NSF Project Description: Exploring boundary conditions on the theory of agreement through the comparative study of African languages - Ken Safir, PI – Rutgers University

1.0 Introduction

The PI proposes to explore the relationship between morphological agreement relations and anaphoric relations in the languages of Africa by expanding the scope of the Afranaph Project.

1.1 Intellectual merit

The central theoretical question this proposal addresses is whether or not the agreement phenomena that occur between nominals and non-nominals, such as subject-verb agreement or concord are driven by the same principles of grammar that determine agreement between two nominals in an antecedent-dependent relation, such as anaphor-binding or bound variable relations. The literature surrounding the empirical issues is overwhelming, but the question is narrowed in this inquiry as to whether, or to what extent, the Agree relation/operation originally proposed and defined by Chomsky (1995), but formulated variously by others, also accounts for the agreement that accompanies antecedency in local anaphora or bound variable relations more generally. Although we will also entertain some new hypotheses, we hope, at minimum, to clarify what the boundary conditions are for any answer to the central question we pose, whether the ultimate answer is based on minimalist syntactic structures or those of another framework. The potential for this outcome constitutes the fundamental intellectual merit of our proposal.

Although we expand upon it below, it is useful to get some idea of the kinds of empirical contrasts we will be paying attention to. Some of the best evidence for distinguishing the two kinds of agreement relations consists of mismatches where A agrees in one way with B and in a different way with C in the same sentence. English conjoined subjects act alike for both local anaphora and subject-verb agreement, e.g., *Those guys/John and Bill are criticizing themselves/each other*. Moreover, comitative constructions employing *with,* as in **John is criticizing themselves/each other with Bill* don't "sum" the subject and the *with*-object to "construct" a plural antecedent in English. By contrast, Lubukusu, a Bantu language, permits singular subject-verb agreement but plural anaphora binding for the equivalent of 'the hunter sees each other with the antelope', which can mean, 'the hunter and the antelope see each other' (Baker, Safir and Sikuku, 2013). Bantu languages are particularly useful for testing antecedency of this kind because nominals that belong to different noun classes cannot straightforwardly be summed to provide for a plural subject-verb agreement (SVA) marker that satisfies the differing plural agreement classes of the conjuncts. Some Bantu languages use a general default form for mixed conjunct plurals, others choose a compromise that shares semantic properties of both conjuncts where possible (e.g., mismatched nominals designating humans would have a different plural from mismatched inanimates), others agree with just the first or last conjunct, or some noun classes win out over others to impose their plural, or comitative splitting of conjuncts leaves only one 'conjunct' to agree. However the quandary is solved, the issue arises as to how anaphors are anteceded by mixed conjuncts and whether or not they match their antecedents morphologically as well as semantically. It appears that reciprocals, as opposed to reflexives, can be anteceded by a sum of the comitative argument and a singular subject, even though SVA for the subject is singular. However, default plural is also possible for SVA when the subject is singular and there is a comitative argument, and in that case, both reciprocals and (plural-interpreted) reflexives are possible, apparently anteceded by the summed singular subject and comitative argument. This phenomenon was explored for Lubukusu in the reference cited, but the PI has found similar phenomena in other Bantu languages that await detailed investigation.

Another phenomenon relevant to the relation between antecedency and Agree that is so far only attested in African languages is complementizer agreement (C-agreement) with an immediately superordinate subject. Although attested before (Kawasha, 2006 and Baker, 2008, among others), Diercks' (2010, 2013) detailed exploration of C-agreement in Lubukusu showed the depth and systematicity of the phenomenon. Apart from Nformi (2017) on Limbum and Diercks and Rao (2017) for Kipsigis, there are no detailed studies of these patterns in other languages except Letsholo and Safir (to appear) for Ikalanga, which is preliminary, but opens new avenues of research. The commonality across all the languages where C-agreement is attested is that person, number and gender (phi-features) on the C

match the phi-features of the subject. The paradigm of C-agreement does not always match matrix SVA, however. Diercks shows that where anti-agreement effects neutralize person agreement, the matrix SVA marker is not the usual class 1 marker (3rd.sg human), but C-agreement shows the same class 1 marker it would if there were no anti-agreement effect. In Ikalanga, it is possible to show that C-agreement still occurs when the agreeing C is in a clause introduced by an infinitive. The understood (interpreted) phi-features of the null (PRO) subject then guide the realization of the agreement features on the agreeing C. This also suggests that antecedent agreement (henceforth, AcA) is responsible for the forms that appear on the C, not SVA. There is no semantic antecedency in C-agreement, however. The agreeing C does not represent a logophoric argument (the perspective-holder), rather it may contribute to the speaker's evaluation of whether or not the proposition represented in the complement clause is true. Moreover, in Ikalanga, there is vestigial tense agreement on the C (which, though it is the same root as the verb 'say', can be shown not to be a verb itself), and additionally there is voice agreement, such that the passive form of the C appears when the matrix verb is passivized. If the agreeing C is introduced by a passivized infinitive, it also shows agreement for passive. These other forms of agreement are clearly not modeled by either AcA or SVA. Letsholo and Safir suggest that phi-agreement and voice agreement are filtered through a Voice head in the matrix clause, but the generality of this approach will depend on how well it matches up with what we still don't know in detail about C-agreement patterns in other languages.

In a broad sense, the theoretical issues are relatively clear. The challenge for Agree-does-it-all theories is to show that the apparent mismatches are expected in certain syntactic environments and/or derived by independently necessary assumptions. On the other hand, if SVA and AcA (local and non-local) are argued to arise from distinct mechanisms, what are those mechanisms and how are they independently motivated? The quest to bring empirical patterns to bear on these central theoretical issues and clarify what is at stake is the central intellectual merit of the proposal.

1.2 Justifying the empirical scope of the work

There is no theoretical reason to limit such an inquiry to empirical patterns in African languages. However, by making African languages our focus, we can use this inquiry as an opportunity to systematically expand our empirical grasp of agreement patterns in understudied languages, some of them endangered, while building on resources that the Afranaph Project has already developed for the study of anaphora in African languages. In particular, we propose to focus on languages for which we have already developed rich data resources on anaphora and expand our data sets with new elicitations designed to explore agreement relations in those languages. This method of working has several benefits.

First, the PI and others participating in the Afranaph Project are already exploring the boundaries of anaphora and agreement, including work on the relation of verb affixes to argument reflexives and reciprocals (Safir and Selvanathan, 2016), object agreement and clitic doubling (Diercks and Sikuku, 2015, and the Object Marking in Bantu elicitation questionnaire for Afranaph, Diercks, 2016), the morphological shape of bound variable pronouns (Safir, 2014), the relation between anaphor binding and verb agreement with mixed conjuncts and comitative constructions (Baker, Safir and Sikuku, 2012), the morphology and interpretation of logophoric agreement (Adesola, 2003, Anaphora Questionnaire, Safir, 2005, and work in progress), prolepsis and long distance agreement (Newkirk, in preparation), and apparent antecedent relations where complementizers agree with immediately superordinate subjects (Letsholo and Safir, to appear, and literature cited there). The direction of Agree operations ('up' or 'down') has been explored in works by Carstens and Diercks (2013), Carstens (2016) on delayed valuation and downward agree. Prof. Mark Baker, a colleague of the PI who is a collaborator on the Clausal Complementation and Selection Afranaph Sister Project, is recognized as a leading figure in debates about the typology of agreement and the directionality of Agree (Baker, 2008). The PI is also the organizer of a workshop on agreement in African languages at the International Congress of Linguists taking place in Cape Town during the summer of 2018 for which the program has just recently been set.

Second, as part of the work of the grant, we will be enriching our data sets both for the languages where we already have extensive data on anaphora, as well as for languages for which we do not have any data at all. All the work just cited, as well as the more specific theoretical focus to be elaborated on in section 2, will guide our new elicitation. We are not starting from scratch, however, since work on

anaphora, and to a lesser extent, agreement, has been part of Afranaph research since the project was founded in 2003. Thus we have extensive resources to build upon, as will be illustrated in section 3. Comparative data from other languages where the pattern of anaphora has been explored provides a much richer context and guide for new elicitations. Plans for new elicitations are described in section 4.

Third, by expanding Afranaph resources, resources that are freely available to any interested party without login barriers, our data could be used not only address the question we pose in this study, but questions that are yet to be posed by researchers with other interests. Indeed patterns that were 'bycatch' of work on clausal complementation and local anaphora led to some of the other inquiries mentioned above, as well as to inquiries that are only tangentially connected to patterns of anaphora and agreement (such as the exploration of verb stem structure in Joola Eegimaa, see Bassene and Safir, 2016 and Safir and Bassene, 2016, or the distribution and interpretation of actuality entailments in Lubukusu and Akan, as in Safir, Baker and Sikuku, to appear, and Owusu and Safir, in preparation). The new datasets generated by research on agreement will complement almost all of the Afranaph Sister Projects, particularly the Anaphora Project, the Object Markers in Bantu Project, Situating the Subject in Bantu Project, DP Positions in African Languages Project, The Morphosyntax of Bantu Nouns Project, and even the Clausal Complementation and Selection Project, where SVA with clauses and C-agreement have emerged as particular areas of interest. In short, developing additional data based on research into agreement-anaphora relations has the potential to enhance the value of Afranaph resources to most of those already involved Afranaph research projects, as well as many users outside the project.

1.3 Broader Impact

The broader impact of our proposal does not extend far beyond linguistics, but it is significant and integral to our manner of research.

One of the underlying goals of Afranaph is to improve the empirical underpinning of linguistic research by providing access to more data than can fruitfully appear in any one article reporting research. When evaluating a proposal in the literature that draws evidence from an unfamiliar or understudied language, the author necessarily presents certain generalizations about the language that are not supported by examples in the article if only to save space. Expanding upon or refuting such a proposal typically requires additional access to data in that language that the evaluator must seek independently, especially if it is not in other published sources. By providing a rich source of data that can be publically accessed, such generalizations can be independently evaluated on the basis of primary data if they are based on languages in the Afranaph Project. It is indeed one of the features of the Afranaph Project that we do not aim for comprehensive coverage of African languages, but rather offer rich primary data of the languages we have been able to find NSLCs for. The WALS database, for example, is more focused on providing typological generalizations based on preliminary analyses of as many languages in the world as they can find any evidence for, and while this is a great service providing a broad view of what is 'out there', much of their property ascriptions are based on primary data that is not accessible in their database. What we lack in breadth we make up for in depth for the languages we cover. Afranaph, and other databases we hope our style of work will encourage, hold empirical research in syntax (in particular) to a higher standard, insofar as our project allows researchers (a) to put claims about Afranaph languages in the context of richer data about the language in question and (b) to access and examine the full range of evidence used to establish any empirical generalization - far more data than there is space to present in any research paper.

The focus on African languages and using the Afranaph Project as a vehicle for that enterprise in turn supports a broad, vibrant and growing research community of African and Africanist linguists for whom Afranaph resources are a focus. Indeed a major goal of the Afranaph Project as a whole is to build a research community around a common public resource. An essential focus of the project is to involve African linguists directly in our research, not only as the source of our data, but also in order to encourage career development of linguists in or from Africa. Over 45 Afranaph native-speaker linguist consultants participate in Afranaph research, many of them from the time they are graduate students (see the statement on mentoring), while others are established scholars, some of whom also direct Afranaph sister projects. Periodically, Afranaph Project Development Workshops bring together our community (in 2010

and 2013) and a third one is being planned for fall, 2018 (not part of this grant proposal). Sister research projects conducted with Afranaph resources currently include 7 projects directed by a total of 13 researchers from North America, Europe and Africa (some are also consultants). The concerns of our consultants, moreover, occasionally involve Afranaph in projects designed to benefit particular speech communities, such as the Kinande Dictionary Project or the Babanki community school, though no NSF funds are used to support these endeavors. Moreover, some of the languages in the Afranaph Project happen to be endangered, so the data that we collect plays a role in the documentation of those languages (see Bassene and Safir, 2017). All of these community-building efforts are part of the broader impact of the Afranaph Project that research based on our proposal will support.

Our proposal also involves developing new features for the Afranaph Database that will help to organize our data and research into the intersection of anaphora and agreement. The development of our database design is another of the broader impacts our work has the potential to have. The Afranaph Database (henceforth, ADB) is an internationally consulted resource that has been used for research (both theoretical and areal) and instruction (as a fieldwork tool, as a resource for student projects) and as a model for other databases, at some institutions in more than one of these ways. For example, the ADB was studied as a model for databases under construction or currently online developed at ZAS in Berlin and at SOAS in London (both venues where the PI has been invited as a consultant). It is being used as a basic resource for a major research grant on the anaphor-agreement effect at the University of Liepzig as well as for instruction at UC Berkeley, and the National University of Singapore. The PI is often invited to demonstrate the ADB at other institutions and has been invited to conferences specifically focused on the development of linguistic databases (most recently, at a conference organized by the Linguistic Data Consortium, 2016, for the NIEUW Project, also supported by NSF). The PI was also invited as a featured speaker for a workshop on linguistic databases at the 2017 meeting of the West African Linguistic Society where the president of the West African Linguistics Society in Winneba, Ghana, specifically endorsed the project and encouraged conference attendees to participate (and several were recruited as a result). The design of the ADB, then, which this proposal will allow us to extend, has made, and will make, a contribution beyond the particular scientific problem we plan to address.

Finally, funding for this scientific inquiry would have the added consequence of assuring the continuity of the Afranaph infrastructure that supports our broader range of research. That infrastructure (described in section 2) is primarily located Rutgers University, but also partially at Georgetown U. and U. of Utrecht, where the server for the ADB is located. Independent efforts to insure the continuity of the project for many years to come are discussed in section 3.1 and in the data management plan.

2.0 More on the theoretical questions at stake

The roles of universal grammar and information structure in discourse in determining relations of coreference and anaphoric dependency has been a central concern in theoretical linguistics at least since Postal (1966). The fact that syntactic structure significantly impacts the patterns of anaphoric interpretation crosslinguistically suggests that anaphora can be a guide to deeper syntactic relations. The distribution and nature of agreement phenomena has always been a concern of both theoretical and descriptive linguistics, but precisely because agreement is not conceptually necessary (and some languages largely lack it), the sensitivity of agreement to abstract lexical, syntactic, semantic or pragmatic factors provides evidence for the structure of those components of grammar. By studying the intersections of agreement and anaphora, this study holds the promise of triangulating to discover deep properties of the human grammar capacity and the components that make it up. The advance in knowledge achieved by close attention to the empirical consequences of theoretical proposals in the literature will, we hope, provide a basis for any future inquiry into the interaction of agreement and anaphora.

Minimalist approaches to agreement are largely based on the Agree operation, which applies as the derivation is being built by the combinatory operation Merge. Agree is usually taken to be a feature transfer that occurs between a head that c-commands a nominal or a nominal that c-commands a head, such that the head's features are 'valued' by or transferred from the nominal in question. On some accounts, a nominal can enter into only one Agree relation, that is to say, a nominal is 'active' until it enters into an Agree relation and is 'inactive' subsequently. There is also an intervention effect, such that

a head with unvalued features (a 'probe') must find the first nominal (goal) it c-commands, so if a probe α c-commands a goal β, α can only have its features valued by β if there is no nominal γ such that α c-commands γ and γ c-commands β. Insofar as it is discussed, it is always assumed that Agree is bounded not only by intervention but by phases. Phases are locality domains in which relations of movement or agreement can occur, but once a phase (e.g., CP, vP) is complete, the complement of the phase head (the complements of C and v, respectively) are no longer visible to further operations of movement, Agree or rules of morphology. Linguists differ considerably on how exactly they formulate Agree, but all formulations include c-command, intervention, phases and activity (see, for example, Baker (2008).

Current theories of anaphora, particularly of local anaphora, feature a special role for agreement relations, either with respect to the agreement that accompanies local anaphora or that which effects bound variable relations more generally. A separate question which we will also address is whether (locally) bound variable anaphora is itself achieved by Agree (e.g., Reuland, 2005, 2011, Heinat, 2006, Hicks, 2009, Kratzer, 2009, Rooryck and van den Wyngaerd, 2011, and Diercks, van Koppen, and Putnam, 2017, but see also Safir, 2013 and Charnavel and Sportiche, 2016, for arguments against deriving binding from Agree), rather than just accounting for agreement that accompanies independently determined anaphor-binding (e.g., Safir, 2014). Agree-based theories of binding face formidable problems where agreement is non-local, as in cases of quantifier-bound pronouns, logophoric pronouns, resumptive pronouns and prolepsis, but systematic differences in the form of agreement (e.g., which features agree and which do not) between local and long distance relations are underexplored. For long distance dependencies or pronouns without antecedents in a sentence, questions arise as to whether reference is fixed by presuppositional matching (e.g. Sauerland, 2004, 2008) or syntactic variable-binding (e.g., Collins and Postal, 2012) or aspects of both (Safir, 2014).

Any claim that local bound anaphora or the agreement that accompanies it is achieved by Agree depends on how Agree is formulated. Proposals vary with respect to all the components of Agree.

Some theories assume that the probe must always c-command the goal (downward Agree) such as Chomsky (2001) and, amongst anaphora theorists, Rooryck and van den Wyngaerd, 2011 (henceforth, R&vdW), some (fewer) argue that the goal must always c-command the probe (upward Agree) (e.g., Zeijlstra, 2012) and some argue that Agree can probe in either direction. Amongst the latter, Baker (2008) argues that upward probing is parametrically available and Carstens (2016) argues that it is possible only when downward Agree cannot succeed (both appeal to evidence from Bantu languages). Some theorists propose that nominals can be probes as well as well as goals. So for example, if nominal anaphors are feature-deficient probes then they could potentially participate in upward Agree, a form of classic antecedent agreement, or, if they c-command their antecedents, as in R&vdW, they can probe downward to value their features. Assumptions like the latter could determine that anaphors must move, overtly or abstractly, in order to be in a position where these elements can value their features (on such 'hoisting' derivations, see Safir, 2013).

Although all Agree theories assume that there are interventions, there are differences about what the interveners are, and in some accounts whether or not a nominal is active or not determines whether or not it is an intervener. For some accounts, a nominal that is not a potential agreement partner (an inherently casemarked nominal) does not count as an intervener, while in other treatments, it counts as a 'defective' intervener. Afranaph data on differences between reflexives and reciprocals being bound across causative arguments are directly relevant to these questions. Interveners for SVA, however, do not always match those for anaphora, an issue that our empirical work will explore. Movement of an anaphor to a position higher than an intervenor might allow an anaphor to avoid intervention. For example, C-agreement involves complementizers in clausal complements that agree with the subject of the immediately superordinate clause. C-agreement appears to require upward Agree, but can be reinterpreted as movement of a C-anaphor above its antecedent, permitting downward Agree (as in Diercks, et. al, 2017). The issue arises as to whether or not an intervening superordinate direct object blocks agreement or if movement of the anaphor jumps over it, avoiding intervention by the direct object (with the effect being 'subject-orientation' of the anaphor). These questions play out, for example, in an empirical

contrast between Lubukusu C-agreement (Diercks, 2013) and Ikalanga C-agreement which is currently under investigation (see Letsholo and Safir, to appear).

In addition to treatments of activity, c-command, and intervention, Agree is taken to be sensitive to the boundaries of locality domains defined in work of the last 20 years by phases (Chomsky, 2001 and many since). Phase heads, C and v, define complement domains and edges, where the edge of the phase is that portion of the maximal projection of the phase head that is not part of the complement of the head. When one phase is embedded in another, only the edge of the lower phase is accessible to the next higher phase. Movement of an element from the phase complement to the phase edge allows for successive cyclic long distance relationships to be formed. A variety of approaches assume that the locality domain for anaphora (traditional Principles A and B of the Binding Theory, Chomsky, 1981) corresponds to vP phase in most cases. If antecedent agreement is achieved by Agree, then it is expected that antecedent agreement should show locality. Antecedent agreement is frequently non-local, as in the case of pronouns bound as variables, resumptive pronouns, logophoric pronouns, and proleptic resumption. While some long distance relations can be reduced to the result of cyclic movement, this is not the case for AcA that crosses phase and syntactic island barriers. If AcA at a distance is inescapably necessary and Agree cannot account for it, then reducing AcA to Agree may be an inappropriate overreach. Then again, in anaphora theories like those of Reuland (2011) and Reinhart (2008), there are posited to be computational preferences for applying Agree when it is available, on the assumption that long distance AcA is formed by another device when Agree is not available.

One can also ask whether Agree must apply where it can, and whether it must succeed where it applies (see Preminger, 2014). Although Preminger is primarily concerned with contexts where Agree plays a role in argument licensing and related morphology, one can also ask if failure to agree in mismatch contexts constitutes a failure of Agree. Patterns of SVA in Bantu include, as mentioned earlier, defaults for agreement when the subject is a mixed conjunct. One issue to resolve is whether there is default agreement when no Agree relation can be formed, or if Agree has applied, but the morphological result is spelled out as a default when there is no way to resolve feature mismatches. In the latter case, some mismatches might be more successful than others. Evidence from mixed conjuncts in Bantu suggests the latter approach, because there are 'compromise' outcomes different from 'default' class 8 if both conjuncts are animate or both human (depending on the language, see de Vos and Mitchley, 2012).

Yet another set of questions about failure of agreement arise for the instances where SVA cannot agree with an anaphor (the anaphor-agreement effect, see, e.g., Woolford, 1999, and more recently Sundaresan, 2011 and Messick, 2017). Afranaph languages, particularly Bantu, include several instances where anaphors appear to agree in SVA contexts, however, specifically where an object-raising analysis is in question. These are cases that also potentially involve relations between object agreement and SVA, as well as AcA for anaphors, they are of particular interest for AcA/SVA mismatches. The PI is a consultant for a project in Leipzig exploring the anaphor-agreement effect headed by PIs Hedde Zeijlstra and Sandhya Sundaresan and this work will be coordinated with them.

There are certain antecedent agreement effects or forms of anaphoric agreement that are always local, a fact exploited in Safir (2014) and Safir and Selvanathan (2017) where local anaphora is (potentially) mediated by agreement with Voice. In recent years, the vP spine has been hypothesized to have an enriched structure of functional heads (see, e.g., Kratzer, 1996, Travis, 2010, Harley, 2013, Legate, 2014, amongst many others). Baker, Safir and Sikuku (2013) appeal to distinct voice morphemes for reciprocity and reflexivity in Lubukusu that play a role in antecedency and Safir and Selvanathan take this a bit further by arguing, on the basis of evidence drawn from Afranaph data, that anaphoric agreement is sensitive to a form of reciprocal voice, such that the form of direct object anaphors is determined by agreement with a distinct Voice head in the vP phase (see also Ahn, 2012, who also argues at length for reflexive voice that mediates local anaphora). Letsholo and Safir argue that C-agreement is sensitive to passive voice, which is less surprising if agreement is mediated by voice heads within the vP phase in a more general way. *The possibility that voice-mediated local anaphora might be expected to contrast systematically with non-voice-mediated non-local anaphora will be will be one of the working hypotheses that will guide our research.*

An additional question about Agree concerns when it applies. It is possible that some of the mismatches between AcA and SVA, for example, may be timing effects, based on differences between stages of derivations. As mentioned earlier, Messick (2017) reports that several African and South Asian languages express de se readings with a long distance 3rd person anaphor or pronoun in subject position and 1st person agreement on the verb. Messick argues that there is indexical shift to account for the *de se* reading, but the agreement mismatch is taken to result from a morphological change to 3rd person that is subsequent to the application of Agree to yield 1st person SVA. In contexts where a logophoric pronoun binds a local anaphor, this suggests that a reflexive anaphor will be in 1st person form, not 3rd, because the change to 3rd person is subsequent to anaphoric binding in the previous phase. This is not testable in Nuer because the anaphor does not have exponents of the relevant phi-features of its antecedent. In languages where we can test, the result, whatever it is, and however general it is across such languages, will reveal whether this sort of ordering approach is viable.

Almost all accounts of Agree assume it is part of the syntactic derivation, but some regard the formation of the Agree relationship as separate from the transfer/valuation/matching of features in that the latter is taken to be achieved only in the morphophonological interface, where lexical items are inserted (as in Distributed Morphology). Thus another of source of mismatched agreement could be based on distinctions between the effects of Agree in different components of grammar. This separation has antecedents in work by Chung (1998), Bobaljik (2008) and the subject of current work by Atlamaz and Baker (to appear), where Agree creates a probe goal relationship that is only spelled out in the morphological component, after morphological operations bring features of the goal in reach of the Agree relation. The separation of the Agree relation and its persistence into the morphophonological component requires a way of preserving the dependency relation between goal and probe. Exactly the same relation preservation is needed to characterize binding for anaphora, where the content of the relationship is identity of discourse referents rather than identity of features. This also suggests an avenue for common ground for agreement and antecedent relations (see Scarcerieau, 2012) and, at the same time, a natural source of surface divergence.

Attempts to unify agreement and anaphora relations by attributing them to a single operation is a natural theoretical move, but if the execution of the idea requires ad hoc extensions to fit the facts, then a more elegant theory might result from subtle conspiracies of independently motivated principles that result in varying agreement patterns. Our best guide to the right direction for theory will be the empirical patterns that emerge from our attempt to find evidence for versions of particular theoretical hypotheses. We believe the intellectual merit of our proposal lies in advances to our understanding of the relations between agreement and anaphora that the data we gather will illuminate.

3.0 The right tool for the job

The Afranaph Project has been gathering data in collaboration with native speaker linguist consultants for 14 years and during that time we have honed our operations to improve the flow of data as well as the richness of the analysis and presentation of the data we collect. Section 3.1 describes our data-gathering methodology, section 3.2 explains how the database is structured to facilitate research and section 3.3 suggests how data already in Afranaph provides a springboard for the study of agreement/anaphora intersections. It should be kept in mind that support for this proposal has a multiplier effect; Funding the infrastructure described below that is needed to support this research proposal will also, as a matter of course, support many other ongoing Afranaph investigations.

3.1 Our current mode of operation

Our current mode of operation is most easily illustrated with an account of how we elicit data for the Anaphora Project, the first of the sister projects in Afranaph.

The Anaphora Questionnaire (AQ) is designed to be usable for any language and its formulation is guided by our current understanding of linguistic anaphora and theoretical attempts to account for it. Thus the AQ is in no way customized to the effects present in any particular language, but instead attempts to elicit relevant information that helps to address as many interesting questions as we are aware of for the domain of grammar we are studying. The design of all Afranaph questionnaires will naturally reflect the perspectives of, and serve the research purposes of, those who design them. We try to insure

our questionnaires are always a little broader than our current research questions. As Afranaph expands, questionnaires have been developed by researchers with different perspectives and interests, including our native speaker linguist consultants (NSLCs). Naturally, we expect that our questionnaires will be revised as commentary and experience show their weaknesses (and indeed this has already been the case).

An important aspect of our methodology is that our questionnaires are designed to be filled out by consultants who are native speakers trained as linguists (at more than rudimentary levels of training). At this point in history, there is an ever-widening pool of potential consultants who fit this description, and so far, the more than 40 linguists who are working with us now as NSLCs are only a fraction of those we hope to attract. Working with NSLCs facilitates our research through electronic communication, since the sorts of data we are interested in can be harvested by collaborators who know how to treat their own intuitions of acceptability as data. Frequently, NSLCs can see what we are trying to find out even if our elicitation is clumsy (either in general or in their language), and so a multitude of shortcomings in our elicitations are ameliorated by the expertise of those whose data we rely on. Thus we are counting on the linguistic skills and training of those who work with us in order to make this means of harvesting data efficient and insightful. Either the PI actively recruits NSLCs or NSLCs become aware of the project and volunteer to participate. The consultant is paid after the second follow-up, though most NSLCs continue to participate thereafter. Sometimes NSLCs become collaborators in work destined for publication on the site or in conferences, collections or journals (about 25% of NSLCs who finish questionnaires).

The latter point is underscored by the fact that our elicitations only begin with the questionnaire (the AQ, as our model case). The NSLC begins to fill out the questionnaire (the AQ or one of our many others) and sends a sample for comment and correction as to glossing, form, completeness and clarity. Glossing is particularly important, since it is one of the fundamental advantages of working with an NSLC who has the training to recognize the distinction between word and morpheme boundaries, though often these analyses are only first approximations that improved in follow-up. When a first draft of the questionnaire response is complete, the PI or Afranaph Sister Project director goes through it not only for glossing, form, completeness and clarity (with an eye to putting it into proper form for input into the database). Next, additional follow-up questions are asked that are designed to explore issues and constructions where one language has a more articulated set of distinctions than others (for example). There are additional elicitations. Thus the presentation of the data contains more than just uniformly complete the answers to the AQ, and in this respect the character of the follow-up process is partially visible from the commentary that appears directly in the AQ responses (AQRs) for each language (particularly in their static .pdf case file versions). In the earlier years of the project, a sketch of the pattern of anaphora in the language in question was then written up (by the project director, the consultant, or in collaboration between them).

The role of the anaphora sketches has declined as our database has become more articulated over the years. As we flesh out the properties for anaphoric markers required by the database, much of the analysis is available in the database itself. More recent practice is to input completed questionnaires to our database before follow-up to help clarify what sorts of comparative data we are still missing and to develop new research questions to pursue through additional elicitation. Data is entered into the database by graduate and undergraduate assistants at Rutgers or (for Afro-Asiatic languages, at Georgetown) guided by the database property attribution guidelines (on our site). The analyses in the database attributions for any given language are reviewed by the PI and/or the RA. Indeed, more energy is now directed to writing up research results to be presented in other venues rather than collating the database information into anaphora sketches for particular languages.

The work is labor intensive on both ends. Although NSLCs are paid for completed work, the payment is primarily intended to recognize the contribution NSLCs make as linguistics professionals. Although the money can sometimes be significant for NSLCs residing in Africa, it is a very low rate of pay for the hours of thought and effort that go into answering our questionnaires and participating in follow-up. Only those driven by genuine interest in the nature of the language they speak and the opportunity to explore it with other linguists ever manage to complete the work. Some NSLCs are very dedicated, active and responsive despite everything else going on in their lives and complete the work in a

timely fashion, but they are a minority. Some who initially agree to participate do not ultimately follow through, and it is often the case that NSLCs begin the work but do not finish it. Sometimes NSLCs complete the work, but it takes several years, during which more important priorities are tended to first. We at Afranaph have learned patience and by our continued existence, we have seen many NSLCs complete work deferred for years (our record is 5 years, start to finish, for one NSLC completing one questionnaire). On the Afranaph staff side, the sister project directors and or the PI spend a great deal of time over email encouraging NSLCs to finish and to respond to follow-up questions and a great deal more time analyzing the thoughtfully prepared data that our NSLCs provide. Often the NSLCs provide insightful commentary in their responses, information that is often crucial to understanding what is going on in matters small and large (e.g., when morphophonology disguises morphosyntactic relations). At times, just correcting and standardizing the glossing takes Afranaph staff (usually the PI) many hours for each questionnaire response. When the data is ready for input, graduate and advanced undergraduate students are employed to input the data, though only those with more training are asked to input answers to analytic questions about entities (on database entities, see section 3.2).

The PI is responsible for the research goals of this particular project within Afranaph, including choosing directions for research, directing or supervising new elicitation documents, following up with the NSLCs, and taking a lead in writing up the results. However, there will be many collaborators and consultants who will contribute to the analyses and theoretical proposals that emerge from the project. Afranaph has, until the last few years been essentially PI-centric, however, and this is changing. In future years, Afranaph will be supervised by a board of directors that will choose new directions for research, encourage grant-writing, and choose a director for the project. The current PI will be the first director under this arrangement, but as others bring in funding, this will change. The PI has invited six Afranaph researchers to form the nascent board (Vicki Carstens, Michael Diercks, Claire Halpert, Ruth Kramer Justine Sikuku, and Jenneke van der Wal, all independently affiliated with Afranaph, as mentioned earlier) and the board will subsequently determine its own membership. The willingness of these scholars to serve in this way shows that Afranaph has had success in building the research community initially envisioned. Moreover, they are also all researchers with particular interests in the theory of agreement (and no two of them agree about Agree). The current project is in part designed to get them further involved in a more specific common project.

3.2 The Afranaph Database

The Afranaph Database (ADB) came online in 2008 and has grown steadily since. The original plan for the Afranaph did not call for it, but as our resources grew, more efficient ways to search, evaluate and analyze the data collected required something of the sort. Finding a database designer with sufficient understanding of the research enterprise was difficult, but in 2007 Prof. Alexis Dimitriadis, who has designed other linguistic databases, began his collaboration with the PI to build the ADB. The ADB has since become a central feature of the Afranaph Project.

Besides allowing searches of all our sentence data, including full morpheme breakdowns, glosses and translations, all of which can be explored with word searches in the different fields for those entries, we also had a need to search for properties of the sentences on a slightly more analytic level (e.g., find all sentences that contain relative clauses), so a form of tagging for sentence properties was added in addition to tagging for languages. We also needed searches to be guided by properties of analytically identified entities. Since the project began as research into patterns of anaphora, the ability to compare and group properties of anaphoric markers (the exponents of, or strategies to form, anaphoric relations) was of prime importance. Prof. Dimitriadis already had formed a database with such an 'analytic entity' that allowed relevant anaphoric marker properties to be compared. When we built it into the ADB the anaphoric marker entity was enriched with additional properties and additional search functions were added. As the project expanded to research questions outside of anaphora, different sorts of analytic entities were needed, but also separate organization of the data for different research priorities.

In 2013 we began to explore how the database design could allow each sister project to direct its own research while the data collected by each project would remain available to all the other projects is a common resource. The PI and Prof. Dimitriadis came up with the multiportal design that the ADB now

has. Each portal permits researchers to enter the data they collect and all the sentence data is glossed according to Afranaph standards and tagged according to general Afranaph requirements for sentence data. However, the portals can also add tagging specific to their project that is visible only in their portal. Each portal can also be customized to have its own 'analytic entities'. Analytic entities are constructs in the database that are empirically important constructions or word types to which properties can be ascribed. For the anaphora portal, as mentioned, 'anaphoric markers' are an analytic entity, e.g., English *himself* is an anaphoric marker, one that is reflexive, sentence-bound, a compound form, etc. These properties can be ascribed to *himself* as an anaphoric marker and can serve to both define language internal or cross-linguistic searches that return sets of sentences, sets of anaphoric markers, or sets of languages (e.g., all sentences with anaphoric markers that are compounds, all anaphoric markers that are compounds and bound within the most local domain, all languages that have anaphors that are ambiguous with respect to reflexive/reciprocal interpretation, etc.). The Clausal Complementation (CC) Project, our first new customized project portal, has different analytic entities and more of them, allowing for much more complex search possibilities and correlations that are specific to that project. Moreover, CC Project researchers can tag sentences collected by other projects according to their own needs within their portal, without disturbing the data as entered by the projects that collected it. Moreover, our understanding of the analytic entities evolves in the course of study and must sometimes be revised (e.g., we realize that a property or a set of properties that an entity or entities has is better coded another way or deleted), but this can then be done without disturbing the basic sentence data or retagging – only the analytic entity properties are revised and the revision only affects the portal where the analytic entity is defined. Other project-specific portals are currently in the works. Researchers whose interests do not touch on anaphora or clausal complementation can use the Generic Portal to search or properties of interest to them.

Even without adding new features, the Anaphora Portal already permits searches that bear on the relationship between agreement and anaphora. For example, we can search for all the anaphoric markers that agree, all the languages that have anaphoric markers that agree (as opposed to those that are invariant), or all the sentences in any set of languages that have anaphoric markers that agree. We can search for all the sentences in one or many languages that have both a particular anaphoric marker and a particular morpheme (e.g., CAUS) in order to see if amongst the sentences in our database with this description there are some that will help us determine if an intervening causee argument is an intervenor for agreement between the causer and the object of the causativized verb. The existing Anaphora Portal needs additional analytic entities to integrate agreement data, as addressed in the next subsections.

3.3 Existing Afranaph data relevant to the task at hand

Data already in Afranaph includes a great deal of information on antecedent agreement. Of the 33 languages entered into the ADB so far, 28 AQ responses have been completed and entered in the database in addition to 2 which are partially entered and 3 in preparation (likely to be finished and/or partially finished). Twenty-one other questionnaire responses are also entered in the ADB with many more in preparation. Although not all of the anaphora data has been analyzed and policed in the database (an ongoing process), many searches that relate patterns of anaphora and agreement are already possible.

For example, we have hypothesized that reflexivity and reciprocity marked on verbs might actually be a voice morpheme. It is possible that when this is so, there is no agreement for person, number or gender on the affix itself unless it is a clitic (as in Romance or some Germanic languages). Of the 18 reflexive or reciprocal verbal affixes in the ADB, not one shows agreement, even though if there are object markers in the same language or argument position reflexives, they have full paradigms (in those languages where the reflexive affixes are distinguishable from clitics). Quite independently of the ADB, reciprocal verbal affixes that agree in person (independent of SVA) do not appear to be attested, although there are argument position reciprocal forms that do, such as in Lubukusu (e.g., Afranaph ID#1301). Some verbal affix reflexives are associated with argument anaphors that agree with the ultimate antecedent in person number and gender, a matter of additional interest, especially since many African languages have argument anaphors that are ambiguous between reflexive and reciprocal meanings unless the addition of a verbal affix distinguishes the readings (an issue partially addressed in Safir and Selvanthan, 2017). Many argument anaphors that are ambiguous in this way can be found in the ADB

(Safir, 2016 provides a guide for such searches). Though the data we currently have is incomplete for exploring the voice hypothesis more generally, the data we already have are an excellent place to start.

As mentioned earlier, we already have some data about the relations between mixed conjunct agreement, comitative constructions, and antecedent anaphor relations, primarily for Lubukusu, but there exist a number of studies of mixed conjunct agreement in Bantu languages in particular (e.g., de Vos and Mitchley, 2012) that provide a road map for the investigation their intersection of anaphor antecedent relations. They include instances of first conjunct agreement, last conjunct agreement, default and 'compromise' agreement (where certain noun classes have something in common, e.g., both conjuncts are human), but they do not usually include information for comitative summing or antecedence relations associated with mixed conjunct antecedents. Although the ADB does not make these distinctions yet, it is possible to find most of the relevant data we already have, e.g., it is already possible to search for sentences with an anaphor that has a conjoined antecedent, for example.

The existence of upward C-agreement mentioned earlier was found incidentally in Afranaph data elicitations designed to explore other properties. A few such sentences led to the ongoing studies of Ikalanga by the PI and Prof. Rose Letsholo. That study will soon extend to Setswana, which also has C-agreement and for which new elicitations are planned. The new focus on antecedent agreement has brought this pattern to prominence in our study, since this is a form of agreement that shows person agreement, just as argument anaphors do, but there is no obvious semantic sense of antecedency.

Primary responsibility for the theoretical and analytic work rests with the PI, but many researchers already involved in Afranaph research will be contributors and collaborators in various ways for this research initiative, particularly Prof. Alexis Dimitriadis of the U. of Utrecht, Prof. Mark Baker of Rutgers, Prof. Vicki Carstens of the U. of Southern Illinois, Prof. Michael Diercks of Pomona College, Prof. Claire Halpert of U. of Minnesota, Prof. Ruth Kramer of Georgetown University, Prof. Troy Messick of Rutgers, Prof. Philip Ngessimo Mutaka of U. of Yaounde 1, Prof. Justine Sikuku of Arap Moi University, and Prof. Jenneke van der Wal of University of Leiden.

4.0 How we plan to expand Afranaph resources relevant to our project

Using the Afranaph Project infrastructure, including its current network of NSLCs, its recruiting, remuneration structure, data-handling operations, the website and the database, we will explore the theoretical questions outlined in the last two sections. In so doing, we will enhance our data sets with additional analytic tagging and analysis of existing data, elicitations of new data using new questionnaires designed for the purpose, and by adding new parameters of presentation and analysis into the ADB.

The data we already have on many of these languages could be better categorized for the purposes of the research questions we are now interested in. For example, data in our notes and published .pdfs include full agreement paradigms for the most typical cases of antecedent agreement, subject verb agreement for most Afranaph languages and for C-agreement in the languages that have it, but the database does not have a display entity for agreement paradigms that allows them to be searched and cross-referenced. Questions like the following are relevant: Are the exponents of AcA the same as those of SVA? Do C-agreement exponents in a paradigm more closely match AcA agreement exponents or SVA agreement exponents (or some other paradigm, such as possessive agreement)? Does a given paradigm have or lack person, or gender/noun class, or number. Are there agreement paradigms for less common forms of agreement, such as tense, voice or definiteness exponents? It is then possible to search for all agreeing C-types that use an SVA agreement paradigm as opposed to those that use an AcA paradigm, or else a paradigm similar to neither. Thus by adding an 'agreement paradigm' (AGR-paradigm) analytic entity to the Anaphora Portal, useful searches become possible without ever altering the sentence data (which, however, can be used as evidence to establish the agreement paradigms).

Using properties of C-type, anaphoric markers and AGR-paradigms, one could determine, for example, whether there are languages where 'say' complementizers agree like verbs, like adjectives, or according to some other paradigm. Divergence between C-agreement paradigms and SVA (or adjectival agreement, etc.) generate new questions for how Agree should be formulated or when it should apply. The AGR-paradigm entry would also allow searches for languages where tense and/or voice participate in agreement paradigms, e.g., for C-types in Ikalanga. As we expand our research to richer C-agreement

paradigms for Limbum, Ibibio, Setswana and Kinande (or for languages where NSLCs are not already recruited) in addition to those for Ikalanga and Lubukusu, searches of this kind will become more significant. The addition of AGR-paradigm is just one possibility among many for improving our search and presentation functions that will be explored. The key point is that the properties of new analytic entities can be populated from our existing data and this is one way in which we can make our present data more easily accessible for searches relevant to our current interests.

At this point it is unclear whether a new portal for the database is necessary for this research or if it is only necessary to enhance the Anaphora Portal with the AGR-paradigm analytic entity and perhaps others. As mentioned, the development of software to build new portals on the common data pool to serve different research projects was an innovation of our last grant, so that capability is available, if needed.

Second, we will be devising questionnaires to elicit new data. For example, thanks to pioneering work by Lydia Newkirk, a Rutgers grad student working on the project, we have a new perspective on long distance agreement for binding by proleptic arguments. Proleptic arguments (see, e.g., Salzmann, to appear) are topical objects of thought or conversation that can be introduced as though they are additional arguments of a verb of saying, believing or some psychological predicates that furthermore must bind a (pronominal) variable in clausal complements. In English, for example, we cannot say "I believe (it) of Bill that the earth is flat" but we can say "I believe (it) of *Bill* that it is impossible that people could like *him*." Newkirk (2018), shows that prolepsis binding in Lubukusu of one kind is sensitive to islands, but another kind of prolepsis binding is not. The distinction seems to reside in whether the proleptic argument in the matrix predicate is introduced by a preposition or applicative marker, or if the proleptic argument is only a verbal affix (an object or reflexive marker) on the matrix verb. In the latter case there is island sensitivity, but not when the proleptic argument is introduced by applicative or a preposition. She shows that in languages where 'long distance agreement' is reported (e.g., Polinsky and Potsdam, 2001, Branigan and MacKenzie, 2002 a. o.), the semantic restrictions on prolepsis appear to be exactly the same as in Lubukusu prolepsis. Moreover, the movement analysis she proposes for affixal prolepsis, closely matches proposals for long distance agreement phenomena in the references cited. At present, we do not know if this sort of binding aligns with resumptive pronouns in relative clauses, but the shape of the bindee for prolepsis binding is different from that found in resumptive relatives. With respect to both resumption and the 'proleptic affix', the role of verb-affix object markers as either clitics or agreement markers is another issue that comes into play (and that is precisely the focus of the Object Markers in Bantu sister project for which we tested, and then posted, a new questionnaire in 2017). It is not year clear how these contrasts play out in a general theory of AcA, in part because we don't have enough information about long distance bound variable relations in these languages, including information on pronouns bound by quantifiers, resumption structures and logophoric binding. Logopohoric binding is widely attested in African languages (see Hagège, 1964, Clements, 1975, Culy, 1994, Schlenker, 1999, Safir, 2005, Adesola, 2005, Anand, 2006 and also Messick, 2017, for richer references). Thus new questionnaires that systematically elicit information on long distance binding relations (particularly resumption, logophoricity, and prolepsis) would be among the new elicitation tools that would be developed and deployed as part of the work of the grant, as well questionnaires on agreement with mixed conjuncts, comitative summing and the relation of these to antecedent agreement, as mentioned earlier. A new questionnaire specifically focused on C-agreement, for those languages that have it, is currently being constructed.

When we devise a new questionnaire, Afranaph procedure is to test it by working with our most responsive consultants to monitor whether the questions and model sentences are user friendly, whether they are enabling us to gather the right sort of data, and whether the data we are collecting is sufficiently comprehensive to provide a good base for comparison once other questionnaire responses come in. In the past, Afranaph postdocs have played this role and this is the most efficient way to get a project underway. Unfortunately, if we are to meet our budget target, we cannot include funds for a postdoc for this grant.

The PI does expect that the project will reveal interesting challenges to the Agree theory of AcA and the hypothesis that many agreement relations are mediated by voice, which may or may not be compatible with the Agree theory of AcA. While other researchers on the project may come to different

conclusions, the PI expects that this research will uncover patterns and puzzles that any theory of agreement or antecedency will have to address.

5.0 How our results are shared

The data gathered by Afranaph, once it is vetted and properly prepared for presentation, is freely available to the public through our website and database without any password protection/barrier. Only entry of new data is password protected. We also provide case files for each Afranaph language that features much of our primary data in .pdf form, so that more NSLC commentary is available in context and some of the back and forth between the Afranaph analyst and the NSLC is also available for outside scrutiny and edification. The case files also include featured information about the NSLC for that language, as well as grammar sketches and anaphora sketches, where these have been prepared by Afranaph researchers or NSLCs. Increasingly we also try to include select bibliographies in the case files for those who want to dig deeper into languages in the project.

The ADB, however, has become our featured resource, and like most databases, there are barriers to use that have nothing to do with passwords and everything to do with how user friendly that database is. The ADB has some design features that are easier to navigate with training, though it is most typically tried out by those who have no training at all. There are guides on our website to using the ADB which are periodically updated to insure new features are explained. In order to provide a tutorial aimed at linguists who might come to the Afranaph Database with a particular research question, the PI has published an essay in *Linguistics Vanguard* (Safir, 2016) that shows how the ADB can be used to search for meaningful generalizations with a particular theoretical question in mind. There is also a guide to the criteria we use to define analytic entities and to attribute properties to analytic entities. Any user can then use these documents to challenge these attributions based on our own data, either with respect to the way our definition is formulated or the way it is applied in a given case. Since all of our sentence data is available to any researcher, it can be disputed or re-evaluated when any claim based upon it is presented in the literature, that is, the primary data is always available and it is richer than what might be discussed or briefly alluded to in any particular article. Finally, the PI periodically gives presentations demonstrating how the site can be used (e.g. at UC Berkeley, SOAS in London, ZAS Berlin, U. of Education in Winneba, Ghana and SUNY Stony Brook).

All of the elicitation documents that have been developed by Afranaph are part of our intellectual contribution to the field. They are posted on our website for free download in .pdf and MS word versions and our NSLCs access them from there. Other websites designed to support fieldwork research post some of our questionnaires (e.g., the Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology, TulQuest of CNRS) or report elicitations for non-African languages that use them (e.g. California Language Archive, see session 19 March, 2015). We expect the new elicitation documents that we develop will be freely available to be used in the same way.

In the past, the Afranaph Project has held two Afranaph Project Development Workshops (programs on our site) which bring together Afranaph researchers, NSLCs, those using Afranaph resources, theoretical linguists interested in African languages and others interested in Afranaph technology or particular languages in the project. These workshops were open to the public and well-attended. There are tentative plans for a third Afranaph Development Workshop to take place in Spring, 2019, though that is not part of our funding request. The Workshop is, however, another way that we make our research available to the public.

The goal of this research is to contribute to the theoretical and empirical understanding of agreement patterns and their origins. The best way to reach the target audience interested in these questions is to publish where the target audience reads. In the past, research conducted by Afranaph affiliated scholars or using Afranaph data has been published in refereed journals (such as *Linguistic Inquiry, Natural Language and Linguistic Theory, Syntax, Lingua, Journal of West African Linguistics*) and presented at conferences (such as the Annual Conference on African Linguistics, West Coast Conference in Formal Linguistics, World Congress of African Linguists, International Congress of Linguistics), sometimes also appearing in the proceedings. One Afranaph-inspired paper appeared in 2017 in an edited volume from Cambridge University Press and another will appear in an edited volume next

year. Several papers on Lubukusu are tentatively destined for (re-)publication in a volume edited by Michael Diercks for Language Science Press. In addition, working drafts of Afranaph research are often published on our website as Afranaph Technical Reports. We expect to present the results of our new research in most, if not all, of these venues.

6.0 Conclusion

To summarize these efforts in other terms, the deliverables of this grant include an empirically enhanced public website, an empirically and structurally enhanced public database, and publications on central questions in all the usual academic venues of empirical and theoretical linguistics research, as well as training, mentoring, and community-building.

The broader impact of our work consists in effectiveness of the Afranaph Project in building a research community of those interested in the structure and comparison of African languages. Our website and database are at once resources and tools of organization for that community. The Afranaph Project provides training and mentoring for many who participate in it, including many native speaker linguists who would otherwise be isolated from contact a broader research community. For our consultants, participating in Afranaph is also a productive way to expose the language(s) they speak (typically underreported or even endangered ones) to interested scholars, and in some cases, these contacts result in projects outside of Afranaph that serve their speech communities.

The intellectual merit of the work consists in the theoretically-driven empirical research that we believe will help to delineate the boundaries between essential operations in grammar, in particular, those that effect agreement and anaphora, and the empirical patterns that are created by or influenced by these interactions. Whether there will be (a set of) essays that change the way we think about these questions will depend on what we find out, but whatever our results turn out to be, the theories of agreement, anaphora and the relations between them will based on a much richer empirical foundation.

Results of Prior NSF Support

The NSF supported Afranaph Project (NSF BCS-0303447, -0523102, -0919086) has used internet access and the expertise of NSLCs to elicit rich and specialized data aimed at discovering new empirical patterns within the African languages that are relevant to research questions in theoretical linguistics. All Afranaph research results and primary data are freely available on the internet for the use of any interested party. Currently, NSF BCS-1324404 supports an initiative to develop new research projects employing Afranaph infrastructure and resources to investigate empirical domains new to the project in a way that enhances the value of our existing resources and returns value to each of the contributing projects. Current funding serves 5 goals simultaneously, the first two new to this grant: (1) To expand our research model and our infrastructure for application to new investigations, some led by Afranaph associates at other institutions and NSLCs. (2) To enable the discovery, documentation, analysis and explanation of new empirical patterns of interest to linguistic theory and. (3) To insure that all new data collection enriches the effectiveness of existing resources, including the ADB, our website, and our elicitation documents. (4) To build an integrated research community around our resources and (5) to provide training and research opportunities for Ph.D. students and for African linguists.

Intellectual Merit: The intellectual merit of the project is embodied not only in the expansion of our empirical resources, but in the research it has produced. Our theoretical results include work on the nature of anaphoric relations (including foundational and typological work), the new proposals about the morphosyntax of complex words, the nature of agreement relations (complementizer agreement phenomena in particular), and some novel discoveries about the nature of clausal selection.

Six Afranaph Sister Projects got underway. These include research teams working on different projects, each with its own elicitation document. Most active were the Anaphora Project, the Tense and Aspect Project, and the Clausal Complementation Project (CCP). The new CCP questionnaire elicited data for 8 languages now entered into the database – other responses are in process. The Object-marking in Bantu Project was initiated by Michael Diercks and its new questionnaire has just completed testing (on Tsonga-Changana with NSLC David Langa). In all, new data has been collected, analyzed and posted for Akan, Esahie, Ewondo, Ikalanga, Joola Eegimaa, Kinyarwanda, Limbum, Lubukusu, Kinande,

14

Makaa, Setswana, Tsonga-Changana, and Yoruba amongst others. Afranaph questionnaires have since also been posted on independent websites devoted to linguistic fieldwork.

The development of the multiportal design for the ADB design that allows researchers with different interests to organize Afranaph data differently according to their needs is a major result of this grant period. The PI was invited to demonstrate the new design both for interested Africanist linguists, and for those interested in natural language database design as mentioned in the project description.

All of the following papers were published or are slated to appear based on work during this grant period: Baker, Safir and Sikuku (2013), Bassene and Safir (2017), Carstens (2016), Carstens and Diercks (2013), Diercks and Sikuku (to appear), Newkirk (2018), Owusu and Safir (in preparation), Safir (2013, 2014, 2016, to appear), Safir, Baker and Sikuku (to appear), Safir and Bassene (2016), Safir and Letsholo (to appear), and Safir and Selvanathan (2017). In addition, six papers were published on the Afranaph site as Technical Reports (TR) #9-14 during this grant period, four of which were subsequently published in the list above and two of which are under review: TR#9 Implications of Xhosa expletive constructions, Vicki Carstens and Loyiso Mletshe and TR#12 Linkers and relators in Kinande, Patricia Schneider-Zioga & Philip Ngessimo Mutaka. New questionnaires are currently under development and may be completed before this grant period ends. Other products of our work include the Afranaph Website, which is continually expanding with new case files and new features and the Afranaph Database as described above.

Broader impacts: The broader impacts of the Afranaph Project mostly concern service to the field of theoretical linguistics and to African linguistics in particular, with some ancillary effects. These impacts are a natural part of the way Afranaph functions and so some of what follows below is drawn from the broader impacts statement in the current proposal description. These goals include (1) To improve the empirical underpinning of linguistic research by (a) putting claims about Afranaph languages in the context of richer data and (b) allowing access to the full range of evidence used to establish any empirical generalization. The focus on African languages and using the Afranaph Project as a vehicle for that enterprise in turn supports a broad, vibrant and growing research community of African and Africanist linguists for whom Afranaph resources are a focus. Over 45 Afranaph NSLCs participate in Afranaph research, many from the time they are graduate students (see mentoring), while others are established scholars, including Afranaph Sister Project directors involving a total of 13 researchers from North America, Europe and Africa and permanent arrangements with Georgetown and U. of Utrecht.. The concerns of our consultants occasionally involve Afranaph in projects designed to benefit particular speech communities, such as the Kinande Dictionary Project or the Babanki community school, though no NSF funds are used to support these endeavors. Moreover, some of the languages in the Afranaph Project happen to be endangered, so the data that we collect plays a role in the documentation of those languages (see Bassene and Safir, 2017). An innovation this year represents a coming of age for the Afranaph community; A board of directors has been organized to which the Afranaph Project Director will report. The new board will oversee current research, current resources, and plan new research and new directions for the project. The board insures continuity after the PI is no longer the project director.

The ADB has lately emerged as an internationally consulted resource used for research (both theoretical and areal) and instruction (as a fieldwork tool, as a resource for student projects) and as a model for other databases. The ADB was studied as a model for databases under construction or currently online developed at ZAS in Berlin and at SOAS in London (the PI was invited as a consultant for both). The PI has been invited to conferences specifically focused on the development of linguistic databases. The ADB is cited as a basic resource for a major research grant on the anaphor-agreement effect at U. of Leipzig as well as for instruction at UC Berkeley, and the National U. of Singapore. Thus the design of the ADB has made a contribution beyond the particular scientific problem it is used to address.