

The Macroparameter in a Microparametric World*

Mark C. Baker

Rutgers University

1. Introduction

The first and most famous parameter to have been proposed is the Pro-drop Parameter (or Null Subject Parameter) of Chomsky (1981) and Rizzi (1982: ch. 4). In retrospect, we might call this a “medioparameter”. First, it was medium-sized in its formulation. It was stated as a general rule of grammar, not as the features of a lexical item, but also not as variation in a core principle of grammar. It was also medium-sized in its effects: it supposedly accounted for a cluster of properties, including subject omission, subject inversion, and the absence of complementizer-trace effects. These properties all concerned the subject of a finite clause, but together they had a fairly significant effect on the overall feel of a language. Finally, it was medium-sized with respect to the sort of comparison that led to its discovery. The Pro-drop Parameter was proposed as a response to differences among the major Romance languages, French versus Italian and Spanish, not as the result of comparing local dialects of those languages, nor as the result of comparing Indo-European languages with languages of some other family.

History has not been kind to the Pro-drop Parameter as originally stated. It is now well-known that the cluster of properties that this parameter claimed to be related fragments in various ways when one looks at a wider range of languages, either dialects or languages from other families (Gilligan, 1987). Moreover, few new proposals for parameters with the same kind of scope as the Pro-drop Parameter have gained currency since. As a result, thought about parameters has gone in two different directions.

A large portion of the field has moved away from looking for classical parameters to looking for microparameters, following the methodology and reasoning championed by Richard Kayne (see, for example, Kayne (2005)), and taken up by many others. This approach uses the same reasoning as classical parameter theory, but applies it at a higher level of magnification. It tends to look for very localized differences with small but nontrivial effects in the grammars of very closely related languages (alias dialects). Larger differences among languages are typically thought to be the result of the accretion of many of these microparameters. This microparametric approach, in contrast with the classical parameter approach, has proven to be very fertile, creating a steady stream of results that shows no signs of slowing down.

Another response, tried by some linguists, is the intuition that perhaps the most important parameters are best found by a larger scale comparison, comparing languages from different language families. Whereas classical parametric theory compared French, Spanish, and Italian, and microparametric theory compares Trentino, Fiorentino, and Piedmontese, macroparametric theory might compare Mohawk and Mayali with Spanish and Swahili. I believe this style of work has also produced results (Baker, 1996). The results have admittedly been fewer than those achieved by microparametric approaches. But it is the thesis of this chapter that this is partly an artifact of the fact that fewer linguists have taken the macroparametric path, for no very strong intellectual reason.

I develop this thesis as follows. First I offer some conceptual remarks about the true differences between microparametric and macroparametric research (section 2). One key point is that the undeniable existence of much microparametric variation does not give us any reason to doubt that there is also macroparametric variation. It does, however, affect what that macroparametric variation will look like, in a way that is not widely appreciated (section 3). I then present briefly two new parameters of a more or less macroparametric sort, discovered by comparing Bantu languages with Indo-European languages (section 4). I show why it seems wrong to reduce them to a series of microparameters, and why they probably could not be discovered by comparing dialects chosen all from one language family. Finally, I comment on preliminary results of my ongoing effort to explore these parameters of agreement over a sample of 100 languages chosen from around the world. My moral: that there is some reason to think that there are macroparameters out there hiding in a forest of microparameters, and the generative linguistic community should be trying harder to seek them out.

2. Three Kinds of Micro- vs Macro- Comparative Syntax

First let us try to gain some additional clarity on what can—and should—be meant by macroparametric syntax as contrasted with microparametric syntax. I believe that there are three distinct issues that these labels can be used to invoke, each touched briefly on in Kayne (2005), among others. It is important both to realize that the three are logically independent of each other, and to see why they tend to be intertwined in practice.

2.1 The locus of variation

The first issue can be called the “locus of variation” question. It concerns the exact ways that one speaker’s knowledge of language can differ from another speaker’s knowledge of language. The

microparametric style of syntax is associated the strong and interesting hypothesis in (1), familiar from many works.

(1) The Borer-Chomsky Conjecture

All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon.

I call this view “The Borer-Chomsky Conjecture” for the following reasons. It was first suggested, so far as I know, in Borer (1984), shortly after the original introduction of the notion of the parameter, when cracks in the classical Pro-drop Parameter were already beginning to show (see also (Fukui and Speas, 1986, Webelhuth, 1992). Chomsky (1995) then adopted it, incorporating it into his Minimalist Program, and many learned of it from there. Kayne has done much to promote it, clarify what it might amount to, and find interesting examples that illustrate it. I refer to (1) as a conjecture because it clearly has that honorable status: it was posited in advance of much of the comparative syntax work of the last 10-20 years, and has guided how much of that work has been done. It is an intriguing generalization over a few suggestive examples that were available in the 1980s and early 1990s, which makes sense and has some attractive conceptual properties. But it is a hypothesis to be proven empirically, not an established result. Simply finding many differences among languages that are consistent with (1) does not at all prove that (1) is true, because (1) is a universal statement. That is particularly so if much of the field is pursuing a methodology that presupposes that (1) is true and is designed to find parameters of this type, but which is not at all likely to find parameters of other kinds—as I claim is the case.

I, for one, am interested in the possibility that the Borer-Chomsky Conjecture may be false. The contrary view is that there can be variation in the grammar proper in addition to variation that can properly be attributed to the properties of particular lexical items. This macroparametric view can be stated as in (2).¹

(2) There are some parameters within the statements of the general principles that shape natural language syntax.

(2) is in fact no different from the classical early 1980s view of the parameter. For convenience, let us refer to parameters consistent with (1) as “lexical parameters” and parameters consistent with (2) but not (1) as “grammatical parameters” because they concern principles of grammar that cannot be localized in the lexicon per se.

The Borer-Chomsky Conjecture has a certain Occam’s razor-like charm to it, which makes it fit well with the Minimalist Program. Given that lexical items clearly vary from one language to another and these must be learned in any case, it is attractive to think that all variation

might be reduced somehow to that (Borer, 1984: 29). But there seems to be a certain global unity to a head-final language as opposed to a head-initial language, or to a polysynthetic language as contrasted with a more isolating language, which seems more pervasive than can be attributed to any particular lexical item, or even to a small class of lexical items (see section 3 for discussion). Nor does the Minimalist Program necessarily rule out the possibility of classical/macroparameters in the sense of (2). It is compatible with the pursuit of paring linguistic theory down to virtual conceptual necessity to think that there could be several locally optimal syntactic solutions to the basic design criteria for language.

Calling (2) a macroparametric view is something of a misnomer. It is perfectly possible that a lexical parameter consistent with (1) could have a substantial impact on the language generated, particularly if it concerned some very prominent item (such as the finite Tense node). Conversely, it is quite possible that a syntactic parameter of the sort envisioned in (2) could have only a small and hard-to-notice impact on the language generated. (The bounding nodes parameter of Rizzi (1982: ch. 2) could be a case in point.) But it is not a total misnomer, as we will see. And in any case the choice between (1) and (2) is the most substantive and theoretically significant question at stake in the cluster of issues surrounding micro- and macroparametric syntax (see also Kayne 2005: 4-5). It should certainly be discussed and debated under one label or another.

2.2 *The extent of variation*

As foreshadowed in the last paragraph, the second macro- versus micro- issue to distinguish is what can be called the “extent of variation” question. We know that English is very different from Japanese, and both are very different from Mohawk. The question is whether these large-scale differences are always the cumulative effect of there being many smaller-scale differences between the languages or not.

The standard microparametric view is that the primitive, scientifically significant differences among languages are always relatively small scale differences, typically tied to (at most) a few closely related constructions. All of these parameters are the kinds of things that even closely related dialects might differ in (Kayne 2005: 10). Large differences between languages always reduce to lots of these small differences. In this respect, the microparametric view is somewhat similar to standard Neo-Darwinian evolution, where the large differences between species are the result of there having been many small mutations from an original common ancestor.

In contrast, the macroparametric view is that there are at least a few simple (not composite) parameters that define typologically distinct sorts of languages. For example, there might be a single parameter in the statement of Merge that induces the core difference between head-initial and head-final languages (Stowell, 1981). There might be a single parameter that lays down the core structure of a nonconfigurational polysynthetic language as opposed to more configurational, isolating languages (Baker, 1996). And so on. The biological analogy for the macroparametric view could be Stephen Jay Gould's notion of punctuated equilibrium, where transitions between species types need not be smooth, but can go rather suddenly from one stable cluster of properties to another quite different cluster.

I agree with Kayne (2005: 5-6) that the extent of variation issue by itself is hard to pin down, and is ultimately not so important for the science of linguistics. First, what counts as a big and impressive difference as opposed to a small localized difference is a judgment call, impossible to make precise. Even more importantly, there is no reason to think that the causes of relatively big, impressive differences among languages are any different in kind from the causes of smaller, easily overlooked differences. Presumably whether a parameter (either lexical or grammatical) has large-scale effects on language has more to do with accidents of frequency and with how different factors happen to interact than with the inherent nature of the parameters themselves (see also Baker (2001: 125-26)). To the scientist, the interesting question is not usually how common something is, but rather how revealing it is about the essence of things.

But I do not think that the extent of variation question is *completely* irrelevant in practice. While there are no guarantees, it is true that lexical parameters consistent with the Borer-Chomsky Conjecture will *tend* to have smaller scale effects than grammatical parameters. If a parameter is essentially a particular assignment of features to a lexical item, it will only affect structures that contain that lexical item. A parameter associated with a very common lexical item (e.g., the finite indicative tense maker) could have a large effect on the observed language. But parameters associated with items of average frequency or less will be much more limited in scope. In contrast, the grammatical parameters envisioned in (2) are not intrinsically limited in this way. They might pertain to all headed phrases, or to all theta-role assignment relationships, or to all functional heads of a certain type, and hence transcend a small number of constructions. Large-impact parameters are not inherently interesting just because they are large-impact. But they might be heuristically significant because they tend to point to loci of variation in the grammar as opposed to the lexicon.

2.3 *The methodology of comparison*

The third micro- versus macro- issue to discuss is the question of what is the most effective methodology for discovering the true parametric structure of language.

The microparametric approach highlights the benefits of comparing very closely related languages for finding genuine parameters (Kayne 2005: 8-10). Kayne (p. 10) conjectures that it may turn out that for any parameter that distinguishes (say) Mohawk from English, there could be two closely related dialects that differ in the setting of that parameter. Indeed, Kayne emphasizes that comparison of closely related languages is likely to be the easiest and most reliable method for finding parameters. The shared history of such languages/dialects means that most parameter values other than the one under study are shared, minimizing the chances that other differences conceal or contaminate the effects of the parameter in question. Kayne compares the value of this methodology to that of designing controlled experiments in physical sciences, where most factors can be held constant and only one is varied at a time.

In contrast, macrocomparative syntax emphasizes that comparing historically unrelated languages may be necessary to discover certain kinds of parameters. By hypothesis, these parameters can have a large impact on the shape of a language. Hence, mutually intelligible dialects will tend not to differ in these respects; if they did, they would no longer be mutually intelligible. Presumably the kind of interaction that one village tends to have with another normally prevents macroparametric differences from creeping into one dialect and not the other, unless some radical and sustained isolation takes place. Comparing dialects is thus typically not a good way to find macroparameters, if there are any.

There is no doubt that the microcomparative method has been productive and should be pursued. But it should be pointed out that there is no guarantee that it will reveal the true parametric structure of language. Kayne's analogy with a controlled experiment assumes that two closely related and similar-looking languages will have almost the same parameter settings. But this is not necessarily the case. The shared history of (for example) Fiorentino and Trentino and their on-going interaction guarantees that the two *E*-languages must be similar—that they have roughly the same strings of grammatical words. It does not, however, guarantee that the *I*-languages (internalized grammars; languages seen as a set of rules and principles) are similar. It is possible that two dialects could differ in a macroparameter, but the differences are largely cancelled out on the surface by a series of microparametric choices that the languages make. In that case, the languages would look similar on the surface, and even be mutually intelligible, but have very different parametric structures. Paradoxically, comparing unrelated languages might show the effects of a parameter difference more clearly, because there is no constraint of mutual intelligibility that might tend to spawn microparameters that cancel out the effects of a

macroparametric difference. I don't know how serious a problem this possibility presents in practice, but it is worth bearing in mind that there is no guarantee that languages with similar-looking grammaticality patterns always have similar grammars. No naturally occurring linguistic situation provides anything that is too much like a controlled experiment.

A more serious concern is that the microcomparative method, if pursued exclusively, presupposes the truth of the Borer-Chomsky Conjecture. It is a methodology designed to find parameters under the explicit assumption that those parameters are small and lexical in nature. If there are macroparameters at all, it is hardly surprising that pursuing this methodology does not reveal them. If we want to get empirical evidence that bears on the truth of the Borer-Chomsky Conjecture one way or another, one needs to pursue a macrocomparative method in conjunction with microparametric inquiries.

Is the macrocomparative method too hard, or too uncontrolled, to be feasible and reliable? My personal experience suggests that the answer in practice is no, at least if macrocomparison is pursued in a certain way. I do not find it much harder to compare typologically similar but historically unrelated languages like Mohawk and Mayali, or English and Edo, or Swahili and Spanish, than it is to compare closely related languages. This is because in point of fact there are probably relatively few macroparameters, and most of syntax is indeed universal. If there were many macroparameters and they interacted with one another in complex ways, then languages could differ crazily in ways that would be hard to pull apart. But if there are relatively few macroparameters, and if they are mostly semi-independent from each other, then this is not such a danger in practice. Perhaps it is hard to start with two arbitrarily chosen languages (e.g. English and Japanese (Kayne 2005: 9)) and work out the parameters that distinguish them from scratch. But there is no need to do so; one can start with clusters of unrelated languages that are similar to one another along some dimensions and are interesting to compare along other dimensions. I feel that people often overestimate how hard it is to do macrocomparative syntactic research—and underestimate how much fun it can be.

3. What evidence for macroparameters should we look for?

We know that microparameters exist and that there are many of them—probably very many of them, as Kayne has shown. Suppose that macroparameters also exist. What evidence would we expect to find for them, given the presence of so many microparameters? And given the all-too-little that we know now, is there any reason to believe that we have some such evidence?

To make the discussion slightly more concrete, I make reference to the Head Directionality Parameter and the Polysynthesis Parameter as two possible macroparameters,

without going into any detail about how they are stated or what their effects are (see Baker 1996 for much discussion of the second). Hopefully some useful general points can be made without getting bogged down in the specifics.

Suppose that the classical vision of parameters had been exactly right, and there were only macroparameters of the kind envisioned in (2). Then for any binary-valued parameter one would expect to find two sharply different parametric clusters, with essentially no intermediate cases. We should observe perfect head-initial languages and perfect head-final languages and nothing else. Similarly, there would be purely polysynthetic languages and purely nonpolysynthetic languages, period. It should always be easy to find out the parameter settings for a particular language, simply by checking the properties associated with the parametric cluster.

We now know beyond any reasonable doubt that this is not the true situation. For any possible parameter, there is typically noise in the data as well as mixed, transitional cases. Typologists report the impression that certain language types might be more or less common for various reasons, but no combination of surface properties is truly impossible (e.g. Bickel in press, Nichols in press). Nor should we be at all surprised by this, now that we know about the ubiquity of microparameters as a result of Kaynian microcomparative research on Indo-European varieties. The character of the typological record falsifies the view that all syntactic variation is macroparametric variation. But no one has that view now anyway.

Next suppose that there were only microparameters, that the Borer-Chomsky Conjecture is true. It is not at all hard to imagine that all putative macroparameters can be reduced to a set of microparameters. For example, instead of stating that heads follow their complements in language X (the Head Directionality Parameter), one could simply state the adpositions follow their complements, that verbs follow their complements, that complementizers follow their complements, and so on for each category (or each item of each category?!?). Kayne (2005: 7) also mentions this possibility explicitly for the Polysynthesis Parameter, pointing out that the nonconfigurational properties of polysynthetic languages could be the result of these languages happening to have all of the clitic doubling/dislocation constructions found in Romances. He writes: "...It could alternatively be the case that the systematic obligatoriness of pronominal agreement morphemes in Mohawk is just an extreme example of what is found to a lesser extent in (some) Romance." I believe that he is absolutely correct about this: the difference between Mohawk and Romance can insightfully be described as Mohawk having more dislocation structures than Romance languages, and having them more systematically.

If it is correct to reduce all macroparameters to a series of relatively independent microparameters in this way, then one would expect to find a relatively smooth continuum of languages. Along the polysynthesis dimension, languages would range from those that have few or no dislocation constructions, to those that have several, to those that have many, to those that have a complete set. Similarly, one could have a wide variety of languages when it came to head-complement order, with pure head-final languages like Japanese and pure head initial languages like English simply being the special cases in which all of the categories happen to be set for the same complement-head order. Many linguists think that this is the case, both within the generative microcomparative syntax tradition and within other sorts of linguistic traditions (construction grammarians, typologists).

Finally, let us imagine what we might expect to observe if natural human language permits both macroparameters and microparameters. We would expect there to be parametric clusters in something like the classical way. But we'd expect these clusters to be much noisier, as a result of microparametric variation. For example, a generally isolating language might have one or two constructions that are (or look like) those found in a polysynthetic language as the result of some kind of microparametric property. Similarly, a generally head initial language might have one or two head-final projections, perhaps as a result of the complement moving to specifier position for some reason. These microparameters could partly conceal the effects of the macroparameter. The macroparameter might then function in practice almost like a default, regulating whatever cases the more specific macroparameters do not determine. On this view, too, there would be intermediate cases as well as the pure types.

Given that both a theory with macroparameters and a theory without them expect a variety of intermediate cases as well as "pure" types, how can we hope to distinguish which theory is correct? We cannot tell by looking at a single language, or at a pair of languages, given that a macroparametric value can always be recast as a series of microparametric values. The truth will have to come from looking at statistical distributions of properties drawn from a significant number of languages. All things being equal, if there are only microparameters and they are logically independent of each other, one would expect a single normal distribution (i.e., a bell curve) of properties.² In particular, there should be many mixed languages of different kinds, and relatively few pure languages of one kind or another. Suppose there are 10 different phrase types, each of which is set for head-initial expression or head-final expression (and similarly for polysynthetic expression, through agreement or incorporation). Simple combinatorics says that we should find lots of languages with four or five phrase types with one order and five or six phrase types with the other order (or about half with isolating expression and

half with polysynthetic expression); indeed, one would about expect 65% of the total number of languages to have that character. In contrast, only about 2% of the languages should be pure or almost pure languages of either kind.

This contrasts with the statistical expectations generated by the view that includes macroparameters as well as microparameters. On this view, one would expect a bimodal (or multimodal) distribution, with one peak at or near the cluster of properties associated with a pure head-final language, and another peak at or near the cluster of properties associated with a pure head-initial language. Similarly, one might expect a peak in the distribution at the pure polysynthetic languages. Microparameters then cause a degree of “noise” around these peaks, creating something like a normal distribution around each one. One could find intermediate cases that are halfway between the pure head-initial type and the pure head-final type. But crucially one expects relatively few of these intermediate types, because it takes a whole series of microparametric choices all set in a certain way to override or disguise the effects of a single macroparametric choice. The essential difference, then, is that the strict microparametric view predicts that there will be many more languages that look like roughly equal mixtures of two properties than pure languages, whereas the macroparametric-plus-microparametric approach predicts that there will be more languages that look like pure or almost pure instances of the extreme types, and fewer that are roughly equal mixtures.

Which of these predictions fits better with our current state of knowledge? I am the first to admit that we don’t know as much about this as we should. But what I know looks much more like the macroparametric vision. Consider first the Head Directionality Parameter, as applied for simplicity only to verb phrases and adpositional phrases. The unalloyed microparametric vision, where direction of V and NP and direction of P and NP are set independently, should expect roughly equal numbers of consistent languages (V-NP and P-NP order, or NP-V and NP-P order) and inconsistent languages. The macroparametric vision, where VP and PP order are set by the same parameter, but some microparametric “noise” is also tolerated, expects more consistent languages than inconsistent languages. According to *The World Atlas of Language Structures (WALS)* (Dryer, 2005), the latest figures are as follows:

- (3) a. V – O and P – NP: 417 languages (consistent languages)
- b. O – V and NP – P: 427 languages
- c. V – O and NP – P: 38 languages (inconsistent languages)
- d. O – V and P – NP: 10 languages

These numbers falsify a simplistic and inflexible macroparametric view, with a single head directionality parameter and no microparametric variation. But they look much more like the

bimodal distribution that the macroparametric view expects than like the normal distribution that a pure microparametric view would expect.³ My impression is that this holds for word order parameters more generally: there are many more almost pure head-initial and head-final languages than there are languages with a nearly equal mix of the two orders.⁴ If so, macroparameters should be taken seriously.

In the last chapter of Baker (1996), I tried to develop the beginnings of a similar argument for the Polysynthesis Parameter. There I compared seven languages that have syntactic noun incorporation with six languages of similar morphological complexity that do not have that type of noun incorporation over some fourteen relevant grammatical properties. What I found was that all of the languages with noun incorporation had the same values for nearly all of the other thirteen properties. In contrast, every language that did not have noun incorporation also deviated from my polysynthetic ideal in at least three other properties. This was intended as the first step toward a statistical demonstration that there are in fact more pure polysynthetic languages than there are near-polysynthetic languages (in my technical use of the term polysynthesis). If correct, this bears witness to the unity of the Polysynthesis Parameter. Kayne is right that a language like Mohawk can be modeled by a series of microparameters involving agreement and dislocation structures of various kinds (plus incorporation). The fact that Mohawk has all those properties is not the real argument for the Polysynthesis Parameter: the real argument is that there seem to be more “pure polysynthetic” languages like Mohawk than the sorts of “near polysynthetic” languages that one might expect if the Polysynthesis Parameter really reduces to a collection of microparameters. I’m not sure I am right about this; my discussion falls well short of a statistical proof, and there are many complexities in this understudied class of languages. But the true form of the argument should be emphasized, because it is the form of argument that has a chance of deciding the crucial theoretical question of the truth of the Borer-Chomsky Conjecture on empirical grounds—one way or another.

Going a little further, I also think that the little evidence we have does tend to support the macroparametric view of at least head directionality and polysynthesis, rather than the contrary. We should not rule out syntactic parameters by theoretical fiat or methodological convenience, but should actively encourage the macrocomparative work that can bear on these questions.⁵

4. Illustrating with two new macroparameters

There is much more that could be about the macroparametric view of head directionality and polysynthesis. But I have little more to say about it that is genuinely new.⁶ Instead, I devote the rest of this chapter to sketching two new parameters that seem to be macroparameters in the sense

of (2), not consistent with the Borer-Chomsky Conjecture. One embarrassing fact for my perspective is that not very many new macroparameters have been proposed in recent years. By showing how new parameters can come to light, I want to encourage the idea that this is a side-effect of the kind of research that linguists are doing, not of the paucity of macroparameters in the true theory of natural language. The example also illustrates why a methodology of purely microcomparative research risks missing what is really going on.

The agreement parameters that I have in mind began emerged as a detailed comparison between Niger-Congo languages and Indo-European languages, partly in collaboration with Chris Collins (see Collins (2004) for his first version of the relevant parametric variation). They can be stated simply as follows (see Baker (to appear) for much fuller discussion):

(4) A functional head F agrees with NP only if NP asymmetrically c-commands F.

(Yes: Niger Congo languages; No: Indo-European languages)

(5) A head F agrees with NP only if F values the Case feature of NP or vice versa.

(No: Niger Congo languages; Yes: Indo-European languages)

These are two distinct syntactic conditions restricting when a functional head F can agree with an NP in its environment in person, number, and gender features. They are to be understood as the variable parts of a general theory of agreement that has universal conditions as well, along the lines of those studied in Chomsky's (2000) theory of Agree. The core idea is that Niger-Congo languages agreement obeys a certain structural configuration, where the agreed with NP must be higher in clause structure than the agreeing head. This requirement does not hold in Indo-European languages; instead, IE agreement is subject to the condition that the two elements match in case. The most familiar instance of this is a finite T agreeing only with the NP that it assigns nominative case to in IE languages, but (5) generalizes this to allow for (say) determiners, adjectives, and participles agreeing with NPs in phi features as a result of their agreeing with the NP in case (case concord). For purposes of presentation, I illustrate (4) and (5) by contrasting the Bantu language Kinande with various IE languages, to show a wide range of effects in a single language.⁷

The first and most obvious domain in which these parameters apply is “subject” agreement—more precisely agreement on the finite Tense node. Both NC languages and IE languages show agreement between the finite verb and the preverbal subject in simple clauses. But a clear difference appears when something other than the thematic subject moves to Spec, TP. Whenever this is possible, the Bantu verb agrees with the phrase that has moved to Spec, TP, whereas IE languages typically do not. Rather, the IE verb agrees with the nominative NP (the thematic subject) regardless of where it appears in the structure. (6) shows examples where the

thematic object moves to Spec, TP, past the thematic subject. In Kinande, the finite verb agrees with the fronted object; in Yiddish it agrees with the postverbal subject (Diesing, 1990):

(6) a. Olukwi si-lu-li-seny-a bakali (omo-mbasa).
 wood.11 NEG-11S-PRES-chop-FV women.2 LOC.18-axe.9
 ‘WOMEN do not chop wood (with an axe).’

b. ...az vayn ken men makhn fun troybn oykh
 that wine can one make from grapes also
 ‘(You should know)...that one can make wine from grapes also.’

(7) shows a similar contrast with locative inversion, when a locative phrase is moved to Spec, TP in a passive or unaccusative sentence rather than the thematic object. In Bantu languages, the finite verb agrees with the fronted locative; in IE languages like English it agrees with the thematic object in nominative case:

(7) Oko-mesa kw-a-hir-aw-a ehilanga. (Baker, 2003)
 LOC.17-table 17S-T-put-PASS-FV peanuts.19
 ‘On the table were put peanuts.’

(8) shows examples in which nothing is moved into Spec, TP; rather this position is either left empty or is filled with an expletive that does not have phi-features of its own. In English, the finite verb still agrees with the postverbal subject. In Bantu it does not; rather the subject agreement slot is filled with a locative prefix, which can be analyzed either as a default or (more likely) as agreement with a null expletive similar to *there*.

(8) a. Mo-ha-teta-sat-a mukali (omo-soko).
 AFF-there-NEG/PAST-dance-FV woman.1 LOC.18-market
 ‘No woman danced in the market.’

b. There arrive three new delegates each day.

Overall, then, Tense in NC languages is systematically different in terms of what it agrees with from Tense in IE languages, in a way that is correctly described by the parameters in (4) and (5). This much is reasonably well known. The next step, then, is to show the same difference applies to all of the other agreeing functional heads in these languages.

Consider next the possibility of agreement on complementizer heads. This sort of agreement is not very common in either NC or IE languages, but when it does occur there is a difference. (9a) shows a Kinande example in which the complementizer *ti* agrees with the subject of the matrix verb. In contrast, (9b) shows an example from West Flemish, where a similar complementizer agrees with the embedded subject, not the matrix subject (Carstens, 2003: 393).

(9) a. Mo-ba-nyi-bw-ire ba-ti Kambale mo-a-gul-ire eritunda.

AFF-2S-1sO-tell-EXT 2-that Kambale AFF-1S-buy-EXT fruit
 ‘They told me that Kambale bought fruit.’

- b. Kvinden *dan* die boeken te diere zyn.
 I-find that-PL the books too expensive are
 ‘I find that those books are too expensive.’

The complementizer in (9b) is not c-commanded by the subject that it agrees with, showing that (4) does not hold in West Flemish. However, the C does arguably participate with Tense in assigning nominative case to the embedded subject (Watanabe, 1993), and it is this element that it agrees with, consistent with (5). In contrast, I assume that the Kinande complementizer in (9a) really agrees with a logophoric operator generated in Spec, CP and controlled by the matrix subject (cf. Koopman and Sportiche (1989)). This operator does c-command C, but is not in a case valuation relationship with it, showing that (4) holds but (5) does not in Bantu. Kinande Cs also agree with overt NPs that appear in their specifiers as a result of focus movement, as shown in (10).

- (10) Eritunda **ry-o** n-a-h-a omukali
 fruit.5 5-FOC 1sS-T-give-FV woman.1
 ‘It’s a fruit that I gave to a woman.’

This sort of agreement is also not found in IE languages, because C does not assign case to an A-bar moved NP in Spec, CP. Note that Kinande’s C is like Kinande’s T in only agreeing with an NP (overt or not) that appears on its left; C in IE languages is like T in IE languages in being able to agree with an NP to its right, as long as there is a case relationship.

Consider next the possibility of determiner-like heads agreeing with the associated NP. In Kinande, elements like *osi* ‘all’ agree in number and gender with the NP they form a constituent with, and that NP appears to the left of *osi*, as in (11).⁸

- (11) aba-kali b-osi; ama-tunda w-osi
 2-woman 2-all 6-fruit 6-all
 ‘all the women’ ‘all the fruits’

This contrasts in two respects with another determiner-like element that has a similar meaning, *obuli* ‘every’. *Obuli* comes before the associated NP, and it is an invariant form that does not agree with that NP in number or gender, as shown in (12).

- (12) obuli mu-kali; obuli ri-tunda
 every 1-woman every 5-fruit
 ‘every woman’ ‘every fruit’

Once again, agreement is highly dependent on linear order, and ultimately on syntactic structure in Bantu. I assume that ‘woman’ is the complement of ‘every’ in (12), and hence does not asymmetrically c-command it; agreement on ‘every’ is thus not licensed by (4). In contrast, ‘women’ is the specifier of ‘all’ in (11), so agreement is legitimate here. Unlike Bantu languages Ds in IE languages can perfectly well agree with their complements, as shown by English examples like *that child* versus *those children*. This shows that (4) does not hold for these languages. Ds must however agree with their NP complements in case in IE languages, as seen in more highly inflected languages like German and Greek; this I take to be a consequence of (5).

Similar considerations apply to the category P, although this category participates in agreement less often than D. In Kinande, P cannot agree with an NP that remains as its complement, as shown in (13a). P can, however, agree with the NP if that NP has moved to the left of (above) P—presumably through Spec, PP—to land in Spec, TP, as shown in (13b) (see Kinyalolo (1991) for the original discussion of this pattern in Kilega).

- (13) a. Kambale a-ka-kanay-a na-(*bo) abanyakulu.
 Kambale 1S-PRES-speak-FV with-2 2.old.people
 ‘Kambale is speaking with the old people.’
- b. Abanyakulu si-ba-li-kan-ibaw-a na-bo.
 2.old.people NEG-2S-PRES-speak-PASS-FV with-2
 ‘Old people are not spoken with.’

Like the other heads we have seen, P can only agree with something to its left, something that c-commands it in Bantu. In contrast, some IE languages allow P to agree with an unmoved DP that remains to its right, such as the Welsh example in (14) (Harlow, 1981: 220).

- (14) Soniais i amdan-o ef.
 talked I about-3sM him
 ‘I talked about him.’

This is possible because (4) doesn’t apply to IE languages, whereas (5) is satisfied since P values the case of its complement. (In Welsh, the P shows agreement only if its complement is a pronoun, but this restriction does not hold in other typologically similar languages.)

Very much the same holds true of “object agreement” realized on the v head associated with transitive verbs. In many Bantu languages, including Kinande, such agreement is ruled out when the object remains in situ, inside the verb phrase, as shown in (15a). Agreement shows up with an overt object only when it has been dislocated to the periphery of the clause, as in (15b).⁹

- (15) a. N-a-(*ri)-gul-a eritunda. (Baker, 2003)
 1sS-T-OM5-buy-FV fruit.5

‘I bought a fruit.’

b. Eritunda, n-a-ri-gul-a.

fruit.5 1sS-T-OM5-buy-FV

‘The fruit, I bought it.’

In contrast, object agreement with an unmoved object should be possible in an IE type-language, as long as the object receives accusative case from *v* (i.e., as long as it did not have dative or some other oblique case). In fact, no IE language has true agreement (as opposed to clitics) on *v* that I know of, but other languages with the same parameter settings do, such as certain languages of Australia and New Guinea (see Baker to appear).

The sixth and final category that I discuss here is the so-called linking particle that appears inside the verb phrase in Kinande, analyzed in detail in Baker and Collins (in press). This particle appears between the two objects of a double object construction, as shown in (16). Baker and Collins argue that either the theme object or the goal object must move to the specifier of this functional category in order to undergo case licensing with the transitive verb. The choice of which NP to move gives rise to the two grammatical word orders shown in (16).

(16) a. Kambale a-seng-er-a omwami y’- ehilanga.
Kambale 1S/T-pack-APPL-FV chief.1 LK.1 peanuts.19

‘Kambale packed peanuts for the chief.’

b. Kambale a-seng-er-a ehilanga hy’- omwami.
Kambale 1S/T-pack-APPL-FV peanuts.19 LK.19 chief.1

‘Kambale packed peanuts for the chief.’

The crucial observation for our purposes is that the linking particle necessarily agrees with whichever object moves to its Spec; it cannot agree with the object that remains to its right, inside the verb phrase. This is yet another testimony to the validity of (4) in Bantu languages.

Conversely, if Baker and Collins (in press) are right about the case-theoretic properties of (16), (5) implies that IE languages could not have a linking particle with the same agreement properties, because the particle does not assign Case to (or agree in case with) the NP in its specifier.

What is the moral of all this for the theory of parameters? I have presented this material in such a way as to emphasize that there is a systematic difference in how agreement works in NC languages as opposed to how it works in IE languages. The difference is not a difference in how one particular lexical item agrees, nor even in how a whole category of lexical items agree (say, all the Ps). Rather, it extends to the full range of functional categories, including at least T, *v*, P,

D, C, and Lk. This looks like a sweeping generalization about languages as wholes, which would be lost if one looked at these facts only through the lens of the Borer-Chomsky Conjecture.

Of course, determined microparametric syntacticians need not give up yet. They can claim that the behavior of Kinande can be captured as a series of microparameters that say how each individual functional category works with respect to agreement. Kinande is then a “pure” language in which all the functional heads happen to work in one way, and IE languages are “pure” languages in which they all work in the other way.

The ultimate test, then, would be the statistical one. The microparametric view should (all things being equal) expect that there will be many mixed languages, in which roughly half the functional heads show the IE behavior and the other half show the NC behavior. In contrast, the view that there are macroparameters at work expects to find many languages of the Kinande type, many languages of the IE type, and only a few cases that are intermediate or hard to classify.

In order to evaluate this, I have a project underway to test the parameters in (4) and (5) against the 100 languages in the core sample of *The World Atlas of Language Structures* (plus or minus a few). The macroparametric prediction is that there should be roughly four kinds of languages: the NC kind, the IE kind, plus perhaps a kind in which both (4) and (5) hold, and a fourth kind in which neither holds. I do not expect to observe a perfect parametric cluster of all six categories for every language, because of the existence of microparametric variation. The reader can already see how this inevitably comes about. For example, not every language has words of category C or words of category P, and only a handful have (overt) particles like Kinande’s linking particle. Even when a category exists, there is no guarantee that it will undergo agreement. The parameters in (4) and (5) say nothing about whether a given category agrees or not; they only determine which NP a category agrees with if it participates in agreement at all. I mentioned that C and P are not agreement hosts in most IE languages or in most NC languages, even though they behave as expected in those few languages in which they do agree. Indeed, *v* never seems to be an agreement bearer in IE languages, for reasons I take to be accidental. Finally, a language might have an agreeing T, but still lack the possibility of moving something other than the thematic subject to Spec, TP the way Kinande and Yiddish can, for other (microparametric?) reasons. In such a language, it might be difficult to tell whether (4) or (5) holds for T, because every parameter setting gives the same observable effects over the range of structures the language permits. Therefore, we can only expect to get partial testimony as to the values in (4) and (5) in most languages. Nevertheless, filtering out this “noise” out as carefully as possible, the macroparametric prediction is that one should find converging evidence for uniform settings of (4) and (5) from different functional categories in most languages.

Having checked about 66 languages out of the 100 language sample so far, I am optimistic that the macroparametric prediction will be borne out. All the Niger-Congo languages I have looked at that have rich agreement show evidence of obeying (4) and not (5). This extends even to non-Bantu languages like Lokaa and Ibibio. Languages of this type from other families include the New Guinean language Arapesh, the South American language Tariana, Berber, Austronesian languages like Fijian and Tukang Besi, and maybe Mapudungun. On the other hand, as far as I know every IE language obeys (5) but not (4), from Icelandic to Hindi. This is also true for Basque, Finnish, Abkhaz, most Australian languages, many Afro-Asiatic languages, many New Guinean languages, and those Dravidian and Khoisan languages that have agreement. I have several promising cases of languages that seem to obey neither (4) nor (5), including Georgian, Burushaski, and Warlpiri. I am least confident of finding languages that obey both (4) and (5), but one gets some intriguing results if one analyzes Turkish, Lango, and perhaps Mohawk in this way. So far I have found only one language (Kewa) where I have not been able to find converging evidence from different functional categories for one parameter setting or another. This is largely for the rather dull reason that T seems to be the only agreement bearer in that language. I will report fuller results in more detail in Baker (to appear); for now, my point is that this sort of research is perfectly feasible and bears on the question of the existence of macroparameters in the sense (2) as opposed to (1) in a clear way.

5. Final Discussion

The case study in section 4 also has implications for the other micro- versus macro- issues identified in section 2. Are (4) and (5) large-scale parameters or not? I'm not sure what to say. They seem to be intermediate cases with respect to the "extent of variation" question, having more impact on the language as a whole than the average microparameter, but not as much as the Head Directionality Parameter or the Polysynthesis Parameter. In languages that have relatively few functional heads, or where few of them are agreeers (like Kewa), they seem like microparameters. This illustrates nicely Kayne's point that the extent-of-variation question is not well defined or theoretically very interesting. The status of the parameters (4) and (5)—or the series of microparameters that could conceivably replace them—is exactly the same whether they happen to have a large and obvious effect or only a subtle and narrow effect.

Finally, and perhaps most importantly, this case study shows why it is not sufficient for the field to do just microcomparative syntax. The parameters in (4) and (5) could probably never have been discovered simply by comparing dialects of Indo-European languages. The reason is simply that (as far as I know) all Indo-European languages seem to have the same values for these

parameters. So do all the Niger Congo languages, and the Australian languages. Indeed, these parameters seem to be very conservative, not varying much within historically related language families. The only fairly clear case of languages in the same family having different values for these parameters that I have found so far is in the Afro-Asiatic family, where Berber has the NC values, but Hausa and Arabic have the IE values. We know that dialects can have different values for microparameters, but we don't know that they can have different values for macroparameters; it would not be surprising if they could not. This underscores the point that pursuing only a microcomparative approach risks presupposes the truth of the Borer-Chomsky Conjecture, perhaps wrongly.

References

- Baker, Mark. 1996. *The polysynthesis parameter*. New York: Oxford University Press.
- Baker, Mark. 2001. *The atoms of language*. New York: Basic Books.
- Baker, Mark. 2003. Agreement, dislocation, and partial configurationality. In *Formal approaches to function in grammar*, eds. Andrew Carnie, Heidi Harley and Mary Ann Willie, 107-134. Amsterdam: John Benjamins.
- Baker, Mark. in press. On zero-agreement and polysynthesis. In *Agreement and argument structure*, ed. Peter Ackema.
- Baker, Mark, and Collins, Chris. in press. Linkers and the internal structure of vP. *Natural Language and Linguistic Theory*.
- Baker, Mark. to appear. *The syntax of agreement and concord*. Cambridge: Cambridge University Press.
- Borer, Hagit. 1984. *Parametric syntax: case studies in Semitic and Romance languages*. Dordrecht: Foris.
- Carstens, Vicki. 2003. Rethinking complementizer agreement: agree with a case-checked goal. *Linguistic Inquiry* 34:393-412.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam. 2000. Derivation by phase, Linguistics, MIT: unpublished ms.
- Collins, Chris. 2004. The agreement parameter. In *Triggers*, eds. Anne Breitbarth and Hank van Riemsdijk, 115-136. Berlin: Mouton de Gruyter.
- Diesing, Molly. 1990. Verb movement and the subject position in Yiddish. *Natural Language and Linguistic Theory* 8:41-80.
- Dryer, Matthew. 2005. Relationship between the order of object and verb and the order of adposition and noun phrase. In *The world atlas of language structures*, eds. Martin

- Haspelmath, Matthew Dryer, David Gil and Bernard Comrie, 386-389. New York: Oxford University Press.
- Fukui, Naoki, and Speas, Margaret. 1986. Specifiers and projections. *MIT Working Papers in Linguistics* 8:128-172.
- Gilligan, Gary. 1987. A Crosslinguistic approach to the pro-drop parameter., University of Southern California: Ph.D. dissertation.
- Harlow, Stephen. 1981. Government and Relativization in Celtic. In *Binding and Filtering.*, ed. Frank Heny, 213-254. Cambridge: MIT Press.
- Kayne, Richard. 2000. *Parameters and Universals*. New York: Oxford University Press.
- Kayne, Richard. 2005. Some notes on comparative syntax, with special reference to English and French. In *The Oxford Handbook of Comparative Syntax*, eds. Guglielmo Cinque and Richard Kayne, 3-69. New York: Oxford University Press.
- Kinyalolo, Kasangati. 1991. Syntactic dependencies and the SPEC-head agreement hypothesis in KiLega, UCLA: unpublished Ph.D. dissertation.
- Koopman, Hilda, and Sportiche, Dominique. 1989. Pronouns, logical variables, and logophoricity in Abe. *Linguistic Inquiry* 20:555-589.
- Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.
- Stowell, Timothy. 1981. Origins of phrase structure, MIT: Ph.D. dissertation.
- Watanabe, Akio. 1993. *Case Absorption*. ms.: University of Tokyo.
- Webelhuth, Gert. 1992. *Principles and parameters of syntactic saturation: Oxford Studies in Comparative Syntax*. New York: Oxford University Press.

Notes

* For comments on earlier versions of this work, I thank Richard Kayne, Chris Collins, and the audience at the workshop on Parametric Variation held at the University of Newcastle in September 2005. Any errors are solely my responsibility.

¹ There is interesting view that is intermediate between (1) and (2), but has not been very clearly distinguished from them in the literature. This is the view that languages can differ in the properties that large classes of lexical items have. For example, one language might have the general property that all Tense nodes trigger verb raising to Tense, while another language has the property that no Tense node does (e.g., French versus English). This view is like the Borer-Chomsky Conjecture in that the variation is attributed to lexical items, but it sees the lexical items as classes. It claims that an important generalization is missed by simply saying that each

member of the class happens to have the same property. This view is thus akin to the macroparametric view in that it can make statements about a language as a whole, not just about individual lexical items in a language. I would not be surprised if most (all?) of the macroparameters that I envision could be cast in this way. It would be worthwhile to articulate this intermediate view more fully, and consider its advantages and disadvantages.

A number of variants of (1) have been explored in the literature, with somewhat different advantages and disadvantages. Space does not permit me to explore these differences here.

² There are two ways that a microparametric syntactician might consider denying that the settings of a series of microparameters are independent of each other. One way is to say that the setting of one microparameter is formally linked somehow to the settings of other microparameters. For example, one might assume that all head directionality microparameters are initially set the same way, and their settings are changed only if explicit evidence is encountered. I take this to be a significant departure from the Borer-Chomsky Conjecture, and to be nondistinct from one implementation of a macroparametric theory. The second way would be to say that all the microparameters are formally independent of one another, but do not behave like they are statistically independent in practice because of functional pressures of one sort or another (e.g., ease of parsing, effects of language change, etc.). This is a perfectly coherent view, and is widely held among functionalist typologists. How this functionalist view can be distinguished empirically from the macroparameters-plus-microparameters view is a topic for another occasion.

³ Kayne himself has a very different approach to word order variation from the classical one that frames my discussion here; see Kayne (2000: Ch. 15) and Kayne (2005: 49-52) for sketches of how he foresees explaining Dryer's results in his terms. I believe that the same macro- versus microparametric issues can be raised within Kayne's theory of word order. (For example, one can ask whether or not it is the very same lexical item "W" that is generated above both PPs and CPs to trigger V-P-NP or V-C-IP word order as opposed to NP-P-V or IP-C-V word order, and whether all instances of "W" have uniform movement-triggering power in each language.) But showing this in detail would take me far a field.

⁴ It is actually quite difficult to confirm this using *WALS*, for reasons that are somewhat instructive. Other tables in *WALS* display things like determiner-noun order and degree word-adjective order. These do not correlate very well at all with verb-object order or adposition-NP order. But we know that demonstratives and degree words are not uniform categories across languages. It is clear that in some languages demonstratives are a subclass of determiner heads, whereas in other languages they have the syntax of adjectives. Demonstrative determiners should

be subject to the Head Directionality Parameter, but demonstrative adjectives (as adjoined elements) should not be. To see if demonstrative determiner-noun order correlates with verb-object order, one would first have to isolate the demonstrative determiners from the demonstrative adjectives—a labor-intensive job that the *WALS* authors have not attempted. Similarly, we know even in English and Dutch that there are two kinds of degree words, true degree heads and adverbs. These would have different word orders in (at least) head final languages, so they need to be distinguished from each other, and again *WALS* has not done the hard work necessary to do this. It is thus not surprising that few crisp correlations come to light in that ambitious work and similar projects. We know from chemistry that one must purify one's ingredients before one can hope to get clear and decisive chemical results; the same figures to be true for syntax.

⁵ I want to record here that Kayne himself, although a vigorous and persuasive proponent of microcomparative syntax, has never actively discouraged macrocomparative syntax work. On the contrary, he actively encouraged my early work on Mohawk, and Kayne (2005: 8) affirms that macrocomparative syntax is also essential. If microparametric work has flourished and macroparametric work has not, it is partly because Kayne has been more effective at inspiring people to attempt his kind of research than I (and people like me) have been—an imbalance that I have fantasies of changing to some degree by writing this chapter.

⁶ See Baker (in press) for some discussion of the attempt to extend the Polysynthesis Parameter to Mapudungun, another language with noun incorporation. My first impression was that this language did deviate from the polysynthetic ideal in certain serious ways that could undermine the macroparametric view of polysynthesis. More detailed work showed that there were more similarities between Mapudungun and (say) Mohawk than were apparent on a casual reading of the Mapudungun grammar, and Mapudungun data ultimately support the core of the Polysynthesis Parameter (with one or two minor corrections).

⁷ The Kinande examples in this section all come from work I have done with Philip Mutaka, who I thank for his generous help. Most of the phenomena mentioned have received insightful discussion in the previous Bantu literature; see Baker to appear for references. The following abbreviations are used in the glosses of the examples: 1sS, 1st singular subject marker; AFF, affirmative; EXT, extended mood marker; FOC, focus particle; FV, final vowel (perhaps a mood marker); LK, linker particle; LOC, locative gender prefix; NEG, negative; PASS, passive; OM, object marker; PRES, present tense; SM, subject marker; T, tense prefix. Arabic numerals between 1 and 19 are used to represent the various Bantu gender/number noun classes; they can

be appended to the gloss of a noun (e.g. fruit.5), to a subject agreement (5S), to an object prefix (OM5), and to certain other elements that agree with nouns in their environment (e.g., focus particles).

⁸ Richard Kayne (personal communication) reminds me that there is reason to think that the initial vowel in Kinande nominals like *e-ri-tunda* (the fruit) and *a-ba-kali* is a determiner head (see Baker and Collins (in press: xx) and references cited there). This appears to agree with its NP complement, even though the order is D-NP. Here, however, the “agreement” can be analyzed as being purely the result of vowel harmony targeting the phonologically underspecified determiner: the determiner is always realized as [a] if the immediately following vowel is [a], as [e] if the following vowel is [i], and as [o] if the vowel in the gender-number prefix is [u]. There is no morphologically governed irregularity in this.

⁹ Object agreement also shows up in cases of pro drop, both in VP and in PP. Since the agreed-with NP is null, one cannot tell by inspection where it is. Given the overall paradigm, I assume that pro must move to Spec, vP and Spec, PP to be licensed by agreement in Bantu.

For reasons to think that *ri* in (15) is an instance of object agreement, not a moved clitic similar to those found in Romance languages, see Baker (to appear).