry about this imagined 15th Century generativist, we do have to worry about thoroughly analogous, synchronic cases of intra-speaker variability, of the sort that we already saw earlier, in the examples in (1-13). What I will turn to now, then, is a discussion of one of these cases in detail. Specifically, I examine *they*-existentials in Appalachian English, a construction which we got a preview of in example (11).

3.1 Appalachian *they*-existentials (Tortora 2006)

The case that I wish to discuss involves existential constructions in Appalachian English. First I will review a few facts, and then I will briefly summarize the theoretical issues in question.⁶

⁶ This section represents a summary of a more detailed description and analysis in Tortora (2006); I refer the reader to that article for a full discussion.

As is fairly well-known, in various parts of the Appalachian region there is an existential construction which uses the morphological form *they* as a so-called “expletive;” this can be seen in the examples in (20) (W&C = Wolfram & Christian 1976; M&H = Montgomery & Hall 2004; DOHP = *Dante Oral History Project*):⁷

(20) a. They is something bad wrong with her. M&H:lxii
cf.: *There is something bad wrong with her.*

b. They is six trees would have made anybody a good dwelling house M&H:xlx
cf.: *There is/are six trees that would have made anybody a good dwelling house.*

Tortora (2006) addresses the challenge that the construction in (20) poses for the robust cross-linguistic generalization discussed by Cardinaletti (1997). Specifically, Cardinaletti studied a number of expletive constructions cross-linguistically which show the following: if the expletive morpheme is unambiguously marked for nominative, then it triggers verb agreement. This is illustrated by French *il* in (21):

(21) Il arrive trois hommes.
il arrive.3sg three men

As can be seen, French *il* triggers singular agreement, despite the fact that the associate is plural. Crucially, this contrasts with case-vague expletives, like German *es*, seen in (22):

(22) Es sind wohl viele Leute angekommen.
es are probably many people arrived

In contrast with the facts in (21), we see in (22) that the case-vague *es* does not seem to play a role in agreement (so that the associate in such expletive constructions is the agreement trigger).

Given these facts, Cardinaletti (1997:526) proposes the Nominative Agreement Hypothesis:⁸

(23) **Nominative Agreement Hypothesis (NAH):** *The verb agrees with the expletive if and only if the expletive morpheme is not ambiguous with an object morpheme; i.e., only those expletives that are unambiguously marked as nominative trigger agreement with the verb.*

Given the NAH, the challenge that the Appalachian *they*-construction poses is clear: why would the expletive morpheme *they* not trigger plural agreement? In other words, why would the verb form in (20a) and (20b) not be *are*, agreeing with the apparently plural *they*? This is what we actually would expect,

⁷ I will refer to this pro-form as “expletive *they*,” to distinguish it from the referential pronoun *they* (as in *They are happy*).

For the reader who may be wondering whether expletive *they* is best synchronically analyzed as a phonologically reduced form of the morpheme *there*, note that several arguments are given in the literature against this hypothesis (notably, Wolfram & Christian 1976 and Montgomery 2006). This is not to say that existential *they* does not diachronically derive from existential *there*; see Montgomery (2006) for a very plausible scenario (which involved phonology, morphology, and syntax), for how the change from *there* to *they* could have obtained historically.

⁸ Cardinaletti’s NAH is highly reminiscent of Henry’s (1995) analysis of singular concord in Belfast English, where it seems that only those plural referential pronouns which are not case-vague (as in (i)) obligatorily trigger plural agreement with the verb; referential pronouns such as *themuns*, which are case-vague, allow for lack of agreement (as in (ii)):

(i) They are happy / *They is happy
(ii) Themuns are happy / Themuns is happy.

given that the expletive morpheme *they* is homophonous with referential *they*, which itself is Nominative, and which itself robustly triggers plural agreement with the verb in Appalachian English (like it does in Belfast English), as can be seen in the examples in (24):⁹

- (24) a. They are happy.
 b. What made you a Democrat? Well, because they're for the poor class of people, that's the reason I'm a Democrat! They believe in helpin' the poor man and the others is big fat cats.
 (DOHP: GAC)
 c. You know when it come a snow they didn't work roads back to like they do now
 (DOHP: GAC)

Thus, in contrast with the referential pronoun *they* seen in (24), expletive *they* does not seem to obey the NAH, as it does not seem to trigger plural agreement (see (20)).

However, rather than take Appalachian expletive *they* to be exceptional in not obeying the NAH, Tortora (2006) proposed the following: contrary to appearances in examples like those in (20), expletive *they* actually does trigger agreement, like all other Nominative expletives cross-linguistically. The question would then be, in what sense does expletive *they* trigger agreement in sentences such as those in (20)? The specific proposal put forth was that expletive *they* in Appalachian lacks a value for its number feature; thus, in contrast with referential *they*, it is not formally plural (where formal marking of plural involves the value minus for the singular feature). This is sketched in (25) and (26):

Appalachian expletive *they* (*They is something bad wrong with her*):

(25) *they*_[sing]

Appalachian referential *they* (*They are happy*):

(26) *they*_[sing]

The idea is that there are two *they* morphemes in Appalachian English: expletive *they* and referential *they*. The two differ in that expletive *they* is “weak” (in the sense of Cardinaletti & Starke 1999 (C&S)), and as such lacks a value for its number feature.¹⁰ This is in contrast with referential *they*, which is “strong”

⁹ Consider the following preliminary piece of evidence supporting the hypothesis that expletive *they* is in fact Nominative: as Tortora, Bernstein, den Dikken & Zanuttini found (see footnote 1), four Appalachian speakers interviewed who accepted (i), rejected (ii):

- (i) Was they a time you'uns/y'all were in school together?
 (ii) *You gotta let they be enough time to plant.

This indicates that speakers who accept expletive *they* reject it in ECM contexts (which in turn suggests that expletive *they* is impossible in a structural position associated with Accusative case).

¹⁰ As argued in Tortora (2006), this is not simply an *ad hoc* solution devised just to account for the Appalachian facts; rather the idea that expletive *they* has an impoverished feature structure is consistent with the more general properties of weak pro-forms across languages, in contrast with their strong counterparts (cf. e.g. C&S's analysis of weak *esse* 'they' in Italian, which lacks a value for a [human] feature). The idea that expletive *they* is not formally plural is reminiscent of other non-referential and/or non-plural uses of the form *they*. L. Horn cites (pers. comm.) in this regard use of “singular indefinite” (and sometimes sex-neutral) *they*, as in the following examples (I thank Horn for these; see also Horn & Kleinedler 2000):

- (i) No mother should be forced by federal prosecutors to testify against their child.
 — *Monica Lewinsky's mother's attorney*
 (ii) I challenge you to find a lesbian who doesn't want to see themselves portrayed on television.
 — *actress on Showtime's The L-Word*

(again, in the sense of C&S). Thus, the nominative expletive in (20)/(25) does, unexceptionally, enter into an agreement relationship with the verb (like all nominative expletives); the verb form nevertheless surfaces as morphologically singular, because the expletive's number feature lacks a value. Note that this proposal predicts that whenever you get expletive *they* in Appalachian, the verb will surface in its 3rd singular form.

At this point, I would like to try to get closer to the main point regarding what the Appalachian expletive *they* facts have to do with the question of intra-speaker variability, and the challenge it poses for clustering theory. Specifically, let us consider something that we have not considered until now: although the use of the expletive *they* in existentials in the Appalachian region overwhelmingly involves the verb form *is* in the present tense (regardless of the number of the associate), we do find some rarer instances attested with apparently plural verb forms. This is exhibited in (27) and (28):

- (27) Are they stories about snakes? W&C:125
 (28) So, I've got the rock upstairs. They've been some fellows still wanting to buy it. DOHP: CC

Since both (27) and (28) have plural associates (*stories about snakes* and *some fellows*, respectively), we might be led to conclude that there is a grammar in which expletive *they* is not in fact capable of triggering verb agreement; in this case, it would be the plural associate that triggers agreement. However, the sentences in (29) through (30), all of which exhibit an apparently plural verb form in the presence of a singular associate (*another one*, *water*, and *a big change*, respectively), suggest that this conclusion might not be on the right track:

- (29) They are another one down the street. M&H:xlix
 (30) It seems like they used to be more water in the streams than they are [water in the streams] now. M&H:xlix
 (31) They have been a big change. M&H:l

These data, together with those in (27) and (28), seem to indicate that the number of the associate does not influence the form of the verb. Instead, the data could lead us to conclude that there is a grammar in which expletive *they* has a value for the number feature, just like referential *they*; in this case, the expletive would trigger plural agreement with the verb (just like referential *they* does).

Before we jump to the conclusion that expletive *they* is marked for plural number for those speakers who allow (27) through (31), however, it is important to consider another fact, namely, that we also find — in the same corpus — examples like those in (32) and (33):

- (32) You're better than he are! M&H:xlix
 (33) The moon were shining bright. M&H:xlix

That is, we find the verb forms *are* and *were* with 3rd person singular subjects. So now, what might *these*

Another use of non-referential *they* which comes to mind is existential *they*, as in (iii) (also discussed for AAE by Green 2002):

- (iii) They have a lot of linguists at the College of Staten Island.
 (= There are a lot of linguists at the College of Staten Island)

These various uses suggest that the form *they* undergoes a progressive “weakening,” or grammaticalization, where at least 4 stages can be identified:

- [1] Definite, referential, animate, 3rd plural *they* (e.g., *They are happy.* (= *the girls*))
 [2] Indefinite, referential, animate 3rd person (gender/number neutral) *they* (i.e., Horn's “indefinite *they*”)
 [3] Indefinite, non-referential, animate, existential *they* (e.g., *They have a lot of linguists at CSI.*)
 [4] Indefinite, non-referential, inanimate, Appalachian existential *they*

examples indicate? It seems that speakers who generate (32) and (33) exhibit paradigm-leveling, whereby the forms *are* (and *were*) are generalized throughout the verbal paradigm. In these grammars, then, *are* and *were* would have to be taken to be 3rd singular verbs. But once we accept this view of things, it then becomes entirely possible that the apparently plural verb forms in (27) through (31) are also simply cases of singular agreement, just like in (32) and (33), and also, just like the examples in (20).

3.2 The problem of finding empirical support for clustering hypotheses in *they*-existentials

The punch-line is this: the theory of agreement in Appalachian *they* existentials put forth in Tortora (2006), which tied the Appalachian data in with the general cross-linguistic facts of nominative and case-vague expletives and also of weak pro-forms, predicts the verb forms in (27) through (31) to be 3rd singular forms, just like those in (32) and (33) (and just like those in (20)). In other words, it predicts that a mono-dialectal speaker who uses (27) through (31) will also only use (32) and (33). The problem, however, is that we are not likely to find such a mono-dialectal speaker in Appalachia who can us give such neat judgments. Rather, what we find are speakers who produce the forms in (27) through (33), in addition to the forms in (20), and in addition to sentences like *He is happy*; that is, we find intra-speaker variability. To make this point clear, let us consider all of the above examples together, repeated here as (34) through (37).¹¹

(34) a. They is something wrong with him.

b. They is not so many there now.

(35) a. They are something wrong with him.

b. They are not so many there now.

(36) a. He is better.

b. They are better.

(37) a. He are better.

b. They are better.

This summarizes the actual forms that one can find within a single speaker; thus, the forms *is* and *are* are possible all around in the 3rd person, regardless of whether the subject is a 3rd singular referential pronoun, or expletive *they*.

In light of this intra-speaker variability, then, how can we know what belongs together? As argued above, there is cross-linguistic and theory-internal motivation to pursue what we can call here “Hypothesis A,” namely, that expletive *they* is unvalued for a number feature; this hypothesis thus predicts that (34) and (36) cluster, and that (35) and (37) cluster, as illustrated here:

Hypothesis A (there are two different grammars, an “*is* grammar” and an “*are*-levelling grammar”; existential *they* is unvalued for number in both):

Cluster 1 (the “*is* grammar”):

(34) a. They is something wrong with him.

b. They is not so many there now.

(36) a. He is better.

b. They are better.

¹¹ As in footnote 2 above, the sentences here are abstractions of the attested examples in (20), (24), and (27) through (33); I abstract to highlight the morpho-syntactic patterns, for the purpose of clarity of the argument, to follow.

Cluster 2 (the “*are*-levelled grammar”):

- (35) a. They are something wrong with him.
 b. They are not so many there now.
 (37) a. He are better.
 b. They are better.

The *Cluster 1* box is intended to illustrate the “*is* grammar”, and the *Cluster 2* box illustrates the hypothesized “*are* grammar.”

Given the discussion above, however, one could easily imagine a competing Hypothesis — call it Hypothesis B — with different theory-internal motivation, which holds that in one grammar (call it “Grammar X”), existential *they* triggers plural agreement (like referential *they*). Like Hypothesis A, Hypothesis B would correctly predict the data in (35) to exist, but it would make different claims about clustering, as illustrated here:

Hypothesis B (in a Grammar X, existential *they* is valued for a plural number feature; in a Grammar Y, it is unvalued for number):

Cluster 1 (“Grammar X,” in which existential *they* triggers plural verb agreement):

- (35) a. They are something wrong with him.
 b. They are not so many there now.
 (36) a. He is better.
 b. They are better.

Cluster 2 (“Grammar Y,” in which existential *they* is unvalued for number):

- (34) a. They is something wrong with him.
 b. They is not so many there now.
 (36) a. He is better.
 b. They are better

In contrast with Hypothesis A, Hypothesis B would take (35) to cluster with (36) (as the plural verb form would be analyzed as plural agreement with the expletive).¹² The problem, though, as the reader might have surmised by this point, is that the intra-speaker variability leaves us with a lack of an empirical means to decide the issue, and we are thus left at an impasse. And crucially, here, we cannot simply claim that the dialect form in (35) “matches up” with the more “non-standard” form in (37a), because in this case, (34) and (35) are both “non-standard.” In other words, in the Appalachian case at hand, neither of the varying

¹² Nothing in theory precludes the possibility of there being a system which has “Grammar X” for existential *they*, but which is also an “*are* grammar.” In this case, (35) would again cluster with (37), but it would then not be clear whether (35a) and (35b) have *are* because they belong to an “*are* grammar,” or because existential *they* triggers plural agreement. This in turn raises the question of whether agreement is triggered at all, in a levelled paradigm. The same question can be asked of the simple past tense of lexical verbs in English, where all varieties have a levelled paradigm (e.g., *walked*). Another way to ask the question is as follows: do speakers take *they walked* to represent “plural agreement,” and *he walked* to represent “singular agreement” (i.e., is the formal feature for number plural in the former, and singular in the latter)? I leave this question open.

forms in (34) and (35) belong to the standard, so we cannot rely on some intuition that some forms go with one dialect/register, and others go with another; we are thus left with precious little to help us to confidently tease apart which forms belong to *Dialect1*, and which belong to *Dialect2* (and furthermore, whether or not there is even a *Dialect3* in the mix here). Note that the problem is not just one of training the relevant speakers to give grammaticality judgments; even a speaker who is good at giving judgments (on the level of a well-trained linguist) is not in a position to definitively determine whether (35) and (37) belong to the same system.

3.3 Some comparative notes on Belfast and African American English

3.3.1 Belfast English: the low subject parameter

And as we saw earlier, knowing what belongs to what system is a pre-condition to determining whether one or the other clustering hypothesis is even viable. It might be useful here to compare the Appalachian case with the Belfast case in (3) and (4), dealt with in Henry (1995), and repeated here:

- (3) a. **The eggs is cracked.**
 b. The eggs are cracked.
- (4) a. **Any country couldn't stand it.** (= 'No country could stand it')
 b. No country could stand it.

As noted in the introduction, Henry (1995) hypothesized that singular concord (as seen in (3a)) and the possibility of NPIs in matrix subject position (as seen in (4a)) were both surface reflexes of a single property of the grammar, namely, a structurally low subject position. To support this hypothesis, she noted that speakers who accepted (3a) also accepted (4a), while speakers who rejected (3a) also rejected (4a). Now, while these judgments cannot serve as empirical confirmation that (3a) and (4a) *cluster*,¹³ they are at least not *inconsistent* with the clustering hypothesis, as they seem to tell us that at the very least, (3a) and (4a) belong to the same system, or dialect. But for Appalachian, we cannot achieve, through introspection, even this basic sorting out the data (which would hold that two forms belong to the same system), given that the variant forms in question are all "non-standard." In contrast, the Belfast English examples can at least be grouped according to what is intuitively taken to be non-standard vs. standard.

However, as we shall see immediately below, the sorting of examples according to the intuition that some are "standard forms" while others are "non-standard forms" can itself lead to trouble.

3.3.2 African American English: the T-to-C movement parameter

The discussion above leads me to a novel problematization of some African American English data discussed by Green et al. (2007) and Roeper & Green (2007) (R&G), illustrated in (1) and (2), and repeated here:

- (1) a. **She can leave?** (= 'Can she leave?')
 b. Can she leave?
 c. **Why she can have that book?** (= 'Why can she have that book?')
 d. Why can she have that book?
- (2) a. **I wonder can she leave.** (= 'I wonder if she can leave')
 b. I wonder if she can leave.

Let us first consider the examples in (1), where we see that some speakers of African American English allow for both lack of inversion (1a,c) and inversion (1b,d) in unmarked matrix interrogatives. Some

¹³ Note that there could be an independent reason why speakers jointly would reject (3a) and (4a): they are both non-standard.

observations made in footnote 3 above regarding the unmarked status of (1a,c) bear repeating. In particular, the term “unmarked” is crucial here: as Green et al. 2007 and R&G note, the example in (1a) is possible as a “real” question in AAE — i.e., a real request for information. The possibility of (1a) as an unmarked sentence in AAE should thus *not* be confused with the fact that other varieties of English (e.g., mainstream) allow the string in (1a) as a “special” question, i.e., one that carries with it a presupposition (or, as characterized by L. Horn, pers. comm., an “incredulity” or “surprise” question). The fact that varieties other than AAE allow (1a) as a “special” question is something which is not relevant to the present discussion. Relevant in this regard is the very fact that (1c) is likewise possible in AAE, which falls under the same rubric of “no T-to-C movement in real questions.” Note that varieties which only allow (1a) as a “special” question disallow (1c) altogether — real, special, or otherwise.¹⁴

Now that we have cleared the air regarding the sentences in (1), let us consider the examples in (2), where we see that the same speakers who allow (1), also allow for both lack of inversion (2b) and inversion (2a) in embedded interrogatives.¹⁵ Taking all the sentences in (1) and (2) together, a question arises regarding how speakers would sort these data. I have asked both native and non-native speakers alike to sort the data according to which sentences would be classified together in terms of “dialect.” Interestingly, I consistently have found amongst my consultants a strong intuition that (1a), (1c), and (2a) should be classified together, and that (1b), (1d), and (2b) belong together.¹⁶ That is, people have a “sense”

¹⁴ I should clarify that Green et al. (2007) observe that (1a) is also possible for AAE speakers as an incredulity question. From the perspective taken here (i.e., the “two grammars” hypothesis), however, this fact is unremarkable, if we take the relevant speakers of AAE to also have a mainstream English grammar. In this case, it follows that the string in (1a) would get the incredulity interpretation it gets for speakers of mainstream English. Here is where I differ from Green et al., who claim that the ability of AAE speakers to allow for both the unmarked *and* the incredulity reading for (1a) shows that AAE “...has a wider range of variation than MAE.” From the perspective offered here, it is not that AAE has a wider range of variation, but rather, that speakers of AAE also use MAE forms.

Under this view, the question arises as to why the string in (1a) should be possible with an incredulity/surprise reading in mainstream English, while the string in (1c) is not possible for the same function. The fact that (1c) does not serve the same incredulity function as (1a) in mainstream English suggests, I think, that the uninverted unmarked string in (1a) in AAE has a different underlying structure than the uninverted incredulity string “She can leave?” in mainstream English.

¹⁵ Again, we must be careful here not to conflate the embedded inversion seen in (2a) with the possibility of embedded inversion for some mainstream English speakers, who may allow it — but with a special interpretation, as L. Horn (pers. comm.) points out. Specifically, Horn notes (citing Gordon & Lakoff 1971) that for some mainstream speakers, a string such as that in (2a) is possible as an indirect request for information (as opposed to “...simply reporting ... internal reflections as to the possibility of her leaving”). It is important to note, in this regard, that in contrast with this special interpretation, the embedded inversion in AAE in (2a) is the pragmatic/functional/semantic equivalent of the non-inverted sentence in (2b), and does not have any special status. See Henry (1995) and McCloskey (2006) for discussion of varieties of Hiberno-English which exhibit the same phenomenon.

¹⁶ The reader who is a native speaker of (any) English (variety) can try this thought experiment on him or herself; it seems that the intuitions follow from what we perceive to be non-standard ((1a), (1c), (2a)) vs. what we perceive to be standard ((1b), (1d), (2b)). And as is the case with the two non-standard Appalachian forms in (34) and (35), the examples in (1e,f) (repeated here) show that sometimes two variants are both non-standard:

- (1) e. **What that gots to do with me?** (= ‘What does that have to do with me?’)
 f. **What does that got to do with me?** (= ‘What does that have to do with me?’)

This raises the question of the status of (1f), which seems to exhibit a form recognizable as “standard” vis-a-vis T-to-C movement, but also a form recognizable as “non-standard” vis-a-vis choice of verb form (*does got*). It is not incoherent to analyze the grammar of T-to-C (which depends on the nature of the C

that the bolded forms go together socio-linguistically.

Importantly, these intuitions go against Green et al.'s clustering hypothesis, which essentially takes the sentences in (1a) and (1c) to cluster with the sentence in (2b) (as they all involve the same grammar of lack of T-to-C raising); likewise, their theory takes the sentences in (1b) and (1d) to cluster with the sentence in (2a) (again, as these all involve the grammar of T-to-C raising). Note that the speaker intuitions, which would put (1a,c) and (2a) together, also contrast with the cross-linguistic facts: there are numerous dialects (such as Appalachian English and Belfast English) which have embedded inversion, as in (2a), but which do not have matrix non-inversion, as in (1a) and (1c). This again suggests that embedded inversion and matrix non-inversion do not “go together,” even in the basic sense of belonging to the same system (and certainly not in the clustering sense).

3.3.3 Final notes on “data sorting” intuitions

The point I wish to make with these cases from Belfast and African American English is that, despite the fact that speakers might have strong intuitions regarding which forms “belong together,” generativists analyzing these structures may have strong theory-internal reasons to hypothesize that they in fact do not even belong to the same grammar. To look at the problem the other way around, we cannot support our clustering hypotheses — or even necessarily trust what are the promising directions of study — simply by relying on native speaker intuitions regarding which phenomena go together; as both the Belfast and African American English cases show, in some cases the intuitions line up with our theories (perhaps coincidentally; see footnote 9), but in other cases they do not. And for the cases at hand, there is no compelling reason to conclude that the intuitions are reliable in one case, but not in the other. Thus, the question of how to use native intuitions to empirically support our clustering hypotheses in the context of intra-speaker variability, or even just to look for promising directions of study, is not a trivial one.

4. A look to diachronic studies for help

Even though it is difficult (if not impossible in some cases) to obtain empirical support for a hypothesis that two phenomena belong to the same system, just by appealing to native-speaker intuitions, maybe we do not have to throw in the towel and conclude that there is no way to tease apart — independent of the theory — which phenomena go together, or even more importantly, which phenomena cluster. In this regard I would like to consider Kroch's (1989;1994) diachronic corpus studies, which might provide us with a way to do an end-run around this problem of intra-speaker variability.

Specifically, let us return to a matter under discussion in section 3 (exx. (16)-(19)), namely, that at an earlier stage of English, both (i) the order of the lexical verb with respect to adverbs like *never*, and (ii) the phenomenon of *do*-support were variable. I repeat the relevant examples here in (38) through (41):

(38) Mary never smokes.

(39) Mary smokes never.

(40) Does Mary smoke?

(41) Smokes Mary?

As Ellegård (1953) and Kroch (1989; 1994) note, in the earliest English texts (Early Middle English), the structures in (38) and (40) were not possible; instead, the (now ungrammatical) forms in (39) and (41) were the rule. However, at some point, (38) and (40) were introduced as apparent options, and subsequently, for a period of hundreds of years, all of the forms seen above were possible. Then, with the passage of time, the grammatical options in (38) and (40) occurred more and more frequently, while the original forms occurred less and less frequently. Of most significance in Kroch's (1989; 1994) study, however, is the following: the increase of *do*-support in interrogatives over time occurred at the same rate

head) as independent from the grammar giving rise to particular non-standard verb forms; however, if our intuitions about the sociolinguistics of these different grammatical strategies are on the right track, then (1f) might be analyzable as a case of code-switching. I leave this matter open.

as the increase of the adverb-verb order, regardless of the fact that at any given time, one construction might have appeared in the text more frequently than the other. This is what Kroch calls the “Constant Rate Effect.” But why would these two apparently unrelated phenomena increase in use at the same rate over time?

First, let us recall that we are assuming these “mixed phenomena” to reflect intra-speaker variability; thus, the two sets of variants ((38)/(39) and (40)/(41)) reflect the juggling — on the part of the speaker — of “two different grammars” (or, “syntactic diglossia” in the sense of Kroch 2001). As such, a speaker exhibiting all forms in (38) through (41) would be bi-dialectal, in that she would have both a weak Infl and a strong Infl in her lexicon:

(42) Lexicon of Speaker exhibiting all forms in (38) through (41):

[A] weak Infl (yields no overt verb movement)

[B] strong Infl (yields overt verb movement)

This would contrast with a mono-dialectal speaker of Modern English, whose lexicon would only contain [A]:

(43) Lexicon of mono-dialectal Modern English speaker:

[A] weak Infl (yields no overt verb movement)

It would also contrast with a mono-dialectal speaker of Early Middle English, whose lexicon would only contain [B]:

(44) Lexicon of mono-dialectal Early Middle English speaker:

[B] strong Infl (yields overt verb movement)

For the bi-dialectal speaker of Middle English, then, the question of whether the lexical verb moves (as in (39)/(41)) or not (as in (38)/(40)) is just a question of whether this speaker (depicted in (42)) chooses [A] or [B]. That is, following Borer (1984), who argued that parametric variation is a reflex of the properties of individual functional heads, Kroch hypothesized that the variants in (38) through (41) reflect (42), which gives rise to the appearance that verb movement was optional. But as Kroch argues, the optionality is only apparent. That is, while true that the speaker has an option regarding which variant to choose from the lexicon (weak Infl or strong Infl), once a speaker chooses strong Infl, verb movement must obtain; of course, if the speaker instead chooses weak Infl from the lexicon, verb movement is forbidden. The appearance of optionality in the grammar thus derives from the fact that speakers have a lexical choice (which Kroch hypothesizes to be governed by sociolinguistic factors); however, once the lexical choice is made, the grammatical consequence is exceptionless. This approach to syntactic variability gives specific theoretical content to the notion of “two different grammars.”

Getting back to the question of the Constant Rate Effect, and why two apparently unrelated phenomena would increase in use at the same rate over time, as Kroch argues, it reflects the increased choice, over time, of a single grammatical option (in this case, “weak Infl”). To look at it the other way around: the increased choice of a single parameter gives rise to the increased use of all of the syntactic phenomena that are the surface reflexes of this choice (and of course, this entails that their increased use will be at the same rate).

Now, it may seem obvious to us that this should be the case; given our theory of functional heads and feature strength (together with a theory of verb movement, adverb placement, and T-to-C movement in interrogatives), we attribute (38) and (40) to a single property of the grammar, and so we expect the diachronic data to reflect this. But this is precisely the point: our theory makes a prediction which can be tested by statistical analysis of a corpus; in other words, we can test a hypothesis that X and Y cluster by looking at the rate at which X and Y increase, over time. In this particular case, the prediction of the theory is borne out with the Constant Rate Phenomenon.

So, what is the relevance of Kroch's study to the current discussion? Well, what his study shows us is that, even in the face of variability and lack of reliable native speaker intuitions, we can tease apart mixed phenomena. And as I illustrated in section 3, in my review of the Appalachian existential construction (as well as in my review of the Belfast and African American English phenomena), variability and lack of reliable speaker intuitions is exactly what we are facing in the synchronic analysis of many dialects. So Kroch's methodology seems like just what we need to resolve the problem that intra-speaker variation presents to parametric theory.

But here, there are some objections which could be levelled against the proposal that we apply Kroch's methodology to synchronic situations. First, and most obviously, it would seem impossible to apply Kroch's methodology to synchronic analysis, since by definition Kroch looks at change over time. In other words, how can we look at different stages of a synchronic situation? A second, and perhaps somewhat less obvious objection that could be levelled against this idea is that the cases of grammar competition studied by Kroch are based on written texts. And as Kroch (1994:fn 6) notes, for all we know, what we are seeing in historical texts is the "competition between the grammar of the spoken language of a given time and an archaic but still influential literary standard." If this is the case, then as Kroch points out, the variability we see in texts has "no purely linguistic significance."

Let me address these two possible objections: regarding the latter one, one could actually look at it as a strong motivation to study intra-speaker variability in living languages using Krochian statistical analysis; as Kroch (2001) notes, "Given the strong possibility that textual data do not give evidence for the process of language change in a vernacular, there is a real need for the study of syntactic innovations in living languages, using sociolinguistic methods to observe unreflecting speech." As such, the study of living languages could give us a way of confirming whether the kinds of variability we find in historical texts mimic the kinds of variability found with individual speakers. And getting confirmation of this would be a positive step towards developing a much-needed trust that analyzing historical texts is relevant to the theory of I-language (as it is entirely possible that what we are seeing in historical texts is not merely the "competition between the grammar of the spoken language of a given time and an archaic but still influential literary standard").

Regarding the former objection, namely, how can we apply Kroch's methods of analysis to synchronic situations if by definition Kroch and colleagues look at change over time? I address this question as the final part of this paper. Allow me to preface the next section by noting (as a way of avoiding a premature raising of the reader's hopes) that what I propose has yet to be enacted; as such, the reader should take the remainder of the discussion as a promissory note, pending the creation of a large scale parsed corpus of Appalachian speech (see section 6).

5. How can we do diachrony in synchrony to simulate the Constant Rate effect?

Looking for diachrony in synchrony seems like a contradiction. How can we plot diachronic change in a synchronic situation? Even though Kroch himself claims (2001) that there is a "real need for the study of syntactic innovations in living languages," he also notes that "Such studies do not at present exist, in part because syntactic change is relatively rare and hard to catch on the fly." And any number of sociolinguists currently attempting to find change in current varieties can tell us how difficult a task this is.¹⁷

That said, here I would like to discuss three ways that we can study dialects synchronically, or quasi-synchronically, using Kroch's corpus methodology to tease apart mixing of syntactic phenomena, so that we can try to get around the problem that intra-speaker variability presents to parametric theory. The idea is to engage in kinds of synchronic studies that allow us to simulate the "Constant Rate" phenomenon, and uncover the kinds of correlations that are like the lock-step changes found in the diachronic studies. This in turn would help us determine, for cases of intra-speaker variability, what clusters. Specifically, I discuss the following three kinds of synchronic or quasi-synchronic studies which can be done to uncover potential clusters of properties:

¹⁷ W. Labov notes (p.c.) that "Change in progress is our main tool for studying the mechanism of change, but no one's been smart enough to locate extensive syntactic change going on right now."

- [I] We can compare closely related dialects as representing different stages of change
- [II] We can do apparent-time studies
- [II] We can do real-time studies

I address each of these in turn.

5.1 Related dialects as stages of change

Let us take seriously an observation made in Wolfram (1984), which is in fact an observation that goes back to Gaston Paris (Paris 1888; see also Labov 1975. I thank Paola Benincà for pointing these out to me): What both Paris and Wolfram note is that stages of language change can be reflected in the contemporaneous grammars of closely related dialects. Wolfram (1984:21-219) uses a particular example from the morpho-phonology of Appalachian English for illustration, so let me review this example here.

We can preface the Wolfram illustration by noting that the English pronoun *it* etymologically derives from the Old English form *hit*; clearly, what has happened in Standard English is that the [h] has been lost over time. There are however American dialects in which the form *hit* is still possible, and Appalachian English is one of them. Some examples can be seen in (45) (taken from Montgomery & Hall 2004:xxxviii; 323); the examples are intended to show the various syntactic contexts in which this pronoun is possible (a perusal of Montgomery & Hall 2004 reveals that for Smokies speech, at least, the only time the form *it* is used at the exclusion of the form *hit* is as an expletive in existentials):

- (45) a. **Hit**'s been handed down to him.
 b. **Hit** must have been in the thirties.
 c. I believe they called **hit** the Cable School.
 d. They had to raise the young one and take care of **hit**.
 e. That dog done **hits** best to break loose.

The [h] of *hit* is slowly disappearing in America now (along with the [h] of *hain't*). As Wolfram notes, the loss of [h] in American dialects appears to have been proceeding in an orderly fashion, such that at first, *hit* variably occurs with *it* in unstressed syllables only, and then at a later stage, *hit* and *it* occur variably in both unstressed and stressed syllables. These two stages are illustrated as Stages 2 and 3 in the Table 1, which is an adaptation of the Table 1 in Wolfram (1984:219):

Table 1: Stages for the loss of [h] in *hit*

	unstressed syllables	stressed syllables
Stage 1: Before Change	hit	hit
Stage 2: Beginning [h] loss	hit / it	hit
Stage 3: General variability	hit / it	hit / it
Stage 4: Ending [h] loss	it	hit / it
Stage 5: Completed Change	it	it

After Stage 3, in which *hit* and *it* co-occur all around, there is a Stage 4 in which the form *hit* disappears completely in unstressed syllables, such that the only place it is found is in stressed syllables. This is followed by the final Stage 5, in which the form *hit* disappears completely from all contexts (as in Standard English).

The point of bringing up this example of successive stages of language change is the following: Wolfram notes that not only do Stages 1 through 5 in Table 1 represent different stages of overall change, but they also represent different related dialects; as such, we could convert Table 1 into Table 2. So, Table 2 is just like Table 1, except instead of representing the successive stages of change, it represents contemporaneous related American dialects:

Table 2: Dialects

	unstressed syllables	stressed syllables
Dialect 1	hit	hit
Dialect 2	hit / it	hit
Dialect 3	hit / it	hit / it
Dialect 4	it	hit / it
Dialect 5	it	it

As can be seen in Table 2, Dialects 2, 3, and 4 have the pronouns *hit* and *it* occurring variably in different surface contexts. In Dialect 2 they occur variably only in unstressed syllables, in Dialect 3 they occur variably in both unstressed and stressed syllables, and in Dialect 4, they occur variably only in stressed syllables.

Now let us see how all of this is relevant to the question of how to apply the Krochian corpus methodology to synchronic situations. If the Tables in 1 and 2 represent both (i) stages of change, and, (ii) contemporaneous dialects, then that means that comparing the different dialects in Table 2 is the synchronic equivalent of comparing different stages of language change. Now, recall that the whole purpose of comparing different stages of language change is so that we can test whether certain syntactic phenomena cluster. So, regarding this specific example of [h]-deletion: imagine you have a theory, sketched in (46), which holds that the presence of [h] in the English pronoun *hit* is linked to another phenomenon, X (which let us say contrasts with phenomenon Y, which you have reason to hypothesize is linked to the pronoun *it*):

(46) *hit* ↔ X
it ↔ Y

Such a theory would predict that if [h] starts disappearing, it should disappear at the same rate at which phenomenon X disappears (given the Krochian Constant Rate effect). Given that Dialects 1 through 5 in Table 2 also represent different stages of change, we can thus test this prediction via a synchronic comparison of corpora from these dialects; to do so, we would have to determine the frequencies of the occurrence of *hit* and of phenomenon X in these dialects; if the hypothesis represented in (46) is correct, what we should find is the following: the decrease in use of the pronoun *hit* should be at the same rate as the decrease in use of phenomenon X, as we move from Dialect 1 to Dialect 5 (even if the pronoun *hit* and phenomenon X themselves do not occur with the same frequencies in each individual dialect).

So, the idea is, by comparing closely related dialects, we can use Kroch's methodology to test for the Constant Rate phenomenon synchronically, which could give us a way to tease apart the mixing of syntactic phenomena within a single dialect (or speaker), of the kind discussed in section 3, and thus get

around the problem that intra-speaker variation presents to parametric theory.

5.2 Apparent-Time Studies

In this section I briefly discuss another possible way of looking for the Constant Rate effect synchronically, so that we can determine whether syntactic phenomena we hypothesize to be the surface reflexes of a single parameter do in fact cluster. Here I have in mind “Apparent-Time” studies, whereby one studies the speech of different generations within the same community as “Stages of Change.” If we take the speech of different generations within the same community to be different dialects — and here, it just becomes a matter of terminology — then the application of the Krochian methodology in an Apparent-Time study would be just as I described in detail for cross-dialectal study, so it is not necessary to elaborate the idea further. However, a few comments are warranted regarding the possibility of change in progress vs. the possibility “age grading” in apparent time studies

As is well known, generational differences may sometimes be due to change in progress, while sometimes they may be due instead to the “Age Grading” phenomenon (where different forms are socially appropriate for different ages), in which case cross-generational comparison does not yield an effect of “apparent time.” However, in the context of the issues raised in this paper, it is not clear to me whether Age Grading vs. Apparent-Time would make any difference to the Constant Rate effect: ultimately, the Constant Rate effect should come about because a speaker chooses a single abstract grammatical option, with that choice resulting in more than one surface syntactic reflex. Given this, the question arises as to whether the Constant Rate effect must, by necessity, only show up in situations of change, where there is movement in one direction. As stated immediately above, if the Constant Rate effect is about a speaker’s choice of a single grammatical option, we might expect to find the Constant Rate effect in closely related dialects — or in different generations within the same community — which have different frequencies of variability for particular phenomena, with no necessary promise of a trajectory of movement. As such, regardless of whether they reflect age grading or change in progress, apparent time studies can still provide us with the data needed to determine whether two syntactic phenomena cluster with one another.

5.3 Real-Time Studies

In this last section, I briefly consider Real-Time studies, which are diachronic in that they involve studies at different stages of particular individuals’ lives. Such studies can yield data on stages of change, if a particular form or forms can be shown to change over the life-span of an individual (and if the changes can be shown not to be due to Age Grading). Again, here, if we take the speech of an individual over the course of his or her life to reflect different stages of change, then the application of the Krochian methodology in a Real-Time study would be just as I described in detail for cross-dialectal study in section 5.1 above. Here it is worth mentioning the study reported in Cukor-Avila (2002), which seems to show that Real-Time studies can indeed uncover the Constant Rate effect (I thank Anthony Kroch for having pointed me in the direction of this work).

Specifically, Cukor-Avila (2002) analyzes the use of quotative *be like*, *go*, and *say* in the speech of African American residents of Springville, Texas; here I will just give examples of quotative *be like*. Two examples can be seen in (44) (taken from Cukor-Avila 2002):

- (47) a. I was like, “What’s your name?”
 b. He was like, “Yeah.”

Cukor-Avila’s study spans a 12-year period, which enables her to observe the changing use of *be like* in the speech of a particular group of girls over time. There are two points from this study that are worth noting in the context of this paper: First, Cukor-Avila finds that quotative *be like* is used far more frequently to quote “inner dialogue” than it is to quote someone’s actual dialogue, and it is used far more frequently to quote first person dialogue than it is to quote third person dialogue. What this tells us is that the surface frequencies are different, depending on grammatical and discourse contexts. Second, though, and more importantly, she finds in this study that as the use of quotative *be like* increases over the 12-year

span of these girls' lives, it increases at a constant rate for the different grammatical contexts (first vs. third person) and the different discourse contexts (inner dialogue vs. actual dialogue).

As Cukor-Avila notes, this suggests that “the constant-rate effect may also hold for real-time change in vernacular varieties of speech” (p. 21) (this addresses a comment made in Kroch 1994:fn 6: “Only work on possible cases of competition in living languages can determine whether it exists in unreflecting vernacular speech, a question to which we do not yet know the answer”). This result is very promising in the context of the points I have tried to make in this paper. For those of us studying the syntax of a variety that exhibits robust intra-speaker variability, the opportunity to test whether two (or more) syntactic phenomena are linked via a single abstract parameter indeed exists. Native speaker judgments do not readily afford the opportunity, since in many cases of variability it is difficult — if not impossible — for a speaker to intuit what forms belong to the same system. But applying the Krochian methodology in the study of closely related dialects, or in apparent-time studies, or in real-time studies, we do have the opportunity to test our hypotheses regarding clustering of phenomena.

6. Conclusions

The discussion in this paper calls for corpus-based studies of Appalachian English (and varieties like it, which exhibit robust intra-speaker variability). To this end, a research team — which consists of Michael Montgomery (University of South Carolina), Beatrice Santorini (University of Pennsylvania), Frances Blanchette (CUNY Graduate Center), and myself — is in the process of creating a 1-million word parsed corpus of Appalachian speech, based on transcripts from Montgomery's *Archive of Traditional Appalachian Speech and Culture*; given the promissory note offered in the second half of this paper, I will take this opportunity to describe the nature of our project, in progress.

Our corpus will be based on the *Archive of Traditional Appalachian Speech and Culture* (ATASC), an unparsed corpus of speech which Montgomery has collected and transcribed from recordings from ten parts of the Appalachian region in six different states.¹⁸ The corpus, which will feature the *Penn Treebank* method of syntactic annotation, will be accompanied by a full set of digitized, text-searchable recordings of the speech from which the corpus is transcribed, in the form of *.wav* files. The *.wav* files on which the parsed corpus is based will be text-searchable using *Praat* (Boersma and Weenink 2011), as a result of aligning the transcripts with the speech signal, using the “text-to-speech” alignment technology developed and made available by Jiahong Yuan the University of Pennsylvania. Our aim is for the final product to afford an unprecedented approach to the analysis of synchronic English dialect data for all kinds of linguists. Syntacticians will be able to search, count, and analyze syntactic configurations of any type. This in turn will allow for large-scale, statistical studies of syntactic data gleaned from the corpus, and for cross-dialectal comparison with syntactic data gleaned from other, extant parsed corpora, such as the *Penn Parsed Corpora of Historical English*. Such studies will allow researchers to address overarching and general theoretical questions regarding dialect variation and language change — such as those discussed in this paper — which would otherwise remain unanswered without such a corpus. But the usefulness of the corpus will not be limited to syntacticians; phoneticians and phonologists will easily be able to do text-searches through the speech signal for phonetic or phonological properties of interest. Furthermore, because the text-to-speech alignment process forces the written recording of every aspect of the utterances with the highest degree of accuracy, discourse analysts will find the corpus useful as well, as it will include every last hesitation, pause filler, and interruption, as well as time-stamps for overlapping speech. And given the text-searchability of the speech signal, the database will be completely transparent; that is, the user can check the accuracy of the transcription of any word(s) for him/herself, using garden-variety

¹⁸ Some examples of collections represented in the ATASC are: (i) the Appalachian Oral History Project housed at Alice Lloyd College in Pippa Passes, KY, and at Appalachian State University in Boone, NC (representing traditional speech from Eastern Kentucky and Northwestern North Carolina); (ii) the Dante Oral History Project housed in the Archives of Appalachia at East Tennessee State University (representing speech from Dante, in Southwestern Virginia); (iii) the Joseph Hall Tapes, on which Montgomery & Hall (2004) is based.

text-searches which allow him/her to jump directly to the desired points in the one-million-word speech signal. This is a feature which will greatly reduce the chance for error in all future work based on the corpus, and greatly enhance replicability.

As perhaps the reader understands, however, the building of a parsed corpus based on transcribed speech is not exactly the same endeavor as the building of one based on written text. In contrast with written texts, for example, it is less straightforward to parse transcribed speech. Historical texts are literary in nature, and are therefore essentially composed of grammatical sentences; speech, on the other hand, is disfluent, a problem the collaborators on this project must overcome.¹⁹ Nevertheless, it should be clear that the existence of multiple parsed corpora of non-standard dialects, which generally speaking, do tend to exhibit robust intra-speaker variability, would serve as invaluable tools for teasing apart mixed phenomena. In the introduction to this paper I stated that intra-speaker variability presents a challenge to our desire, within a theory of parameters, to understand which phenomena cluster. However, I believe that such syntactic variability also presents an opportunity for getting at the problem of clustering in a potentially fruitful way.

All of this said, a final word is in order regarding the idea of using corpora for the study of syntactic variation: I do not want this paper to suggest that analyses of corpora can replace the need for native speaker judgments, because of course they cannot. Certain facts regarding corpora will always remain true: putting aside the obvious and well-known fact that, just because something does not appear in the corpus, does not mean it is not a possibility in the grammar in question (and just because something *does* appear in the corpus, does not mean it is grammatical), corpora present fundamental problems regarding ambiguities. For example, if we find the string in (48) in an Appalachian English corpus, how can we know whether the speaker intended for negation to take scope over the quantified subject?

(48) All of his friends isn't bald.

Yet for theoretical reasons we might wish to know what the possible readings of this string are. For example, if it is true that presence of the form *isn't* is the surface reflex of the subject's appearance in a structurally low position (as discussed in Henry 1995 and Tortora & den Dikken 2010), one might expect the instance *isn't* in (48) to only allow for a reading in which an abstract negative operator in a NegP (higher than the low subject position) takes scope over the quantifier in the low subject position (thus, it might be that (48) in Appalachian English can only mean *It is not the case that all of his friends are bald*). Trying to ascertain this subtlety of interpretation by just looking at a corpus proves difficult (and may prove impossible).

In a similar vein, examination of a corpus of African American English cannot readily reveal particular semantic interpretations associated with certain constructions; consider in this regard Green, Roeper, and Terry's work on Negative Inversion, where it is revealed that Negative Inversion structures in African American English such as those in (49) have a different interpretation than the corresponding non-inverted case in (50) (examples from Green, Roeper, & Terry 2007):

(49) Nobody don't be working on the 10th floor.

(50) Don't nobody be working on the 10th floor.

Without a native speaker of the variety to consult, one could erroneously conclude that (49) and (50) are variants, just by looking at a corpus.

I do hope, however, to have shown how corpus studies of the type I talked about here could be invaluable *supplements* to the native speaker judgment methodology, because sometimes, we are dealing with native speakers who simply have not been trained to give judgments, and in any case, even if we can train speakers to give grammaticality judgments, they still may not be in a position to tell which variant

¹⁹ Methods for syntactic annotation of disfluent speech will be based on previous work done by Martins et al. (2010) and Hindle (1983).

forms may “go together” (in terms of Dialect1, Dialect2, etc.), as discussed in section 3 (especially 3.3.2). But the fact remains that, no matter what speakers can be trained to intuit, we still cannot use these intuitions to tell us which phenomena cluster.

References:

- Borer, H. 1984. *Parametric Syntax: Case Studies in Semitic and Romance Languages*. Studies in Generative Grammar no. 13. Dordrecht: Foris.
- Boersma, Paul, and David Weenink. 2011. Praat: doing phonetics by computer, version 5.2.32. <http://www.fon.hum.uva.nl/praat/>
- Cardinaletti, Anna. 1997. “Agreement and Control in Expletive Constructions,” *Linguistic Inquiry* 28.3: 521-533.
- Cardinaletti, A. & M. Starke. 1999. “The Typology of Structural Deficiency: A Case Study of the Three Classes of Pronouns,” in Henk van Riemsdijk (ed.) *Clitics in the Languages of Europe*, pp. 145-233. Berlin: Mouton de Gruyter.
- Cukor-Avila, P. 2002. “*She say, she go, she be like*: Verbs of quotation over time in African American Vernacular English,” *American Speech* 77.1: 3-31.
- Dante Oral History Project* (DOHP), Archives of Appalachia, Center for Appalachian Studies and Services, East Tennessee State University, Johnson City, TN.
- den Dikken, M., J. Bernstein, C. Tortora, & R. Zanuttini. 2007. “Data and Grammar: Means and Individuals,” *Theoretical Linguistics* 33.3: 335-352.
- Ellegård, A. 1953. *The Auxiliary do. The Establishment and Regulation of Its Use in English* (Gothenburg Studies in English II). Stockholm: Almqvist & Wiksell.
- Emonds, J. 1976. *A transformational approach to English syntax*. New York: Academic Press.
- Green, Lisa. 2002. *African American English: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Green, L., T. Roeper, & J.M. Terry (2007) “A Features-Based Account of Variation in Question Inversion and Negative Inversion in African American English: Stable Nodes and Unstable Features,” talk given at the Formal Approaches to Variation in Syntax Conference, York, England.
- Gordon, David & George Lakoff. 1971. “Conversational postulates,” *Chicago Linguistic Society* 7: 63-84.
- Henry, A. 1995. *Belfast English and Standard English. Dialect variation and parameter setting*. Oxford/New York: Oxford University Press.
- Hindle, D. 1983. “Deterministic parsing of syntactic non-fluencies,” in Proceedings of the 21st Annual Meeting of Association for Computational Linguistics, pp. 123-128.
- Horn, L. & S. Klienedler. 2000. “Parasitic Reference vs. R-based Narrowing: Lexical Pragmatics Meets He-Man,” talk given at the 2000 meeting of the *American Dialect Society*.
- Kroch, A. 1989. “Reflexes of Grammar in Patterns of Language Change,” *Language Variation and Change* 1.3:199-244.
- Kroch, A. 1994. “Morphosyntactic Variation,” in K. Beals et al. (eds.) *Proceedings of the 30th Annual Meeting of the Chicago Linguistic Society* (Parasession on Variation and Linguistic Theory.), vol 2, pp.180-201.
- Kroch, A. 2001. “Syntactic Change,” in M. Baltin & C. Collins (eds.) *The Handbook of Contemporary Syntactic Theory*, pp. 699-729. Malden, MA and Oxford: Blackwell Publishers.
- Labov, W. 1975. “On the use of the present to explain the past,” in L. Heilmann (ed.) *Proceedings of the 11th International Congress of Linguists*, pp. 825-851. Bologna: Il Mulino (also in A. Makkai et al.(eds.) *Linguistics at the Crossroads*, pp. 226-261).
- Labov, W. 1994. *Principles of Linguistic Change*. (vol. 1, *Internal Factors*). Oxford: Blackwell.
- Labov, W. 1995. “The two futures of linguistics,” in I-H Lee (ed.) *Linguistics in the Morning Calm* 3 Seoul: Hanshin Publishing Company.

- Labov, W., P. Cohen, C. Robbins, & J. Lewis. 1968. *A Study of the Nonstandard English of Negro and Puerto Rican Speakers in New York City*. Final Report, Cooperative Research Project No. 3288, United States Office of Education.
- Martins, Ana Maria, Ernestina Carrilho, et al. 2010. CORDIAL-SIN: Syntax-oriented Corpus of Portuguese Dialects, 2010.
<http://www.clul.ul.pt/en/researchteams/212cordialsinsyntaxorientedcorpusofportuguese-dialects>.
- McCloskey, James. 2006. "Questions and Questioning in a Local English," in R. Zanuttini, H. Campos, E. Herburger and P. Portner (eds.) *Crosslinguistic Research in Syntax and Semantics: Negation, Tense and Clausal Architecture*. Georgetown University Press.
- Montgomery, M. 2006. "Notes on the Development of Existential *they*," *American Speech* 81: 133-146.
- Montgomery, M. & J. Hall. 2004. *A dictionary of Smoky Mountain English*. Knoxville: University of Tennessee Press.
- Paris, G. 1888. "Les parlars de France," *Revue des patois gallo-romans*: 161-175.
- Pollock, J.-Y. 1989. "Verb movement, universal grammar, and the structure of IP," *Linguistic Inquiry* 20.3: 365-424.
- Roeper, Tom & Lisa Green. 2007. "Node labels and features. Stable and unstable dialects and variation in acquisition," *Linguistic Variation Yearbook* 7: 1-26.
- Tortora, C. 2006. "The Case of Appalachian Expletive *they*," *American Speech* 81.3: 266-296.
- Tortora, C. & M. den Dikken. 2010. "Subject Agreement Variation: Support for the Configurational Approach," *Lingua* 120:1089-1108.
- Wolfram, W. 1984. "Is there an Appalachian English?" *Appalachian Journal* 11:215-226.
- Wolfram, W. & D. Christian. 1976. *Appalachian Speech*. Arlington, VA, Center for Applied Linguistics.