ESRC
Polaris House, North Star Avenue, Swindon, Wiltshire,
United Kingdom SN2 1UJ
Telephone +44 (0) 1793 413000
Web http://www.esrc.ac.uk/

COMPLIANCE WITH THE DATA PROTECTION ACT 1998

In accordance with the Data Protection Act 1998, the personal data provided on this form will be processed by ESRC, and may be held on computerised database and/or manual files. Further details may be found in the **guidance notes**

Research Grants PROPOSAL

Document Status: With Council
ESRC Reference: ES/N018389/1

Research Grants (Open Call)

Organisation where the Grant would be held

Organisation	University of Edinburgh	Research Organisation Reference:	EngThaiVietKithU20
Division or Department	Sch of Philosophy Psychology & Language		

Project Title [up to 150 chars]

Connecting cognitive biases and typological universals in syntax

Start Date and Duration

a. Proposed start date	01 July 2016	b. Duration of the grant (months)	36

Applicants

Role	Name	Organisation	Division or Department	How many hours a week will the investigator work on the project?
Principal Investigator	Dr Jennifer Culbertson	University of Edinburgh	Sch of Philosophy Psychology & Language	7
Co-Investigator	Professor David Adger	Queen Mary, University of London	School of Languages Linguistics and Film	3
Co-Investigator	Dr Klaus Abels	University College London	Linguistics	4.32

Objectives

List the main objectives of the proposed research [up to 4000 chars]

A classic question in linguistics is the extent to which universal features of our cognitive system determine our linguistic behaviour. For example, we know than there is extensive variation in the linguistic systems of the world, but are certain patterns more learnable than others? Is natural language syntax shaped by the human cognitive system, and if so how? Linguistic typologists have identified an intriguing source of evidence: cases in which particular syntactic patterns appear to be over- or under-represented in the world's languages. These are often called typological universals. However, the extent to which typological universals reflect cognitive constraints or biases, rather than external factors like genetic or areal relationships among languages, remains heavily debated. This project aims to significantly advance this debate, by accomplishing the following objectives:

1) We will use artificial language learning experiments to test cognitive explanations for a theoretically important set of typological universals in syntax: Greenberg's Universal 20, word order harmony, and the suffixing preference. For each universal, we use previous work in linguistics to formulate biases hypothesized to underpin it. We then create artificial languages designed to compare how learners treat patterns which conform to or violate these hypothesized biases. If they are treated differently, we have evidence that cognitive biases indeed affect language learning, and therefore may shape natural language syntax.

2) We will test for the presence of cognitive biases by comparing whether speakers of typologically distinct languages master artificial languages in the same or distinct ways. This represents an important step forward in attempts to connect typology to cognitive biases; it allows us to assess whether these biases are truly universal, and the extent to which they are modulated by the language one has already learned. In addition to English, we target speakers of Thai, Vietnamese and Kitharaka (a language of the Bantu family, spoken in Nairobi, Kenya).

3) We will assess the strength of cognitive biases hypothesized on the basis of strong versus weak typological differences. Much of the evidence for typological universals comes from comparing the frequencies of alternative linguistic patterns across a sample of languages. When such frequency differences have been subjected to statistical testing, some are robust while others are relatively weak (or statistically non-significant). We test whether the strength of hypothesized biases is modulated when typological differences are larger or smaller.

4) We will compare results across two artificial language learning paradigms which exploit, respectively (i) learners' treatment of variability in the input and (ii) learners' inferences under conditions of impoverished input. These paradigms have been used in previous work investigating biases in phonology and syntax, but they potentially tap into distinct mechanisms for uncovering bias. The extent to which they show converging results has not been investigated systematically.

5) More generally, the vast majority of research using artificial language learning methods to investigate how cognitive biases might shape language has been undertaken in phonology. We aim to bring these methods further into the mainstream in syntax by:

-Significantly increasing the number of studies which use artificial language learning methods to test theoretically important issues in this domain.
-Conducting a week-long tutorial for researchers interested in using these methods in syntax.

Importantly, our project will thus expand the empirical base of syntax, and feed back into theoretical proposals and computational models of constraints on human grammar. More broadly, our results will provide insights into what types of patterns are preferred by the human cognitive system and how prior (linguistic) experience shapes these preferences.

Summary

Describe the proposed research in simple terms in a way that could be publicised to a general audience [up to 4000 chars]

Languages can be very different from each other. For example, just focussing on the order of words, languages like English put adjectives before nouns ('red house') while languages like Thai put them afterwards ('house red'). Similarly, languages like Vietnamese put Numerals before nouns ('three houses'), while others, like the Kitharaka (spoken in Kenya), put numerals after ('houses three'). If word ordering was simply due to happenstance, we would expect to see all different orders appearing in equal proportion across languages, but we don't find that. In fact, some orders are very common, some are very rare, and some don't seem to appear at all. For example, many languages are ordered like English ('three red houses'), and many are also ordered like Thai, which is exactly the reverse ('houses red three'). But the Kitharaka order ('houses three red') is much rarer, and its mirror image ('red three houses') never seems to occur. Why is this?

One of the major controversies in the language sciences is whether we need to appeal to the basic set-up of the human mind to explain the ways languages can vary, or whether these properties are instead a result of cultural differences in communication and social interaction. A great deal of recent work coming from the perspective of psychology assumes the latter: that the properties of language can be boiled down to communication, interaction and the vagaries of history, while most work in linguistics assumes the former: there must be biases in the human mind that allow us to learn languages of particular types more easily than others. This project seeks to resolve that issue.

In order to do this, we test how well people learn languages of various types, to see whether their behaviour follows the general tendencies we see across real languages. Importantly, we use artificially constructed languages, rather than natural languages, in order to make sure that they only differ in the crucial respects. For example, we present English speakers with artificial languages that use word orders from Thai and Kitharaka. If Thai orders are more common across languages than Kitharaka ones because the former are easier to learn, then we should see this reflected in the behaviour of learners in our experiments. We can also see whether such patterns are always harder to learn, or if speaking a language which uses them-like Kitharaka-makes them easier to pick up in a new language. To do this, our experiments compare English, Thai, Vietnamese and Kitharaka speakers. If our learners all show the same kinds of patterns in how they learn our artificial languages that we find across real languages, that will suggest that the way languages vary is not random, nor is it entirely a product of historical facts. Rather it would suggest that there are universal cognitive biases at play.

We plan to look at not just the basic question of what orders appear, but also two other well-known cases where languages don't seem to vary randomly. The first relates to how words like adjectives and numbers are placed relative to the nouns they modify: most languages place them both before or after (like English and Thai), rather than putting them on opposite sides (e.g., 'two houses red', like Vietnamese). We will test whether this type of pattern is always easier to learn in a new language. Second, we will look at whether people prefer to learn languages with suffixes (e.g., 'cat-s') rather than prefixes (e.g., 'un-happy'). Both types are present in English, but most languages have (more) suffixes. Our project we will shed light on whether there are universal cognitive biases in language learning, if such biases are at play for the particular phenomena we look at, and how people's native languages affect these biases.

Academic Beneficiaries

Describe who will benefit from the research [up to 4000 chars].

Relevant academic communities.

Our key stakeholders are researchers in cognitive science, among whom this research seeks to build new bridges. A central assumption of generative linguistics is that language variation is shaped by constraints on the language faculty. However, in the wider cognitive science community this assumption has met with skepticism. This project takes such skepticism seriously, and aims to provide direct behavioral evidence of the kinds of constraints proposed by linguists (though we formulate them as cognitive biases). Interdisciplinary collaboration like this is crucial; linguistics benefits from methods, explanations, and ideas coming from the broader psychological community, and that community in turn benefits from the detailed knowledge about language structure which is the purview of linguistics. It is also timely, given recent high

profile debates centered around these issues (Evans & Levinson, 2009; Dunn et al, 2011; a.o.).

The results of this project will also be of independent interest to researchers in a number of fields. The specific constraints targeted here will be of interest to syntacticians. The general issues and methods used also makes the research relevant to phonologists. The kinds of biases we investigate may contribute to shaping how languages change. Our focus on learning as a window into biases makes our results of interest to language acquisition and language development researchers. Any biases we uncover can be formalized in models by computational cognitive scientists. Finally, the project will identify patterns that pose special challenges to adult learners, and show how prior language knowledge affects biases-with clear implications for bilingualism, second language learning and applied linguistics.

Dissemination to those academic communities.
Our plan for dissemination seeks to engage the interdisciplinary community identified above through (1) conference presentations and journal publications, and (2) a continuing development workshop.

Journal publications: Publications represent the main route to dissemination. We will target high profile journals in linguistics and cognitive science (Language, Syntax, Cognition) as well as generalist journals (e.g., PNAS).

Conferences: We plan to attend two conferences per year, with a mix of UK and international venues: LAGB and the LSA meetings, which reach a broad audience of linguists in the UK and USA; AMLaP, attracting researchers in psycholinguistics, language acquisition, and cognitive modeling; BUCLD, a top international conference for language acquisition; Cogsci, the premier venue for multidisciplinary work in the cognitive sciences; Linguistic Evidence, a conference intended to bring together linguists interested in improving the empirical adequacy of linguistics.

Workshop on ALL methodology. One of our key aims is to promote the use of artificial language learning (ALL), thus our plan for dissemination to the academic community involves capacity building. To accomplish this, we will hold a 3-day long workshop on ALL design, implementation and data analysis. The PI has relevant teaching experience (a related course at the Advanced Core Training in Linguistics Summer School, run by Co-I Klaus Abels). We will invite two internationally recognized researchers with complementary expertise to teach part of the workshop (one from the US, and one from the UK). The workshop will be held in year 2 of the grant, at Edinburgh, organized by the PI and Co-I David Adger with assistance from knowledge-exchange and Bilingualism Matters staff. It will be open to researchers at any level, and will be video-recorded for wider dissemination. A number of resources at Edinburgh are in place to help us achieve this, e.g., resources for recording/editing video, and a blog and YouTube channel hosted by the Linguistics and English Language department.

Staff Duties

Summarise the roles and responsibilities of each post for which funding is sought [up to 2000 characters]

PI and Co-Is will together manage the project, design the experiments, supervise data collection, and disseminate the results. The PI will be responsible for experiment implementation and data analysis, with assistance from the postdoc where appropriate. More specifically:

Jennifer Culbertson (PI)
- Coordination between Edinburgh, QMUL and UCL teams
- Supervision of postdoc
- Supervision of field research in Thailand and Vietnam (with postdoc)
- Design of experiments (with CO-Is)
- Programming/implementation of experiments (with postdoc)
- Analysis of data collection (with postdoc)
- Writing up results for conferences and journal articles (with Co-Is and postdoc)
- Presenting at conferences (with Co-Is and postdoc)

- Organization and running of continuing development workshop (with Co-I David Adger)
- Organization of Bilingualism Matters event (with support from department knowledge-exchange officers, and Bilingualism Matters research coordinator Dr. Madeleine Beveridge)
- Coordination and participation in other knowledge exchange activities related to the grant (with Co-Is and postdoc)

David Adger (Co-I)
Duties shared with PI as listed above:
-Design of experiments
-Write-up and dissemination activities
-Knowledge exchange activities including continuing development workshop

Klaus Abels (Co-I)
Duties shared with PI as listed above:
-Design of experiments
-Write-up and dissemination activities
-Knowledge exchange activities (particularly events in Kenya)
And:
- Supervision of field research in Kenya (with postdoc)

Postdoctoral researcher (supervised by PI)
Duties shared with PI as listed above:
- Design of experiments
- Programming/implementation of experiments
- Analysis of data collection
-Write-up and dissemination activities
And:
- Recruiting and running participants

Impact Summary

Impact Summary (please refer to the help for guidance on what to consider when completing this section) [up to 4000 chars]

This project investigates how cognitive biases shape language learning, and thus ultimately the distribution of linguistic patterns across the languages of the world. The approach we use is distinctly interdisciplinary: we target hypotheses put forward by linguistics using methods developed by psychologists and psycholinguistics in order to answer questions that have broad implications for the cognitive science of language. Beyond its relevance to this academic community, the project speaks to questions that are of interest to the general public. For example, what are the deep common features of human psychology that shape the unique trait of language that we share? How do these features interact with our experience, culture, and history to produce the rich variety we see across the world's languages?

In addition to these big picture questions, our research also has important implications for language learning-in particular how cognitive biases and our prior experience together impact how we learn a new language. For example, our results will show how prior language experience-coming from both first and second language knowledge-affects adults' ability to acquire linguistic patterns hypothesized to present a learning challenge. These findings will have clear implications for language education practitioners who work with learners from a range of linguistic backgrounds (e.g. English for Speakers of Other Languages (ESOL) support workers, or primary school staff teaching modern languages to children with English as an Additional Language). The research also speaks directly to an important strand of current research which seeks to uncover the effects, both positive and negative, of multilingualism. Bilingualism Matters is a Centre at the University of Edinburgh, based in the same school as the PI. The Centre is dedicated to research and knowledge exchange around language education and bilingualism and has a proven track record in shaping public policy in this field (e.g., advising Scottish Government on their 1+2 policy for language learning in primary schools). The proposed grant will benefit from the

Centre's existing networks (including language practitioners, key organisations such as Education Scotland and Scotland's National Centre for Languages, and members of local and national government). These non-academic stakeholders will be invited to contribute to the project from the outset (e.g., by discussing how the research and tools developed in the project might feed into Continued Professional Development for practitioners), as well as being key audiences for the dissemination of research results.

Ethical Information

Has consideration been given to any ethical matters raised by this proposal ? Yes

Please explain what, if any, ethical issues you believe are relevant to the proposed research project, and which ethical approvals have been obtained, or will be sought if the project is funded? If you believe that an ethics review is not necessary, please explain your view (available: 4000 characters)

Ethical approval from University of Edinburgh will be sought for the proposed experiments. The department of Linguistics and English Language at the University of Edinburgh complies with the standards set forth by the University following the UK Research Integrity Office's Code of Practice for Research. Adults participants will be provided with an information sheet describing the nature of the research and will be given the opportunity to ask questions. They will then provide written informed consent. All documents will be in the native language of the participant (English, Thai, Vietnamese, Kitharaka), and the experimenter present will be a native speaker of this language. All experimenters will be provided with training and supervision from the PI and/or postdoctoral research fellow in ethical research practices. All participants will be assigned an anonymous code number. Data will be stored with this code number only: no identifying information such as name and address will be recorded. See Data Management Plan for additional details.

Summary of Resources Required for Project

Financial resources

Summary fund heading	Fund heading	Full economic Cost	ESRC contribution	% ESRC contribution
Directly Incurred	Staff	132988.00	106390.40	80
	Travel & Subsistence	25272.00	20217.60	80
	Other Costs	23856.00	19084.80	80
	Sub-total	**182116.00**	**145692.80**	
Directly Allocated	Investigators	83531.67	66825.34	80
	Estates Costs	37304.00	29843.20	80
	Other Directly Allocated	17980.00	14384.00	80
	Sub-total	**138815.67**	**111052.54**	
Indirect Costs	Indirect Costs	177520.00	142016.00	80
Exceptions	Staff	0.00	0.00	100
	Other Costs	0.00	0.00	100
	Sub-total	**0.00**	**0.00**	
	Total	**498451.67**	**398761.34**	

Summary of staff effort requested

	Months
Investigator	14.00
Researcher	36
Technician	0
Other	0
Visiting Researcher	0
Student	0
Total	50

Other Support

Details of support sought or received from any other source for this or other research in the same field.
Other support is not relevant to this application.

Staff

Directly Incurred Posts

			EFFORT ON PROJECT							
Role	**Name /Post Identifier**	**Start Date**	**Period on Project (months)**	**% of Full Time**	**Scale**	**Increment Date**	**Basic Starting Salary**	**London Allowance (£)**	**Super-annuation and NI (£)**	**Total cost on grant (£)**
Researcher	Postdoctoral Research Assistant	01/07/2016	36	100	UE07	01/08/2017	33242	0	8606	132988
									Total	132988

Applicants

Role	Name	Post will outlast project (Y/N)	Contracted working week as a % of full time work	Total number of hours to be **charged** to the grant over the duration of the grant	Average number of hours per week **charged** to the grant	Rate of Salary pool/banding	Cost estimate
Principal Investigator	Dr Jennifer Culbertson	Y	100	924	7	56555	31671
Co-Investigator	Professor David Adger	Y	100	396	3	103820	24917
Co-Investigator	Dr Klaus Abels	Y	100	570	4.3	77996	26944
						Total	83532

Travel and Subsistence

Destination and purpose		Total £
Outside UK	Bangkok, Thailand (data collection trip for two researchers for a period of 2 weeks): £800 airfare + £100 local transport + £1000 accommodation + £200 subsistence = £2100 each	4200
Outside UK	Nairobi, Kenya (data collection trip for two researchers for a period of two weeks): £700 airfare + £100 local transport + £1000 accommodation + £200 subsistence = £2000 each	4000
Outside UK	Hanoi, Vietnam (data collection trip for two researchers for a period of 2 weeks): £1000 airfare + £100 local transport + £1000 accommodation + £200 subsistence + £500 government research fee = £2800 each	5600
Outside UK	Linguistic Society of America annual meeting, USA (dissemination trip): £190 registration fee + £700 airfare + £500 accommodation + £175 subsistence = £1565	1565
Within UK	Linguistic Association of Great Britain annual meeting, UK (dissemination trip): £155 registration fee + £120 train + £200 accommodation + £70 subsistence = £545	545
Within UK	Architectures and Mechanisms of Language Processing conference, Europe (dissemination trip): £230 registration fee + £300 airfare + £200 accommodation + £70 subsistence = £800	800
Outside UK	Boston University Conference on Language Development, Boston, USA (dissemination trip): £148 registration fee + £700 airfare + £500 accommodation + £175 subsistence = £1523	1523
Within UK	Cognitive Science Society Annual Meeting, Europe (dissemination trip): £412 registration fee + £300 airfare + £200 accommodation + £70 subsistence = £982	982
Within UK	Linguistic Evidence conference, Germany (dissemination trip): £37 registration fee + £300 airfare + £200 accommodation + £70 subsistence = £607	607
Within UK	Edinburgh, UK (yearly project meetings, 2 CO-Is to travel from London): £120 train + £200 accommodation + £70 subsistence = £390 each, in each year of the grant	2340
Outside UK	Edinburgh, UK (continuing development workshop, Co-I to travel from London, two experts to travel from UK and USA): £120 train + £200 accommodation + £70 subsistence for Co-I; £700 airfare + £300 accommodation + £105 subsistence for each expert (£1,105)	2600
Within UK	Edinburgh, UK (Bilingualism Matters event, 2 CO-Is to travel from London): £120 train + £100 accommodation + £35 subsistence = £255 each	510
	Total £	25272

Other Directly Incurred Costs

Description	Total £
Participant payments; 330 English speakers, 200 Thai speakers, 160 Kitharaka speakers and 160 Vietnamese speakers. 850 participants in total @ £7 each	5959
Laptops (3 MacBook Airs) @ £999 each	2997
Catering and promotional materials for Bilingualism Matters event: £1500 for research briefing (£400 editing, £200 design, £200 printing, £600 postage), £1375 for catering	2875
Catering and promotional materials for continuing development workshop: £1500 for promotional materials (£400 editing, £200 design, £200 printing, £600 postage); £525 for catering	2025

Ad-hoc field payments to field workers: an assistant each in Thailand, Vietnam and Kenya to help with data collection. Hourly pay is an average of £10 per hour for 300 hours work = £3,000 each researcher	9000
Recruitment and relocation costs for the Research Assistant post	1000
Total £	23856

Other Directly Allocated Costs

Description	Total £
Infrastructure Technicians	71
Pool staff costs	16259
Total £	16330

Research Facilities/Existing Equipment

Description	Total £
Lab costs	1650
Total £	1650

Timetable estimates of the number of months after the start of the project to reach the following stages:

Stage	Number of Months
Completion of all preparation and design work	5
Commencement of fieldwork or material/information/data collection phase of study	6
Completion of fieldwork or collection phase of study	30
Commencement of analysis phase of study (substantive phase where research facilities are involved)	8
Completion of analysis phase of study	32
Commencement of writing-up of the research	10
Completion of preparation of any new datasets for archiving	34
Completion of writing-up	36

Data Collection

If the research involves data collection or acquisition, please indicate how existing datasets have been reviewed and state why currently available datasets are inadequate for this proposed research. If you do not state to the contrary, it will be assumed that you (as principal applicant) are willing for your contact details to be shared with the affiliated data support service (UK Data Service) working with the Research Councils.	No experimental studies have been carried out which address the research questions set forth in the Case for Support. Our experiments are carefully designed to answer these questions and differ substantially from previous related research.
Will the research proposed in this application produce new datasets?	Yes
Will this data be:	✔ Quantitative Qualitative
Please give a brief description of the datasets.	Data will consist of participant responses, specifically, utterances produced or typed and/or buttons pressed in experimental sessions.

It is a requirement to offer data for archiving. Please include a statement on data sharing. If you believe that further data sharing is not possible, please present your argument here justifying your case.	Anonymized participant data will be archived and will be made publicly available in cases where participants give consent.
Who are likely to be the users (academic or non-academic) of the dataset(s)?	The most likely users of the data are academic.
Please outline costs of preparing and documenting the data for archiving to the standards required by the affiliated data support service (UK Data Service) working with the Research Councils.	Preparing and documenting of data for archiving will be undertaken by the post-doctoral fellow under supervision of the PI.

OTHER INFORMATION

Academic Reviewers

1	Name	Address	Town	Email Address
Professor Jeff Lidz		University of Maryland	Linguistics	jlidz@umd.edu

Academic Reviewers

2	Name	Address	Town	Email Address
Professor Charles Yang		University of Pennsylvania	Linguistics, Computer Science	charles.yang@ling.upenn.edu

Classification of Proposal

(a) User Involvement

The nature of any user engagement should be indicated

Design	
Execution	
Dissemination	x
Training	
Not applicable	

Proposal Classifications

Research Area:

Research Areas are the subject areas in which the programme of study may fall and you should select at least one of these. Once you have selected the relevant Research Area(s), please ensure that you set one as primary. To add or remove Research Areas use the relevant link below. To set a primary area, click in the corresponding checkbox and then the Set Primary Area button that will appear.

Please select one or more Research Areas

Subject	Topic	Keyword
Linguistics	Linguistic Theory	
Linguistics	Linguistics (General)	
Linguistics	Psycholinguistics	
Linguistics	Syntax [Primary]	
Psychology	Cognitive Psychology	

Qualifier:

Qualifiers are terms that further describe the area of study and cover aspects such as approach and geographical focus. Please ensure you complete this section if relevant.

To add or remove Qualifiers use the links below.

Type	Name
Approach	Experimental
Approach	Technique/Method Development
Collaboration location region	Africa
Collaboration location region	Far East (China/Japan)
Collaboration location region	UK & Ireland
Public Engagement Audience	Audiences for diversity
Public Engagement Audience	Families
Public Engagement Audience	Government/policy-makers
Public Engagement Audience	Primary and secondary schools
Public Engagement Audience	Scientists/engineers/academia
Public Engagement Audience	Street audience/passers-by
Public Engagement Audience	Teachers
Public Engagement Audience	Young people (outside school)
Public Engagement Methodology	Lecture
Public Engagement Methodology	Publication
Public Engagement Methodology	Social networking
Public Engagement Methodology	Training for researchers
Public Engagement Methodology	Workshop

Free-text Keywords:

Free-text keywords may be used to describe the programme of study in more detail. To add a keyword, you first need to search existing Research Areas by entering the keyword in the Search box and selecting the Filter button.

If the keyword is adequately reflected by one of the terms displayed below, click in the corresponding checkbox then select Save. If no potential matches are displayed, or none of those displayed are suitable, select the Add New button followed by the Save button to add it as a descriptor.

To add or remove those previously added use the links below.

Free-text Keywords
Artificial Language Learning
Cognitive biases
Language universals
Typology

Pathways to Impact

This project investigates how cognitive biases shape language learning, and thus ultimately the distribution of linguistic patterns across the languages of the world. As outlined in detail above, our primary academic stakeholders are members of the interdisciplinary community engaged in research on language science. The research questions set forward in our case for support have broad implications for this community, and will foster, we hope, greater collaboration among researchers. To reach this audience we have proposed, in addition to publications and conferences, a continuing professional development workshop aimed at educating researchers in the use of the methods we ourselves have developed and used. The workshop will also be advertised to ESOL practitioners in order to increase collaboration between researchers and practitioners, with a view to developing future partnerships, and to facilitate evidence-based practice. In addition to this, our project will have an impact outside of academia in three ways: (1) through public engagement and knowledge exchange about bilingualism and second language learning, (2) through public engagement about linguistics and language variation, and (3) through engagement with local language groups.

Public engagement and knowledge exchange about bilingualism: The proposed project will generate data which speak directly to a number of issues central to research on bilingualism. First, our experiments will shed light on how L1 experience might alter how we learn a second language (L2) and our susceptibility to cognitive biases. Second, by incorporating measures to explicitly quantify L2 experience, our project will also provide insight into how the particular features of an L2, *or* knowledge of an L2 in general, might help learners overcome biases when learning novel linguistic patterns. Third, previous research has found that in some cases (but not others), L2 experience can impact the use of the L1 (the most obvious case is language attrition, but there are also cases in which L1 processing or phonology are affected in quite subtle ways). This project will allow us to make predictions about the circumstances under which this might be more likely to happen; if the L1 incorporates properties that go against a cognitive bias, then it may be more susceptible to impact from an L2 which does not. Finally, because of the particular methods we use, results from our project will also contribute to the question of how much learning can be achieved on the basis of limited input, depending on the characteristics of the target language and its relationship with the L1. A better understanding of this has clear implications for second language learning in the classroom, where teachers must distribute the precious resource of training time on where it is most needed. Therefore, there are two key audiences for knowledge exchange: those involved in teaching modern languages to students (who may or may not speak English at home), and those involved in teaching English to Speakers of Other Languages (ESOL).

Both modern language teaching and ESOL are key policy areas for the Scottish Government. Scotland's refreshed ESOL strategy was launched in April 2015 to focus on a learner-centred approach, in which individual circumstances and needs are given more weight: knowledge of how particular language backgrounds might affect language learning therefore provide a clear link between research and practice in this field. Recent policy changes mean that Scottish and UK governments are implementing new language teaching requirements in primary schools. This project can contribute to the evidence base about language learning on the basis of limited input, which will be of use to teachers, school management and to policy officials.

To involve relevant stakeholders and enhance impact, we will work with Bilingualism Matters, a Centre at the University of Edinburgh, founded by Professor Antonella Sorace (www.bilingualism-matters.ppls.ed.ac.uk) with 16 branches around the world. The Centre carries out and shares high-quality research into bilingualism and language-learning, and places partnership working at the heart of its impact strategy. By working with Bilingualism Matters, the PI will be able to involve a wide range of non-academic stakeholders in the project, allowing their input to shape impact where possible. For example, the PI is already liaising with the City of Edinburgh's ESOL strategy leader to discuss which aspects of the research she anticipates would be

Pathways to Impact

most useful for ESOL practitioners. The project will be able to build on previous work at the Bilingualism Matters Centre, benefitting from existing partnerships with schools, community groups and cultural institutions, local authorities and key government agencies, ESOL and English as an Additional Language (EAL) services, the Scottish Refugee Council and Children in Scotland. In addition to its international branches, Bilingualism Matters has also collaborated with the European Commission's "Piccolingo" campaign, has obtained funding as a partner in two EACEA Lifelong Learning projects, and is part of the EU FP7 funded Advancing the European Multilingual Experience project, on which Co-I David Adger is also named. Thus, the PI would have access to a wealth of contacts beyond Scotland and the UK in order to facilitate impact.

Impact activities include dissemination of information and awareness-raising initiatives, creation of resources and training tools for parents, teachers and other professionals, and dissemination of information in policy contexts. Through the impact activities of Bilingualism Matters, our project will reach a large (non-academic) audience both during and after the research phase, allowing interested stakeholders (e.g., ESOL practitioners) to contribute their knowledge as well as to take away learning from the findings themselves. In addition, through this service, our research is more likely to play a role in shaping language policies. To begin disseminating our research, in the third year of the grant we will plan and host an event at the University of Edinburgh for relevant Bilingualism Matters stakeholders including families, teachers, relevant agencies like SCILT (Scotland's National Centre for Languages), and interested members of the Scottish Government. We expect that this event will also help us to identify groups most interested in our findings and thus lead to future events beyond the lifetime of the grant.

Engagement with the general public about linguistics and language variation: The field of linguistics suffers from a general lack of public awareness; most people do not know what linguistics is or what types of issues linguists study. Despite this, linguistics is in fact of broad interest to the public–it is the scientific study of a unique aspect of our species which reflects both deep common features of human psychology and cultural differences that result from transmission and evolution of languages over generations. Both these aspects are central to the proposed research, and thus we are in an excellent position to engage with the public on issues surrounding the richness of language variation and the ways we can understand it. The PI is a member of the Language Evolution and Computation research group at the University of Edinburgh, which has been active in conveying research like this to the general public. Co-I David Adger has appeared on Radio 4 and Australian Radio News to discuss linguistics in general, and Culbertson and Adger (2014) featured in the New Scientist and other popular media outlets including a piece in The Conversation which we wrote. Co-I Adger is currently the President of the Linguistics Association of Great Britain which is setting up a series of public lectures aimed at improving general understanding of language sciences. We will make use of this experience to disseminate our results via press-release, twitter, blogs (e.g., Language Log), and by writing articles for popular press sites. These activities will also be supported by Bilingualism Matters communication channels.

Engagement with local educators and language groups: The proposed research will create links between the UK institutions involved and the educators and local language groups in Kenya, Thailand, and Vietnam. These groups will benefit from (1) the infrastructure created as a result of the project, (2) new knowledge about the local languages generated by our research, and (3) increased prestige for local languages. The latter are particularly important for Kitharaka, a minority language of Kenya for which there is relatively little linguistic documentation. We will therefore speak with local press in Nairobi about our project. In interacting with speakers and local linguists we will also document and make publicly available additional data and analysis of this language.

Pathways to Impact

Introduction. A classic question in cognitive science is how languages are shaped by the human cognitive system. Intriguing evidence for this comes from *typology*, specifically, cases in which particular linguistic patterns appear to be over- or under-represented in the world's languages. Such cases suggest that linguistic patterns may differ in cognitive accessibility–some may, for example, be more easily learnable than others–and these differences are often built into influential linguistic theories. However, the link between typological frequency and cognitive accessibility is problematic for several reasons. Foremost among these is the inability to tease apart the factors that have historically influenced any particular distribution of language types–including genetic and areal relationships among languages, accidents of history, and other factors external to the cognitive system (Cysouw, 2005; Evans & Levinson, 2009; Piantadosi & Gibson, 2014). In recent years, psychologists and linguists have turned to experimental methods–in particular, artificial language learning (ALL) studies–to test whether differences in typological frequency are better explained by cognitive biases or by these cognition-external factors (e.g., Wilson, 2006; Finley & Badecker, 2008; Moreton, 2008; Culbertson et al., 2012). This body of research has made important contributions to models of language acquisition and adult grammar, but the majority of studies are in the domain of phonology–despite the fact that many of the most controversial claims about universal cognitive (or linguistics) biases come from syntax. This project therefore targets a set of so-called typological universals in syntax with the aim of investigating the cognitive biases hypothesized to underpin them.

Three typological universals of syntax:

Universal 20, concerning how nominal modifiers (demonstratives like 'this', 'that', adjectives and number words) tend to be ordered relative to each other, and to the nouns they modify (Exp. 1, 2).

Word order harmony, the tendency to order syntactic heads (e.g. the noun in a noun phrase like "that big house") consistently before or after modifiers and other dependent elements (Exp. 3, 4).

Suffixing preference, the tendency for morphological affixes (like plural "–s") to follow rather than precede lexical roots (Exp. 5).

For each universal, we use previous work in linguistics to formulate a bias (or set of biases) hypothesized to underpin it. **The key idea** is that *all* speakers should show these biases when they learn a new language, whether or not the critical features are actually present in their first language. Our studies therefore contrast typologically dissimilar languages, to see whether speakers learn new languages based on universally shared biases (i.e., learn in the same way), or based on their first language (i.e., learn in opposite ways). This goes well beyond previous work, that has used only on English-speaking participants (e.g., St. Claire et al., 2009; Culbertson et al., 2012; Culbertson & Adger, 2014). For example, Thai and Vietnamese differ in how they order elements in the noun phrase: Thai is *harmonic*, placing adjectives, numerals and other modifiers consistently after the noun; Vietnamese is *disharmonic*, it places some modifiers before, but others after. If harmony reflects a cognitive bias, speakers of both languages should nevertheless show evidence of a preference for harmony. The extent to which Thai speakers show this preference more *strongly* than Vietnamese speakers then tells us how learners' biases change as a result of language experience.

Research methods. We use two primary artificial language learning methodologies in this work. First, the **Mixture-shift paradigm** (MSP; Culbertson et al., 2012), which exploits learners' treatment of variability in the input. Second, the **Poverty-of-the-Stimulus** paradigm (POS; Wilson, 2003), which tests learners' inferences under conditions of impoverished input. Note that these paradigms potentially tap into distinct mechanisms for uncovering bias–the MSP tests what patterns participants find easier to learn in the face of noise, while the POS paradigm taps into what participants find most natural when they have been given no direct evidence from which to learn.

Case For Support

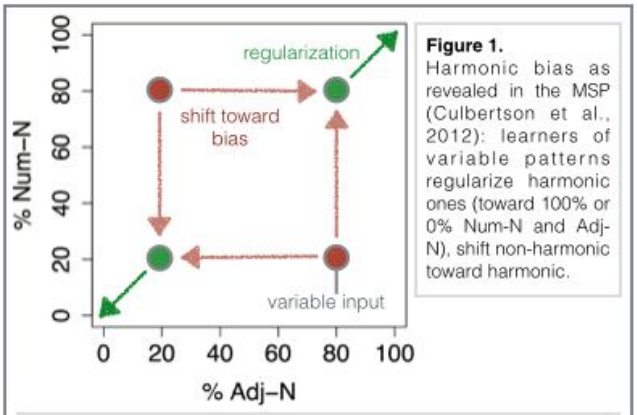

Figure 1. Harmonic bias as revealed in the MSP (Culbertson et al., 2012): learners of variable patterns regularize harmonic ones (toward 100% or 0% Num-N and Adj-N), shift non-harmonic toward harmonic.

The **MSP** was originally used by Hudson Kam & Newport (2005) to investigate regularization–when learners produce less variation than was present in their input. We use this paradigm to see whether regularization is modulated by hypothesized biases; learners are predicted to regularize a pattern more if it aligns with their biases. Previous research has successfully used the MSP to show that when the input contains variation in word order, learners regularize only when the dominant pattern present is harmonic (Culbertson et al., 2012; Culbertson & Newport, 2015). This paradigm is of particular interest in linking cognitive biases to language typology through language change; for example, Culbertson et al., (2012) and Culbertson & Newport, (2015) show that learners actually shift non-harmonic patterns toward harmonic ones.

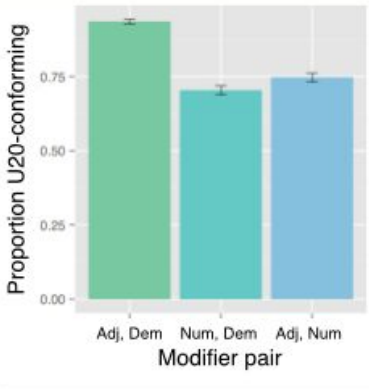

Figure 2. Inference of U20-conforming orders in the POS paradigm (Culbertson & Adger 2014): learners infer post-nominal orders with modifier position conforming to U20 rather than English.

The **POS** paradigm is extensively used in phonology to test for biases hypothesized based on typological data (see references in Moreton & Pater 2012a,b). In this paradigm, learners are presented with examples from a new language in a way that withholds critical evidence about its structure. At test, learners must generalize to held-out data that will disambiguate alternative hypotheses. The first use of this paradigm in syntax was our previous work (Culbertson & Adger, 2014). We trained English speakers on a miniature artificial language with N-Adj, N-Num, and N-Dem order, withholding phrases containing more than one modifier. At test, learners were asked to choose what they thought the relative order among modifiers was in the language. They were overwhelmingly more likely to choose relative orders of modifiers conforming to Universal 20, as we discuss in more detail below.

Strand 1. Greenberg's Universal 20.

The first set of experiments focus on Greenberg's (1963) Universal 20 (U20), which has generated significant interest in theoretical syntax. (Adj=adjective, Num=numeral, Dem=demonstrative)

U20	In pre-nominal position, nominal modifiers Dem, Num, and Adj (or any subset) are ordered *Dem-Num-Adj-N*. In post-nominal position the relative order is the same: *N-Dem-Num-Adj* or the mirror order: *N-Adj-Num-Dem*.

Since Greenberg's original formulation, additional typological data has revealed that alternative orders are in fact attested, but recent research has identified the following frequency asymmetries:

1	Orders with Adj closest to N, Dem most distal, and Num in between, are most common.
2	*Dem-Num-Adj-N* and *N-Adj-Num-Dem* are the most frequent of all.
3	Orders violating 1 are more likely to involve pre-nominal than post-nominal modifiers.

For each of these asymmetries we can formulate a bias based on extant theoretical work. The first we call **isomorphism**. This bias connects the semantic relations among elements in the noun phrase to their linear order. Several lines of research (e.g., Partee, 1987; Adger, 2003; Rijkhoff, 2004) suggest that among nominal modifiers, Adj has the tightest semantic relationship to N (modifying properties inherent to it), while Dem has the most distant relationship (connecting nominal material to the outside discourse). This conceptual nesting can be seen schematically in Figure 3. Linear order can either preserve this nesting–that is, maintain an isomorphism with the underlying semantic relations–or not. For example, N-Adj-Num-Dem is isomorphic (can be read directly off Figure 3) but N-Dem-Num-Adj is not. The second bias, already mentioned above, we call **harmony**. This bias states that heads (here, N) should be consistently before *or* after modifiers and other dependent elements. This may reflect simplicity: harmonic patterns have a single general

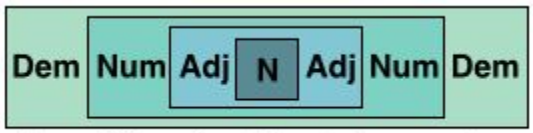

Figure 3. Semantic relations in the noun phrase

Case For Support

rule ordering modifiers, while non-harmonic patterns do not (Culbertson & Newport, 2015). Harmony is relevant beyond U20, and is in fact one of the most well-studied typological universals. The third bias we call the **N-first** bias. In general, noun-first orders are more common than noun-last orders. For example, there are 6 possible N-first orders and all of them are attested, including N-Dem-Num-Adj which violates the isomorphism bias (e.g., Kitharaka, a Bantu language spoken in Kenya). By contrast, very few languages have any pre-nominal pattern other than Dem-Num-Adj-N. *No* languages have been found with the N-last order violating isomorphism, i.e., Adj-Num-Dem-N. (Specific mechanisms are proposed in Cinque, 2005; Abels & Neeleman, 2012)

The three biases outlined here have been proposed (in various forms) in the theoretical linguistics literature on the basis of pure typological frequency data, from relatively small samples, in most cases with no statistical testing. In Culbertson & Adger (2014), we took the first steps toward investigating the cognitive underpinnings of the isomorphism bias using the POS paradigm. We showed that English speakers, whose first language (L1) is isomorphic (Dem-Num-Adj-N), choose an isomorphic order of *post*-nominal modifiers (N-Adj-Num-Dem) when learning a new language. Surprisingly they chose this over a non-isomorphic order that was more surface similar to English (N-Dem-Num-Adj). These results are consistent with two hypotheses: (i) speakers have learned from their L1 that isomorphism holds, and extend this to new languages, or (ii) the isomorphism bias is active in learners, regardless of the properties of their L1. This is a crucial difference, since only the latter hypothesis points clearly to a *universal* bias. In Experiments 1 and 2 we compare English, Kitharaka and Thai speakers. Kitharaka has a non-isomorphic order (N-Dem-Num-Adj), allowing us to test if the isomorphism bias is active in speakers of a language which violates it. To our knowledge, no previous work connecting laboratory learning results to language typology has attempted to provide this crucial kind of evidence. Thai is isomorphic and harmonic, but has post-nominal modifiers (N-Adj-Num-Dem), allowing us to test the N-first bias by comparing the strength of the isomorphism preference in English and Thai speakers.

Testing across L1s. Comparing across L1 populations is key to answering our research questions. In order to carry out the proposed research, we will test each non-English-speaking population with the help of local collaborators and research assistants as outlined in the Technical Appendix §2: materials will be prepared in advance of two-weeks visits by the team, testing will begin during the visit and continue until complete. While our primary goal is to manipulate L1 experience, one possibility, which has not yet been shown, is that second language (L2) knowledge might affect participants' learning outcomes. We will therefore control for L2 proficiency in all the populations we test (measured using a proficiency questionnaire, see Technical Appendix §3 for details), and additionally assess effects of L2 proficiency in the combined dataset.

Experiment 1a. The first experiment will replicate our English results in the lab (the original was a web-based study), and validate the paradigm for Kitharaka speakers. *Participants:* English and Kitharaka speaking adults (N=120; 20 per condition). *Procedure*: POS (for details see Technical Appendix §4, 5). During training, participants see a phrase in their L1, and then hear a translation in the new language. Training data are phrases with a noun and a single modifier—Adj, Num, or Dem—with no evidence about the relative order of modifier combinations. Conditions differ in having a subset or all three modifiers present during training; these led to different strengths of response bias in our previous study. English learners are taught post-nominal modifiers; Kitharaka speakers learn pre-nominal modifiers (see tables). At test, they again see a phrase to be translated, and then must choose the phrase "most likely to be used by a speaker of the language." The held-out data are phrases with multiple modifiers, thus requiring learners to make an inference about their relative order. Choices are distractors not consistent with training (e.g., Adj-Dem-N) and critical test items which follow isomorphism (e.g., N-Adj-Dem and Dem-Adj-N) or violate it (e.g., N-Dem-Adj, Adj-Dem-N). The session lasts about 20 mins. *Stimuli:* L1 lexical items (following Culbertson & Adger, 2014). Note that using L1 lexical items has several consequences, all positive: it makes the task very simple, is in keeping with the use of the paradigm to investigate

Case For Support

Language: *Kitharaka*				Language: *English*			
Training	Testing (hold-out)	Critical choices	Bias predicting	Training	Testing (hold-out)	Critical choices	Bias predicting
Dem-N, Adj-N	{N, Adj, Dem}	Dem-Adj-N	← **isomorphism**	N-Dem, N-Adj	{N, Adj, Dem}	N-Adj-Dem	← **isomorphism**
		Adj-Dem-N	←L1 similarity			N-Dem-Adj	←L1 similarity
Dem-N, Num-N	{N, Dem, Num}	Dem-Num-N	← **isomorphism**	N-Dem, N-Num	{N, Dem, Num}	N-Num-Dem	← **isomorphism**
		Num-Dem-N	←L1 similarity			N-Dem-Num	←L1 similarity
Num-N, Adj-N	{N, Num, Adj}	Num-Adj-N	← **isomorphism**	N-Num, N-Adj	{N, Num, Adj}	N-Adj-Num	← **isomorphism**
		Adj-Num-N	←L1 similarity			N-Num-Adj	←L1 similarity

naturalness rather than learning, and increases conservatism since it would influence participants only towards choosing more *native-like* patterns. The isomorphism bias predicts that both populations of learners will prefer isomorphic patterns, however for Kitharaka, simple surface transfer from L1 is confounded with the bias of interest. Our previous results suggest that L1 similarity is not a strong factor here, however it is possible to de-confound them by creating a slightly more complex training regime. In **Experiment 1b** Kitharaka speakers (N=40) are trained on phrases with a single post-nominal modifier (N-Adj, N-Num, and N-Dem), and *one combination*, N-Adj-Dem (opposite of L1 order, thus isomorphic). They are tested on hold-out combinations involving the remaining modifier, Num. If the bias is active even when the L1 violates it, Kitharaka speakers will prefer isomorphic placement of Num. Crucially, isomorphism is de-confounded from L1 similarity which predicts a preference for non-isomorphic orders (e.g., N-Adj-Dem-Num over N-Adj-Num-Dem). See Technical Appendix §7 for full table of conditions.

Experiment 2. Typological evidence suggests that post-nominal modifier orders are preferred to pre-nominal ones: this is the **N-first bias**. For example, violating isomorphism appears possible so long as the order is N-first. This predicts speakers of a language with post-nominal modifiers—e.g., Thai—learning a new language with pre-nominal modifiers should show a *stronger* isomorphism bias than English speakers learning a post-nominal pattern. Such a difference between L1 populations may be subtle compared to the isomorphism bias alone: it is hypothesised based on a relatively small frequency difference (rare vs. unattested). We thus test the N-first bias using the POS paradigm and the MSP. The latter is potentially a more sensitive measure, since bias can be revealed through subtle shifts in frequency of competing variants. Whether these two paradigms provide converging results is of general interest since they have not before been directly compared. Why might the results differ? Isomorphic and non-isomorphic order could in principle be equally well-learned but nevertheless isomorphic orders could be more natural to produce (or chose) spontaneously. By contrast, pre- and post-nominal orders violating the isomorphism bias could be equally unnatural, but one could be easier to learn under noisy conditions.

In **Experiment 2a** we uses the same POS design used for Kitharaka speakers in Exp1a for Thai speakers. *Participants:* Thai speakers (N=60; 20 per condition). The **isomorphism bias** predicts that Thai speakers will prefer Dem-Num-Adj-N (or a relevant subset). The alternative choice is more surface similar to participants' L1 (maintains surface order of modifiers), thus choice of the isomorphic pattern cannot be due to simple L1 similarity. Critically, the **N-first bias** additionally predicts that the isomorphism preference will be stronger in Thai speakers (who are choosing among *pre-nominal* patterns) than English speakers (in Exp1a). **Experiment 2b** uses the MSP to test whether learners are more likely to regularize variable input in which the dominant pattern violates isomorphism post-nominally compared to pre-nominally. We use only Adj and Dem as these provided the clearest results in Culbertson & Adger (2014). *Participants:* English and Thai speakers (N=80; 20 per condition). *Procedure*: During training, participants see pictures of objects modified with Adj or Dem (see Technical Appendix §6 for examples) and hear a phrase consisting of a noun and either a single or two modifiers. English learners hear post-nominal modifiers; Thai learners hear pre-nominal modifiers. Each language has a dominant pattern used in 70% of phrases (30% use the opposite, following Culbertson et al., 2012), and conditions differ in whether the dominant order is isomorphic or not (see table). During testing, participants see similar pictures and must produce phrases describing them. The session lasts approximately 30 mins. *Stimuli:*

Case For Support

Pseudo-L1 lexical items (as in Culbertson & Newport, 2015). The **isomorphism bias** predicts that all learners will regularize more if the dominant pattern is isomorphic. By contrast, L1 similarity predicts the opposite (as in Exp1a). The **N-first bias** predicts orders with post-nominal modifiers should be regularized more, independent of isomorphism.

Language:	***English***	***Thai***	predicted
Condition	Training	Training	regularization
isomorphic dominant	N-Adj, N-Dem N-Adj-Dem (70%) N-Dem-Adj (30%)	Adj-N, Dem-N Dem-Adj-N (70%) Adj-Dem-N (30%)	← more
non-isomorphic dominant	N-Adj, N-Dem N-Dem-Adj (70%) N-Adj-Dem (30%)	Adj-N, Dem-N Adj-Dem-N (70%) Dem-Adj-N (30%)	← less

Strand 2. The harmony bias.

In the second set of experiments we return to harmony, discussed above in the context of U20, but typically treated as a general tendency applying across phrases in a language: it is one of the most well-cited universals. Theoretical work formalizing harmony typically assumes the validity of the typological data (e.g., Chomsky, 1988; Baker 2001). However, critics have questioned whether the data would survive more controlled statistical analysis (e.g., Dunn et al., 2011; Ladd et al., 2014). Further, the reliability of harmony seems to differ across phrase types (Dryer 1992). For example, there is a stronger tendency for harmony across elements in the same domain: e.g., modifiers may *precede* the noun, but verbal dependents may *follow* the verb (as in English, modifier-noun, but verb-object order). We test the harmony bias across L1 groups to determine (i) if speakers with and without harmonic noun phrase orders in the L1 both show evidence for a harmony bias, and (ii) whether that bias differs when participants must generalize order across domains.

Experiment 3a uses the POS paradigm with Thai (harmonic: N-Adj-Num-Dem) and Vietnamese (non-harmonic: Num-N-Adj-Dem) speakers. Participants (N=80; 20 per condition) are trained on order of N and two modifiers (Num, Dem), and tested on a third held-out modifier (Adj). If harmony is a universal cognitive bias, learners in both L1 groups should override their L1 order (N-Adj) to maintain harmony (Adj-N). If L1 structure influences learning, the preference in Vietnamese may be weaker. However, even if Vietnamese speakers prefer the non-harmonic L1

Language:	Training	Testing (hold-out)	Bias	predicted choice
Thai	Dem-N, Num-N (both non-native)	(N, Adj)	**harmony →**	**Adj-N**
			L1 similarity →	**N-Adj**
Thai	Dem-N, N-Num (Dem non-native)		no harmony prediction	
			L1 similarity →	**N-Adj**
Vietnamese	Dem-N, Num-N (Dem non-native)		**harmony →**	**Adj-N**
			L1 similarity →	**N-Adj**
Vietnamese	Dem-N, N-Num (both non-native)		no harmony prediction	
			L1 similarity →	**N-Adj**

Language: *Vietnamese*		predicted
Condition	Training	regularization
harmonic dominant	**Dem-N, Num-N** (70%) N-Dem, N-Num (30%)	← more
	N-Dem, N-Num (70%) Dem-N, Num-N (30%)	← more
non-harmonic dominant	**N-Dem, Num-N** (70%) Dem-N, N-Num (30%)	← some (L1)
	Dem-N, N-Num (70%) N-Dem, Num-N (30%)	← less

order, our design still allows us to find evidence of a harmony bias: we compare L1 preference where there is clear harmony pressure (training modifiers are harmonic) and one where harmony makes no predictions (training modifiers are non-harmonic). The latter gives us a pure measure of L1 bias, namely how much Vietnamese learners prefer N-Adj all things equal. If their preference for N-Adj is lower when harmony pressure is present–providing a pressure to override an N-Adj preference–then we will have evidence of harmony at work in Vietnamese. Comparing with Thai allows us to use the exact same conditions. We will pilot first with English (N=20; training on N-Dem, N-Num). **Experiment 3b** uses the potentially more sensitive MSP, following Culbertson et al. (2012), which found a harmonic bias in English speakers. Vietnamese speakers (N=80; 20 per condition) are trained on one of four patterns of N with Dem, and N with Num, and must produce those phrases at test. As in Exp2b, each condition has a dominant pattern (see table). A harmony bias would lead Vietnamese learners to regularize dominant harmonic patterns more than non-harmonic ones (particularly non-harmonic, non-L1). In Culbertson et al. (2012), English learners regularized a fully non-native (post-nominal) harmonic pattern over a partially non-native disharmonic pattern, thus a harmony bias here is not likely to be due to partial L1 overlap.

Case For Support

Experiment 4 uses the POS to test whether the harmony bias is active across domains *and* whether its strength mirrors typological evidence. Dryer (1992) investigates which phrases show a strong tendency to match order of verb (head) and object (dependent). Languages with head-final OV tend to have head-final N-Postposition (e.g., "school in") and N-Genitive (e.g., "Jane of"), but there is no strong tendency for order of Dem and Adj (dependents) with N (head) to follow order of V and O. More recently, Biberauer et al., (2014) argue that rules governing harmony are present within but *not* across nominal and verbal domains. Here we test generalization of head order from nominal to verbal domain, varying whether the nominal phrases show weak or strong correlation with OV order according to Dryer (1992). English and Thai speakers (N=80; 20 per condition) are trained on phrases in minimally different carrier sentences. For example, participants might see "Jane is in school", and hear a translation of the PP, e.g., "...school in". We use Thai (completely harmonic, head-first) and English (primarily head-initial, nominal modifiers head-final, i.e., Dem-N) to see the effect of some L1 disharmony. If harmony holds across domains, then we predict that learners trained on harmonic head-final ordering in the noun phrase will generalize this to OV order. However, based on Dryer (1992), we predict generalization from nominal modifiers Adj, and Dem to be weaker.

Language:	Training	Testing (hold-out)	Critical choices	Bias predicting
English	N-Post, Gen-N	{Object, V}	O-V	←**harmony** (strong)
			V-O	←L1 (similarity)
	N-Post, Dem-N		O-V	←**harmony** (weaker)
			V-O	←L1 (similarity)
Thai	N-Post, Gen-N		O-V	←**harmony** (strong)
			V-O	←L1 (similarity)
	Dem-N, Adj-N		O-V	←**harmony** (weakest)
			V-O	←L1 (similarity)

Strand 3. The suffixing preference.

We propose to test one additional hypothesized bias: the suffixing preference. Typological data show that suffixes are more common than prefixes in the world's languages (very few use predominantly prefixes). This is argued to stem from a preference to have lexically individuating information earlier in a word than category determining information (Hawkins & Cutler, 1988). This is consistent with findings that children and adults are more tolerant of word form variation at the ends of words (Bruening et al., 2012). As with the previous universals targeted here, experimental studies of the suffixing preference use English speakers (St. Claire et al., 2009; Hupp et al., 2009; Bruening et al., 2012), despite the fact that English uses predominantly suffixes. We therefore propose to compare English and Kitharaka speakers, whose language uses predominately prefixes in the noun phrase. **Experiment 5** replicates a study comparing category learning across conditions which mark categories with a suffix or prefix (St. Claire et al., 2009). *Participants:* English and Kitharaka speakers (N=80; 20 per condition). *Procedure:* Stimuli are a set of nonce roots divided into two categories each marked by a nonce affix. For each language group, half of participants are exposed to roots followed by a category affix (suffixing condition), while the other half hear roots preceded by it (prefixing condition). Participants are trained on root-plus-affix pairs and tested on grammatical (matching affix), and ungrammatical (non-matching affix) pairs. If the suffixing preference reflects a universal cognitive bias, then we predict better category learning in both populations, but if it reflects only prior language knowledge, then we will see it only in English participants. Importantly, if the difference in learning between conditions is smaller for Kitharaka compared to English, this will still provide evidence of a suffixing bias in Kitharaka.

Summary. The proposed experiments investigate a theoretically important set of hypothesized cognitive biases to determine whether they are present in learners with typologically distinct L1s. What we uncover will impact linguistic theory: only if the biases we test are shown to be active across L1 populations should they be built into theories of syntax. Further our results will shed light on issues of broad interest in cognitive science: what types of patterns are preferred by the human cognitive system and how does prior (language) experience shape these biases. This work is also a necessary first step toward further interdisciplinary research investigating whether and how the biases uncovered could themselves be learned (even if they are universal), and whether they are specific to the domain of language or reflect general cognitive mechanisms.

Case For Support

Case For Support

Justification of Resources
Directly Incurred
Staff Costs
Postdoctoral research fellow: We request a postdoctoral research fellow (grade 7) for the duration of the grant. A postdoctoral fellow is needed due to the fact that the field research proposed here involves intensive periods of data collection (in UK, Thailand, Kenya, and Vietnam), the need to maintain close control over differences among research sites, and supervision of local ad-hoc workers at those sites. We will seek to recruit a postdoc with extensive experience conducting field research on language in order to maximize our ability to complete the proposed research. In addition the postdoc will be expected to help with programming experiments, analyzing data, and disseminating research, therefore solid prior training and skill in these areas is key.
Travel and Subsistence
International research trips:
The PI and postdoc will travel to Thailand and Vietnam, and Co-I Klaus Abels and the postdoc will travel to Kenya. Trips must take place in the summer since the PI and Co-Is are full time faculty with teaching duties during the semester. They will stay two weeks in each location in order to meet with local collaborators and assistants, finalize experimental materials, set-up all testing hardware and software, and supervise initial running of participants.
Dissemination trips to conferences: We request funds for one of the project staff to attend the following conferences (6 in total over the course of the grant, estimated 2 per year): *Linguistic Society of America* (LSA) and *Linguistic Society of Great Britain* (LAGB) annual meetings, Architectures and Mechanisms of Language Processing (AMLaP), Boston University Conference on Language Development (BUCLD), Cognitive Science Society (Cogsci) annual meeting, and Linguistic Evidence. These conferences attract top researchers across the interdisciplinary audiences targeted. Cogsci and BUCLD also produce widely read conference proceedings.
Yearly project meetings: Our team will meet on a regular basis via Skype. In addition, we request travel funds for both Co-I's to attend this yearly meeting in Edinburgh to assess progress and plans.
Workshop on artificial language learning methods: We request funds for Co-I David Adger to travel to Edinburgh for the workshop we will host. Prof. Adger will help the PI in running the workshop and his involvement will likely serve to increase interest from the syntax community. To attract participants and add educational value we ask for funds to invite two experts in artificial language learning to speak at the workshop, one from the UK and one from the US.
***Bilingualism Matters event*:** We request funds for Co-Is to travel to Edinburgh for this event.
Other Directly Incurred Costs
Participant Payments: We request funds to compensate adult participants taking part in our experiments (8 experiments in total plus 2 planned pilots). Experiments with English speakers will require a total of 330 participants (280 planned plus 50 extra to account for any additional pilots and occasional loss of participants (e.g., due to equipment failure). Experiments with Thai speakers will require a total of 200 participants (180 planned plus 20 extra). Experiments with Kitharaka speakers will require a total of 160 participants (140 planned plus 20 extra). Experiments with Vietnamese speakers will require a total of 160 participants (140 planned plus 20 extra). We will pay participant at the rate of £7 per session, with 850 participants total = £5950.
Laptops: 3 MacBook Airs (13", 1.6GHz Processor, 256 GB Storage, £999.00 each) are required in order to conduct field research; since we are conducting this research in locations which do not guarantee us computers, this is critical to our ability to conduct the proposed experiments.
Bilingualism Matters impact event: We request funds for a knowledge exchange event with BM (see Pathways to Impact). The event will be held at the University of Edinburgh, therefore we request funds for catering and promotion materials. This event will be organized with the help of BM Research Coordinator Dr. Madeleine Beveridge and we will aim to attract approximately 100 attendees. To encourage participation of educators we plan to hold an evening event.

Justification for Resources

Workshop on ALL methods: One of the main aims of this project is to promote the use of artificial language learning in syntax. To achieve this we will hold a three day workshop on ALL methods for interested academics and ESOL practitioners at the University of Edinburgh. Funds are requested to provide catering for 50 attendees and video recording for wider dissemination.
Ad-hoc field payments to field workers: We request funds for three assistants to recruit and run participants. In Thailand and Vietnam, these assistants will be postgraduate students identified by our collaborators, who are native speakers of the local language with high English proficiency. In Kenya, the assistant will be a teacher at the local primary school with high English proficiency. We will hire each assistant for 300 hours of work commensurate with local pay scales, which on average we estimate at £10/ph. More participants will be run in Thailand compared to Vietnam and Nairobi, however we expect recruitment to be easier in Thailand. While the full team is present, participants will be run concurrently on two laptop computers to maximize use of the assistant's time. After than, participants will be run individually by RAs until data collection is complete.
Recruitment: Funds are requested for expenses associated with recruitment and relocation of a postdoctoral research fellow. These are standard amounts used at Edinburgh, which will allow us to interview candidates and attract the best postdoc possible.

Directly Allocated costs

PI/Co-I staff costs:

Dr. Jennifer Culbertson (PI): The PI is an expert in language acquisition with a speciality in learning biases across development. We request 7 hours/week of her time, averaged over the course of the grant. The PI's duties include experimental design, supervision of data collection, data analysis, and dissemination activities. The PI will travel to Thailand and Vietnam for research trips (estimated 2 weeks each), during which time she will work exclusively on the project.
Dr. David Adger (Co-I): The Co-I is an expert in theoretical syntax and its relation to cognitive science, including how theories of syntax account for language variation within and across speaker populations. We request 3 hours/week of his time, averaged over the course of the grant to cover involvement in experiment design, supervision of data collection, and dissemination activities.
Dr. Klaus Abels (Co-I): The Co-I is an expert in theoretical syntax, with specialties in word order typology and the syntax-semantic interface. Additionally, he has expert knowledge of the structure of Kitharaka. We request 4 hours/week of his time, averaged over the course of the grant to cover involvement in experiment design and dissemination activities as well as a research trip to Kenya during which time (estimated 2 weeks) he will be working exclusively on the project.

Other Directly Allocated costs:

Computing support: 5% (£8,245) of a computing officer's time to provide local hardware and software support (including installation, upgrade, maintenance, user assistance), maintenance of file storage and back-up systems and help with making the data widely available via online and web-based systems.
Knowledge exchange: 3% (£4,007) of a Knowledge Exchange Officer to provide dedicated support for the knowledge exchange and impact activities identified in the Pathways to Impact attachment. The time included is an average as it's anticipated the level of support needed will vary over the duration of the grant.
Bilingualism Matters Research Coordinator**:** 3% (£4,007) of the Bilingualism Matters Research Coordinator's time to provide dedicated help to organise the BM impact event. The time included is an average as it's anticipated the level of support needed will vary over the duration of the grant.
Research facilities;Lab costs: We request funds to cover the cost of using the Language Evolution Lab, a Small Research Facility within the School of PPLS at the University of Edinburgh, not funded via the University's estates and indirect costs. It will be used by the PI and postdoc for all experiments involving English-speaking participants. The lab houses multiple booths and allows testing of up to 4 individuals simultaneously. We estimate requiring 50 hours access to this facility, at the hourly rate of £33. (N=330 x 30 min=150/3 rooms=50hr x £33/hr=£1650)

Justification for Resources

Dr. Jennifer Culbertson

Education

2005 – 2010	Ph.D. (and M.A.) in Cognitive Science, Johns Hopkins University, Baltimore, MD
2000 – 2004	B.A. (Hons) in Linguistics, Brown University, Providence RI

Employment

2014 – present	Chancellor's Fellow, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh
2012 – present	Adjunct Assistant Professor, Cognitive Science, Johns Hopkins University
2012 – 2014	Assistant Professor, Program in Linguistics, George Mason University
2010 – 2012	Postdoctoral Research Fellow, Center for Language Sciences, University of Rochester

Selected Publications

Culbertson, J. and Newport, E.L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, 139:71–82.

Enochson, K. and Culbertson, J. (2015). Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLoS ONE*, 10(3):e0116946.

Culbertson, J. and Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, 111(16):5842–5847.

Legendre, G., Culbertson, J., Zaroukian, E., Hsin, L., Barrière, I., and Nazzi, T. (2014). Is children's comprehension of subject-verb agreement universally late? Comparative evidence from French, Spanish, and English. *Lingua*, 144:21–39.

Culbertson, J. and Legendre, G. (2013). Prefixal agreement and impersonal 'il' in Spoken French: Experimental evidence. *Journal of French Language Studies*, 24(1):83–105.

Culbertson, J. and Wilson, C. (2013). Artificial grammar learning of shape-based noun classification. In M. Knauff et al., (eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, pp. 2118–2123. Cognitive Science Society, Austin, TX.

Culbertson, J., Smolensky, P., and Legendre, G. (2013). Testing Greenberg's Universal 18 using the Mixture Shift Paradigm for artificial language learning. In S. Kan et al., (eds.), *Proceeding of NELS 40*, pp. 133–146. GLSA, Amherst, MA.

Culbertson, J., Smolensky, P., and Wilson, C. (2013). Cognitive biases, linguistic universals, and constraint-based grammar learning. *Topics in Cognitive Science*, 5(3):392–424. [pdf]

Culbertson, J. and Smolensky, P. (2012). A Bayesian model of biases in artificial language learning: The case of a word-order universal. *Cognitive Science*, 36(8):1468–1498.

Culbertson, J. (2012). Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass*, 6(5):310–329.

Culbertson, J., Smolensky, P., and Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122:306–329.

Culbertson, J., Smolensky, P., and Legendre, G. (2012). Statistical learning constrained by syntactic biases in an artificial language learning task. In A.K. Biller, et al., (eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development*, vol. 1, pp. 139–151. Cascadilla Press, Somerville, MA.

Culbertson, J. and Legendre, G. (2011). Investigating the evolution of agreement systems using an artificial language learning paradigm. In D. Bailey and V. Teliga (eds.), *Proceedings of the 39th Western Conference on Linguistics*, pp. 46–58.

Gross, S. and Culbertson, J. (2011). Revisited linguistic intuitions. *British Journal for the Philosophy of Science*, 62:639–656.

Culbertson, J. (2010). Convergent evidence for categorial change in French: from subject clitic to agreement marker. *Language*, 86(1):85–132.

Culbertson, J. and Gross, S. (2009). Are linguists better subjects? *British Journal for the Philosophy of Science*, 60:721–736.

Demuth, K., Culbertson, J., and Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, 49(2):137–174.

Awards and Grants

2012	Robert J. Glushko Prize for Outstanding Doctoral Dissertations in Cognitive Science
2010	NIH Postdoctoral Research Fellow (Training grant awarded to Center for Language Sciences, University of Rochester)
2006	National Science Foundation Graduate Research Fellow

Prof David Adger

Education

1990 – 1994	Ph.D. in Linguistics and Cognitive Science, University of Edinburgh
1989 – 1990	MSc in Cognitive Science, University of Edinburgh
1985 – 1989	MA in Linguistics and Artificial Intelligence, University of Edinburgh

Employment

2006 – present	Professor of Linguistics, Queen Mary University of London
2002 – 2006	Reader in Linguistics, Queen Mary University of London
1993 – 2002	Lecturer in Linguistics, University of York

Selected Publications

Adger, D. and Svenonius, P. (2015). Linguistic Explanation and Domain Specialization: a case study in bound variable anaphora. *Frontiers in Psychology*, 6.1421.

Culbertson, J. and Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, 111(16):5842–5847.

Adger, D. (2014). Variability and Grammatical Architecture. in C. Picallo, ed. *Variation in a Minimalist Perspective*. Oxford: Oxford University Press. 179-196.

Adger, D. (2013). *A Syntax of Substance*. Cambridge: MIT Press.

Cheshire, J. Adger, D. and Fox, S. (2013). Relative *who* and the Actuation Problem. *Lingua,* 126: 51-77.

Harbour, D., Watkins, L. and Adger, D. (2012). Information structure, discourse structure, and noun phrase position in Kiowa. *International Journal of American Linguistics* 78: 97-126.

Mesoudi, A., McElligot, A. and Adger, D. (2011). Integrating genetic and cultural evolutionary approaches to language. *Human Biology*, 83.2, 141-151.

Adger, D. and Smith, J. (2010). Variation in Agreement: a lexical feature-based approach, *Lingua* 120: 1109-1134.

Adger, D. (2010). “A minimalist theory of feature structure” in A. Kibort and G. Corbett (eds) *Features: Perspectives on a Key Notion in Linguistics*. Oxford: Oxford University Press. 185-218.

Adger, D. Harbour, D. and Watkins, L. (2009). *Mirrors and Microparameters: Phrase Structure beyond Free Word Order*. Cambridge University Press: Cambridge.

Adger D. and Harbour, D., (2007). The Syntax and Syncretisms of the Person Case Constraint. *Syntax,* 10:2-37

Adger, D. (2006). Combinatorial Variation. *Journal of Linguistics*, 42:503-530.

Adger, D. (2006). Remarks on feature theory and movement. *Journal of Linguistics*, 42:663-674

Adger, D. (2006). Post Syntactic Movement and the Old Irish Verb. *Natural Language and Linguistic Theory,* 24:605-654

Adger, D. and Ramchand, G. (2005). Merge vs Move: wh-dependencies revisited. *Linguistic Inquiry,* 36: 161-193.

Adger, D., de Cat, C., and Tsoulas, G. (eds). (2004). *Peripheries: syntactic edges and their effects.* Kluwer: Dordrecht.

Adger, D. (2003). *Core Syntax*. Oxford University Press: Oxford.

Recent Awards and Grants

2015-2019	A Scots Syntactic Atlas. AHRC £800,000 (Co-I)
2015-2017	Clausal Selection: Integrating theoretical and experimental approaches. EU Marie-Curie €185,000 (PI)
2013-2016	Advancing the European Multilingual Experience. EU Framework 7 consortium partner €5,000,000 (Co-I)
2009-2012	Atomic Linguistic Elements of Phi. AHRC £345,000 (PI)

Dr. Klaus Abels

Education

1998 – 2003	Ph.D. in Linguistics, University of Connecticut
1998 – 2001	MA in Linguistics, University of Connecticut
1992 – 2001	MA in Translation, Humboldt Universität zu Berlin

Employment

2012 – present	Reader in Linguistics, Department of Linguistics, University College London
2008 – 2012	Lecturer in Linguistics, Department of Linguistics, University College London
2004 – 2007	Lecturer in Linguistics, Department of Linguistics, Universitetet i Tromsø
2003 – 2004	Postdoctoral Researcher, Department of Slavonic Studies Universität Leipzig

Selected Publications

Abels, K. (2015). Syntax. In N. Braber, L. Cummings and L Morrish (eds.), *Exploring Language and Linguistics*. pp. 137-167. Cambridge University Press, Cambridge.

Abels, K. (2015). Word Order. In A. Alexiadou and T. Kiss (eds), *Syntax – Theory and Analysis. An International Handbook*, pp. 1400–1445. Mouton de Gruyter, Berlin.

Abels, K. (2013). Comments on Hornstein. *Mind and Language*, 28(4):421–429.

Abels, K. (2012). *Phases: An essay on cyclicity in syntax*. Linguistische Arbeiten 543, DeGruyter, Berlin.

Abels, K. (2012). The Italian Left Periphery: A view from Locality. *Linguistic Inquiry*, 43(2):229–254.

Abels, K. and Neeleman, A. (2012). Linear Asymmetries and the LCA. *Syntax*, 15(1):25–74.

Abels, K. (2010). Factivity in Exclamatives is a Presupposition. *Studia Linguistica*, 64(1):141–157.

Abels, K. and Martí, L. (2010). A unified approach to split scope. *Natural Language Semantics*, 18(4):435–470.

Abels, K. and Neeleman, A. (2010). Nihilism masquerading as progress. Lingua, 120:2657–2660.

Abels, K. (2009). Some implications of improper movement for cartography. In J. van Craenenbroek (ed.). *Alternatives to Cartography*. pp. 325-359. deGruyter, Berlin.

Abels, K. and Bentzen, K. (2009) A note on the punctuated nature of movement paths. Catalan Journal of Linguistics 8:19–40.

Abels, K. and Neeleman, A. (2009). Universal 20 without the LCA. In *Merging Features: Computation, Interpretation and Acquisition*. Brucart, Josep M, Gavarró, Anna, and Solà, Jaume (eds.) pp. 60–79 . Oxford University Press, Oxford.

Abels, K. and Muriungi, P. (2008). The focus particle in Kîîtharaka: Syntax and Semantics. *Lingua*, 118(5):687-731.

Abels, K. (2007). Towards a restrictive theory of remnant movement. *Linguistic Variation Yearbook*, 7:53–120.

Abels, K. (2007). Deriving selectional properties of 'expletive' predicates. In *Interfaces and Interface conditions*. Späth, Andreas (ed) pp. 115-140. DeGruyter, Berlin.

Abels, K. (2006). Successive Cyclicity, Anti-locality, and Adposition Stranding. In *Minimalism – The Essential Readings*. Bošković, Željko and Lasnik, Howard (eds) pp. 219-223. Malden, Blackwell Publishing.

Abels, K. (2005). 'Expletive Negation' in Russian: A Conspiracy Theory. Journal of Slavic Linguistics 13(1), pp. 5–74.

Awards and Grants

2012	Electronic Learning Development Grant, UCL
2011	Arts and Humanities Research Council funded research leave: "Linguistic Universals and the Order of Verbs in Germanic" (Grant AH/I020292/1)□
2008	University of Connecticut 40-under-40 laureate□
1999–2003	Predoctoral Fellowship at the University of Connecticut□
1999	Quadrille Ball Scholarship of the American Germanistic Society□
1998–1999	Fulbright Grant□

References

Abels, K., & Neeleman, A. (2012). Linear asymmetries and the LCA. *Syntax*, *15*, 25–74.

Adger, D. (2003). *Core Syntax*. Oxford: Oxford University Press.

Baker, M. (2001). *The atoms of language: The mind's hidden rules of grammar*. New York, NY: Basic Books.

Biberauer, T., Holmberg, A., & Roberts, I. (2014). A syntactic universal and its consequences. *Linguistic Inquiry*, *45*, 169–225.

Bruening, P., Brooks, P., Alfieri, L., Kempe, V., & Dabašinskienė, I. (2012). Children's tolerance of word-form variation. *Child Development Research*, .

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. Cambridge, MA: MIT Press.

Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry*, *36*, 315–332.

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*, 5842–5847.

Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, *139*, 71–82.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*, 306–329.

Cysouw, M. A. (2005). Quantitative methods in typology. In *Quantitative Linguistics: An International Handbook* (pp. 554–578). Berlin: Mouton de Gruyter.

Dryer, M. (1992). The Greenbergian word order correlations. *Language*, *68*, 81–183.

Dunn, M., Greenhill, S., Levinson, S., & Gray, R. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*, 79–82.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–448.

Finley, S., & Badecker, W. (2008). Substantive biases for vowel harmony languages. In J. Bishop (Ed.), *Proceedings of WCCFL 27* (pp. 168–176).

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of Language* (pp. 73–113). Cambridge, MA: MIT Press.

Hawkins, J. A., & Cutler, A. (1988). Psycholinguistic factors in morphological asymmtery. In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 280–317). Oxford: Blackwell.

Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation. *Language Learning and Development*, *1*, 151–195.

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes*, *24*, 876–909.

Ladd, D. R., Roberts, S. G., & Dediu, D. (2014). Correlational studies in typological and historical linguistics. *Annual Review of Linguistics*, *1*.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (leap-q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, *25*, 83–127.

Moreton, E., & Pater, J. (2012a). Structure and substance in artificial-phonology learning, part I: Structure. *Language and Linguistics Compass*, *6*, 686–701.

Moreton, E., & Pater, J. (2012b). Structure and substance in artificial-phonology learning, part II: Substance. *Language and Linguistics Compass*, *6*, 702 – 718.

Partee, B. H. (1987). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, & M. Stokhof (Eds.), *Studies in discourse representation theory and the theory of generalized quantifiers* (pp. 115–143). Dordrecht: Foris.

Piantadosi, S. T., & Gibson, E. (2014). Quantitative standards for absolute linguistic universals. *Cognitive Science*, *38*, 736–756.

Rijkhoff, J. (2004). *The Noun Phrase*. Oxford: Oxford University Press.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 44–449.

St. Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science*, *33*, 1317–1329.

Wilson, C. (2003). Experimental investigation of phonological naturalness. In G. Garding, & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 101–114). Somerville, MA: Cascadilla Press.

Wilson, C. (2006). An experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945–982.

Data Management Plan

Assessment of existing data
Explain the existing data sources that will be used by the research project (with references), and any gaps between the currently available and required data for the research
No experimental studies have been carried out which address the research questions set forth in the Case for Support. Our experiments are carefully designed to answer these questions and differ substantially from previous related research.

Information on new data
Please provide a brief description of new data which you envisage creating.
Data will consist of participant responses, specifically, utterances produced or typed and/or buttons pressed in experimental sessions.

Quality assurance of data
Please briefly describe the procedures for quality assurance that will be carried out on the data collected
Data collected will be standardized across experiments and participant groups by software customised by the project team. Data will coded (or scored) either automatically or by a coder blind to the experimental condition.
Back-up and security of data
Please describe the data back-up procedures that you will adopt to ensure the data and metadata are securely stored during the lifetime of the project.
Data collected on testing computers will be stored locally and transferred daily to a server maintained by the PI. Access to the server will be restricted to the PI, Co-I's, and postdoc. Access to the testing computers will be extended to the ad-hoc field workers at each research site.

Expected difficulties in data sharing
If you expect any obstacles to sharing your newly generated data please explain their causes and possible measures you are going to apply to overcome those.
Data collected will be anonymized, and therefore we foresee no issues in sharing the data. Participants will provide written consent prior to participation and will be informed that their anonymized data will be used for research and may be shared or made publicly available.

Copyright/Intellectual Property Right
Please state who will own the copyright and IPR of any new data that you will generate.
The data will not be covered by a copyright.

Responsibilities
Outline responsibilities for data management within research teams at all partner institutions
The postdoctoral research fellow and ad hoc field workers will be responsible for collecting the data. The postdoc will be responsible for transferring the data to the server daily. The PI will ensure the postdoc has access to the relevant software, is trained in the process, and has time in his or her work plan for data organization and management.

Preparation of data for sharing and archiving
State the plans for management and archiving of collected data
For each study, a summary of experimental design and data and coding collection methods will be created. Individual participant data will be documented with an anonymized participant

Data Management Plan

identification code, an identifier for the experiment and condition, and any demographic data collected.

Other issues

Please indicate if there are there any other issues relating to data management or sharing.

NA

Data Management Plan

Technical Appendix

1. Linguistic issues.

Classifiers. Both Thai and Vietnamese are numeral classifier languages, thus typically phrases including a numeral or a demonstrative would also include a numeral classifier (a small marker that is determined by the semantic properties of the noun). The artificial languages we teach speakers of these languages will not include classifiers since it is reasonable to assume such speakers can acquire languages without them. This will allow us maintain the highest level of similarity between the stimuli for all speaker groups. However, if this proves to be a problem, the experiments here can be run including classifiers. Because this is something Thai and Vietnamese share, we will be able to compare across these languages straightforwardly.

Agreement. Kitharaka has noun classes (similar to gender in languages like French or Spanish) which trigger agreement on nominal modifiers like adjectives. The artificial languages we teach to Kitharaka speakers will not include agreement–again under the assumption that they can learn such a language without undue problems–to maintain similarity across groups. While we don't expect this to disrupt learning, we will pilot a scaled-down version of Exp1a (with English speakers) in which we omit plural morphology (e.g., using "shoe two blue" instead of "shoes two blue") to confirm that the results replicate.

2. Recruitment plan.

Testing speakers of three languages other than English is relatively ambitious, however we have a plan in place for recruiting participants which we believe will allow us to complete the research according to our proposed timeline. For each populations we propose to test, we have identified collaborators who have agreed to provide us with guidance in developing **experimental materials**, **space for testing**, **access to a participant population**, suggestions for **ad hoc field assistants**.

In all cases, we will prepare most materials in advance of trips with the help of our collaborators. The PI and postdoc will travel to Thailand and Vietnam, and Co-I Klaus Abels and the postdoc will travel to Kenya. There they will meet with local collaborators and assistants, prepare final experimental materials, set-up all testing hardware and software, and supervise initial running of participants. Our collaborators are:

In **Thailand (Bankok)**:

- Dr. Peter Jenks (UC Berkeley), an expert in Thai syntax and fluent Thai speaker
- Syntacticians Dr. Nattaya Piriyawiboon and Dr. Sugunya Ruangjaroon (Linguistics, Srinakharinwirot University)

In **Kenya (Nairobi)**:

- Dr. Peter Muriungi (Chuka University, Tharaka County), an expert on the syntax of Kitharaka and other African languages, and a native Kitharaka speaker

In **Vietnam (Hanoi)**:

- Dr. Marc Brunelle (Ottowa), an expert in Vietnamese phonology and fluent Vietnamese speaker with years experience doing research in Vietnam
- Syntactician Dr. Nguyễn Hồng Cổn (Linguistics, Hanoi University of Social Sciences)

In Thailand and Vietnam, we will be visiting **large universities** where we can recruit substantial numbers of students relatively quickly. In Kenya, our collaborator works with a **large local school** where we can recruit parents of the school children as participants and we will likely have a English teacher at the school as our research assistant. Recruitment in Kenya will begin before we travel there. We also plan to speak with local press in Nairobi as part of our impact plan, and can use this to advertise our research to a wider audience of potential participants.

3. Language proficiency test.

We will use the LEAP-Q questionnaire (Marian et al., 2007) to assess other language proficiency and usage. This relatively short questionnaire was developed for use in multiple speaker populations and has been shown to be internally valid and to accurately predict performance on other behavioral measures of language proficiency. A Vietnamese version of the questionnaire already exists. We will develop the Thai and Kitharaka versions with the help of our collaborators. This questionnaire allows for participants to enter information for up to five languages, therefore in addition to English proficiency we will also have data about any other languages participants have knowledge of. We expect to be able to recruit Thai and Vietnamese speakers with varying levels of English proficiency and little or no knowledge of other languages. We expect to be able to recruit Kitharaka speakers will little or no knowledge of English, but potentially knowledge of other Bantu languages which do not differ from Kitharaka in terms of the patterns we are testing (e.g., Swahili).

4. Pilot work on the isomorphism bias in English speakers.

As discussed in the Case for Support, we have conducted pilot research investigating English speakers' inferences about the relative order of modifiers when they come after the noun (Culbertson & Adger, 2014). The first experiment tested modifier ordering for each of the pairs of modifiers in a between-subjects design as in Table 1. In the absence of any input regarding relative ordering of modifiers, participants were very likely to choose the isomorphic order (Figure 4A). The second experiment involved training on all pairs of modifiers within subjects, and testing on the relative order of all pairs (2a), and all pairs plus all three modifiers in a single phrase (2b.). Again, the results indicate very strong preference for isomorphic orders (Figure 4B).

Experiment	Training	Testing (hold-out)	Critical choices	Bias predicting
1	N-Dem, N-Adj	{N, Adj, Dem}	N-Adj-Dem	← isomorphism
			N-Dem-Adj	←L1 similarity
	N-Dem, N-Num	{N, Dem, Num}	N-Num-Dem	← isomorphism
			N-Dem-Num	←L1 similarity
	N-Num, N-Adj	{N, Num, Adj}	N-Adj-Num	← isomorphism
			N-Num-Adj	←L1 similarity
2a	all three	all pairs of two	all three as above	
2b	all three	all pairs of two and all three mods	all three as above and:	
			N-Adj-Num-Dem	← isomorphism
			N-Dem-Num-Adj	←L1 similarity

Table 1. Summary of Experiment 1 conditions.

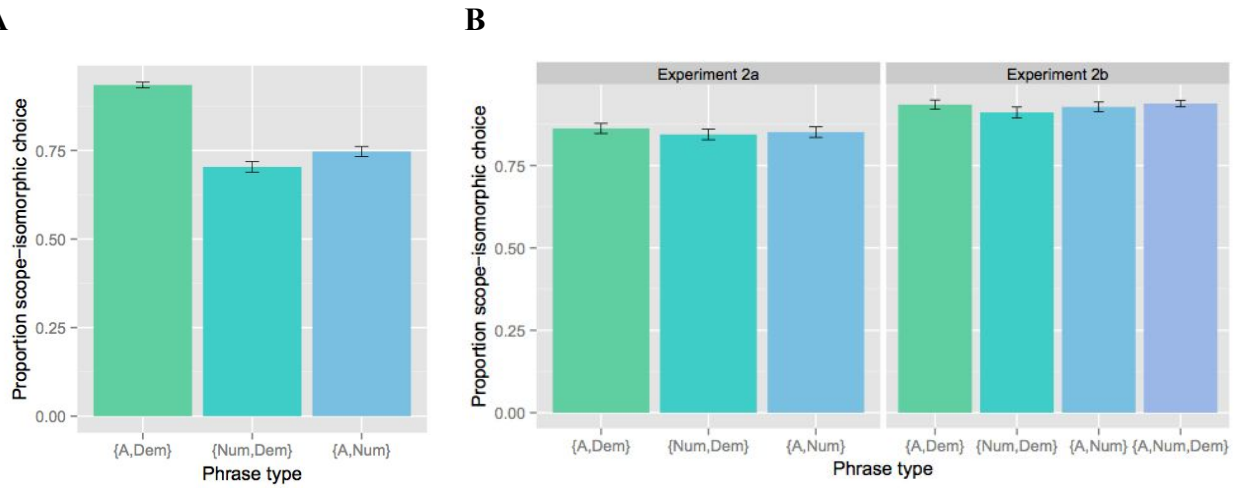

Figure 4. A: Choice of scope-isomorphic order in critical phrase types across conditions in Experiment 1. **B**: Choice of scope-isomorphic order in phrases with a subset of two modifiers (Exp. 2a and 2b), and all three modifiers (2b).

5. Additional information for poverty-of-the-stimulus (POS) tasks.

Example stimuli for the task are shown in Figures 5 and 6. Stimuli will be modified for Thai, Vietnamese and Kitharaka in consultation with our local informants.The procedure is as follows: During training, participants see a phrase in their L1, and hear synthesized audio of the "translation". Participants must click on what they heard. During testing, they see a phrase in their L1 to be translated and must click what they think a speaker would be most likely to say. Choices are either incompatible with training (distractors) or use training-compatible orders that are differ in a critical feature–e.g., isomorphism or harmony.

A **B**

Figure 5. Example training (**A**) and testing (**B**) trials used in POS experiments.

6. Additional information for mixture-shift paradigm (MSP) tasks.

Example stimuli for the task are shown in Figure 6 and Table 2. Stimuli will be modified for Thai, Vietnamese and Kitharaka in consultation with our local informants.The procedure is as follows: In the first phase of training, participants see grayscale individual objects and hear the object labels (synthesized audio is used). They are then tested on their ability to produce the correct label given the object. In the second phase of training, participants see objects modified by Adj or Dem (Figure 6) and hear phrases describing them. They are then tested on their ability to choose the picture (from a set of four) which matches a phrase they hear. During the critical testing phase, participants see objects modified by Adj or Dem (as in Figure 6), and must produce a phrase to describe them.

A **B**

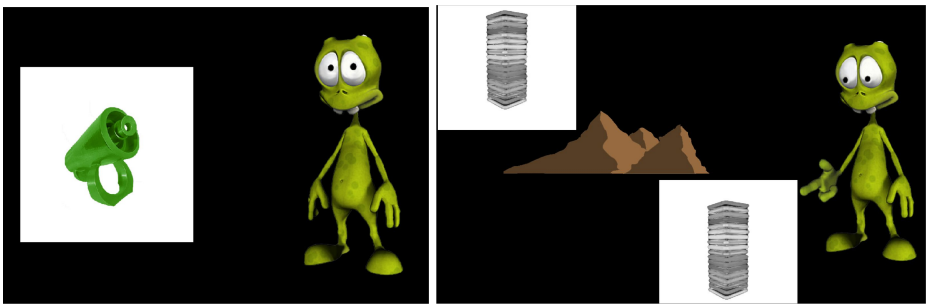

Figure 6. Example adjective (**A**) and demonstrative (**B**) trials used in MSP experiments.

Adjectives		Numerals	
blu0	'blue'	dof	'two'
sprat	'spotted'	0rɛz	'three'
flərf	'furry'	fɔrtʃ	'four'

Table 2. Example "pseudo-English" words.

7. Additional details for proposed experiments.

Figure 7 shows the table of conditions for Exp.1b, which uses the POS paradigm to test whether Kitharaka speakers are biased in favor of modifier orders conforming to the isomorphism bias. They are trained on single modifier phrase N-Adj, N-Num, and N-Dem (as in Exp1a) *and* the combination N-Adj-Dem. They are then tested in one of two hold-out conditions involving a combination of either Adj or Dem with the remaining modifier Num. No evidence for relative placement of Num has been provided. Critical choices are between isomorphic (N-Num-Dem, N-Num-Adj) and non-isomorphic (N-Dem-Num, N-Num-Adj) modifier orders. Isomorphism is de-confounded from L1 similarity here since Kitharaka uses N-Dem-Num-Adj; this makes the non-isomorphic orders of the hold-out phrases more similar to Kitharaka than the isomorphic ones.

Language: *Kitharaka*

Training	Testing (hold-out)	Critical choices	Bias predicting
N-Adj, N-Num, N-Dem, **N-Adj-Dem**	{N, Num, Dem}	N-Num-Dem	← **isomorphism**
		N-Dem-Num	←L1 similarity
	{N, Adj, Num}	N-Adj-Num	← **isomorphism**
		N-Num-Adj	←L1 similarity

Figure 7. Conditions for Exp1b.